Structure Determination of Amyloid-β_{1-42} Fibrils by cryo-EM

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Carla Theresa Schenk
aus Aachen

Jülich, März 2018

aus dem Institut für komplexe Systeme (ICS-6)
des Forschungszentrum Jülich

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Jun.-Prof. Dr. Gunnar Schröder
2. Prof. Dr. Dieter Willbold

Tag der mündlichen Prüfung:
12.04.2018

Acknowledgment

First and most importantly, I want to thank my supervisor Gunnar Schröder for sharing his immense knowledge, giving advice and motivation, proof reading my thesis and being the best supervisor in the world!

I would also like to thank Prof. Dieter Willbold for giving me the opportunity to work in his institute and being co-examiner on my thesis.

I am thankful to all staff members from the secretary's office who helped organizing everything.

My thanks also goes to all collaborators that I had the pleasure to work with. A special thank goes to Lothar Gremer for his great knowledge on amyloid fibrils and his constant motivation throughout the project.

I thank all (former) group members, especially Zhe Wang, Dennis Dellacorte, Michaela Spiegel, Amudha Duraisamy and Tatjana Braun for the great time we had together and, lastly, everyone in the institute for creating such a friendly atmosphere to work in.

Abstract Carla Schenk

In 2010 it was estimated that around 35.6 million people are suffering from dementia with a total cost of 6 billion USD for medical treatment and care and it is assumed that every year 7.7 million new patients are diagnosed. Alzheimer's Disease (AD) is the most common of all neurodegenerative diseases accounting for 60-80 % of all cases. It has been a subject of interest and research since it was first described in 1906 but 100 years later the disease is still far from being fully understood.
Though it is known that in AD fibrils and fibrillar plaques made from the amyloid β (Aβ) protein play an important role and therefore their structure is of particular interest. Structure determination of Aβ fibrils, however, is very complicated as they are usually very heterogeneous.

In this work $A\beta_{1-42}$ fibrils were investigated with cryo-electron microscopy (cryo-EM). The fibrils were produced from recombinant $A\beta_{1-42}$ at low pH with organic co-solvent. Three different types of polymorphs are observed in the fibril sample. For one of the polymorphs a high resolution density map was obtained in which the complete backbone and almost all side chains are visible. The map allowed for building an atomic model of the structure. The model reveals that the fibril consists of two protofilaments which interface is formed by the hydrophobic C-termini. The individual subunits adopt an "LS"-shape and also the N-termini are part of the structure which has not been described before. The subunits are not planar and arranged in a staggered fashion along the helical axis which leads to different fibril ends.
Density maps for the other polymorphs are determined at medium resolution and they show that the second polymorph consists of similar "LS"-shaped subunits but with an interface formed by the N-termini of the protein. The third polymorph observed in this sample is made out of two fibrils of type one and has most likely the same N-terminal interface that was described for the second polymorph.

Furthermore a method to detect ligand binding sites was developed in this work. Only a protein-ligand nuclear magnetic resonance (NMR) titration spectrum and the protein structure is needed to predict the binding site automatically. The method is based on comparing the experimental titration with several predicted spectra and determines the binding site to be the one corresponding to the best matching predicted spectrum. This methods works on many different protein-ligand systems and could also be used to study the interaction between Aβ fibrils and binding partners.

Zusammenfassung

Im Jahr 2010 wurde geschätzt, dass circa 35,6 Millionen Menschen an Demenz leiden und die Gesamtkosten dafür waren etwa 604 Millionen USD für Behandlung und Versorgung der Betroffenen und es wird angenommen, dass jedes Jahr etwa 7,7 Millionen neue Fälle dazukommen. Die Alzheimersche Krankheit ist die am häufigsten vorkommende neurodegenerative Erkrankung und für 60-80 % aller Fälle verantwortlich. An der Alzheimerschen Krankheit wird geforscht seit sie 1906 das erste Mal beschrieben wurde aber auch 100 Jahre nach ihrer Entdeckung ist sie noch lange nicht vollständig verstanden.
Fibrillen und fibrilläre Plaques, die aus dem Protein Amyloid β (Aβ) bestehen, spielen in der Krankheit eine wichtige Rolle und deshalb ist es von besonderem Interesse ihre Struktur zu kennen. Die Strukturbestimmung von Aβ Fibrillen ist jedoch sehr schwierig, da die Fibrillen im Allgemeinen sehr heterogen sind.

In dieser Arbeit wurden $A\beta_{1-42}$ Fibrillen mit Kryo-Elektronenmikroskopie (Kryo-EM) untersucht. Die Fibrillen wurden aus rekombinantem $A\beta_{1-42}$ bei niedrigem pH und unter der Verwendung eines organischen Kosolvens produziert. Drei verschiedene Arten von Polymorphen konnten in der verwendeten Probe beobachtet werden. Für einen Polymorph konnte eine hochaufgelöste Dichtemap erzeugt werden in der das vollständige Proteinrückgrat und fast alle Seitenketten sichtbar sind. Mithilfe der Map konnte ein atomares Modell der Struktur erzeugt werden. Das Modell zeigt, dass die Fibrille aus zwei Protofilamenten besteht, deren Interface aus den hydrophoben C-termini geformt ist. Die einzelnen Proteinmoleküle in der Fibrille haben eine "LS" Form und auch die N-Termini sind Teil der Struktur, was zuvor noch nicht beobachtet wurde. Die Monomere sind nicht planar und gegeneinander versetzt entlang der Fibrille angeordnet, was zu unterschiedlichen Fibrillenenden führt.

Dichtemaps mit mittlerer Auflösung konnten für die anderen beiden Polymorphen erzeugt werden und sie zeigen, dass der zweite Fibrillentyp aus ähnlichen, "LS"-geformten Monomeren wie die erste Fibrille besteht. Das Interface in dieser Fibrille wird jedoch von den N-Termini des Proteins geformt. Der dritte Polymorph in der Probe besteht aus zwei Fibrillen von Typ eins, die wahrscheinlich das gleich N-terminale

Interface wie die zweite Fibrille haben.

Außerdem wurde in dieser Arbeit eine Methode zur Vorhersage von Ligandbindungsstellen entwickelt. Dafür werden nur ein Titration-Kernspinresonanzspektrum (NMR) von Protein und Ligand und die Proteinstruktur benötigt. Die Methode basiert auf dem Vergleich eines experimentellen Titrationsspektrum mit verschiedenen vorhergesagten Spektren. Die Bindungsstelle des Liganden entspricht der Stelle, die zu dem am besten passenden vorhergesagten Spektrum gehört.
Die Methode kann auf viele verschiedene Protein-Ligand Systeme angewendet werden und könnte auch dafür benutzt werden Interaktionen von Aβ Fibrillen mit Bindungspartnern zu untersuchen.

Contents

List of Abbreviations

Aβ ..... Amyloid β
ACN ..... Acetonitrile
AD ..... Alzheimer's disease
ADAM ..... A disintegrin and metalloproteinase domain
AFM ..... Atomic force microscopy
AICD ..... APP incracellular domain
APP ..... Amyloid precusor protein
CA ..... Correspondance Analysis
CC ..... Cross-correlation
CSP ..... Chemical shift perturbation
CSF ..... Cerebrospinal fluid
cryo-EM ..... Cryo-electron microscopy
CTF ..... Contrast transfer function
DDD ..... Direct electron detector device
DQE ..... Detective quantum efficiency
EDC ..... 1-ethyl-3-(3-dimethylaminopropyl)carbodiimid
fAD ..... Familial AD
FBP ..... Filtered back projection
FSC ..... Fourier shell correlation
HSQC ..... Heteronuclear single-quantum correlation
IHRSR ..... Iterative helical real space reconstruction
LCO ..... Luminescent conjugated oligothiophenes
PBS ..... Phosphate buffered saline
PDB ..... Protein Data Bank
NHS ..... N-Hydroxysuccinimid
NMR ..... Nuclear magnetic resonance
RCT ..... Random conical tilt
PCA ..... Posterior cortical atrophy

sAD ..... Sporadic AD
SPR ..... Single particle reconstruction
SNR ..... Signal-to-noise ration
SSNR ..... Spectral SNR
ssNMR ..... Solid-state NMR
RMSD ..... Root mean square deviation
RP-HPLC ..... Reverse phase high-performance liquid chromatography
TEM ..... Transmission electron microscope
TFA ..... Trifluoroacetic acid

List of Figures

1 Introduction

1.1 Alzheimers Disease

Alzheimer's disease (AD) is a neurodegenerative disorder leading to cognitive impairment and progressive dementia. It results from a combination of pathological changes in the brain that lead to neuronal and synaptic loss. AD is the most common form of dementia, accounting for approximately 60 - 80 % of all cases. According to estimations 35.6 million people worldwide suffer from dementia and it is expected that numbers double every 20 years. Despite extensive research effort no effective cure has been found and current treatments are symptomatic only [52].
The neuropathological characteristics of AD are amyloid plaques from amyloid-β (Aβ) fibrils in the brain, neurons containing neurofibrillar tangles made from hyperphosphorylated tau protein, vascular damage from the plaque load and neuronal cell death [42][66]. Clinically, progressive dementia and other cognitive impairments are the main features of AD [90].
Alzheimer's disease looks uniform with regard to the morphological end stage in brain and clinical features but the disease does not have a single cause. There is evidence that a smaller proportion of about 5 - 10 % of all cases is caused by mutations resulting in early-onset AD. The greater proportion (95 - 90 %) of all Alzheimer cases is of late-onset and sporadic origin [46].

1.2 Amyloid-β

Production of amyloid-β

Extracellular plaque in the brain made from fibrils of the amyloid-β protein are a main characteristic of AD. Amyloid-β is derived from the amyloid precursor protein (APP),

a large transmembrane protein, by cleavage of two different secretases: γ-secretase and β-secretase. The first cleavage is carried out by β-secretase immediately at the N-terminal part of Aβ creating an amino fragment called secreted APPβ (sAPPβ) and a carboxyterminal fragment, called C99 wich is still embedded in the membrane. These remaining 99 amino acids are cleaved by γ-secretase at different positions creating Aβ, that is released into extracellular space and the amino-terminal APP intracellular domain (AICD). Depending on the exact points of the cleavage Aβ is made up of a different number of amino acids the main forms being 38, 40 and 42 amino acids long [103].
α-secretase is another processing enzyme for APP but its cleavage site is at a different position than the one from β-secretase. It lies within the Aβ sequence thus preventing Aβ formation. Not much is known about the products from α-secretase cleavage but their activity is regulated by enzymes from a disintegrin and metalloproteinase domain proteins (ADAM), see figure 1.1 [10].

From genetical and biochemical data as well as animal models Aβ is strongly suspected to play a central role in the pathogenesis of AD [88][103]. However Aβ peptides are a normal byproduct of the cellular metabolism and are detected as circulating peptides in the plasma and cerebrospinal fluid (CSF) of healthy humans [66]. Also, the fact that specific pathways exist to generate as well as uptake, breakdown and clear Aβ is a hint that it has a physiological function in the nervous system. Furthermore, there is an indication that Aβ is involved in synaptic activity and neuronal survival [68].

Aβ aggregation

Aβ molecules tend to self-aggregate and form oligomers, protofibrils and mature fibrils and it is these aggregates that are suggested to cause neuronal dysfunction and cell death [66]. Among the three main forms of the Aβ protein that are 38, 40 or 42 amino acids long, the longest, 42 amino acid protein, is most prone to aggregate. Consequently, the relative amount of $A\beta_{1-42}$ is of significance for AD progression [66].

Experimental kinetic studies, in which the aggregated mass is measured over time, can be used to characterize the aggregation process of amyloid fibrils [50]. Different ways for aggregation exists, for example fibrils can be formed from monomers through primary nucleation or from existing fibrils through fragmentation. Secondary nucleation is a combination of monomeric and fibrillar aggregation. Filamentous growth observed over time typically shows a sigmoidal curve: A lag phase is observed before a phase of rapid

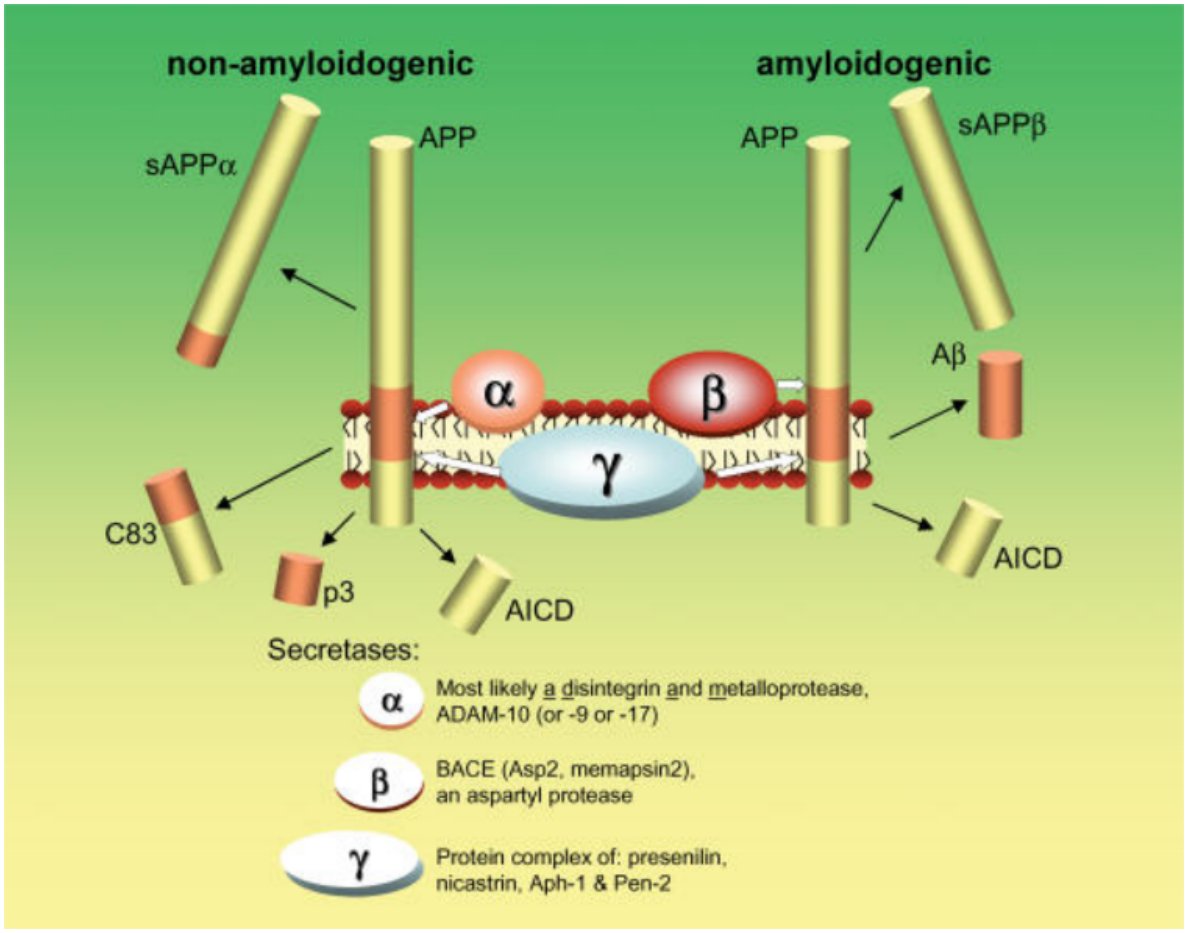

Figure 1.1: Schematic drawing of APP cleavage. It can occur in at least two different pathways. The non-amyloidogenic pathway includes cleavage by α- and γ-secretase. The cleavage site for α-secretase lies the Aβ sequence and its formation is thus prevented. The cleavage products are sAPPα p3 and AICD. In the amyloidogenic pathway cleavage takes place by β-sectretase which cleaves at Aβ N-Terminal site and γ-secretase, that cleaves at different positions at the C-terminal site of Aβ . The other products of this process are sAPPβ and AICD. Figure taken from Pearson 2006.

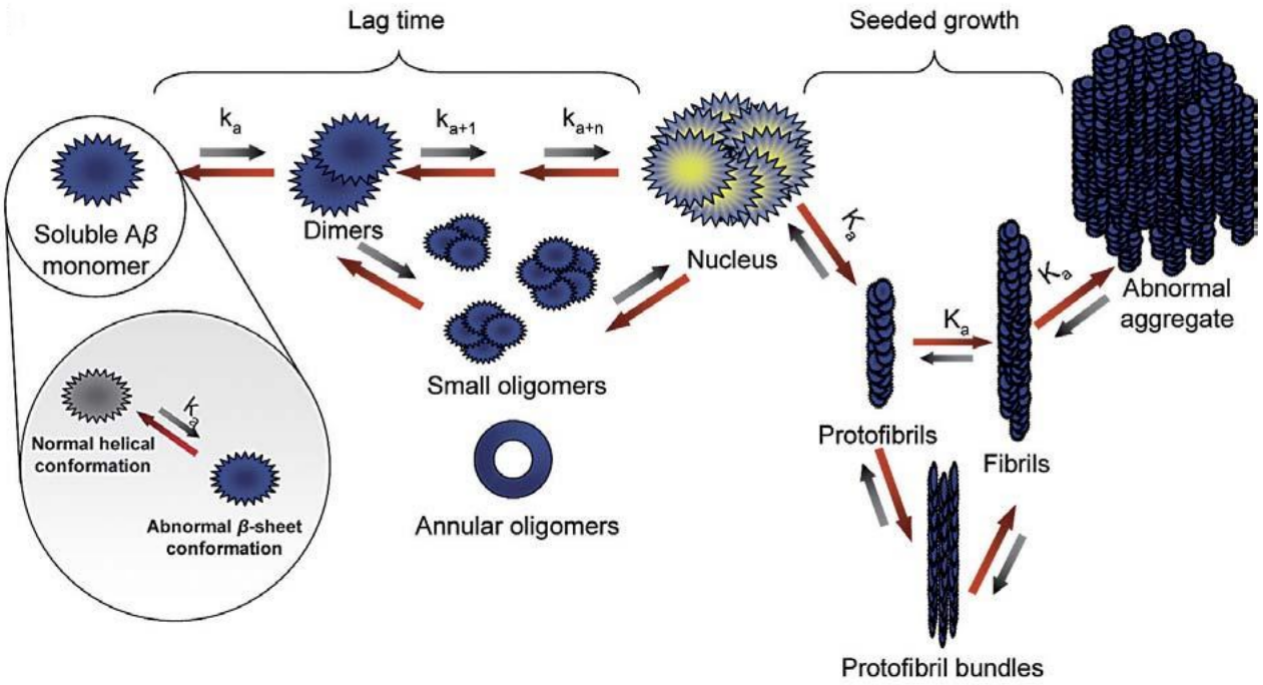

Figure 1.2: Schematic of the Aβ fibrillization process. Aβ monomers enter the aggregation pathway and form oligomers, protofibrils, fibrils and large plaque deposits. The reaction includes a lag-phase followed by rapid fibril growth. Small oligomers are formed on- as well as off-pathway and they are supposed to be the most toxic species. Figure taken from [31] with kind permission from S.Karger AG.

growth. When the total amount of Aβ molecules is limited the growth phase is followed by a plateau due to depletion of available soluble molecules. Seeding is the limiting factor in this process but the lag phase can be overcome by adding preformed aggregates to the solution. For the specific case of Aβ seeding was found to be specific and other amyloidogenic seeds are not able to initiate Aβ fibrillization [31]. This aggregation model also includes the formation of off-pathway oligomers [31].

The "Amyloid Cascade Hypothesis"

A well known hypothesis about the formation of AD was first formulated in 1992 by Hardy and Higgins [42] and is called the "Amyloid Cascade Hypothesis". It suggests that the amyloid fibrils made from Aβ are the cause for neurofibrillar tangles and cell death. Since its first formulation in 1992 the hypothesis has been changed and

updated because newer findings about Aβ fibrils and plaques were in contradiction to the original hypothesis. For example, it was found that there is actually a weak correlation between the severity of dementia and plaque load [20]. Another insight came from APP transgenic mice which showed memory impairment and changes in neuron form and function before the occurrence of amyloid deposition [21]. In addition to that various recent studies using different methods suggest that prefibrillar forms of Aβ are more injurious than the actual fibrils or plaque [66].
Examples are a study by Lue *et al* in which they showed that in human brain there is a correlation between the level of soluble Aβ and the extent of synaptic loss and cognitive impairment. For fibrillar Aβ such a correlation does not exist [21]. A study by Lambert *et al* showed that fibril free Aβ aggregates, called Aβ derived diffusible ligands, lead to synaptic dysfunction [51]. Experiments with transgenic mice suggest that different oligomeric species are responsible for decreased spine density, impairment on long-term potentiation (LTP) and decrease of spatial memory [103]. A variety of other studies using synthetic Aβ cell culture models, cerebrospinal fluid (CSF) and postmortem brain tissue support the assumption that oligomeric species have a higher toxic effect than amyloid fibrils [103].

Although it seems to be evident that oligomers are damaging the mechanism of toxicity is not clear. Experiments revealed that oligomers from different proteins can be toxic regardless of whether the protein forms amyloids or not. Thus, it is proposed that a generic structural motif common to all oligomers might be responsible for their toxicity [28] and it could be based on the interaction of oligomers with cell membranes. Oligomers have been reported to perturb membranes and electrophysiological measurements showed an increase in ion permeability [28].

Many of the new findings have been included in a modified version of the "Amyloid Cascade Hypothesis" that is now described following the formulation by Selkoe and Hardy [88], see also figure 1.3.

One distinguishes between dominantly inherited forms of AD that have mutations in APP or γ-secretase genes. This leads to a life-long increased $A\beta_{1-42}$ production. On the other hand there are non-dominant forms of AD, including sporadic AD, where the failure of Aβ clearance mechanisms marks the beginning of the disease leading to an increase of $A\beta_{1-42}$ concentration in the brain. Either way it comes to accumulation and oligomerization of $A\beta_{1-42}$ in the brain. These oligomers directly injure synapses or

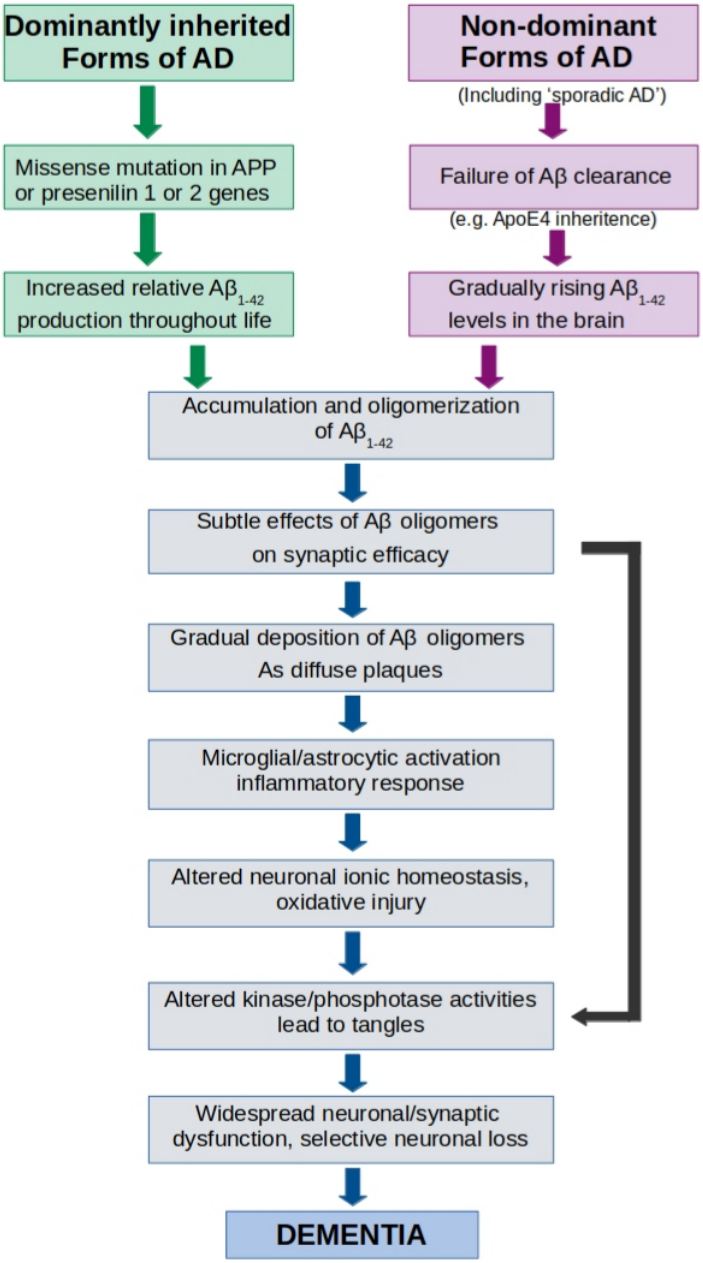

Figure 1.3: The modified "Amyloid cascade hypothesis". In both Alzheimer cases, familial and sporadic, accumulation of Aβ marks the beginning of the disease and initiates the other pathogenic mechanisms of the disease.

activate more complex mechanisms in the brain including microglial activation, inflammatory response and oxidative stress. This ultimately also leads to tangle formation, neuronal and synaptic dysfunction and therefore dementia.

1.3 Amyloid fibril and oligomer structures

Aβ and its different aggregates are expected to be highly involved in the pathogenic process of Alzheimer's disease as fibrils composed of the Aβ protein are found as plaques in the brain of AD patients. Thus, their structures are of great importance for understanding the disease especially in view of the aggregation process.
Aβ, however, is not the only protein forming fibrillar aggregates that are associated with a disease. In fact, more than 20 proteins have been identified to form amyloid fibrils that are involved in diseases. The most prominent example is AD that is characterized by fibrillar aggragates from both Aβ and Tau protein in the brain. Other examples are Parkinson's disease with α-synuclein as the aggregating protein and Amytrophic lateral sclerosis and superoxide dismutase 1 [50].
Surprisingly, there is no evidence for a sequence or native structure similarity between the disease associated proteins. Additionally, in the past years it could be shown that a wide range of different proteins and peptides can assemble into amyloidogenic aggregates and proteins in the amyloid state are highly stable [50][49]. This leads to the assumption that under adequate condition any polypeptide chain can convert into an amyloid structure but for most proteins protective mechanisms exist that hinder proteins to form amyloids in the living cell [50][49].
Key features of amyloids are their close packing and high degree of order. They can differ considerably from the native state of the protein and have a generic structure rich in β-sheet. In contrast to that, the structures of the native proteins are highly diverse ranging from mainly α-helical to predominantly β-sheet. The overall structure of amyloid fibrils from different proteins is strikingly similar: they appear as unbranched filaments with a few nanometer in diameter but often micrometer long. A fibril usually consists of multiple protofilamets that are twisted around each other. X-ray fibre diffraction shows that the core of each protofilament exists in a cross-β structure in which the β-strands form continuous hydrogen-bonded sheets that run along the fibril axis [50] while the residues extend perpendicular to it [29]. β-sheets often appear in pairs and the side chains of each sheet interdigitate with the side chains of the partner sheets. Because of its appearance from the top this motif is called steric zipper [25].

In addition to amyloids beeing protein assemblies with a quaterny cross-β structure their histological definition is fibrillar deposits in cell or tissue that show red/green birefringence under cross-polarized light after staining with Congo red [49]. Despite the cross-β spine amyloid fibrils the of same protein can exist in structurally different morphologies which makes their structure determination difficult [29][60]. This heterogeneity reveals itself in varying fibril widths and cross-over lengths and originates from different types of polymorphism which are variations in protofilament number, variations in protofilament arrangement and different polypeptide conformations.
Protofibrils are closest to mature fibrils regarding their structure. They do not show a similar high order and periodic symmetry as fibrils. Also, they are thinner, shorter and more curvilinear, but they contain a high amount of β-sheets [28]. Their difference to oligomers is their elongated linear shape.
Oligomers are the earliest intermediates in amyloidogenic aggregation and they can be composed of very few up to hundreds of monomers [28][25]. Characterizing their size and structure is difficult because of their heterogeneity and tendency to rapidly aggregate to stable fibrils. At a molecular level there are very few suggestions for structures and most often they are described as spherical particles with a typical diameter of 4 to 25 nm [44][59][28].

Symmetry in amyloid fibrils

As outlined above amyloid fibrils are highly symmetric objects that all share a common cross-β spine but β-strands in the structure can still adopt different arrangements. This section briefly describes the different β-sheet geometries and introduces their terminology.
This first feature of the β-sheet geometry is the β-strands beeing parallel or antiparallel with respect to their N- and C-terminal ends. Parallel sheets are defined as strands beeing aligned with the same polarity so that N-termini are on one side of the sheet and C-termini on the other, see figure 1.4. Consequently, antiparallel sheets have an alternating arrangement of N- and C-termini on both sides.
In addition to that differences in the structure may arise from a different registration of the β-strands in a sheet. Generally, an in-register can be defined as an arrangement where a line drawn between a pair of identically oriented chains is perpendicular to the β-strand as shown in figure 1.4. In parallel β-sheets this means that the backbone amide and carboxyl group of a residue form hydrogen bonds with the same residue

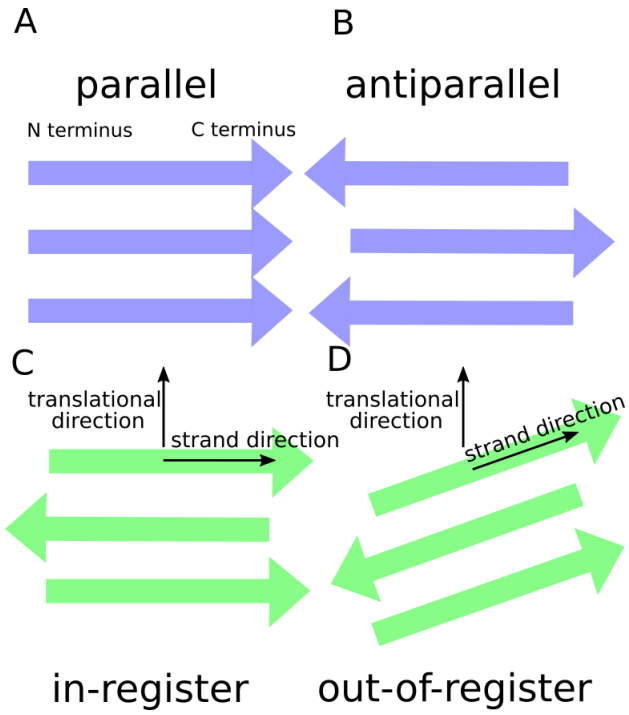

Figure 1.4: Schematic depiction of parallel and antiparallel β-sheets (**A** and **B**) and in-register and out-of-register β-sheets (**C** and **D**).

in the neighboring strands. This way identical residues are stacked along the fibril axis. There seems to be a bias of strands adopting this kind of arrangement as most of the available high resolution structures show parallel in-register alignment. This ladder-like stacking of residues may be energetically favorable for uncharged residues. Additionally, these structures have sharp turns leading to distinctive kinks often with glycines at theses positions in the chain. For out-of-register structures such kinks are unlikely to form because the alignment of the residues in the kink would be disrupted [25].

Structures of amyloid β_{1-40} fibrils

Three different fibril structures of Aβ_{1-40} from ssNMR have been described and all of them show parallel and in-register β-sheets. Still, they are notably different from each other. The first two models were derived from fibrils produced under the same conditions regarding peptide concentration, pH, ionic strength, buffer composition and temperature and only differ in growing conditions, that were agitating for one and quiescent for the other fibril [74][67]. In both fibrils the molecules adopt a "U"-shape of parallel β-sheets with one arm comprised of the residues 30-40 and the other of

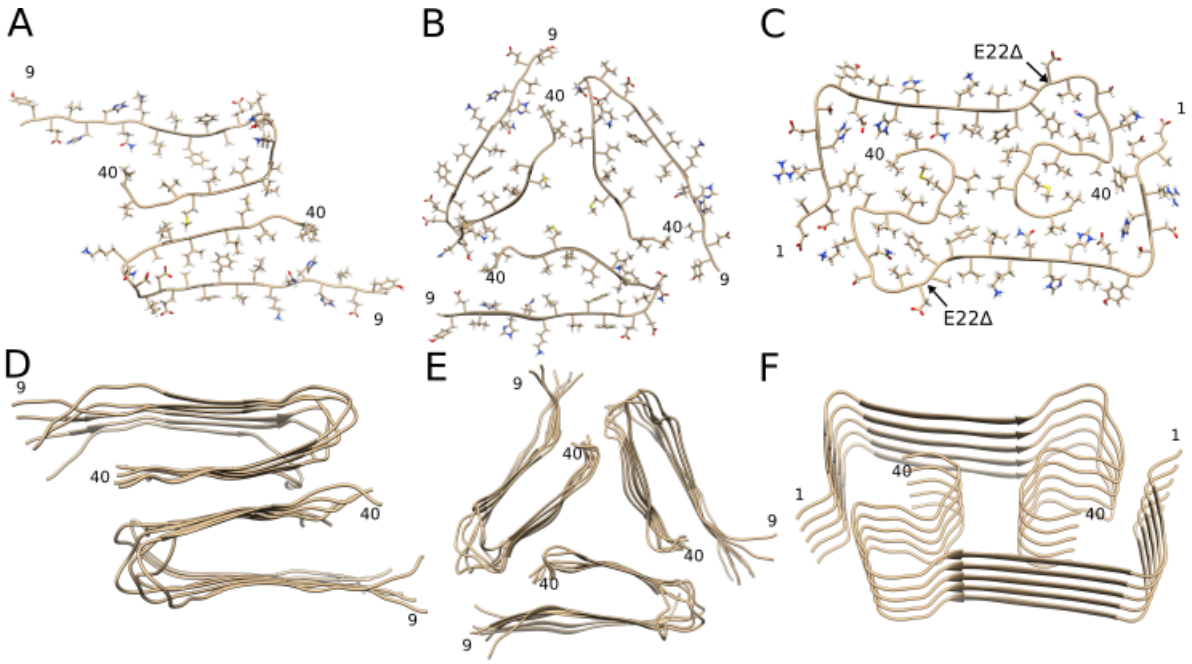

Figure 1.5: Three atomic models determined by ssNMR. A-C show the cross section of the model and D-F the the arrangement of monomers in the fibril. A (*PDB code 2LMN*)and B (*PDB code 2LMQ*) are related structures of wild type $A\beta_{1-40}$ and are a dimer, respectively trimer of the same "U"-shaped subunit. C (*PDB code 2MVX*) is a structure of the Osaka mutant that is again a dimer but has a highly kinked monomer conformation.

residues 10-22, see figure 1.5. The first growing condition leads to a fibril composed of dimers with a 2-fold symmetry. The second condition leads to a trimer where the molecules are arranged as a triangle leading to a 3-fold symmetry.

The third ssNMR structure presented is a fibril from the Osaka mutant ($\Delta 22$) [87]. It is again a dimer but the molecule shape is completely different as it does not have a "U"-shape but several sharp kinks, see figure 1.5. The deletion of residue 22 is situated in a turn region and the authors state that the overall structure can be adopted by wild-type $A\beta_{1-40}$ and other mutants [87]. Also, unlike the other two models the N-terminal part is part of the structure.
$A\beta_{1-40}$ fibrils have also been studied by cryo-EM [82]. In the EM images taken by Sachse *et al* the fibrils show a fixed width of about 190 Å but varying twists. Thus, for 3D reconstruction a more homogeneous subset of fibrils with cross-over distances between 1300 and 1500 Å is selected. The averaged power spectrum of the fibrils shows a clear peak at 4.8 Å supporting the cross-β structure of the fibrils. The cross-section of the reconstruction reveals a 2-fold symmetry. Two "U"-shaped traces with head-to-

head orientation fit into the cross-section. This indicates that the structure consists of two protofilaments and their polypeptide chains are situated opposite to each other bounded by the N-termini, see figure 1.7 B.

Structures of amyloid β_{1-42} fibrils

Much effort has been put into structure determination of Aβ_{1-42} fibrils as they are the main component of the plaques observed in the brain of AD patients.
Recently, two high resolution structures, determined by ssNMR, have been reported [104][11]. The atomic models were derived independently but describe basically the same structure: A dimer is formed from two S-shaped monomers comprising the residues 15-42. Three [11], respectively five [104] main parallel and in-register β-strands are found within each monomer. The most hydrophobic residues are hidden in the fibril core while hydrophilic residues are exposed to the solvent and form the main turns. The first 14 residues are more dynamic and therefore not part of the fibril core. This polymorph shows reactivity with antibodies which points to the fact that it might be relevant in AD [104].

Other polymorphs have been described cryo-EM. Schmidt *et al* [85] used fibrils made from synthetic Aβ peptides that were incubated at pH 7.4 in buffered solution for cryo-EM imaging. X-ray diffraction, infrared spectroscopy, Congo-red green birefringence and thioflavin T binding experiments performed ahead show the amyloid-like properties of the fibrils and their cross-β architecture. 29 fibrils were used for 3D reconstruction to a final resolution of 7 Å. The reconstruction does not resolve the typical 4.7 Å spacing along the fibril axis and cannot determine the specific symmetry of the fibril. However, it shows that the structure consists of one 2-fold symmetrical protofilament with a cross-over distance of 1100 Å. The cross-section is divided into two domains: A central domain (C) and two peripheral domains (P) at the outer region of the cross-section, see figure 1.7 C. The resolution in the C domain is slightly better than in the P domain and consists of two "S"-shaped densities. The C-terminal parts form a β-sheet structure in this domain and its zig-zag pattern is reminiscent of a steric zipper motif. Molecular modeling was done to interpret the density in this domain. Residues 31 to 36 fit into the zipper like fragment and the residues Glu22 and Asp23 hinder the chain to adopt a more favorable packing structure at the interface. Interestingly, mutations at these residues are associated with familial AD and a removal of their charges can accelerate the fibrillation process [85].

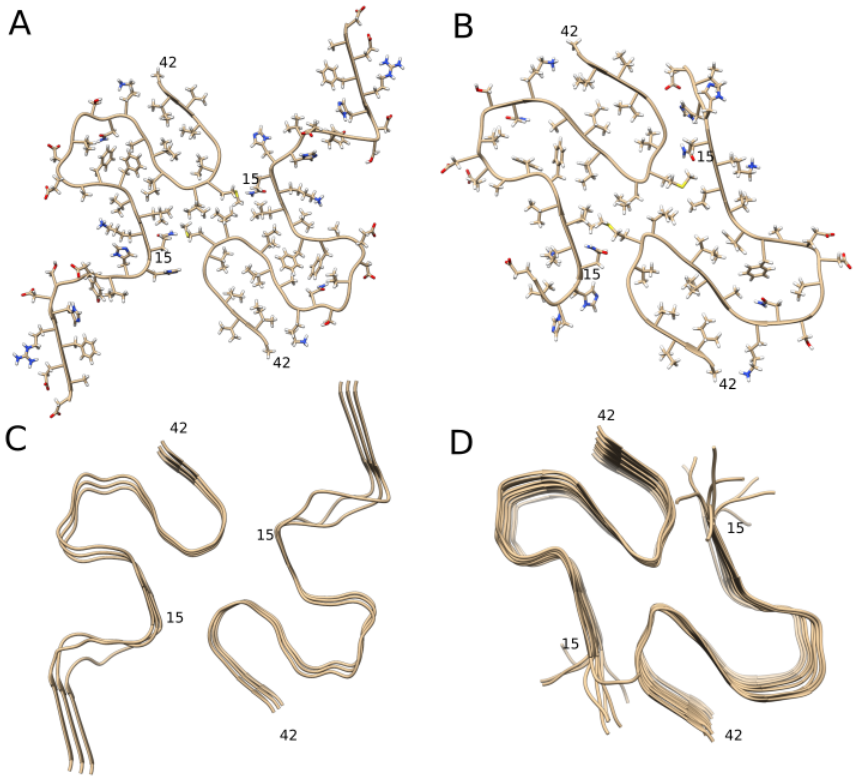

Figure 1.6: The two Aβ_{1-42} ssNMR models have a similar S-shaped dimeric structure but were derived independently. Model A (*PDB code 2NAO*) was determined by Wälti *et al* [104] and model B (*PDB code 5KK3*) by Colvin *et al* [11].

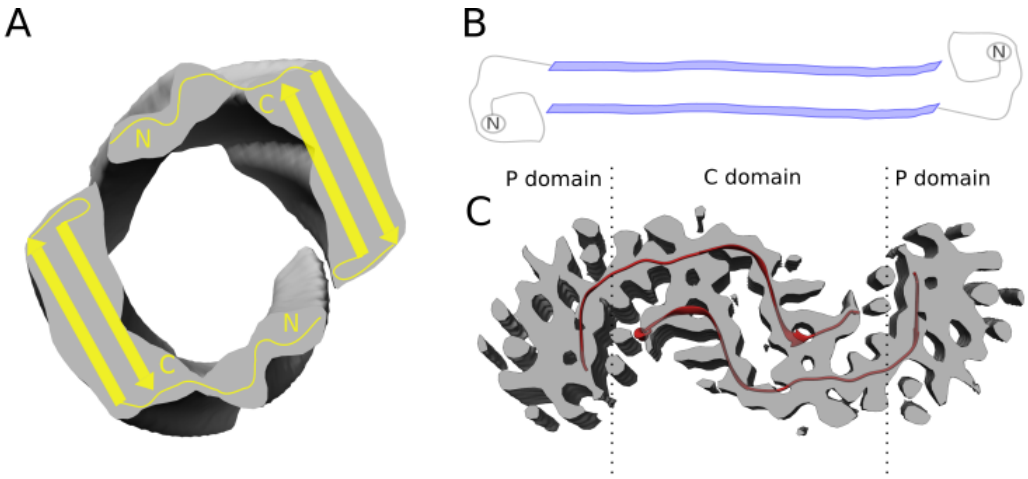

Figure 1.7: (**A**) is a 10 Å density map of an Aβ_{1-42} fibril with a previously derived NMR model with a hairpin shape and an unordered N-terminus [56]. (**B**) is a proposed model for one protofilament of an Aβ_{1-40} derived from a cryo-EM density map. The two opposite chains form a β-sheet core. (**C**) shows the density and related model from an Aβ_{1-42} fibril derived by cryo-EM. The model in the C domain shows a β-sheet structure with a steric zipper motif.

Another cryo-EM structure of Aβ_{1-42} was reported by Zhang *et al* [105]. Synthetic Aβ proteins were incubated at 37 °C and acidic pH to form fibrils over a time period of several weeks. Their aggregates show the typical characteristics of amyloid fibrils, that is thioflavin T fluorescense and a typical β-sheet CD spectrum. AFM images show a population of long and straight fibrils with helical twist and constant width. However, variations in the helical pitch can be observed in the images even within a single filament.
A 3D reconstruction was calculated using images from a homogeneous subclass with a pitch of 460 Å. The resolution of the map is around 10 Å and reveals two smooth protofilaments twisting around a hollow core. The cross section shows that each protofilament consists of a large high density part neighbored by a smaller weak density region, see figure 1.7 A. The authors suggest that two β-sheets in a hairpin like conformation fit into the large density part consistent with a previously published NMR model [56]. The flexible N-terminal part fits into the smaller, weak density.
All discussed models have different structures suggesting that they are different polymorphs.

Structural variability was also reported for amyloid fibrils from the brain of AD pa-

tients. Luminescent conjugated oligothiophenes (LCO) dyes can be used to describe the structural differences between different fibrils extracted from brains [79]. The dyes bind to the cross-β structure of the amyloid and reflect differences in the twist by different spectral properties. Amyloids from different familial AD (fAD) cases, sporadic AD (sAD) and sporadic posterior cortical atrophy AD-(PCA-AD) from different cortical parts (frontal, temporal and occipital neocortex) were investigated.
The spectral properties of amyloids from fAD with specific underlying mutations were the most different from the other fibrils. A striking variability could be observed in the group of sporadic AD. The authors suggest that the great phenotypic variability in AD has probably numerous reasons. The age of disease onset, the location of the first abnormal aggregates as well as the characteristics of spread, the inflammatory response and potential presence of comorbic conditions are expected to be contributing factors. In a previous study Tycko *et al* could show that brain extracts from different patients can induce synthetic Aβ to aggregate into corresponding structural "strains" which is a hint to the fact that a specific Aβ structure dominates in AD brains [77]. These observations can be transferred to the more recent study from Rasmussen *et al* as they observe that the mean LCO spectral emission is similar in the different cortical regions of each patient [79].

Structures of other amyloid fibrils

Tau protein is a microtubule associated protein and filamentous Tau in the brain is a characteristic of many neurodegenerative diseases with Alzheimer's disease being the most prominent one. In AD two different types of filaments can be found in the brain and these neurofibrillary lesions are described as paired helical filament (PHF) and a straight filament (SF). Both filaments consist of two protofilaments in a double C shape and are surrounded by a fuzzy coat [14]. Recently, structures of both filaments derived form AD patients brain have been reported [32]. From cryo-EM images high resolution density maps could be reconstructed that allowed for atomic model building. The structures confirm that the two filaments are made out of C shaped subunits and it could be shown that the structural core is made out of residues 306 to 378 of Tau protein. The structural core is composed of eight β-sheets where three sheets are arranged in a triangular fashion and the others in two regions opposite to each other in a regular cross-β pattern. The hydrogen-bonding pattern is regular and in-register in the hole protofilament. The protofilament structures of PHF and SF are similar which

makes them ultrastructural polymorphs but they differ in the interface between the protofilaments.

α-synuclein is the main component of Lewy bodies that are characteristic for Parkinson's disease. Only few information are available on the fibril structure of α-synuclein. But it is known that the middle domain of amino acids 61 to 95 of the protein, called NAC, is involved in aggregation of α-synuclein [38]. A structure for the residues 44 to 97 could be determined from ssNMR [97]. It shows that only one molecule is present in each layer that has several sharp kinks. The structure is held together by hydrogen bonds in parallel and in-register β-sheets.

Structures of amyloid oligomers

Oligomers are usually described as small spherical particles. Many of them are described to have a high β-sheet content [28] but random coil like conformation have also been reported [83]. Studying oligomers is difficult because of their transient nature and thus, few structural information about them exists.
A proposed model for an amyloid oligomeric structure is the structure from a fragment of αB-crystallin, an amyloid forming protein. The fragment is eleven amino acids long and was crystallized so its structure could be determined by X-ray crystallography. It consists of six β-sheets, that are antiparallel and out-of-register in contrast to most reported amyloid fibrils structures. The sheets are arranged in a cylindrical shape, see figure 1.8 A. A tandem-repeat version of this fragment with 25 residues has the same structure and binds to A11, an antibody that specifically binds to oligomers, but not fibrils or monomers. Out-of-register β-sheet interactions are weaker than the in-register sheet interactions of most amyloid fibrils and this could be an explanation for the transient behavior of oligomers [25]. Based on the observation that A11 binds to a large variety of Aβ oligomers it was proposed that they have a common structural motif [44], like amyloid fibrils. But until now such a general structure element could not be identified [28].
In contrast to this structure the structure of an oligomer from $A\beta_{1-42}$ appears very different. It consists of 15-24 molecules and infrared analysis indicates that it has mostly β-sheets. The X-ray diffraction pattern suggests that the oligomers is formed by two steric zippers that together form a double helix with a central pore that varies in size [91].
In another study crystal structures from cyclic Aβ fragments were investigated [75].

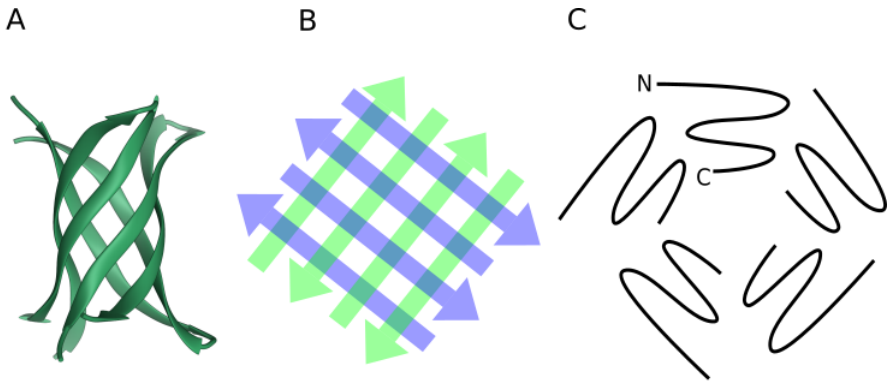

Figure 1.8: Different models for oligomers are presented. A shows a cylindrin structure like it is formed by a fragment of αB-crystallin. B is a representation of the cruciform motif adopted by a cyclic Aβ fragment. C depicts the arrangement of Aβ monomers in disc-shaped pentamers.

The residues 15-23 from Aβ were part of a macrocycle that also contained a molecular template to prevent the formation of fibrils. The authors find that these cyclic peptides form dimers with hydrogen bonds that further arrange in different cruciform tetramers, see figure 1.8 B and some of them assemble to triangular dodecamers. Based on molecular modeling experiments the authors state that these assemblies can be adopted by natural Aβ peptides.
Ahmed *et al* studied pentamers of $A\beta_{1-42}$ with AFM and TEM and showed that under low salt and low temperature conditions these pentamers form discs [3]. These disc-shaped oligomers have a width between 10 and 15 nm and a height of approximately 2 nm. The authors suggest that the monomers in the disc have a diameter of about 5 nm and height of 2 nm and that their C-termini point to the center of the disc, see figure 1.8 C.
EM images from oligomers of the arctic mutant β_{1-42} (E22G) and the A30P and A53T mutations of α-synuclein show that these species form ring-shaped aggregates. The ring comprised of the arctic mutant is around 7-10 nm large in diameter and approximately 2 nm wide. One ring comprises 40 to 60 $A\beta_{arc}$ molecules [53]. The α-synuclein rings have a size of 8 to 12 nm in diameter and approximately 2 nm as the inner diameter. They consist of 20 to 25 monomers. These ring shaped oligomers might explain the enhanced permeabilization of the α-synuclein mutants on acidic phospholipid vesicles [102]. For Aβ channel-like behavior has been reported when it was reconstituted in lipid layers [55] which could be explained by the ring shape of the arctic mutant.

Amyloid fibrils play a key role in neurodegenerative diseases making their structure

of great interest. Much effort has been put into studying Aβ fibrils because of their relevance in AD.
Determining high resolution structures of fibrils is difficult of because of their heterogeneity. Due to poor understanding of the cause of AD and the development of the disease few efficient therapeutic strategies exist. More structural knowledge about Aβ fibrils and oligomers could help understanding the disease and be a guide for efficient drug design.
The aim of this thesis is to build a structural model of an Aβ fibril from cryo-EM experiments. EM images of Aβ fibrils can be used to reconstruct a high resolution 3D density map of the fibril and an atomic model can be build based on that.

Different polymorphs are present in the images and helical reconstruction algorithms are used as well to learn about their structure. In another EM experiment oligomers are imaged to gain information about their size and structure by doing a 3D reconstruction. The determined Aβ structures may help in understanding the progression of AD and be a starting point for efficient drug design. Additionally, the structures can be used to unravel the binding site of known interaction partners, such as Congo-red. In general, the determination of ligand binding sites is an important step in drug development. Therefore, in Chapter 5 an efficient and automated program to screen ligand binding positions is presented. The method is applicable to a large variety of different protein-ligand systems and could also be used to study fibril interactions.

2 Methods

2.1 Cryo-electron microscopy

For many years X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been the dominating techniques for protein structure determination. Recently, also cryo-electron microscopy (cryo-EM) establishes to become a mainstream technology for studying cell architecture, viruses and protein assemblies [61]. Cryo-EM includes a wide range of experimental methods, that all share the basic principle of imaging radiation-sensitive samples in transmission electron microscopes (TEM) at cryogenic conditions [61]. In biology it finds application to study tissue sections, plunge-frozen cells, bacteria, viruses and protein molecules [61]. Sub-disciplines of cryo-EM are electron tomography, single particle cryo-EM and electron crystallography which have differing suitability for studying just mentioned systems [61]. The develoment of the techniques began in the 1970's but it attracted attention in the last decade when obtained structures reached near-atomic resolution [9]. Great progress has been made mainly by the development of direct electron detector device (DDD) cameras [30] and increasing computer power enables the use of more sophisticated image processing methods making better and more reliable density maps possible [9].
This chapter briefly explains the fundamental principles of cryo-EM with a focus on single-particle cryo-EM and helical assemblies.

In electron microscopy the radiation used for imaging are electrons emitted from a source held in high vacuum and accelerated along the microscope column at voltages typically between 80 and 300 kV. The incident electrons are scattered by the sample and focused by electromagnetic lenses afterwards, see fig 2.1 A. For thin samples the obtained images are linear 2D projections of the samples' Coulomb potential [9].

Biological samples are easily damaged by electron radiation as it is breaking chemical bonds and consequently the electron dose used still needs to be low, typically < 20

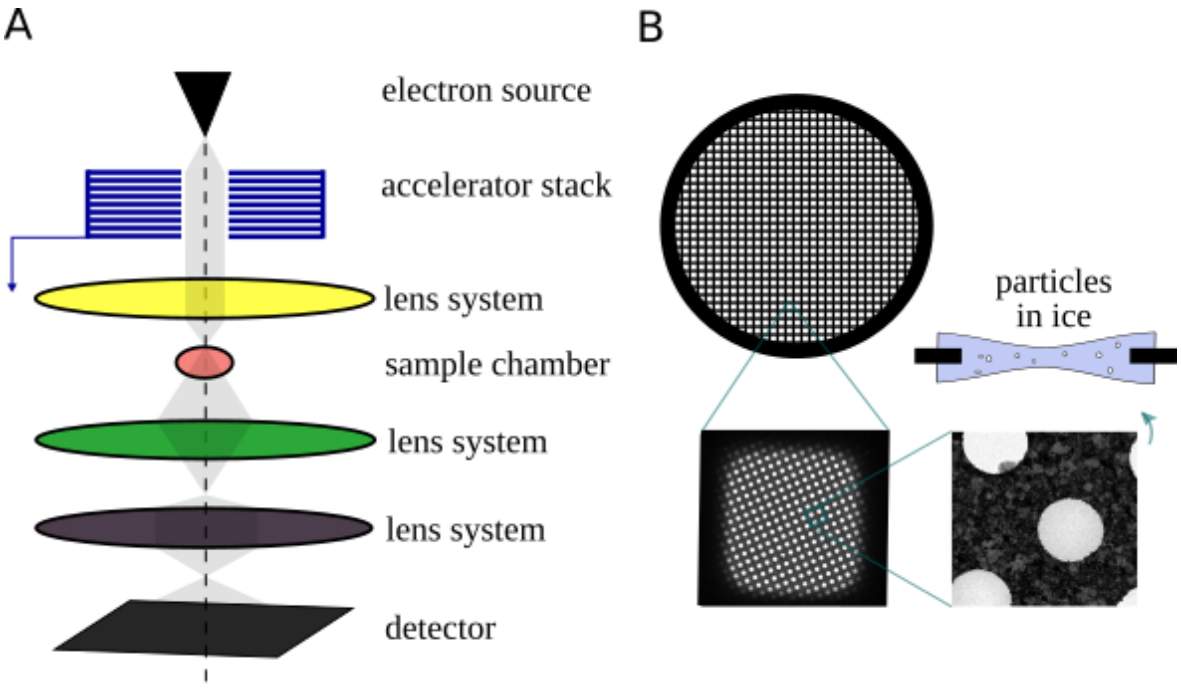

Figure 2.1: (**A**) Schematic drawing of a transmission electron microscope. (**B**) Supporting grid used for sample preparation.

$e\,\text{Å}^{-2}$ [7]. A way to circumvent this problem is to image the sample with negative stain: The molecule surface is coated with a reagent containing heavy atoms, usually uranyl acetate [61]. Negative stain images produce high contrast but at the expense of loosing internal structure information [33]. For cryo-EM images the biological specimen in solution is applied to a grid, usually made out of fine copper mesh or other metals such as gold or nickel. The excess sample is removed by blotting with filter paper. Following, the grid with the sample on is plunge-frozen in liquid ethane. The rapid freezing prevents the formation of crystalline ice. The sample can than be imaged at liquid nitrogen temperature in a layer of amorphous ice [61][15]. This method has the advantages that the biomolecule is kept in a hydrated environment and the greater resistance of frozen organic material to radiation [33]. Desirably, the sample has an ice layer thick enough to accommodate all particles but not much thicker so that they are clearly visible. Also, the particles should be evenly distributed on the grid and show a broad range of different orientations.

As mentioned above, the radiation damage on biological samples is severe and the electron dose needs to be low which results in images with low contrast [7]. For this reason, images are taken in underfocus [9]. In the image formation process the

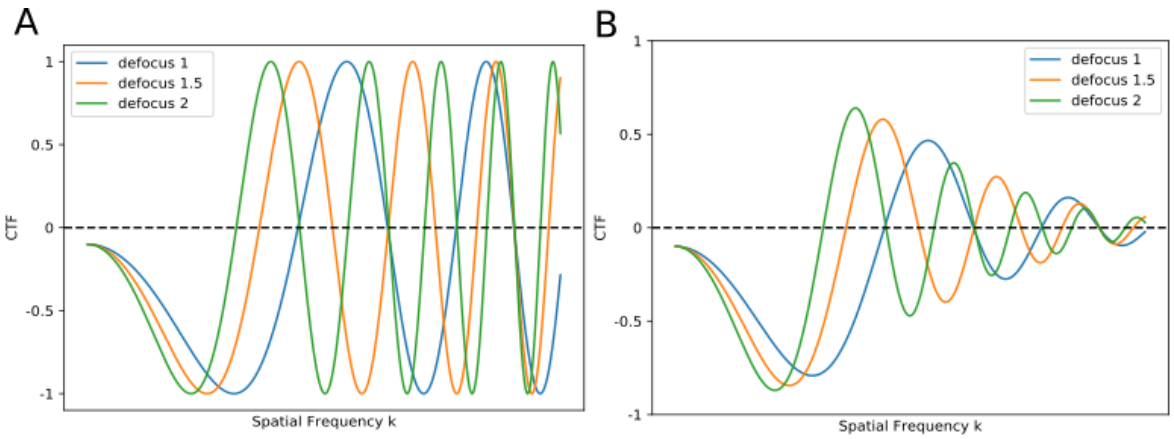

Figure 2.2: (**A**) shows the CTF function at different defocus values (arbitrary units) and the effects of envelope functions on it (**B**).

scattered electrons are effected by lens abberations and the applied defocus. This effect is transfered to the final image and described by the contrast transfer function (CTF). The CTF is a quasi periodic sine function with multiple zero crossings which exact locations depend on the defocus value, see figure 2.2 A. The small but finite spatial and energy spread of the electron source in the microscope leads to a dampening of the CTF [9]. In a good approximation both effects can be described by envelope functions that can be combined to one compound function $E(k) = E_i(k)E_e(k)$. The function $E_i(k)$ describing the partially coherent illumination depends on the defocus value but $E_e(k)$ that describes the energy spread does not [33]. The effect of the combined envelope function on the CTF is presented in figure 2.2 B.

Image restoration methods aim to compensate for the contrast alterations, introduced by the oscillating nature of the CTF and the damping of high resolution components. At the zero crossings of the CTF no information is transferred and to sample the whole Fourier space images are taken at different defocus conditions [9]. To summarize, a high defocus gives high contrast at low resolution but weakens the contrast at higher resolution. Practically, this means that one should use the smallest possible defocus values where the particles still can be seen so that high resolution information is still present in the images [9].
With the new DDD cameras data is usually collected in movie mode: The electron dose is fractionated into a series of images that are aligned and averaged after data collection. Thus, it can be compensated for specimen drift and beam-induced motion [9]. Theoretically, the achievable resolution is dictated by the Nyquist theorem and

is limited to twice the pixel size. In practice, however, one finds that the resolution is lower which is due to interpolation errors and the low detective quantum efficiency (DQE) near the Nyquist frequency and a good approximation to the maximal possible resolution is three times the pixel size [9].
To properly correct for the CTF effects a precise estimation of the parameters that it depends on is needed. Apart from the defocus the CTF depends on a range of other parameters that are acceleration voltage, spherical aberration, astigmatism and amplitude contrast. Voltage and aberration are instrumental parameters and the amplitude contrast is assumed to be constant and for cryo-EM images between 5 and 10 %. The defocus is set during imaging but can only be controlled imprecisely and more accurate values for defocus and astigmatism are obtained by fitting a CTF function to the circular intensity oscillations that can be seen in the power spectrum, called Thon rings [9].
In single particle reconstruction (SPR) the structure determination is based on averaging thousands of identical particles. Thus, the next step is to extract particles from the recorded micrographs. Particle picking can be done either manually or automatically and once particles were chosen they are windowed and combined in a stack.
The first step in the structure determination process is usually analysing the data set by generating subsets that contain a more homogeneous set of particles that can be aligned and averaged to eliminate noise. Alignment of two images is usually done using cross-correlation functions that give the optimal rotational and translational parameters to bring images into register. For the classification of particles in different groups two main strategies exist that differ in the aspect of using references for classification or reference free techniques. In a data set that contains different views of the same and/or different molecules itself the multireference technique aligns a set of N images with L references. N x L correlation coefficients are calculated and each image is put into one of the L groups according to the highest cross correlation. After all images are sorted into a group, averages are calculated. The averages have improved signal-to-noise ratio (SNR) and substitute the initial references. The next iteration improves the values found for rotational and translational parameters and images can move from one group to another. This procedure is repeated until the parameters converged and the migration of images stops.
Reference based techniques have the disadvantage that the initial reference can bias the outcome of the classification. To avoid this reference-free methods were developed and their classification procedure is based on multivariate statistical analysis. A well-

suited method is Correspondance Analysis (CA). Like Principal Component Analysis it is based on the idea that many statistical variables are approximated by less mutually orthogonal linear combinations without losing much information. The reason to use CA for analyzing EM images is that multiplicative factors between images are not taken into account while calculating distances. This way, it is possible to analyze images taken at different exposures. In practice, CA on EM images can be carried out using a matrix notation as outlined in van Heel *et al* [99]. The images are aligned and all pixels of all images are arranged in a matrix. Distances are calculated between the rows or columns of this matrix and these values are put in a new matrix. The eigenvalues and eigenvectors, or factors as they are often called, of the new matrix are calculated. Usually the first two factors are sufficient to describe meaningful variation in the images. Projection of the data into a map of the first two factors clusters the images into groups [34][99].

Structure determination in cryo-EM is based on the principle of 3D reconstruction of an object from its 2D projection images, given that the relative orientations of the projections are known [69]. The orientations are described by five parameters: three Euler angles and two translational parameters. As the position and orientation of the particles in the sample is, more or less, random their orientation parameters cannot be known. There exist different strategies to determine these parameters. If there is some prior knowledge about the structure, for example from a low-resolution negative stain EM 3D reconstruction, an X-ray model or an EM map from a homologue they can be used to deduce an approximate guess of the particle orientations [9]. For systems without prior knowledge one of the most frequently used methods is the random conical tilt (RCT) approach [78]. In this approach one collects pairs of images. One image is recorded at a tilt angle, usually around 60 °and the same area is imaged untilted. The known rotation between the image pairs gives additional information that can be used to determine the particle orientations.

A computational approach to determine the particle orientation is based on the central section theorem which says that the Fourier transforms of 2D projections of a 3D object are central slices through the Fourier transform of the 3D object. From this follows that angular assignment can be obtained in real space by finding the common lines if there are at least three central sections that are 2D projections of the 3D structure [98].

When orientation parameters are set, an initial 3D reconstruction can be carried out. Various reconstruction methods exist and they can be categorized in two main groups:

Algebraic and transform methods. Algebraic methods start from discretizing the equation describing the relationship between a 3D object and its projection

$$g(\boldsymbol{x}_{\boldsymbol{\tau}}) = \int d(\mathbf{r}) d\boldsymbol{\tau}, \boldsymbol{\tau} \perp \boldsymbol{x}, \tag{2.1}$$

with g($\boldsymbol{x}$) being the projections of the object d($\mathbf{r}$) and $\boldsymbol{\tau}$ a continuous distribution of projection directions [69]. The resulting set of linear equations is typically solved iteratively. While this method is relatively easy to implement it has the disadvantage that it is computationally expensive [69].
Transform methods are based on the central section theorem and can be divided in further subgroups: In the first group only a filter is created and applied in Fourier space that accounts for the particular distribution of projections. Reconstruction is carried out in real space often using a back projection algorithm [69]. Back projection was the first method to be used to reconstruct a 3D object from its projections [69]. 2D images are translated into the direction normal to the image plane and the summation of all translations gives the original object [35]. Together with a preceding filtering step this method is known as filtered back projection (FBP).
The second subgroup is performed only in Fourier space and interpolation between the coordinate system of the projection data and the coordinate system of the reconstruction needs to be carried out. These methods are called direct Fourier inversion methods and the most accurate among them is the "gridding" technique. "Gridding" converts data from an irregular to a regular grid and this is done by convolution with appropriate "gridding weights" [72].
Using a reconstruction method and the approximate orientation parameters an initial map of the structure can be calculated. Afterwards the initial map can be refined using a projection matching method: each individual particle image is compared to reprojections of the 3D map in real or Fourier space to optimize orientation parameters [61].
For interpreting the final 3D map assessment of the resolution is of great importance. When near-atomic resolution is achieved secondary structural elements (helices, β-sheets and side chains) can be used to evaluate correctness and the approximate resolution of the structure. In most cases, however, an extra tool to determine the quality of the map is required.
The theoretical resolution of a microscope defined by the Abbe criterion would be $\sim$ 0.12 nm for 300 kV electrons. But reconstructions from biological specimen typically reach much lower resolutions. The reasons for that include instrumental limitations

such as the wavelength of the electrons and the quality of the electron optics that influence astigmatism and the envelope functions. Specific to biological samples is that they produce low contrast images because the density of the molecules and the surrounding ice are very similar. Radiation damage is known to be a problem but even though the electron dose is kept low the sample is damaged to some extent. As a direct consequence of the low dose condition the signal-to-noise ratio (SNR) of the data is low. At the stage of image processing this introduces more errors because the low SNR limits accurate alignment of the images [71].
A common definition of resolution is based on evaluating the point spread function of the microscope by imaging an object with known spacing. In cryo-EM the main resolution limiting factors are instability of the specimen and low contrast images which makes this method not applicable for cryo-EM.
Instead, resolution measures in cryo-EM test the self-consistency of the results. The approach most commonly used is the Fourier Shell Correlation (FSC)

$$FSC(u,v;s) = \frac{\sum\limits_{||s_k|-s|\leq\epsilon}^{k_s} U(s_k)V^*(s_k)}{\Big\{\big(\sum\limits_{||s_k|-s|\leq\epsilon}^{k_s} |U(s_k)|^2\big)\big(\sum\limits_{||s_k|-s|\leq\epsilon}^{k_s} |V(s_k)|^2\big)\Big\}^{1/2}}.$$

U and V are the Fourier components of two 3D reconstructions, s is the spatial frequency and k_s is the number of voxels in the shell with thickness 2ϵ. The FSC is a one dimensional function of the magnitude of the spatial frequency. Its values are correlation coefficients between the Fourier components of two volumes within a shell of approximate equal frequencies. Values of one correspond to a perfect similarity between u and v whereas values of zero mean that u and v are uncorrelated and have no similarity. Typically, the FSC decreases with increasing spatial frequency. If the resolution is reported as a single number this is the maximal spatial frequency up to which information in the 3D map is considered to be reliable. Different cut-off levels have been reported [100][81].

2.1.1 Structure determination of helical assemblies

There is a large variety of biomolecules forming helical assemblies including viruses, cytoskeletal filaments and amyloid fibrils which makes helical reconstruction a relevant part of cryo-EM [19][36]. Structure determination of helical assemblies exists since the

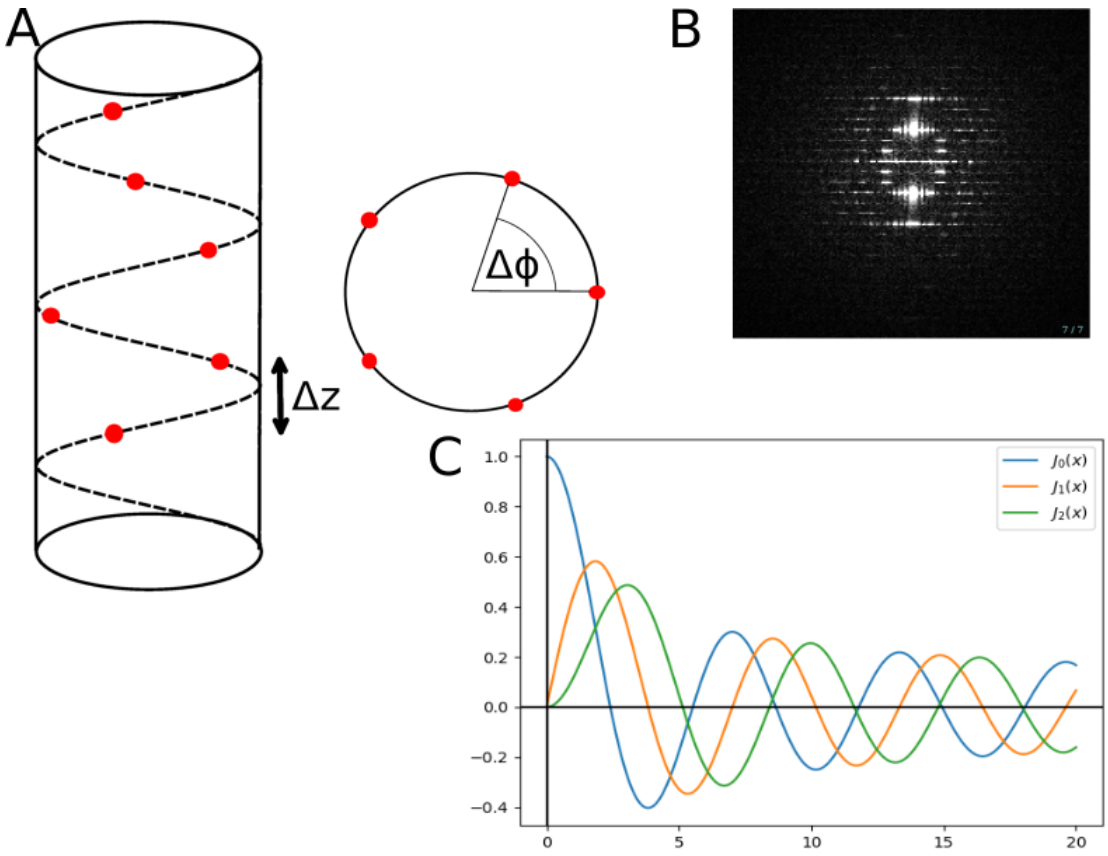

Figure 2.3: (**A**) The helical symmetry of a filament can be described using two parameters. The axial rise Δz is the distance in z-direction between two subunits and $\Delta\Phi$ is the angular increment between them.(**B**) shows layer lines of a helical assembly and (**C**) displays Bessel functions $J_N(x)$ with orders n = 0, 1, 2.

early beginnings of cryo-EM as they provide advantages over non-helical molecules. A single asymmetric unit is repeated many times and their different orientations along the helix make different projections [24]. Thus only few images are sufficient to create a 3D reconstruction because a helix provides all necessary views.
The architecture of helical assemblies can be described by two symmetry parameters: the angle $\Delta\Phi$ that each subunit must be rotated to align with the next subunit and the axial rise Δz that it must be translated along the helix axis, see figure 2.3 [36].

For the last 30 years helical reconstruction was dominated by Fourier-Bessel methods [24] which use the horizontal "layer lines" that are characteristic for Fourier transforms of helical assemblies. The layer lines are analogous to the diffraction pattern of a 2D lattice. The δ-function that describes the diffraction spots in a 2D crystal is convoluted with a Bessel function $J_n(x)$ due to the cylindrical shape of the helix. Bessel functions

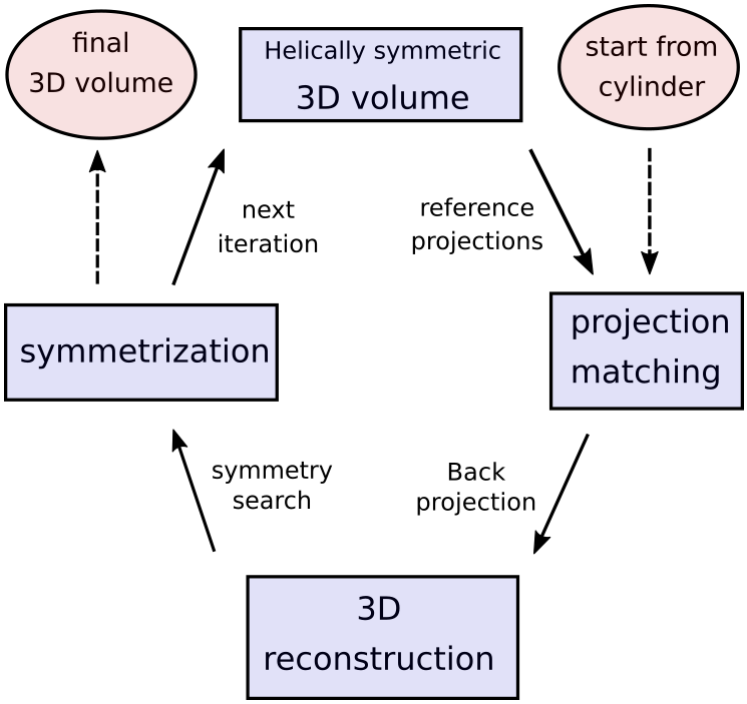

Figure 2.4: Schematic depiction of the IHRSR approach. It alternates between projection matching to determine the orientation parameters, 3D reconstruction and symmetrization of the 3D volume.

are also called cylindrical functions [36]. Once the correct value of the order n is assigned to all layer lines in the helical Fourier transform the 3D structure can be deduced by applying the inverse Fourier-Bessel transform to the layer line data [19].

But in the past 10 years single particle like methods emerged and they are now the main methods used. They circumvent some of the problems that arise from real polymers in the Fourier-Bessel approaches: If the helical filaments are not perfectly ordered the layer lines in the Fourier transform are easily misinterpreted. For weakly scattering filaments the same problem would occur [24]. The new developed methods can overcome some of the problems mentioned earlier and the Iterative Helical Real Space Reconstruction (IHRSR) is the main methodology [24].

In the IHRSR approach projections from a reference volume with different azimuthal rotation angles are generated. These projections are cross-correlated with the helix segments. The highest correlation for each image gives its azimuthal orientation, the in-plane rotation and the translation needed to be aligned to the reference volume. After the orientation parameters are determined the images can be used in a back projection

algorithm to make a new 3D reconstruction. The volume is searched for the helical symmetry it fits best to and the volume is helicised with this symmetry. Subsequently, this volume is the new reference volume in the next iteration. To start with IHRSR there is no prior knowledge of the 3D structure needed as it works with a cylinder as the first reference structure. However, one needs an estimate of the helical symmetry which can come from layer lines, mass per length measurements, from measurements in the micrographs or trial and error [24]. The IHRSR process is continued until convergence is obtained which can be judged by monitoring the following parameters: 1) The orientation parameters are expected to converge, 2) the azimuthal angle should be uniformly distributed and 3) the helical symmetry parameters Δz and $\Delta\Phi$ adopt fixed values. Normally, IHRSR will converge according to the criteria mentioned above but this does not necessarily mean that the obtained structure corresponds to the true structure of the sample. This can, for example, be the case when the initially guessed symmetry parameters deviate significantly from the true symmetry [5]. The basic principle of the IHRSR approach is also presented in figure 2.4.
The IHRSR was originally implemented by Egelman using the SPIDER package [23] but independent implementations of this approach followed. For example, Volkmann used the IHRSR methodology together with the software packages EMAN and CoAn [101]. Grigorieff *et al* implemented a similar method in their software Frealign [40] and the helical reconstruction procedure in SPARX uses a modiefied IHRSR approach, as well [5].

2.1.2 IHRSR implementation in SPARX

Just like in single particle reconstruction (SPR) methods filamentous samples are prepared on a supporting grid and imaged with the direction of the beam being perpendicular to the grid plane. But the obtained projections in this case are very long, that means many times longer than the helical rise Δz. For the IHRSR procedure the filament projections have to be windowed into square units called segments. Each segment comprises a few copies of the asymmetric unit. While windowing, the filaments are oriented such that the filament axis coincides with the $y^{'}$-axis which is the axis that comes from the the z-axis in the 3D coordinate system by making 2D projections, see figure 2.5. To maximize the amount of data that can be used for alignment and reconstruction the segmentation is carried out in such a way that the obtained segments overlap extensively. An important consideration to make while doing the window-

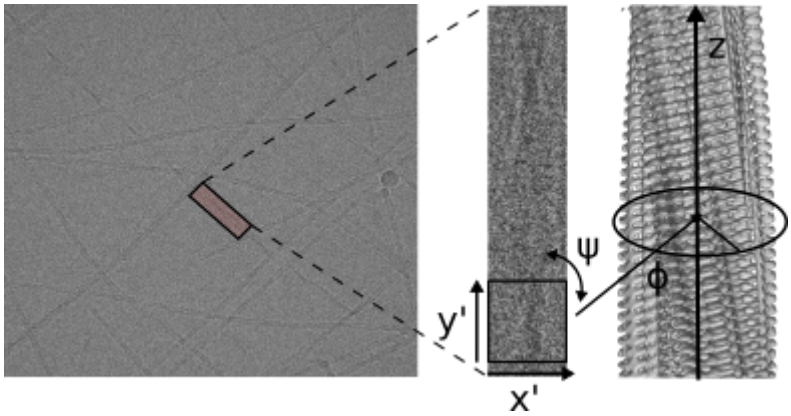

Figure 2.5: Geometrical relationship between 2D images/projections and 3D volumes and orientation parameters (out-of-plane tilt (Θ) not shown).

ing/segmenting step is how to choose the length of the segments and the radius of the cylinder that encompasses the filaments. The optimal parameters for that depend on the rise of the subunits Δz and the flexibility of the filament. Each individual segment should comprise at least a few copies of the subunit and longer segments provide more signal and reduce error during alignment. The optimal length of segments is limited by the flexibility of the filaments because shorter segments have a more coherent ordering. The optimal segment length needs to be found for each system individually.
The reconstruction method used in SPARX is a direct Fourier inversion algorithm that is based on the central section theorem, see section 2.1. The CTF correction is integrated in the reconstruction algorithm via a Wiener filter. It sets frequency regions to zero where the CTF, envelope or the spectral SNR (SSNR), that describes the SNR in Fourier space, are zero. The Wiener filter enhances amplitudes in regions where the CTF and envelope functions are small but only if SSNR is high and reduces the enhancement in frequency regions with low SSNR [70]. Reprojections of a 3D structure are computed using a reverse gridding technique. The projection matching used for angular assignment is based on the comparison between the EM images and reprojections of the current structure. This is done by resampling both images to polar coordinates. 1D FFTs of concentric rings are computed. The inverse FFT of sums of their products gives a cross correlation function and the location of its maximum gives the angle between two images [5].

3 Aβ fibrils

3.1 Aβ sample and fibril preparation

Sample preparation was carried out by Lothar Gremer [39]. Recombinantly produced uniformly labeled $[^{13}C\,{-}^{15}\,N]\mathrm{A}\beta_{1-42}$ was purchased from Isoloid GmbH with a purity of 95 %. Purity was increased to 98.5 % by applying Reverse-phase high performance liquid chromatography (RP-HPLC). HPLC is a common technique to separate different components in a mixture. The sample in a liquid solvent, called mobile phase, is pumped through a column containing adsorbent material, the stationary phase. Each component of the sample interacts slightly different with the adsorbent material, thus leading to different flow rates and a separation when the components leave the column. In RP-HPLC, in contrast to regular HPLC, the stationary phase is non-polar which makes polar molecules elute first. For Aβ purification the mobile phase was 30 % (v/v) acetonitrile (ACN) and 0.1 % (v/v) trifluoroacetic acid (TFA) in water. It was shown before that Aβ molecules elucidate as monomers under these conditions [6]. The collected fraction of monomeric Aβ was incubated at room temperature within several weeks and monomers converted to fibrils.

3.2 Atomic force microscopy and X-ray diffraction

The fibrils were first investigated with atomic force microscopy (AFM) [39]. In AFM the sample is scanned with a cantilever and its deflection, which is characteristic of the interaction between the sample and the tip, can be measured. The AFM images of the fibrils show long, straight and unbranched fibrils that seem to have a regular twist as can be seen in figure 3.1. AFM experiments were performed by Elke Reinartz.

For X-ray diffraction experiments [39] centrifuged fibrils in solution were dried between glass capillary supports and afterwards attached on a pin-holder. Wide angle X-ray

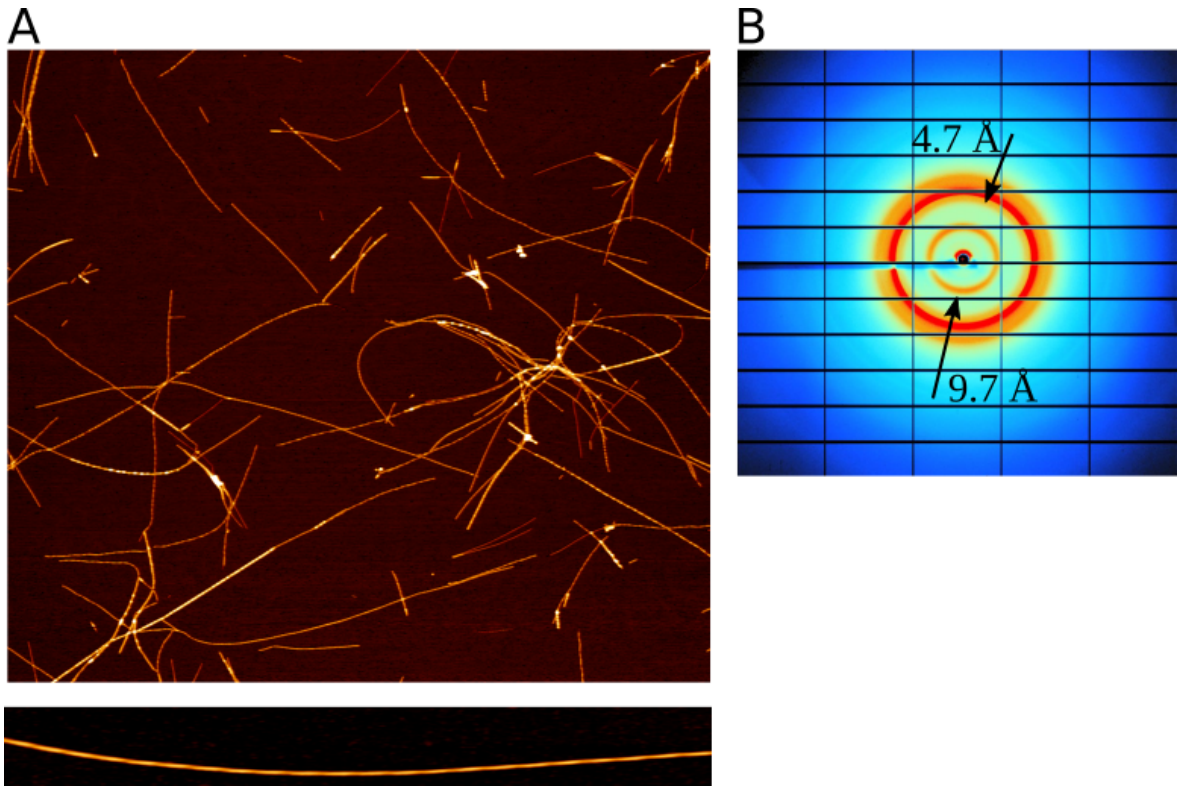

Figure 3.1: (**A**) AFM from Aβ fibrils. An overview and a representative fibrils are shown. (**B**)X-ray diffraction images of Aβ fibrils.

diffraction experiments were performed using a wavelength of 1.0332 Å and a beam size of 50 µm. The data show amyloid-characteristic rings [63] at 4.65 Å emerging from the β-strand separation and at 9.72 Å produced by parallel β-sheets, see figure 3.1. X-ray diffraction experiments were done by Jörg Labahn.

3.3 Electron microscopy

The first step in electron microscopy was to record images of negatively stained Aβ fibrils for more detailed assessment of the sample [39]. The fibrils were prepared on 400 mesh carbon-coated copper grids and stained with 1 % uranyl acetate. Imaging was done on a Zeiss Libra120 microscope at 120 kV. Negative stain images are consistent with AFM images presenting a homogeneous sample with long and straight fibrils as illustrated in figure 3.2.

For cryo-EM images the sample was prepared on glow-discharged 300 mesh gold grids (UltrAuFoil R 1.2/1.3, Quantifoil). 2.5 µl of sample were applied to the grid and after

blotting for 2.5 s the grid was plunge-frozen using a Vitrobot (FEI).

On a Tecnai Arctica microscope operated at 200 kV and using a field emission gun, 2026 micrographs were collected with a magnification of 110,000. Images were recorded with a Falcon III direct electron detector using the software EPU. Each micrograph consists of 80 frames that were aligned and averaged by the internal microscope software. The exposure time was 2 s and an integrated flux of $24\,e\,\text{Å}^{-2}$ was used. The applied defocus range was 0.5 to 1.5 µm and the pixel size 0.935 Å, calibrated from gold diffraction rings in the powerspectra of a cross grating grid. EM experiments were performed by Gunnar Schröder and Raimond Ravelli.

A first inspection of the micrographs revealed long, straight and unbranched fibrils recapitulating what was found earlier by AFM and negative stain images, see figure 3.2. Also consistent with observations from AFM and negative stain, the sample emerged to be fairly homogeneous. About 90 % of the sample consists of thin filaments with a fixed diameter of approximately 7 nm. Two more species were found in the sample which total amounts is around 5 %. One of these has an alternating diameter between 4 and 10 nm, clearly revealing its twisted nature. The other type of fibril is characterized by a relatively thick diameter of around 14 nm. All described species are investigated further in the following sections.

All image processing tasks were performed using the software package EMAN2/Sparx [92][45]. The first step was to determine defocus and astigmatism parameters and assess the micrograph quality using the program *cter*. Only micrographs that were determined to contain significant signal beyond 6 Å were chosen for further analysis. This information is provided by the *cter* output which determines the frequency at which the signal is only 50 % as a result of the defocus error. This resulted in 995 micrographs from which fibrils could be picked. As mentioned before mainly three types of fibrils could be detected in the micrographs, a straight and a twisted filament and a filament with a significantly larger diameter, with the first being the most abundant species in the sample. The three types of fibrils were processed separately and the most frequent, the straight fibril, is described first.

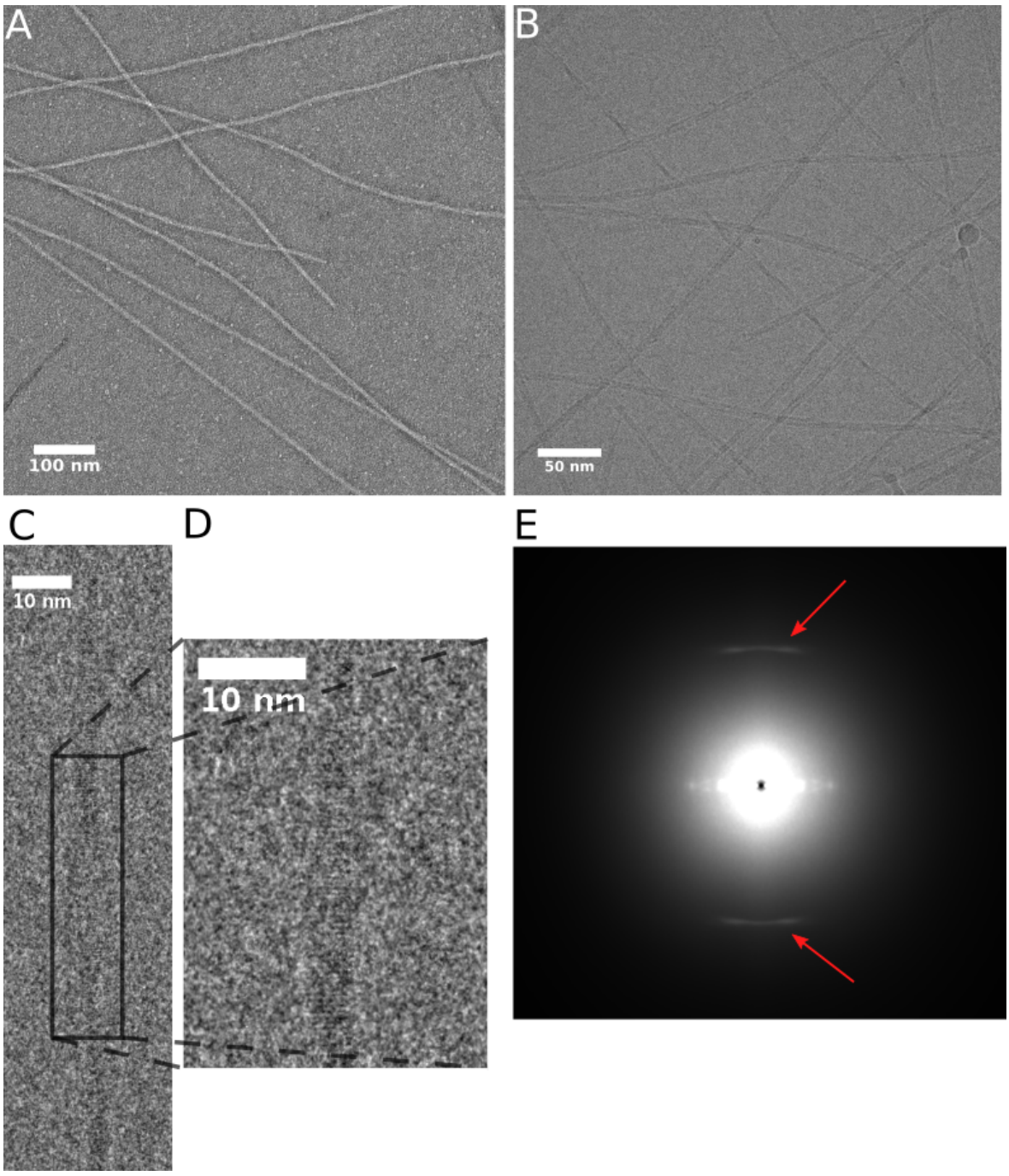

Figure 3.2: (**A**) A negative stain raw image. (**B**) An exemplary cryo-EM micrograph. (**C** and **D**) A representative filament with 4.67 Å β-strand pattern. (**E**) averaged power spectrum of filament stack showing a 4.67 Å peak (red arrows).

3.3.1 Straight filaments

3.3.2 Image processing

After sorting out bad micrographs, fibrils were picked manually using the program *helixboxer*. In total, 3179 fibrils were picked, from which individual segments were extracted with 90 % overlap. Each segment has a size of 216 x 216 pixels which yields a total of 127,765 segments that are combined in an image stack. In the power spectrum of all segments a clear peak at 4.67 Å is visible which can be seen in figure 3.2. Remarkably, a stripy pattern that originates from the cross-β structure can already be seen in some fibrils in the raw micrographs, see also figure 3.2.

In the raw micrographs the cross-over length of some exemplary fibrils is measured to be approximately 550 Å. Assuming two monomers per layer in a C_2 symmetric arrangement would result in an angular increment of $-1.52°$ per subunit, taken into account the measured cross-over length and a rise of the subunits of 4.67 Å.
To determine the symmetry parameters more accurately, 11 different parameter combinations were tested. For each combination a 3D reconstruction is calculated using *sxhelicon* applying the particular symmetry. The map is filtered to 20 Å and an additional run of *sxheliconlocal* is carried out. In a projection matching step cross-correlation (CC) coefficients are calculated to determine the best orientation parameters of each segment. For each segment the highest CC coefficient that determines the orientation of the segment is stored in a file. For all tested symmetry parameter combinations the average cross-correlation coefficient is calculated and used to determine the best symmetry parameters. This is illustrated in figure 3.3 where CC coefficients are plotted against symmetry parameters. As can be seen from this graph the best symmetry parameter combination is 4.67 Å/$-1.45°$ which is only slightly different from the previously measured symmetry of 4.67 Å/$-1.52°$.

A 3D reconstruction calculated with *sxhelicon* and *sxheliconlocal* using the parameters 4.67 Å/$-1.45°$ can bee seen in figure 3.4. In the cross-section of the map it is possible to trace the complete amino acid chain. However, the resolution is not high enough for clear β-strand separation and side chain placement, which can be seen figure 3.4 B.

Further improvement of the map should be reached by discarding bad segments. Bad segments are identified using the same principle that was used to find the best symmetry parameters. Now the symmetry parameters 4.67 Å/$-1.45°$ are used to calculate a 3D

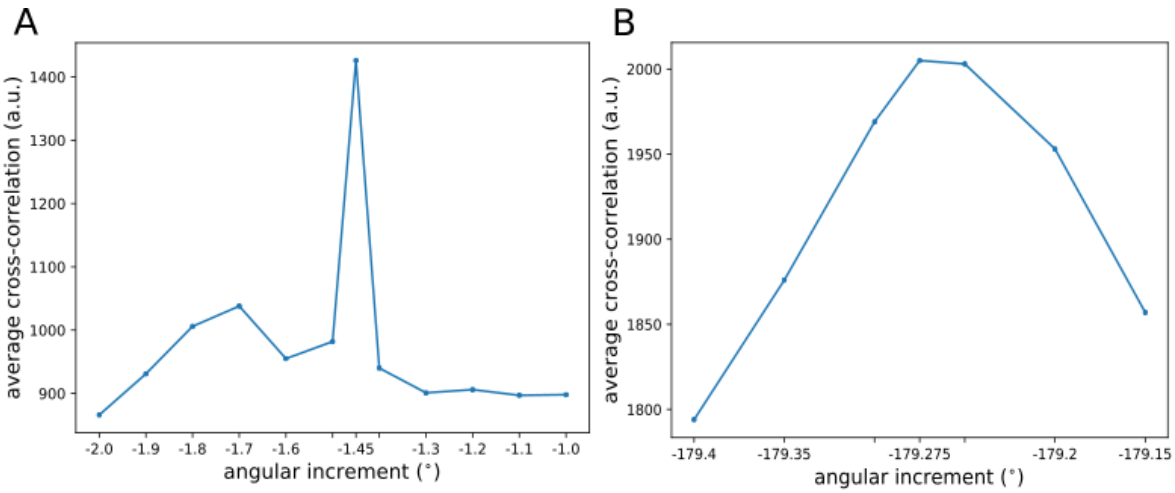

Figure 3.3: (**A**) symmetry search with original image stack and applied C_2 symmetry. (**B**) symmetry search with optimized image stack and C_1 symmetry.

reconstruction and cross-correlation coefficients for each segment in the same way as described above. An average CC coefficient is now calculated for each fibril. The CC coefficient strongly depends on the defocus as indicated by a correlation coefficient of 0.855. The average fibril cross correlation is plotted against the defocus. Different image stacks with varying amount of fibrils are created by selecting fibrils that lie above a line defined by $y = 0.855 * \text{defocus} + n_i$, which selects the fibrils with highest CC coefficient values, as illustrated in figure 3.5.

3D reconstructions using different numbers of fibrils were calculated. The best reconstruction was found to be the one using the best 813 fibrils, which is approximately 1/3 of the original image stack. This stack was created selecting all fibrils above the line $y = 0.855 * \text{defocus} + 0.3$. The selected 813 fibrils contain 27,132 segments. Applying the helical symmetry results in approximately 295,000 asymmetric subunits that are used for further processing.
This image stack was used to generate 2D class averages. First the segments are sorted into 10 groups using a k-means clustering algorithm in the program *sxk-means*. The resulting averages of the clusters appear blurry and for this reason the images in the clusters are aligned again using *sxali2d*. The final averages show high contrast images of the fibril segments with a clearly visible β-strand pattern. Also apparent from the averages is the arrangement of the subunits which are not C_2 symmetric but arranged in a staggered fashion. The averages are presented in figure 3.6.

With this finding the symmetry search is repeated, now testing different symmetries

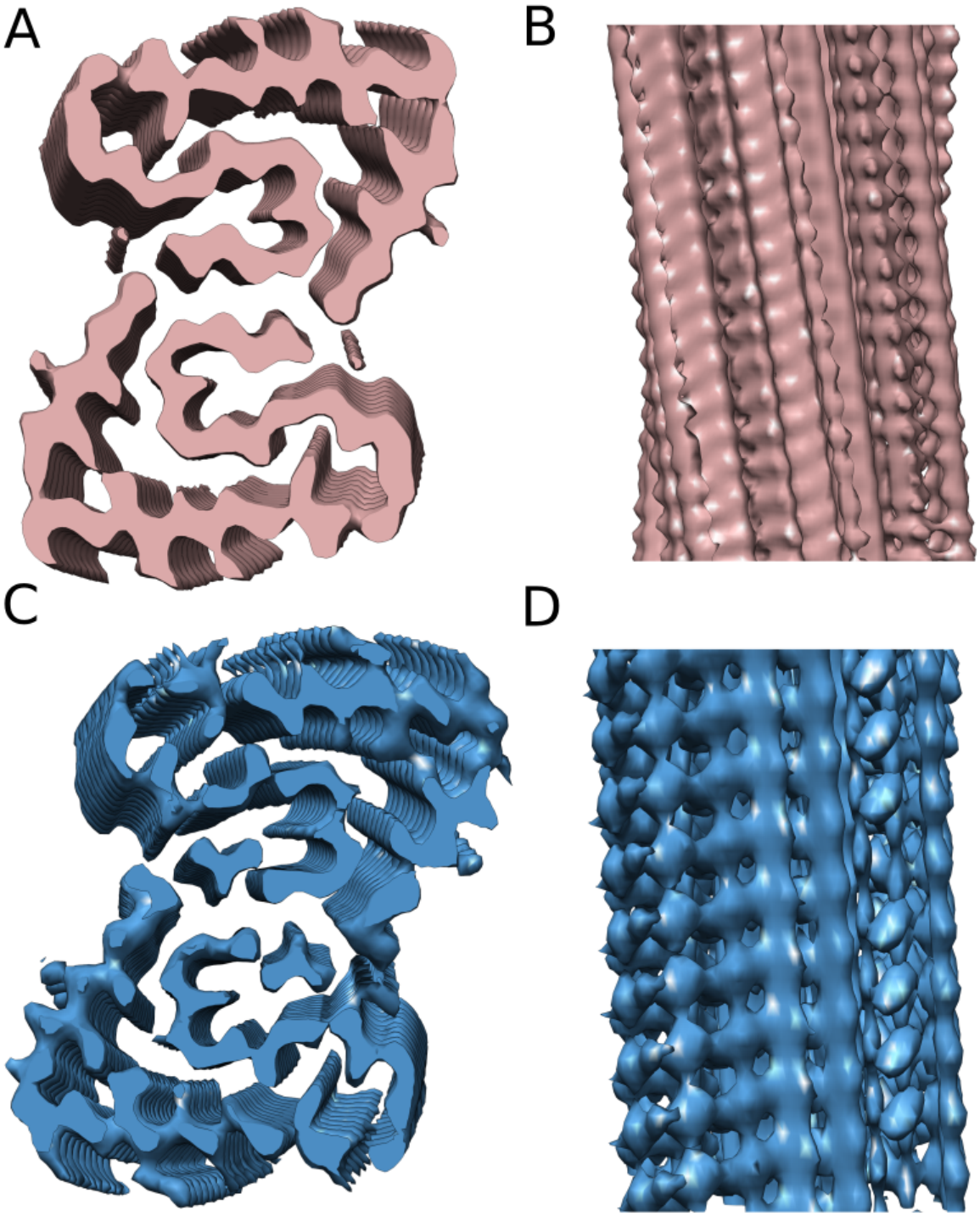

Figure 3.4: (**A**-**B**) 3D reconstruction from original image stack with helical rise 4.67 Å and angular increment $\Delta\Phi$ −1.45° and applied C_2 symmetry. Shown are a cross section and a side view. (**C**-**D**) 3D reconstruction from optimized image stack with helical rise 2.335 Å and angular increment $\Delta\Phi$ −179.275°. A cross section and a side view are shown.

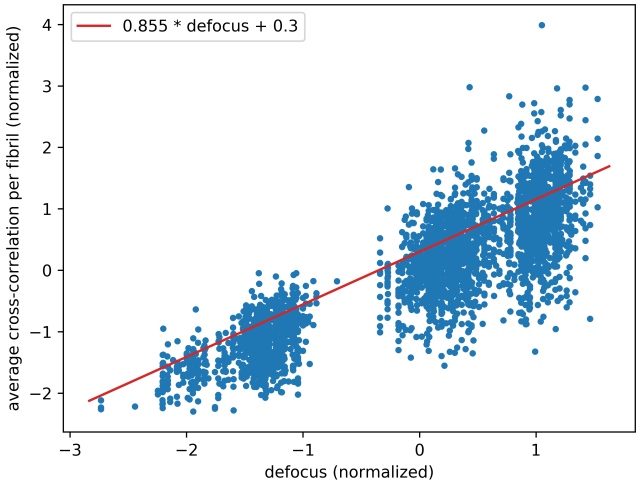

Figure 3.5: Averaged filament cross correlation coefficients against defocus. Averaged filament cross correlation coefficients are plotted against defocus. All filaments above the red line are selected for further processing.

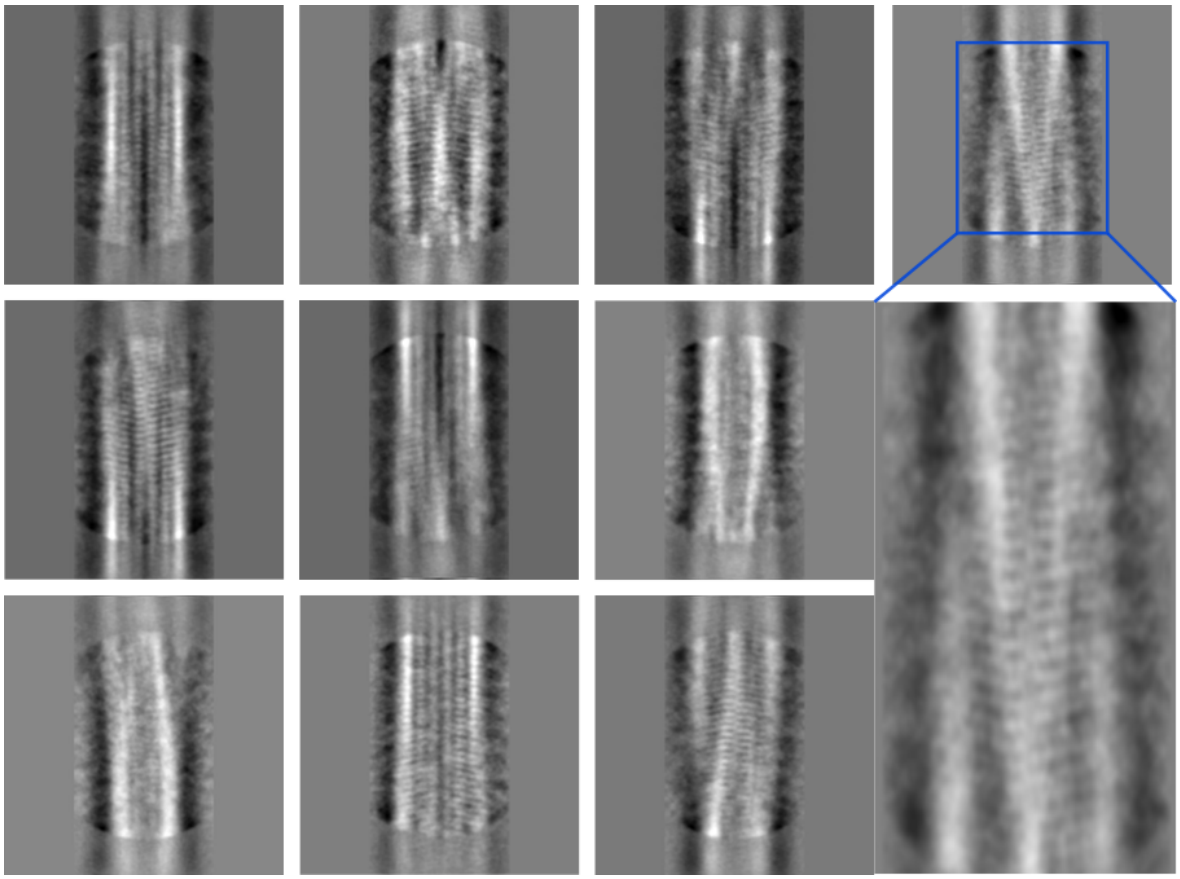

Figure 3.6: Class averages generated from the final image stack.

with an axial rise of 2.335 Å and angles between −179.0° and −179.0°. This corresponds to an approximate 2_1 screw symmetry with the axial rise still being 4.67 Å, consistent with the power spectrum.
7 different symmetry pairs were tested in the same way as described above and the symmetry giving the highest correlation is −179.275°/2.335 Å. The corresponding plot is shown in fig 3.3 B. A 3D reconstruction was calculated using this symmetry and the final image stack. The cross section is similar to that of the previous reconstructions but β-strands are now clearly separated and side chains are visible. This reconstruction is shown in figure 3.4 C-D.

3.3.3 FSC calculation

For FSC calculation, after refinement the data set is split in half (even and odd) using whole fibrils to make sure not to have overlapping segments in the other half set of the data. Both half sets are reconstructed using the program *recons3dn*. For FSC calculation even and odd reconstruction are masked using a soft cylindrical mask with 40 Å radius. FSC values are computed between even and odd map and plotted against spatial frequency. The plot reveals that the FSC curve has some undesired artefacts mostly due to the high symmetry of the map. An increase of the FSC curve at high frequencies can be removed by adding a small amount of high frequency noise to the half maps. The sharp peaks that occur at a few spatial frequencies are eliminated by making slices of the half map of different widths and FSC curves are calculated separately for each slice pair. Afterwards the curves are combined by taking the lowest value from all curves at a particular spatial frequency. As this is not the "gold standard way" of calculating the FSC, it is suggested to use the 0.5 criterion as resolution cut-off, which gives an overall resolution of 4.8 Å. However, the visual inspection of the map and the good agreement with the atomic model refined to this map suggest a higher resolution of around 4 Å. This matches the 0.143 criterion, which gives a resolution of 4.0 Å. Additionally, the FSC between the model and the masked map was calculated, which also estimates the resolution of the map to be 4 Å. The different FSC curves are shown in figure 3.7.

The final map was processed with the EMAN2 option *filter.lowpass.autob* and a B-factor of $-501\,\text{Å}^2$. Subsequently a B-factor of $-50\,\text{Å}^2$ was applied to further sharpen the map. This map was finally filtered to 4 Å.

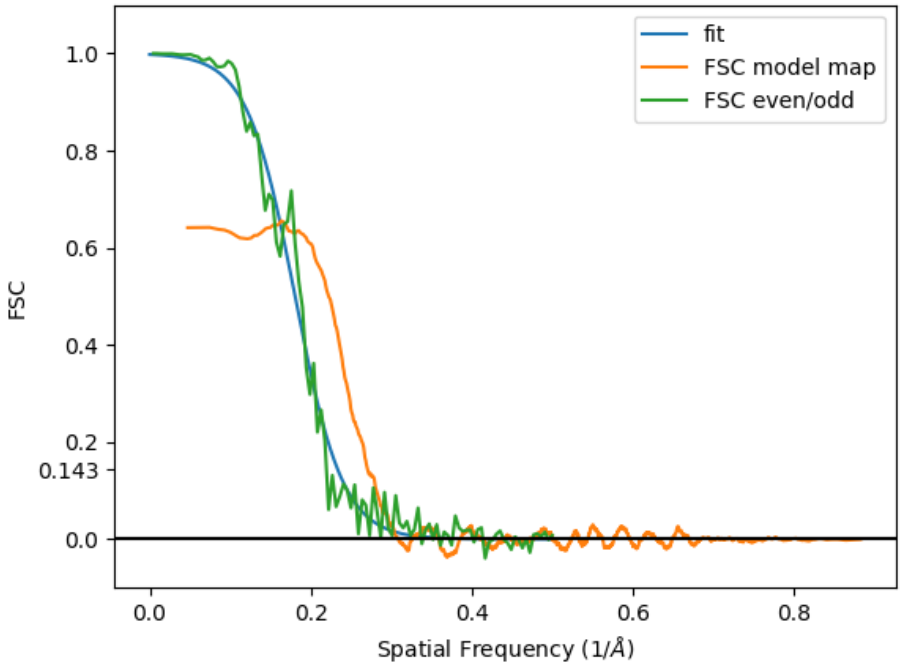

Figure 3.7: FSC curves of half maps and model/map FSC.

3.3.4 Model building and refinement

The atomic model was build in *coot* [26]. Because of the high resolution of the map side chain densities are visible and β-strands are clearly separated in almost all regions of the map. It is therefore evident how to place the individual residues into the map. Only the β-strand between Ser26 and Lys28 could not be separated completely and in this region the sequence could be placed in two different ways along the helical axis. The chain could either go further up entering the next layer or go slightly down again which is shown in figure 3.8. As the latter model fits better in the overall density map it was chosen to be the correct model. However, a second less populated conformer with the first described arrangement could exist.

Refinement of the model against the density map was done using the PHENIX [1][2] real-space refinement tool. A fibril fragment of nine subunits was used for this purpose. The resolution cut-off was set to 4 Å and NCS restraints were applied between all subunits. In the process of refinement, additional restraints between backbone hydrogen bonds between chain i and $i+2$ were added. Also, grouped B-factor refinement was done that yielded an average B-factor of $-88.6\,\text{Å}^2$. The reported map cross correlation

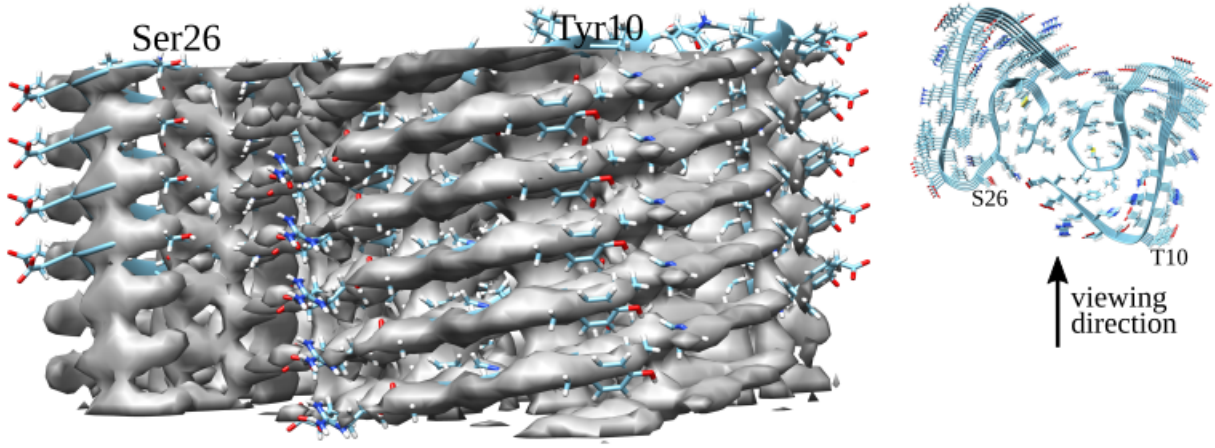

Figure 3.8: The β-strands comprising the residues 1-9 are clearly separated. At position 26-28 the density is not clearly separated and two different placements of the chain are possible.

value around the atoms for the final model was 0.683. For model validation, bond length and bond angle RMSD were analyzed and reported to be 0.014 Å and 1.73 Å, respectively. The all-atom clash score, which is the number of clashes found per 1000 atoms is 7.41 for this model. No Ramachandran outliers were found for this model and 80 % of the residues are in favored regions. Most of the residues that are found in unfavored regions of the Ramachandran plot, such as Tyr10, Asp23, Gly25, Leu34, and Gly37 are forming kinks in the model, see figure 3.9.

Also, no rotamer outliers were found. The Molprobity [8] score which combines the analysis of the clashscore, rotamer, and Ramachandran evaluations into a single score was calculated to 2.67, supporting the high quality of the model. To assess the agreement between model and map the EMRinger [4] score was calculated. A score of 2.1 at 4 Å resolution is an indicator for a good fit.

3.3.5 Discussion

In the final EM density the complete backbone and almost all of the side chains are well resolved. Both the density map and the atomic model show that the fibril consists of two protofilaments winding around each other. The individual $A\beta_{1-42}$ molecules in the filaments are stacked in parallel, in-register β-sheets as found for most amyloid fibrils, see figure 3.10. The monomers are not flat but the two N-terminal β-sheets comprised of the residues 1-9 and 11-21 are tilted by 10° with respect to the long axis

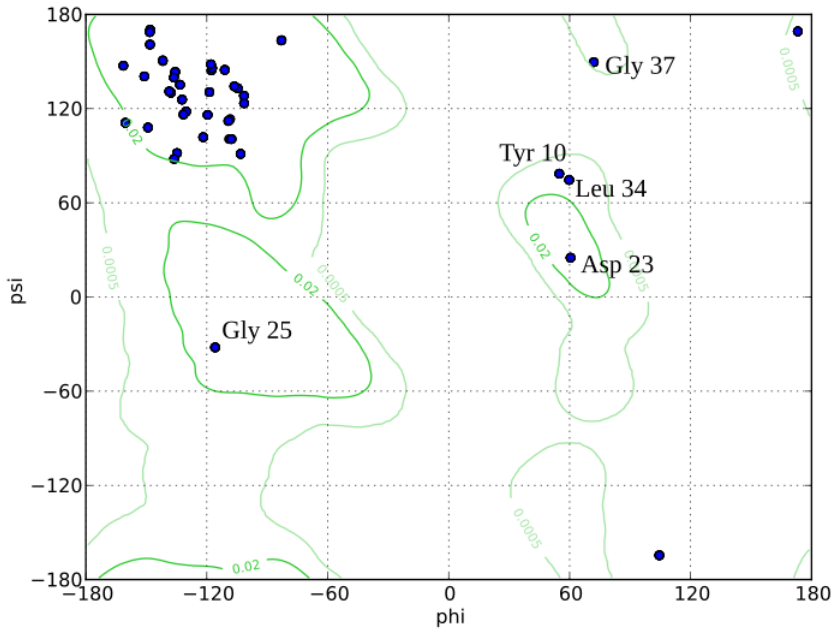

Figure 3.9: Ramachandran plot of the atomic model.

of the helix which is illustrated in figure 3.11 B. The symmetry of the fibrils can be described best with an approximate 2_1 screw symmetry. The rise in each protofilament is 4.67 Å and the angle between subsequent monomers is −179.275°. Because of this particular symmetry subsequent subunits from different protofilaments are not in the same plane but shift by 2.335 Å along the helical axis. The arrangement of the subunits is illustrated in figure 3.12 A. A similar symmetry has been described recently for the paired helical tau filament (PHF)[32].

The arrangement of a single subunit can be portrayed as an overall LS-shape. The L-part is formed by the N-terminal side and comprises approximately amino acids 1-16. The S is formed by the C-terminal part which ranges from amino acid 17-42. The binding interface of two opposite subunits is formed by the C-termini. A cross section of the density and model is shown in figure 3.11 A to demonstrate the subunit arrangement.

Solid-state NMR experiments were carried out on fibrils from the same sample preparation. Secondary chemical shifts were used to predict regions of β-strands, that are very similar to those determined from the atomic model with DSSP [94] as shown in figure 3.12. Also, long-range contacts that were determined from NMR experiments

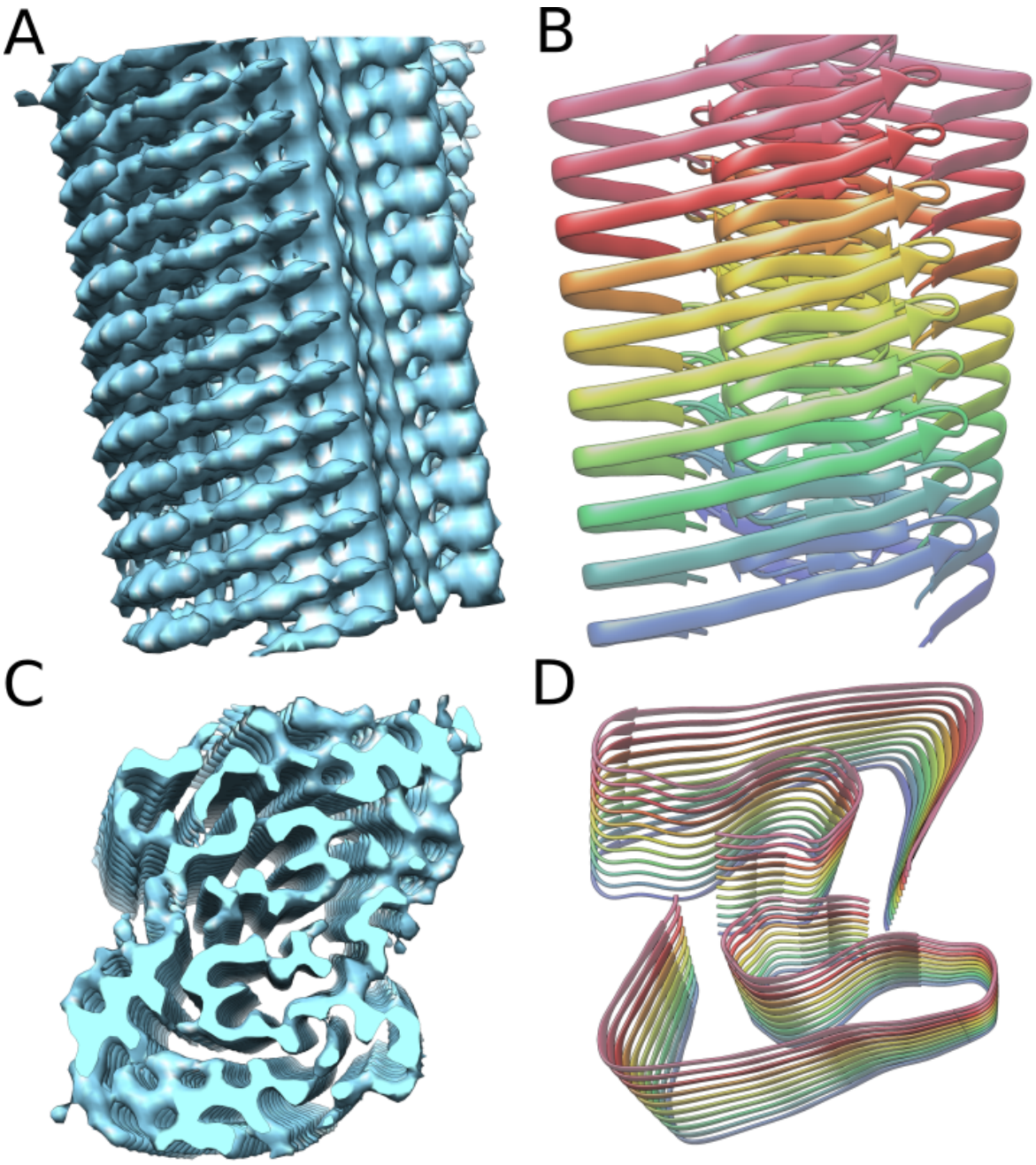

Figure 3.10: Aβ fibril. (**A**) Cryo-EM density of the fibrils showing a clear β-strand separation. (**B**) Atomic model of the fibril revealing parallel β-strand. (**C** and **D**) Cross sections of (**A**) and (**B**).

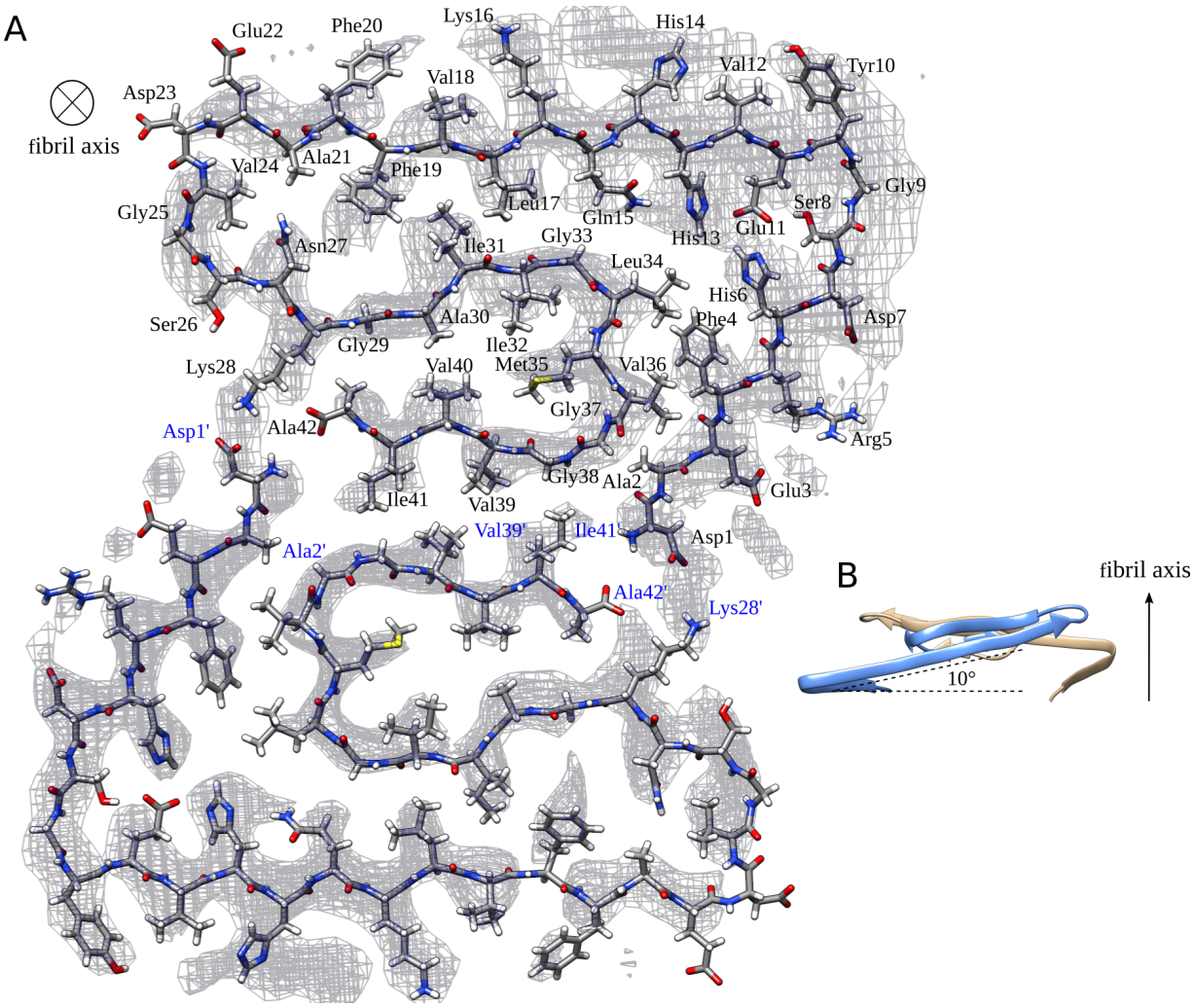

Figure 3.11: (**A**) Cross section of the fibril. The model in the density map is shown. Two subunits from the two protofilaments are opposing each other. Each subunit has an LS-shape. (**B**) Two subunits shown from the side, demonstrating the rise of each subunit

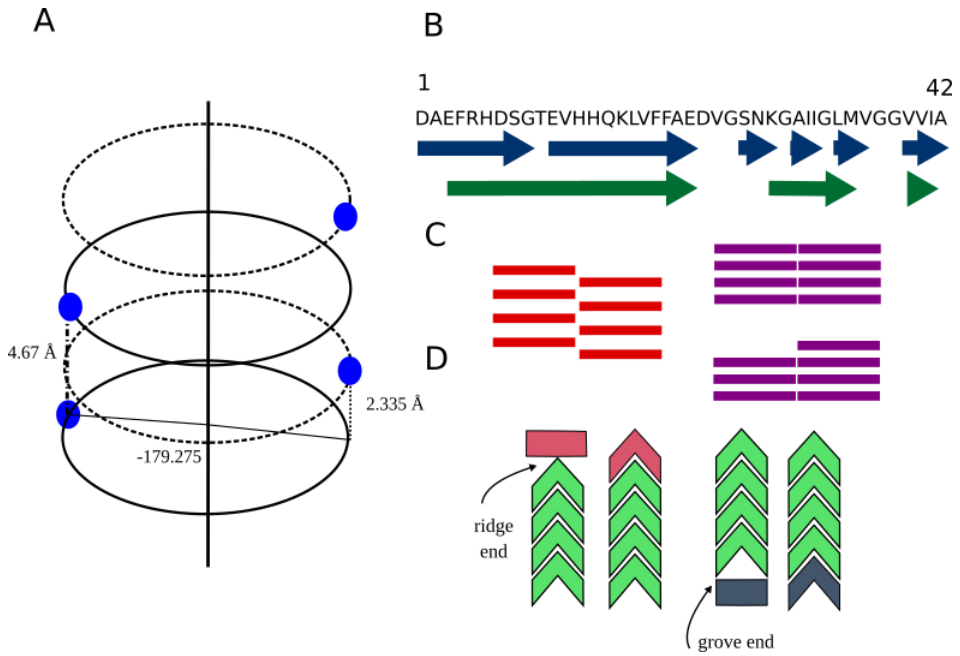

Figure 3.12: (**A**) Depiction of the approximate 2_1 screw symmetry with helical rise of 2.335 Å and angular increment −179.275°. Subunits are represented by blue circles. The spacing within one protofilament is 4.67 Å. (**B**) Aβ_{1-42} amino acid sequence with β-strands predicted from NMR experiments (blue) and determined with DSSP (green). (**C**) Different arramgement of subunits in filaments with C_2 (purple) and 2_1 screw (red) symmetry. (**D**) Depiction of possibly different binding pathways at fibril ends.

are in good agreement with the atomic model.

In addition to the β-sheet hydrogen bonds the fibril structure is stabilized by hydrophobic interactions and salt bridges. Each subunit has three main hydrophobic clusters that are the following: One cluster comprises the amino acids Ala2, Val36, Phe4, Leu34 another cluster is made out of Leu17, Ile31, Phe19 and the third cluster is made out of the amino acids Ala30, Ile32, Met35 and Val40. Theses clusters expand along the fibril axis and are probably essential for fibril stability. Also, the protofilament interface is formed by hydrophobic interactions between the residues Val39 and Ile41 in opposing subunits. A fragment of the fibril colored according to its hydrophobicity is shown in figure 3.14 B.
The EM density map and protonation states determined from NMR experiments are indicators for salt bridges between Asp7 and Arg5, Lys28 and Asp1, His6 and Glu11 and His13 and Glu11. More detailed views of these particular regions are presented in figure 3.13 A-C. The salt bridges with Glu11 stabilize the kink around Tyr10. A

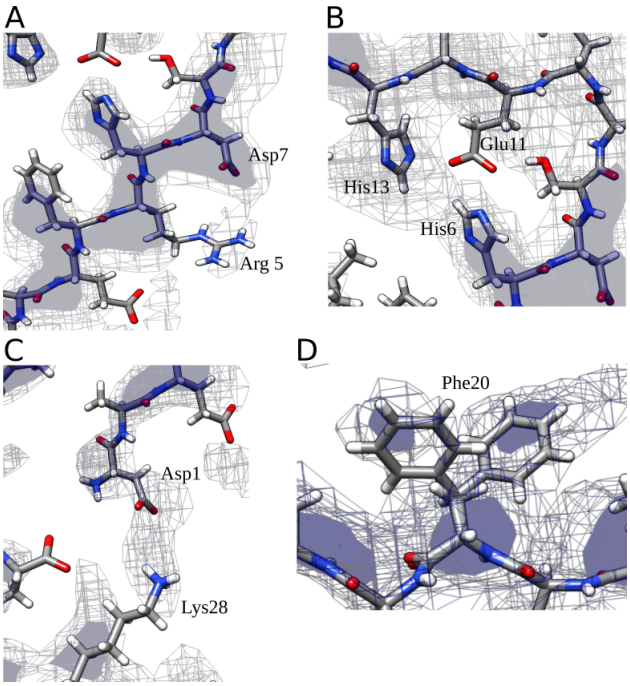

Figure 3.13: (**A**) Salt bridge between Arg5 and Asp7. (**B**) Salt bridge between Glu11 and His6 and His13. (**C**) Salt bridge between Lys28 and Ala1. (**D**) The two rotamers of Phe20.

similar kink was found before in the structure of the Osaka mutant derived from NMR experiments alone, which was described in Section 1.3. The authors state that the turn region between Gly9 and His13 is stabilized by salt bridges [87].
Besides the salt bridges that arise from both cryo-EM and NMR experiments both techniques indicate that Phe20 exists in two different rotamers. Density for both rotamers is visible in the reconstructed map, see figure 3.13 D. Additionally, a resonance doubling for Phe20 and Ala21 can be observed in 2D NCACX correlation spectra.

Due to the staggered arrangement of the subunits and their rise from the N- to the C-terminus, the N-terminus of the subunit i is close to the C-terminus of subunit i-3. This means that the previously described salt bridge between lysine and asparagine involves the subunits i (Asp1) and i-5 (Lys28). For a detailed description of the subunit labeling refer to figure 3.14 A.

The previously described S-shaped C-termini of the model were also reported for the two $A\beta_{1-42}$ models from NMR experiments [104][11] that were described in Section 1.1. However, the structures are still significantly different, as the N-termini in the NMR models are not visible and in the presented cryo-EM structure the N-terminus is part of the cross-β core. Also, the interface for the NMR models is different as it is

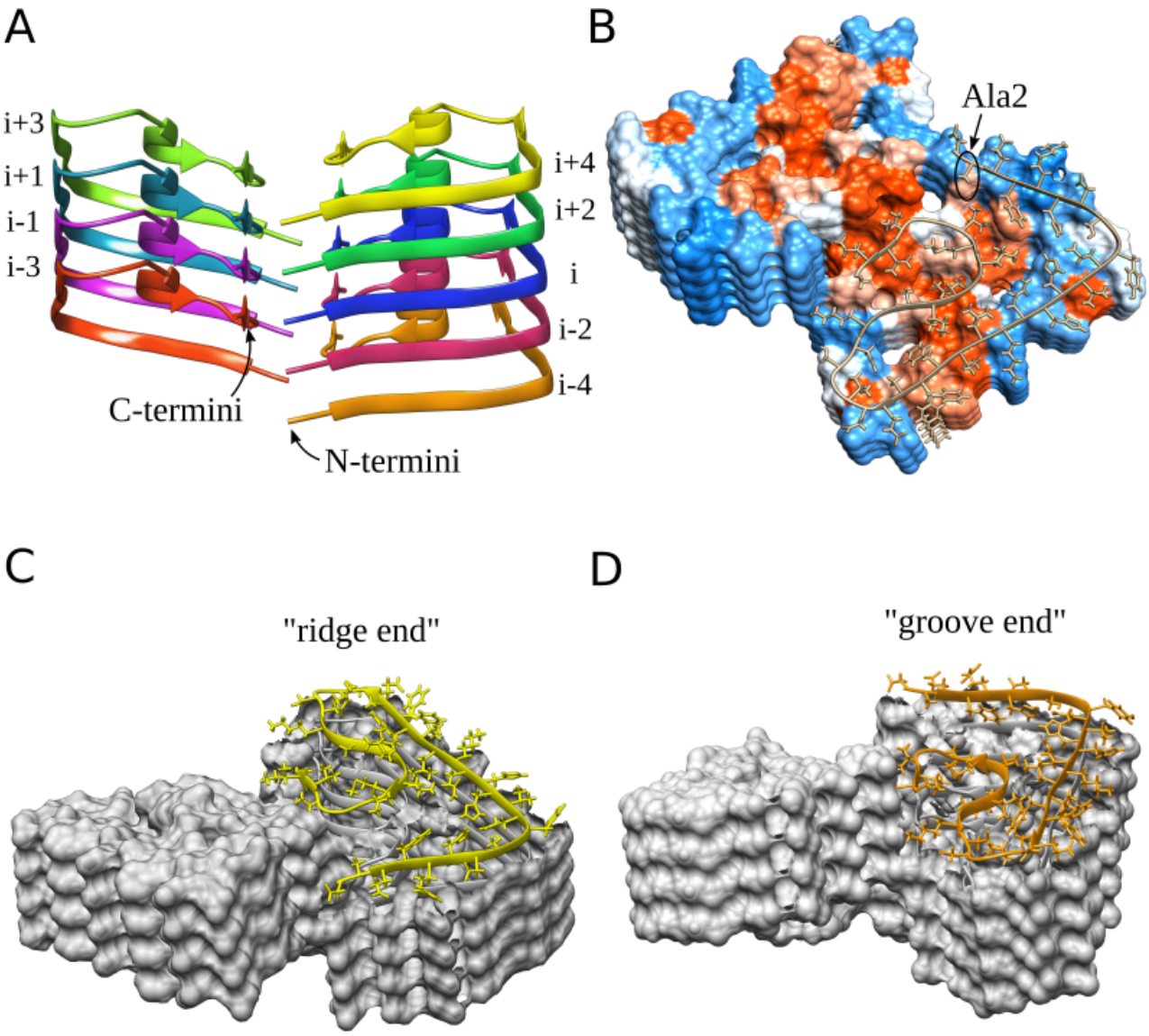

Figure 3.14: (**A**) Staggered arrangement of the subunits. The N-terminus of subunit i is close to the C-terminus of subunit $i-3$. (**B**) A fibril fragment colored according to hydrophobicity in the Kyte-Doolittle scale. Colors range from blue (most hydrophilic) to white and red (most hydrophobic). (**C-D**) Presentation of different fibril ends.

formed by the residues Gln15, Leu17, Leu34, and Met35 whereas the presented model forms an interface between the hydrophobic C-termini, mainly Val39 and Ile41. In addition, different side chains are exposed to the solvent, see figure 3.15 A-B. Another difference between the NMR models and the presented cryo-EM model is that in the NMR models the subunits are flat while in the discussed model the subunits rise from N- to C-terminus. But it should be noted that the arrangement in the direction of the long helical axis cannot be determined by NMR experiments, because intermolecular contacts can not be assigned to a specific layer.

The overall structure of the presented fibril agrees well with a lower resolution EM density map which fibrils were produced under similar acidic, low pH conditions [105]. The authors originally stated that their fibrils fits a hairpin like NMR-derived model. However, if the structure factor determined by Zhang *et al* is applied to the here presented map, the maps looks very similar to each other, which can be seen in figure 3.16. The map of $A\beta_{1-42}$ fibrils derived by Schmidt *et al* [85] is clearly different from the here presented map, see also figure 3.16 and must be a different polymorph.

The 2_1 screw symmetry of the fibrils has consequences for the growth of the fibril. All monomers that are binding to the fibril end are presented with the same interface due to the regular, staggered arrangement. This is in contrast to fibrils with C_2 symmetry, where monomers alternatingly see two identical binding sites or a stepped end formed by the latest added monomer, see figure 3.12. Because of the subunit-rise from N- to C-terminus the fibril ends are different from each other. The difference in height within one subunit is largest at the β-strand formed by residues 27-33. At one side of the fibril these residues protrude at the fibril end whereas on the other end of the fibril this region forms an indentation. The first end is termed "ridge" describing how residues 27-33 stick out at this end. The latter end is called "groove" because the same residues form a groove on the other end. The two fibril ends are shown in figure 3.14. The different end of the fibril might lead to a polar growth of the fibril. Although the binding energy is the same at both ends the binding pathway, that means energy barriers that have to be crossed, could be different. If one side would be favored for monomers to bind, this would result in a polar growth of the fibrils. Polarity in fibril growth has been described before for fibrils of the yeast Sup35 prion protein [18][47].

The next section discusses a few known mutations of Aβ and how they can be related to the presented structure. Two mutations at position Ala2 with opposing effects in regard to the disease have been reported. The mutation A2T, also called Icelandic might be

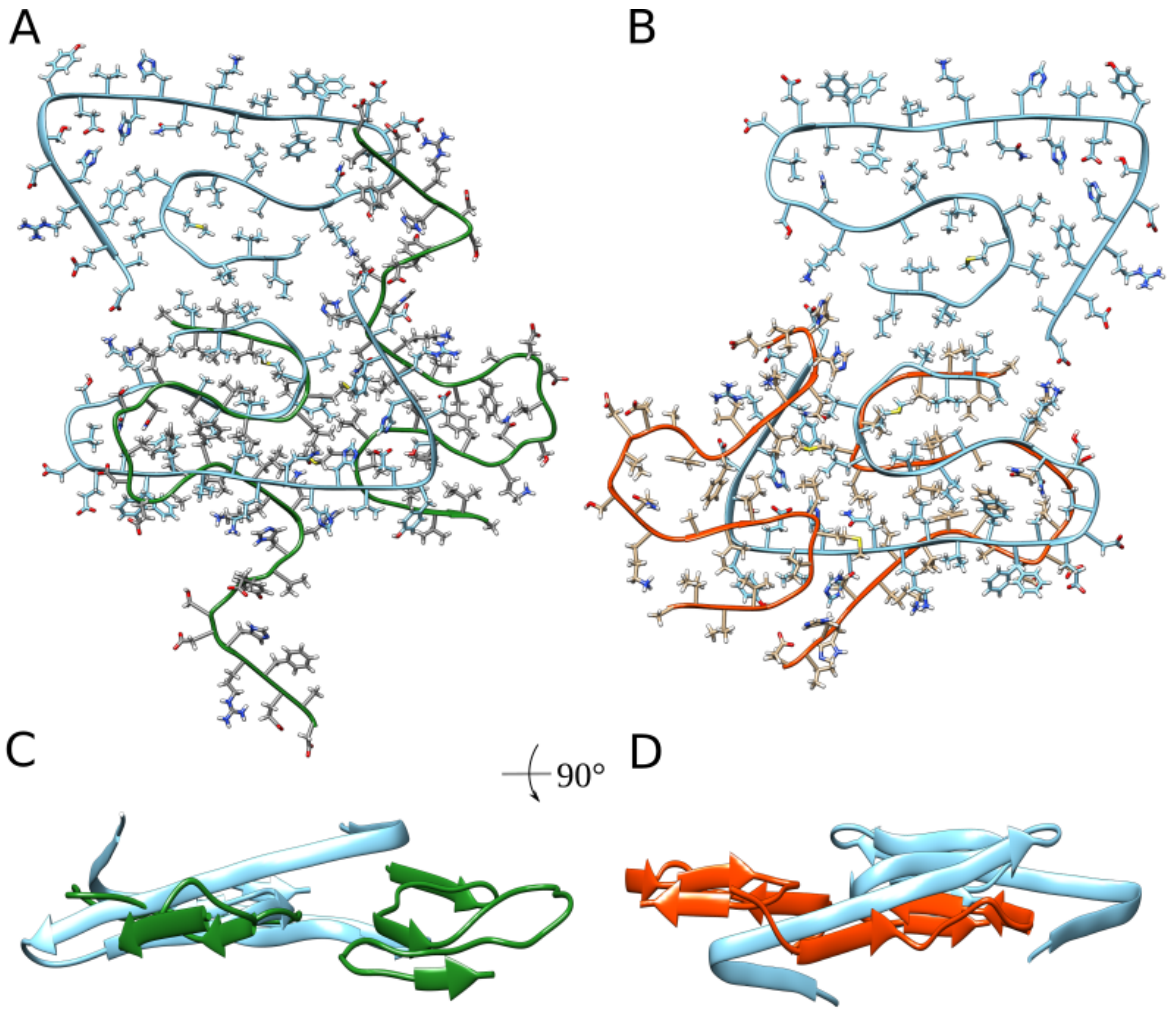

Figure 3.15: Comparison between Aβ fibrils models from NMR experiments and the presented cryo-EM model in a cross section (**A**-**B**) and from the side (**C**-**D**). The green model was proposed by Wälti *et al* [104] the red model by Colvin *et al* [11]. The presented cryo-EM model is shown in blue.

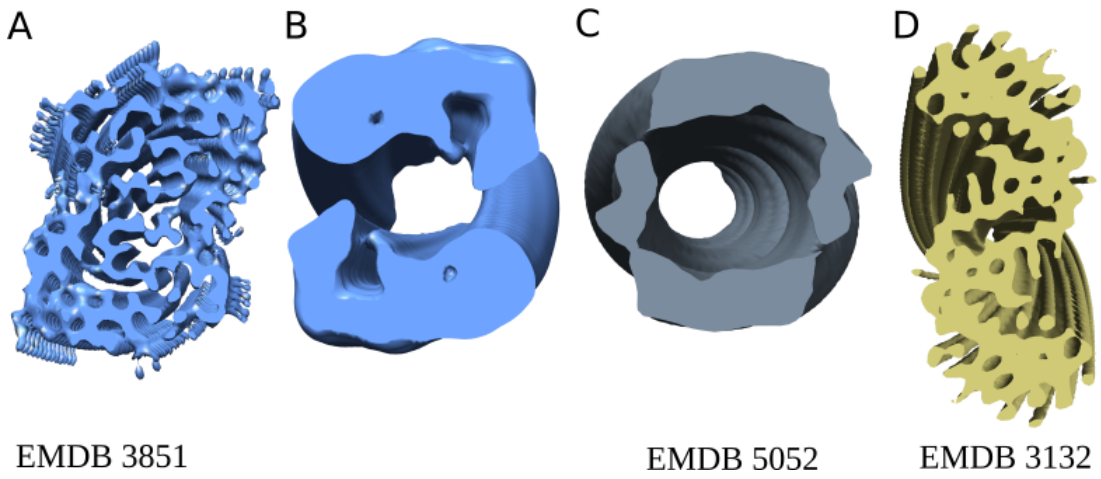

Figure 3.16: Comparison of different density maps of $A\beta_{1-42}$ fibrils. (**A**) The presented density filtered to 4 Å. (**B**) The same density with the structure factor of Zhang *et al* applied. (**C** and **D**) The density maps reported by Zhang *et al* and Schmidt *et al*, respectively.

protective against AD while A2V is pathogenic. Ala2 is part of one of the hydrophobic clusters described above, which are thought to be important to fibril stability. If alanine is replaced by the amino acid threonine at this position, the hydrophobic cluster could be weakened because threonine is more polar than alanine. As a consequence, fibril stability could be decreased. On the other hand, in the pathogenic mutation alanine is replaced by valine which is even more hydrophobic than alanine. This could enhance the hydrophobic cluster and increase fibril stability further.

Several pathogenic familial mutations affect the residues Glu22 and Asp23, for example the Arctic (E22G), Dutch (E22Q) Italien (E22K) and Iowa (D23N) mutation. These residues are part of a region that forms one of the edges of the protofilament structure. Another distinctive feature between amino acid sequence and structure is not related to a mutation but a difference between human and rat and mouse Aβ. Rat and mice are species that are known not to get AD and in their sequence, His13 is replaced by arginine. As described earlier, His13 forms a salt bridge with Glu11 in the kink region around Tyr10 and could be essential to form this kink. A replacement by arginine at this position could prevent the formation of this kink.

3.3.6 Twisted and double filaments

3.3.7 Image processing

As mentioned above the most abundant fibrils in the sample are the straight filaments with a constant radius of 7 nm. Another, less frequent population that can be observed in the images are fibrils showing a prominent twist by a variation in the radius between 3 nm at the thinnest and 10 nm at their widest part. The third observed species is characterized by a clearly thicker diameter, of around 12 nm, compared to the other two types of fibrils.
110 micrographs with twisted fibrils were found and one exemplary micrograph with twisted fibrils is shown in figure 3.17. While the radius among all of these fibrils is the same, their cross-over distances vary roughly between 45 nm and 65 nm. A variation of the cross-over distance could be found even within the same filaments which could be problematic for 3D reconstruction. However, in a first attempt all twisted fibrils present in the micrographs were picked with *helixboxer* which yielded 226 fibrils in total. From these fibrils, fibril segments are extracted with 90 % overlap. This results in 7633 individual segments, each 224 pixel long. Assuming the helical symmetry is similar to the straight filament the complete data set comprises $\sim$ 16,000 asymmetric units which is only around 5 % of the segments used for the final reconstruction of the straight filament.
In the average power spectrum of all segments a peak at $\sim$4.7 Å is clearly visible and, like the straight filament, some of the fibrils show the β-strand pattern in the raw micrographs, which can be seen in figure 3.17.

According to the observed cross-over distances, the expected angle $\Delta\Phi$ describing the helical symmetry ranges from $-1.3°$ to $-1.9°$ if the fibrils comprise a C_2 symmetry and an axial rise of 4.67 Å. If the twisted filaments have a 2_1 screw symmetry similar to the straight filaments, $\Delta\Phi$ would be expected to be in the range between $-179.45°$ and $-179.05°$ with an axial rise of 2.335 Å. Following from this, symmetries are tested between $-1.1°$ and $2.0°$ as well as $-179.50°$ and $-179.00°$ in steps of $0.1°$ and $0.05°$, respectively. The best symmetry is again determined based on a mean cross-correlation coefficient from projection matching between the fibril segments and the latest reference structure as described in more detail in section 3.3.2. It can be seen that segments in C_2 symmetric maps have a higher mean cross correlation. Visual analysis of the maps also suggests that a C_2 symmetry fits the data better than the previously found approximate

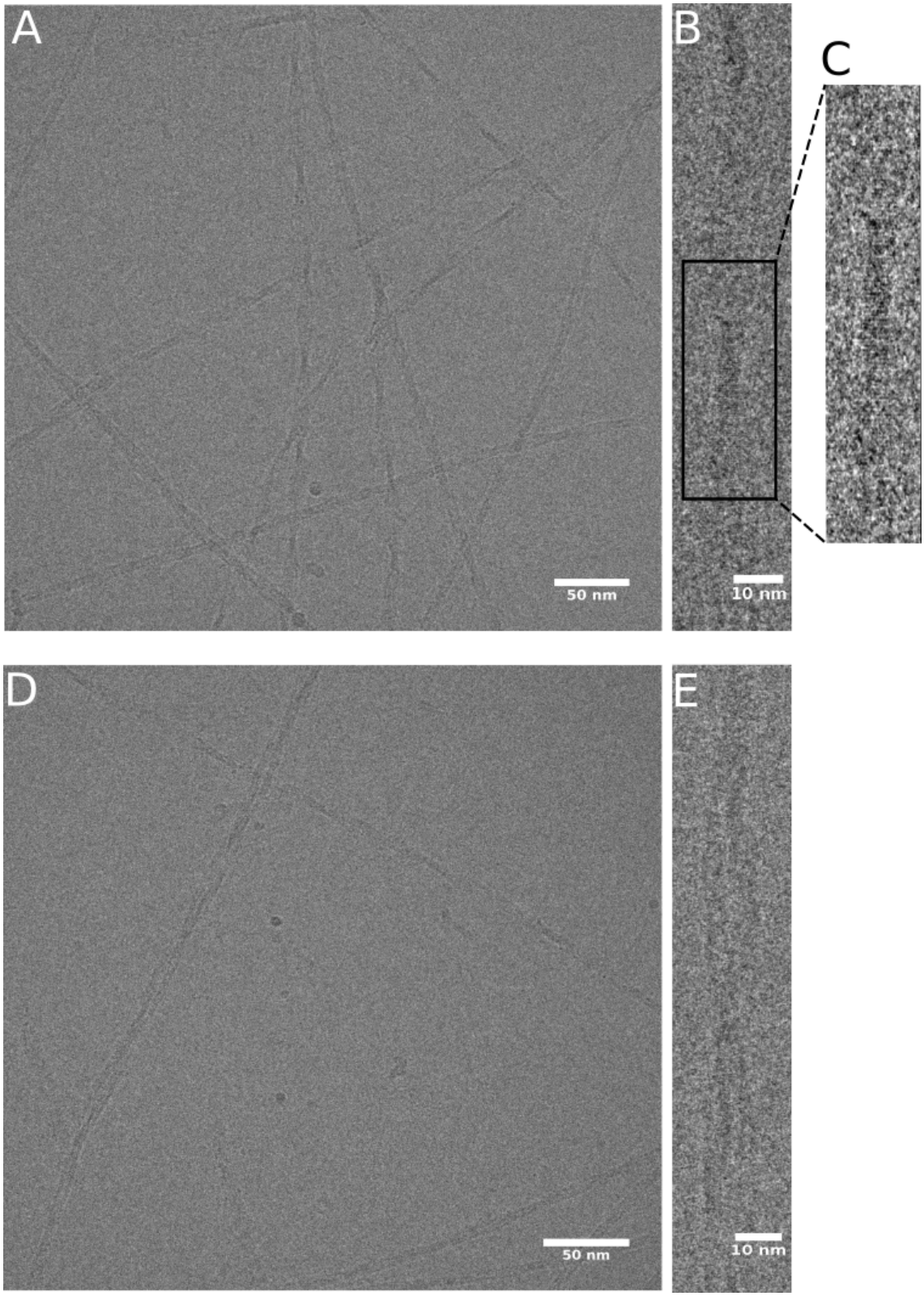

Figure 3.17: Two micrographs containing twisted (**A**) and double fibrils (**D**) and extracted filaments from these micrographs (**B** and **E**).

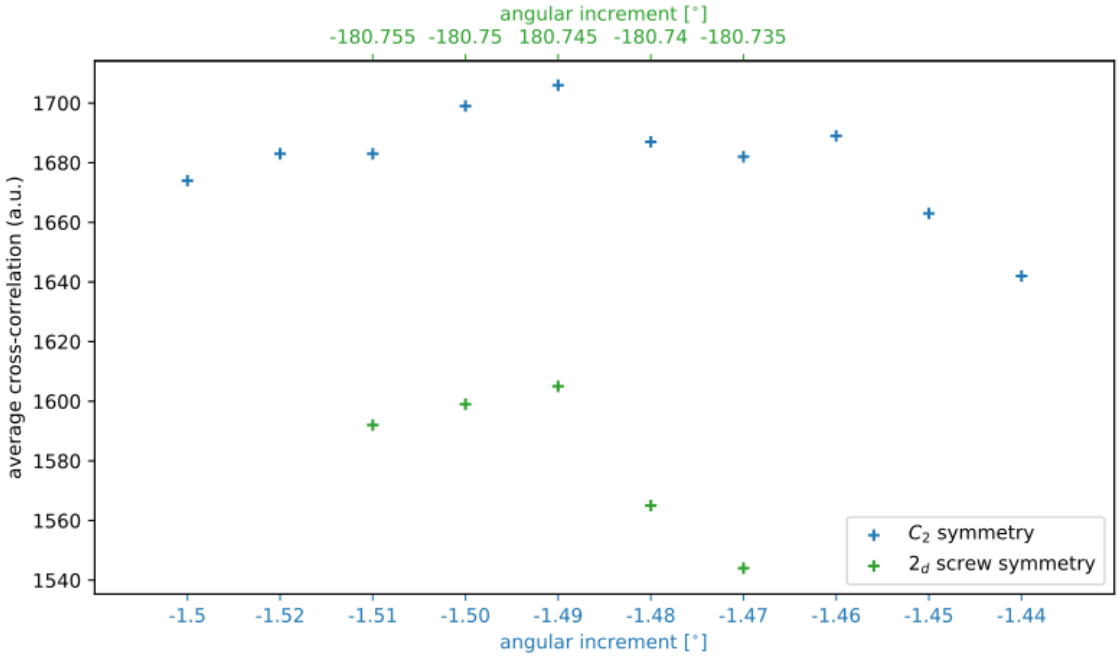

Figure 3.18: Average CC coefficients of all segments used in the reconstruction plotted against the symmetry parameters that were used.

2_1 screw symmetry. The best symmetry combination appears to be $-1.49°/4.67\,Å$, see figure 3.18, which is used for the following reconstructions.

A 3D reconstuction from all fibril segments using these parameters is calculated with *helicon* and *heliconlocal*. It reflects the observation that its cross-section is more elongated than the cross-section of the straight filament, which makes for the prominent twisted appearance in the micrographs. More specifically, the map shows two intertwined protofilaments in which the individual subunits adopt the same "LS"-shape that was described for the previous fibril. The different cross-sections arise from the fact that the interface between the protofilaments is formed by the N-terminal part, that is residues 1-10, in contrast to the straight fibril.

For the straight filament, selecting fibrils based on a cross-correlation coefficient from projection matching (described in section 3.3.2) improved the 3D structure, so the same approach is attempted for the twisted filaments.

It emerges that this is not the best strategy for the twisted filaments because the density does not improve significantly in this way. Thus, the approach is slightly modified: Instead of average cross correlation coefficients per fibril the CC coefficients of the individual segments are plotted against the defocus which is shown in figure 3.19

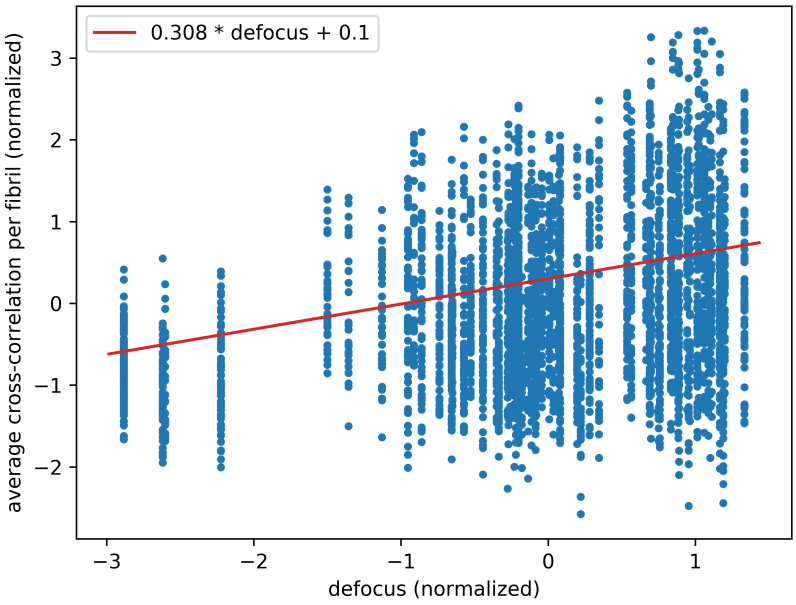

Figure 3.19: CC coefficients of all segments plotted against defocus. All segments above the red line are selected for the final reconstruction.

.

The correlation coefficient between cross correlation and defocus is 0.307. Again, reconstructions with a different number of segments are calculated. The best map was found to be using approximately half the amount of original data, precisely, 2083 segments. They were chosen by selecting all segments that lie above the line defined by $y = 0.307 * \text{defocus}$. The quality of the map was judged by visual inspection. For all reconstructions, the previously determined symmetry parameters $-1.49°/4.67\,Å$ were used. In the final reconstruction β-strands are not completely separated but still visible in the map.

The double fibrils were processed using the helical reconstruction algorithm implemented in Relion [43]. An exemplary micrograph showing such a polymorph is presented in figure 3.17. From 29 micrographs in total fibrils are selected. Individual segments have a size of 272 x 272 pixel with an overlap of 252 pixel. The image stack comprises 3357 segments in total. In refinement local search for symmetry is used to find the best parameter for the angular rise. It is determined to be $-179.527°$. In a later iteration of refinement a cut-off according to the Loglikelihood value was chosen to use only the best half of the segments for the final reconstruction.

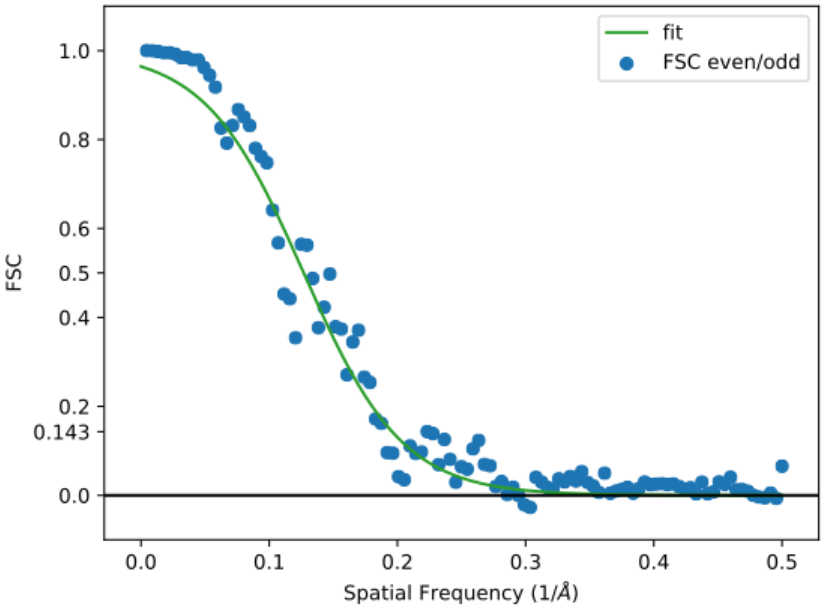

Figure 3.20: FSC calculation. The FSC between odd and even half maps and the function $1/(1 + exp(s - \beta)/\alpha)$ fitted to the curve.

3.3.8 FSC calculation and model building

For the twisted filament the FSC is calculated between half maps as described in Section 3.3.3 for the straight filament. A function of the form $1/(1 + exp((s - \beta)/\alpha))$ is fitted to the data and the resolution is estimated to 7.5 Å according to the 0.5 criterion. This resolution appears contradictory to the fact that β-strands are appearing in the map. But it could be possible that resolution along the helical axis is slightly better than in the other directions. As mentioned in the beginning a variation in the cross-over length was observed in the fibrils. Although segments were sorted, different twists could still be present in the data set and smear out the map in x-y direction. The FSC curve is shown in figure 3.20.

For model building an "LS"-shaped subunit of the straight fibril is used in a local refinement docking search in Rosetta [57] to determine the correct N-terminal interface. The docking was done at pH 2 and Rosetta determined the residue Asp7 to be protonated. For model building, a fragment of the straight atomic model is optimized into the twisted fibril density using DireX [86]. Subsequently, a single subunit is cut out from the model and the previously determined symmetry is applied to it to create a larger model segment with the correct symmetry. This fragment is minimized in the

density with NCS and Ramachandran restraints applied, again using DireX.
For the double filament the segments were split in half before starting the refinement and refined separately for FSC calculation. The resolution of the final reconstruction is estimated to 25 Å. The model was also optimized using Rosetta.

3.3.9 Discussion

The density map and the atomic model of the twisted fibril, shown in figure 3.21, reveal that the straight and twisted filament share the same "LS" subunit shape while the interaction between the subunits is different. In contrast to the straight fibril, the twisted fibril was found to be C_2 symmetric and the rise of each subunit from the N- to C-terminus is less pronounced for the twisted fibril, see figure 3.21. For the twisted fibrils, the protofilament interface is formed by the N-terminal parts as presented in figure 3.22 A. The interface, which was determined with Rosetta, is stabilized by hydrogen bonds between protonated asparagines at position 7. NMR experiments of the sample showed that amino acids Asp1, Arg5 and Asp7 are deprotonated, but it should be noted that the twisted fibrils only exist in small amounts compared to the straight fibril. Thus, it could be that a small portion of Asp7 could be deprotonated but not show up in the NMR data. Considering the pH of the sample, protonation of these amino acids could be possible.
A similar N-terminal interface is also possible when the mentioned amino acids are fully deprotonated. In the straight fibrils salt bridges were described between Arg5 and Asp7. For the twisted fibril a small shift of the protofilaments against each other would allow salt bridges of the same amino acids between opposing subunits. Currently, the resolution of the map is not high enough to discriminate between these binding options and more calculations need to be done to determine the interface exactly.
This N-terminal interface described for the twisted fibril would mean that for the same monomer conformation as in the straight fibril, the C-terminus would be exposed to the solvent. This would be unexpected as C-termini are hydrophobic and in the straight fibril their hydrophobic interaction stabilizes the protofilament interface. However, the density map of the twisted fibril is not unambiguous at the C-terminus and it would also be possible that Gly29 performs a slightly different kink that finally hides the hydrophobic C-terminus in the core of the fibril. Starting from Gly29 the amino acid chain could go along the density in the reverse direction and Ala42 would come close to Gly29 from the core of the protein.

The cross-section of the double fibril density map fits four individual subunits, which indicates that this fibril is composed of four protofilaments. This map could only be determined to low resolution, but it suggests that two straight fibrils are twisting around each other and the interaction between them is between the N-terminal parts of one protofilament each, see fig 3.22 B. In the atomic model for the double fibril, Arg5 from both protofilaments are opposing each other. This is not a likely interface because two positive charges would be directly opposite two each other. Thus, it is more plausible that in the actual interface the fibrils are slightly shifted against each other leading to the same interface that is present in the twisted fibril. On the basis of the density map, the interface cannot be determined reliably and more calculations are needed to describe it precisely.

3.3.10 Conclusion

In this Chapter it was explained how $A\beta_{1-42}$ fibrils were investigated with cryo-EM. Three different polymorphs were observed in the sample and their appearance differs in the diameter that was found to be around 70, 100 and 120 Å. The 100 Å fibril shows a distinctive twist emerging from its elongated cross-section. The most frequent polymorph is the thin, straight filament and a high resolution (4 Å) density map could be reconstructed from the images using an IHRSR approach. The complete backbone and almost all side chains are visible and ordered in the density map. With the high resolution map an atomic model could be build. Both model and density map show that the straight fibril consists of two protofilaments with parallel, in-register β-sheets. In contrast to previously published ss-NMR models the N-terminus of this model is fully ordered and part of the cross-β-core. Individual subunits in the model adopt a "LS"-shape and it can be seen that subunits are rising from the N- to C-terminus, which has not been observed before. This atomic model can help to understand differences in familial mutations as well and the mechanism of fibril growth and might be a starting point to develop ways to inhibit fibril formation and growth.
Models for the two other polymorphs could not be build with atomic details because the resolutions of the reconstructed maps are not high enough. However, it can be seen that the twisted fibril consists of similar shaped "LS" subunits with their interface formed by the N-terminal parts of the monomers. This is an unexpected finding, as until now the N-termini were thought to be mostly unordered.
A third polymorph was found in the sample and its reconstructed density map suggests

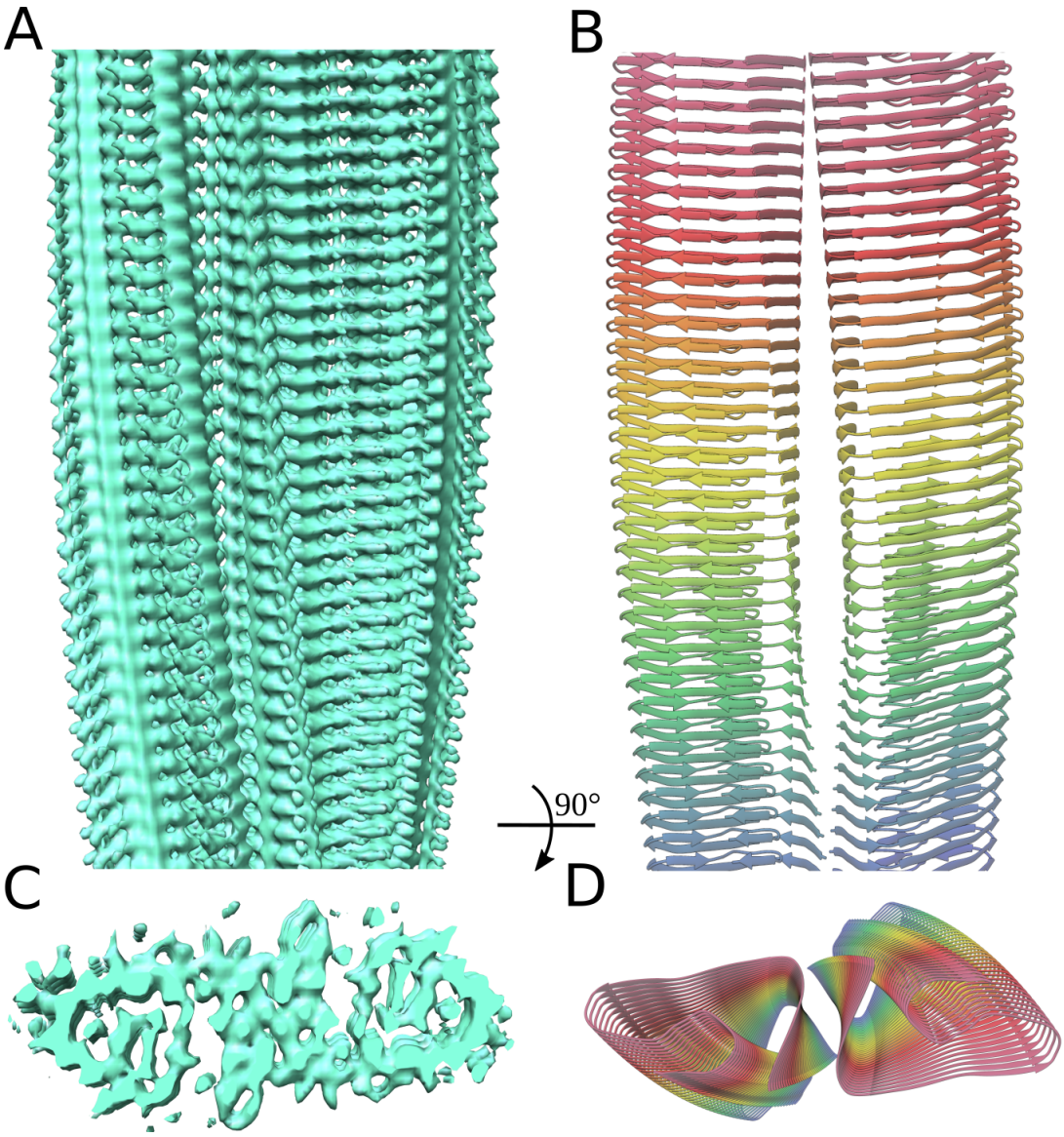

Figure 3.21: (**A** and **C**) Density map of the twisted fibril from the side and the top.(**B** and **D**) Model of the twisted fibril.

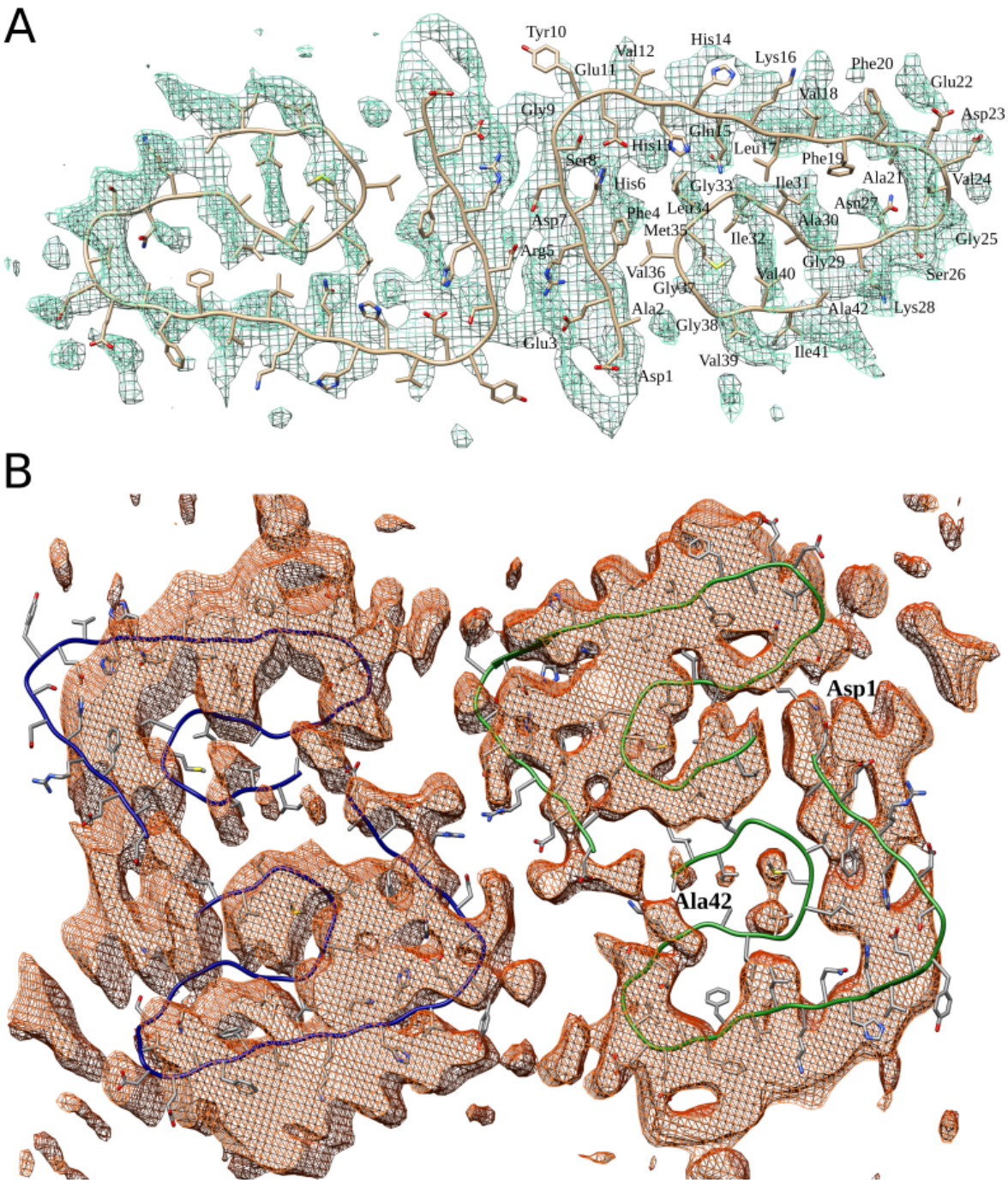

Figure 3.22: Cross-section of density and model of the twisted (**A**) and double (**B**) fibril.

that it is made out of two straight fibrils twisting around each other. Surprisingly, their interface is again formed by the N-termini. The resolution of the map is not high enough to identify the atomic interactions in the interface but it is suspected that it could be the same as for the twisted fibril. More calculations are needed for both interfaces to determine them reliably at atomic resolution.

4 Aβ oligomers

New insights from the field of AD research point towards the fact that small oligomers are the most toxic species of amyloid β aggregates. However, investigation of oligomeric species is difficult, because of their heterogeneous and unstable nature and not much is known about their structure. The following chapter aims to gain structural insight into cross-linked Aβ oligomers with TEM.

4.1 Oligomer production

For structural analysis the investigated molecules need to be stable and in this case the oligomers are cross-linked for this purpose. The oligomers were produced by Katja Kuehlbach and a more detailed description of the production protocol can be found in (Kuehlbach 2014). $A\beta_{1-42}$ molecules oligomerize in PBS for 45 min. The aggregates vary in size so they are size separated via centrifugation in a density gradient and stabilized via cross links with EDC and NHS. For cross linking EDC and NHS are used. The density material is saccharose in concentrations between 5 and 50 % (w/v). Preincubated A β is added to the 5 % saccharose fraction and EDC and NHS are added to the 10 - 50 % saccharose fractions. After centrifugation cross linking is stopped by adding ethanolamin hydrochloride. Everything besides the Aβ oligomers is removed from the sample and the fraction containing the smallest aggregates is used for TEM.

4.2 Image processing

Particle picking and CTF correction

Figure 4.1 shows an exemplary micrograph. The images contain mostly small round shaped particles with a diameter of roughly 70 Å but the images also contain a few

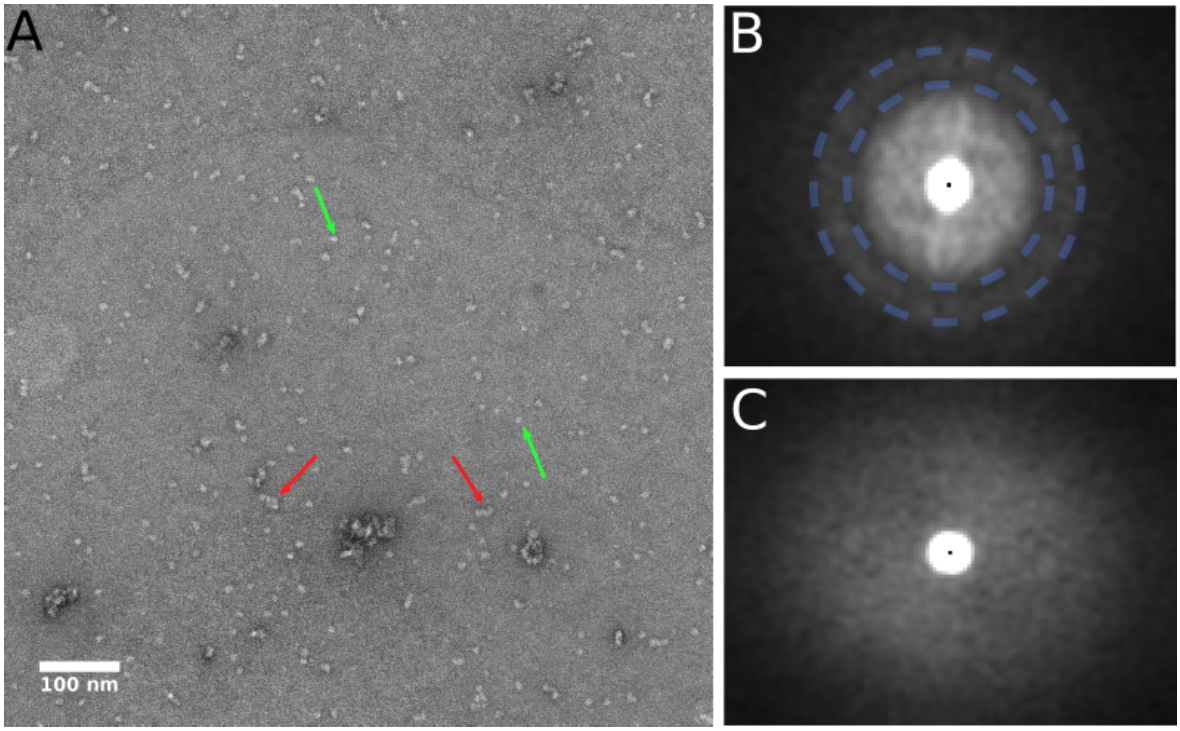

Figure 4.1: (**A**) An exemplary micrograph. The images contain mostly small oligomers indicated by the green arrows but also some larger aggregates, marked with red arrows. (**B** and **C**) Powerspectra of two micrographs. In (**B**) Thon rings are marked by blue rings.

aggregates that are much larger, see figure 4.1. All image processing was done in EMAN2. From the 328 micrographs that were taken particles were picked with a box size of 88 x 88 pixels. Particles were picked manually to make sure that only small particles with approximately the same size were chosen. In total 10592 particles were picked.

As the images were not taken in focus but an underfocus of approximately 0.5 µm they need to be CTF corrected which in done with the program *e2ctf*. First, an automated fitting procedure is carried out for all micrographs followed by a manual inspection of the determined defocus values. Some of the micrographs do not show Thon rings and these images were probably taken in focus as the requested underfocus was small. For these images the fitting fails and so their defocus value is manually set to 0 µm. Subsequently all individual particles are combined in a particle stack.

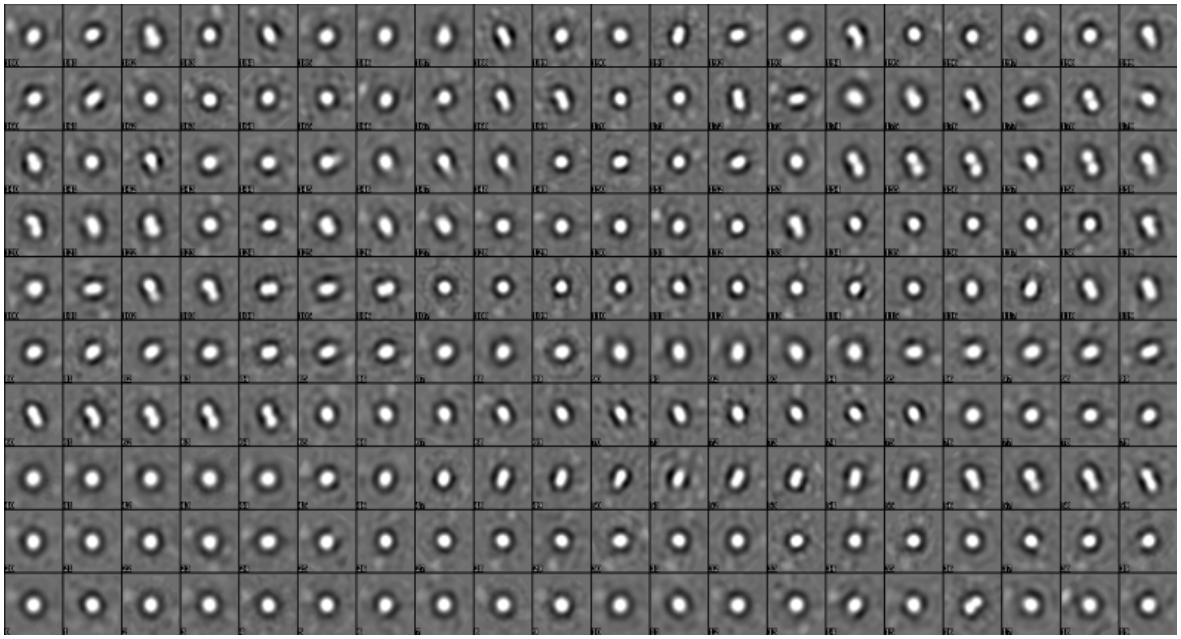

Figure 4.2: 2D class averages of the particles.

2-D class averages

This particle stack is used for generation of 2-D class averages. Similar particles are aligned and averaged together so the resulting class averages have a better SNR than the original particles and the overall shape of the particles appears much clearer. The class averages are produced with the program e2refine2d and the 10592 particles are clustered in 200 classes.

All classes are shown in figure 4.2. Most of the class averages, around 60 %, have a round shape with an approximate size of 70 to 100 Å but also also rod like and dumbbell shaped averages can be seen.

These shapes can be different views from one type of particle but also different views from different particles. As discussed later the different shapes do not seem to origin from one common structure and it seems to be more likely that the different shapes belong to different structures possibly a spherical a rod- and a dumbbell like shape. The round class averages could of course be top views from the other to shapes but as their are by far the most frequently occurring shape in the class averages they probably also stem from a distinct subset of particles.

Determining the exact size of the class averages is difficult because the images are still blurry but there appears to be a difference in the size of the particles in the sample. For better comprehension three representative classes of each mentioned shaped are

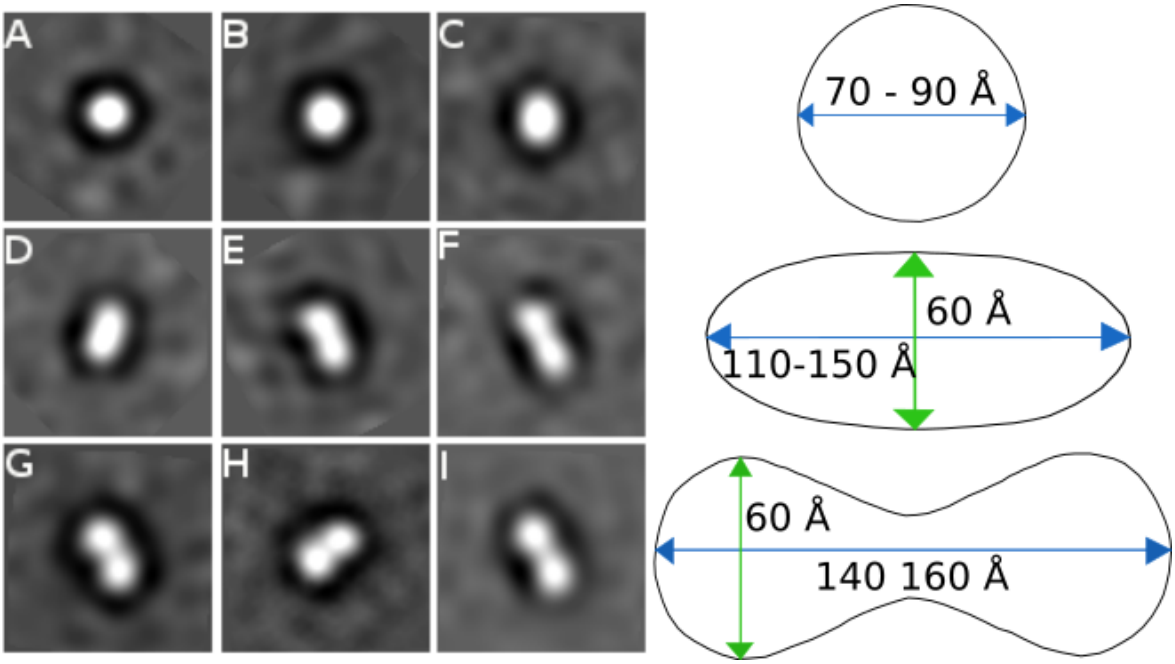

Figure 4.3: (**A**-**I**) Representative shapes of the 2D class averages. (**J**-**L**) Schematic drawings of these shapes.

enlarged in figure 4.3. The difference in size can especially be seen for the rod like particles which also express a great variety of the ratio between their length and width. The shortest and most narrow ones are approximately 110 Å long and 60 Å thick while the largest are up to 150 Å long with the same thickness. A wide range of values for the length of these particles can be found in between. Also, the spherical and dumbbell shaped particles vary in size although it is more subtle. As mentioned earlier the spherical particles have a diameter of about 70 to 90 Å. The dumbbell particles are about 140 to 160 Å long and approximately 60 Å60 thick, at their widest part. This is also demonstrated in figure 4.3.
The particle variation in size may come from the fact that during production of the oligomers aggregates consisting of a different number of monomers formed and they could not be perfectly separated during centrifugation. The potential structural variability needs to be examined further.

Initial models

Besides gaining more information about the particle structure the 2-D class averages are also used to generate an initial model that is needed for the iterative 3-D refinement. The best possible way to this is to select as many different views of the particles as

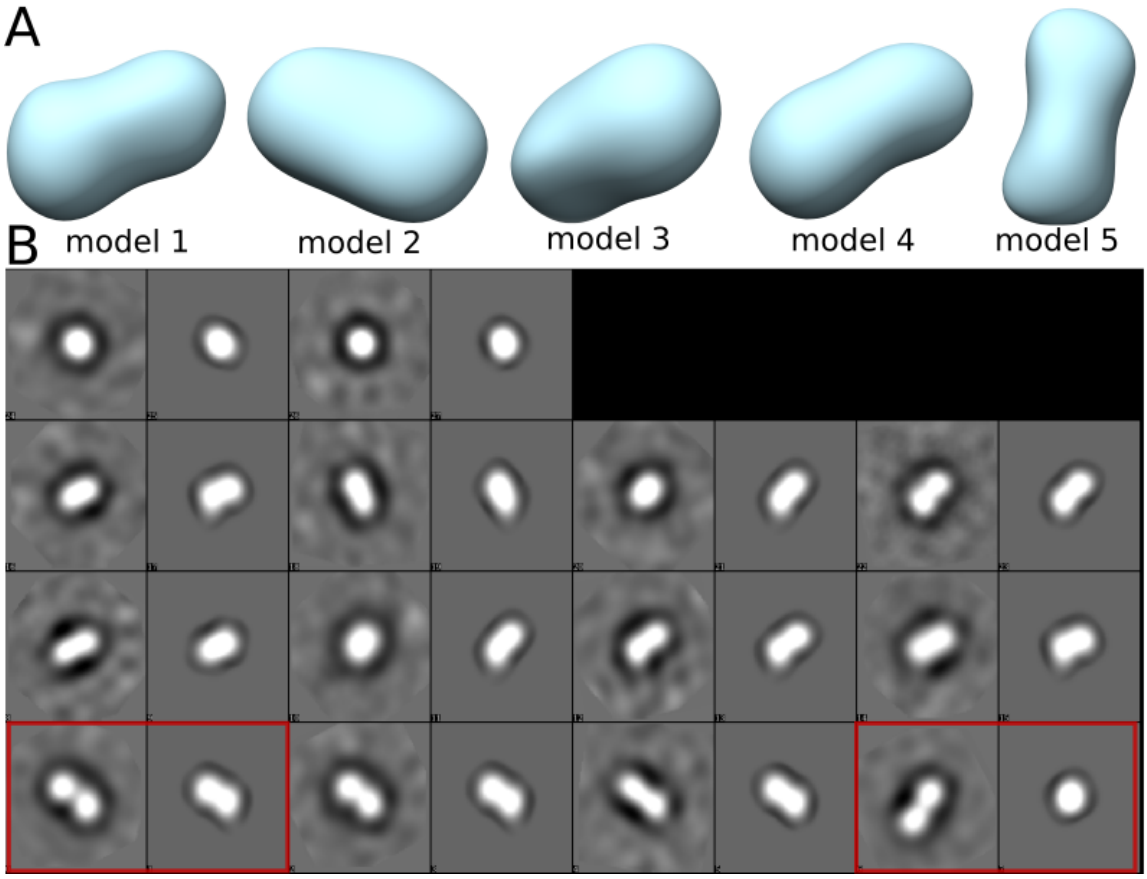

Figure 4.4: (**A**) Different initial models. (**B**) Comparison between selected 2D class averages and backprojections of model 1. Agreement between them is poor, especially for the red marked class averages.

possible from the class averages without including views from a different structural composition of the particles. In this case it is not clear which class averages come from the same structure so instead just a great variety of different views is chosen as input for the generation of an initial model. 15 initial models are generated with *e2initialmodel* and a selection of them is shown in figure 4.4. All models have the same size of about 130 Å in length and 80 Å in diameter and also their shapes look very similar. A good initial model should reflect the features that are also visible in the 2-D class averages. For all the models that were generated the dumbbell feature is underrepresented which can also be seen in the comparison between the selected class averages and the corresponding projections of an initial model, see figure 4.4. The similarity between the class averages and the projections is poor for especially these kind of shapes. Another run of *e2initialmodel* did not generate different models with better agreement suggesting some sort of heterogeneity in the sample.

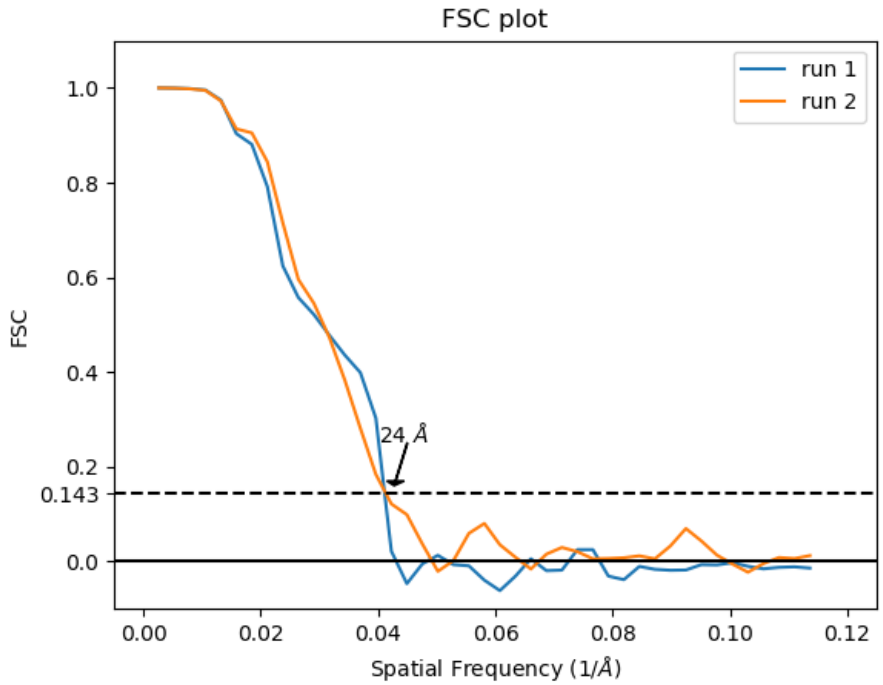

Figure 4.5: FSC curves of two independent refinement runs.

3-D refinement

Since an initial model is needed for 3-D refinement two models with the best agreement of projections are chosen to be the initial model of two independent 3-D refinement runs. Their final 3-D maps are shown in figure 4.6. The maps look similar to each other but different to what one would expect from the 2-D class averages and initial models. The reconstructed particles are approximately 110 Å long and have a diameter of 80 Å that slightly narrows at both ends. The map does not clearly show any of shapes that were present in the 2D classes, e.g. round, rod- or dumbbell shapes but looks more like an average of round and rod-likes shapes.
The backprojections of the 3D reconstruction agree well with the classes used for the reconstruction but different from the 2D classes generated for assessment of the sample. The FSC in EMAN2 is computed according to the "gold standard way" meaning that at the beginning of the refinement the particle data is split into halves (even/odd) and to make the starting models different from each other they are perturbed independently. According to the 0.143 criterion the resolution of the final map for both refinement runs is about 25 Å see fig 4.5. Additional 3-D refinements with different starting models confirm the described maps and indicate that the obtained 3D reconstructions do not

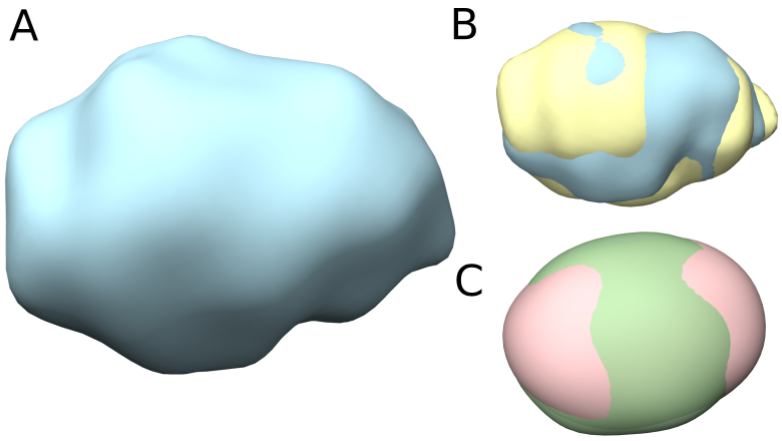

Figure 4.6: (**A**) Final reconstruction from refinement run 1. (**B**) Comparison between reconstructed maps from run 1 and 2. (**C**) Results from multi-modal refinement with two starting models.

have an initial model bias.

Multimodal refinement

The deviation between 2-D class averages, initial models and refined maps suggest that the particle population in the sample is heterogeneous and can not converge to a single solution. A multimodal refinement process, implemented in the program *e2refinemulti* was used to assess this. A normal refinement starts with the projections from one initial model to determine the particle orientations. For the multimodel refinement N starting models are needed and their projections are used to determine the particle orientations and the model each particle matches best followed by a normal refinement. Subsequently one needs N starting models for the multimodal refinement which should not be identical. For this specific case the initial model that has been used before in refinement is perturbed by adding different noise to it and independent multimodal refinement runs with two and three starting models, respectively, were carried out. Figure 4.6 shows the results from the two-model refinement. Surprisingly, the models look very similar to each other and also similar to the maps obtained from the normal refinement besides being smoother. This may come from the fact that less particles contribute to each produced map and different filtering compared to the normal refinement. Also in the multimodal refinement with three staring models the apparent heterogeneity could not be retrieved. It is possible that a variability in the particles is not detectable as they do not show many different features. The relatively small amount of particles may contribute to the fact that different structures can not be

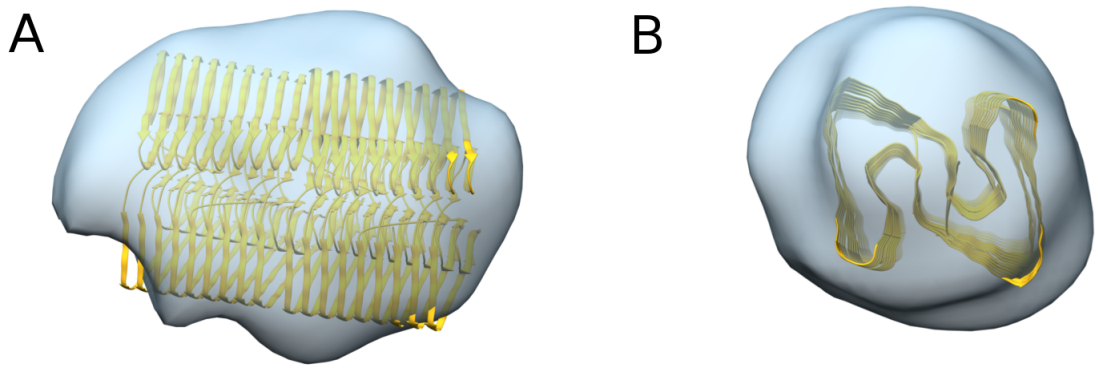

Figure 4.7: Final map from refinement 1 containing 36 subunits from the straight fibril model, from the side (**A**) and the top (**B**). The model of the fibril is superimposed only to get a better impression of size, it is not implied that the oligomer structure looks like this fibril model.

extracted from the particles.

Although the multimodal refinement did not give meaningful results, it is still assumed that the particles have some sort of heterogeneity, which is supported by the 2-D class averages. The reconstructions from the normal refinement, in this case, would be a mix of all the different structures present in the sample. From looking at the 2-D class averages theses different structures could be spherical particles, more elongated rod-like particles and particles with a dumbbell shape. For the latter it would be conceivable that these particles are made out of two smaller spherical particles as these particles have approximately twice the size of small, round particles, see figure 4.3 G-I. As mentioned earlier, the particles have probably some variability in size but the refined maps can still be used to estimate the number of monomers in an oligomers. For this purpose a fragment of the amyloid β fibril was fitted into one of the maps of the normal 3-D refinement. The fragment, consisting of 36 monomers in total, loosely fits into the density, see figure 4.7. From that, one can conclude that the oligomers most likely consist of 40 to 50 monomers.

4.3 Conclusion

In total 1059 particles were picked from 300 micrographs. After CTF estimation the particles were classified into 2-D classes. These classes have a better SNR than the particles in the raw images and their different features can be seen more clearly. 3 main shapes are present in the class averages: Spherical particles, elliptical and to

propapbly two small spherical particles bound to each other, see figure 4.3. These shapes are present in different sizes roughly varying between 70 Å and 140 Å.
Independent 3-D refinements with different starting models were carried out. They gave similar final maps at a resolution of about 24 Å. However the reconstructed maps differed from what can be seen in the class averages, so it can be assumed that the particles are heterogeneous and the reconstructed maps are a mix of the actual oligomer structures. Consequently, a multimodel refinement was attempted but the obtained reconstruction were not meaningful. The reason for that is most likely the small number of particles and the fact that the variability is not detectable in the data because the particles do not show sufficient distinctive features. Considering the 2-D class averages, the particles seem to exist in different structures and with more particles and a sample with less variability in oligomer size it would probably be possible to obtain higher resolution reconstructions and learn more about Aβ oligomeric structures and aggregation.

5 Determination of ligand binding sites using unassigned NMR titration spectra

5.1 Introduction

Interactions between proteins and small molecules are an important aspect in drug design and discovery and NMR spectroscopy is a powerful tool to study protein structures as well as dynamics and interactions [65],[58]. This wide range of different applications was made possible by advances in different sub-disciplines of NMR spectroscopy such as isotope-labeling [106], rapid data collection [84] and automation of different aspects in the experiments [27][96]. Furthermore new experimental strategies have been designed for the different application possibilities of NMR. A prominent example in the field of high affinity ligand binding is the so called SAR by NMR technique [89]. This technique is a combination of finding ligand binding fragments and combining them to produce ligands with extremely high affinity. Binding is monitored by detecting chemical shift changes in heteronuclear single-quantum correlation (HSQC) spectra during successive addition of ligand to the protein. In fact, this is one of the most widely used techniques for studying protein-protein or protein-ligand interaction [95] and for this reason different automated procedures for analyzing chemical shift perturbation (CSP) experiments have been described before.
The software package MUNIN [16] incorporates an automated analysis of large sets of HSQC spectra to detect binding ligands. The principle of this method is based on comparing NMR spectra recorded with only the protein and spectra where a mixture of potentially binding ligands were added to the protein. An automated procedure detects spectra with differences in the peak positions and thus binding ligands. Peak picking is not required for this type of analysis, because it only analyzes whether the

molecules are binding. To additionally locate the site of a binding ligand, amide signals need to be assigned and the 3D structure of the receptor has to be known.
A different approach to analyze HSQC protein-ligand spectra was introduced by Peng *et al* [73]. They focus on the task of peak tracing between the non-perturbed, protein only and the perturbed, protein and ligand spectrum and present an algorithm to find the best matching assignment. Their fundamental assumption is that for correct assigned peaks the overall chemical shift difference between these spectra should be minimal.
Also in the software package NMR ViewJ, which is used for visualization and analysis of macromolecular NMR spectra, a new feature for automated analysis of perturbation spectra was introduced [37]. Clearly, perturbation spectra are an important part of studying protein-ligand interaction and a variety of different tools to help analyzing these type of spectra has been developed.
New procedures were also created in the design of experiments that investigate protein-ligand interactions to facilitate data analysis. One example is the mapping method described by Reese *et al* [80]. Protein interfaces are mapped without peak assignment but using a series of spectra from a specific labeling scheme. Multiple spectra are recorded in which one binding partner is differently labeled and the other binding partner remains unlabeled. The spectra of the complex and the free protein spectra can be compared to identify the number of a certain type of amino acids being part of the interface. This information can be combined with the protein structure to reveal the binding interface.

From the presented approaches it can be seen that determining the binding site of a ligand peak assignment or more complex experiment design is required. Protein resonance assignment is a time-consuming step and usually several 2D and 3D spectra have to be analyzed [54]. Although several methods for automated peak assignment exist, [76][54] skipping this step would facilitate and accelerate the analysis of titration spectra.

Here, a method to identify the ligand binding site is presented that exclusively uses a protein-ligand HSQC titration spectrum and the protein structure. It is based on a quantitative comparison of different predicted and binding site specific spectra with a HSQC spectrum of the protein alone.
A similar scoring method was introduced by Courtney *et al* [13] to verify protein structures. From the protein which structure is to be determined, a single, unassigned

2D spectrum is recorded. A series of models is generated and a 2D spectrum is predicted for each of them. To identify the correct model the predicted and the recorded spectra are compared and scored using a modified Hausdorff distance [22]. The below presented method to identify binding sites uses a similar scoring method to efficiently predict ligand binding sites.

5.2 Method

The general idea of the method is illustrated in Figure 5.1.

To start, a titration spectrum of the protein and the ligand as well as the structure of the protein is needed. The basic assumption of the method is that peaks from the titration spectrum with a large chemical shift perturbation (CSP) are in the vicinity of the ligand binding site. While it is true that large CSP occur at the binding site there can be other reasons that induce large CSPs at other locations of the protein, for example a conformational change of a domain. This distorts the results and is a limitation of the method.
The prediction of the ligand binding site relies on comparing a large amount of potential binding sites with the experimental NMR titration spectra and find the best matching position based on a distance depending score.

The first step is to modify the experimental NMR titration spectrum. As the method uses peaks with large CSP, those peaks should be extracted from the spectrum. In a titration spectrum the peaks can be traced from lowest to highest ligand concentration and the distance between them is calculated for every peak. Here, a peak is considered to have a high CSP, if its CSP is 2.5 times the average CSP of the spectrum. This value was determined experimentally and was chosen because it proved to be a reasonable threshold for different protein-ligand systems. However, there still might be peaks in there, that are not associated with the binding site and to emphasize the peaks with the largest CSP, all selected peaks are assigned a weighting factor that is the distance between the peaks in the lowest and highest ligand concentration spectra. The peaks from the spectrum without ligand and a significant CSP are stored separately and referred to as the "reduced spectrum".
The next step is to produce a large number of potential binding sites. Different methods for efficient binding site generation have been tested. One possibility is to use a docking program such as AutoDock4 [62]. Another possibility is to use the coordinates of the

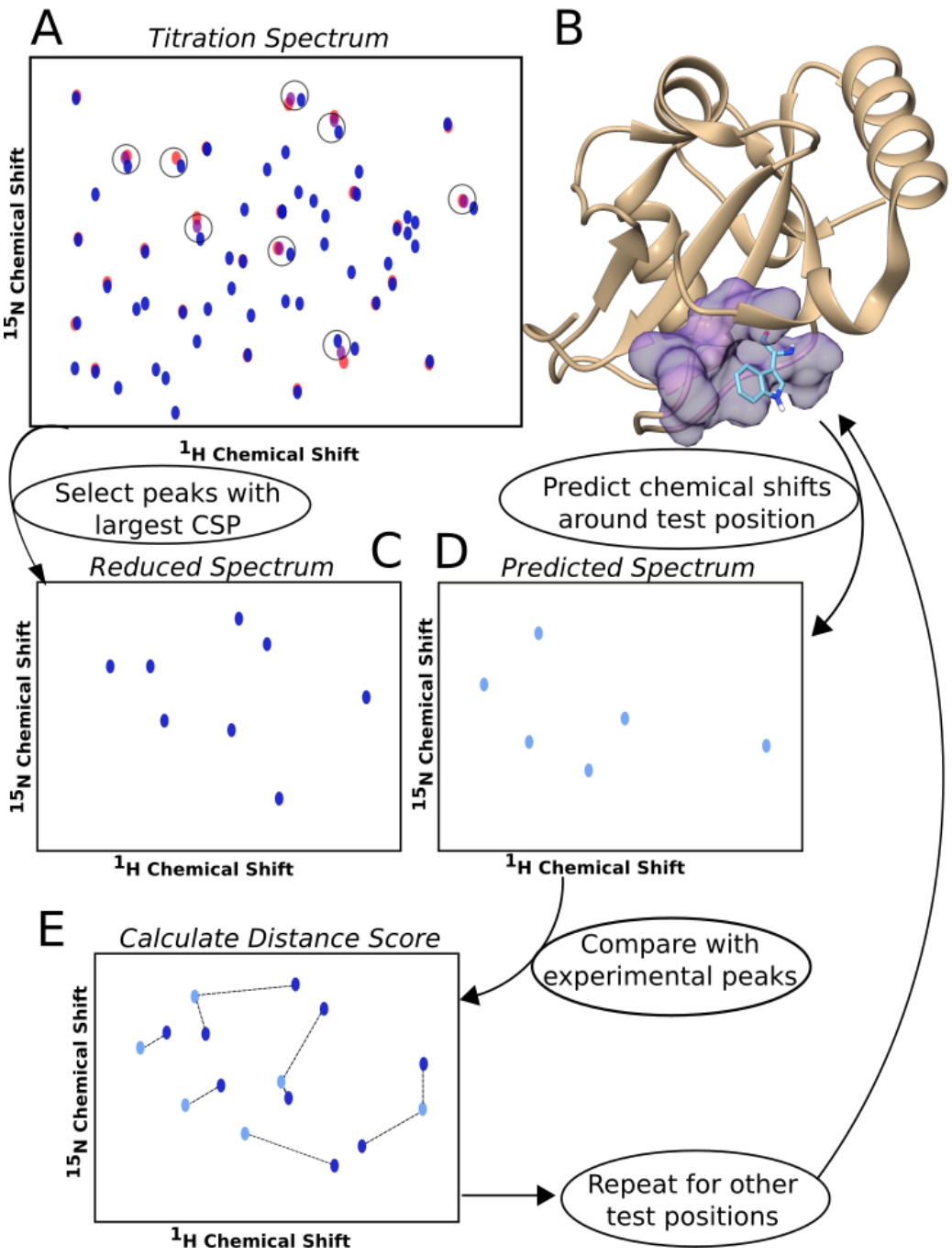

Figure 5.1: A flowchart of the general procedure of the method: Peaks with large CSP are extracted from the titration spectrum (**A**-**B**). Predict chemical shift values for each potential binding position (**C**-**D**). Calculate distance-score between nearest peaks (**E**).

receptor atoms as binding poses and a similar method is to place a regularly spaced 3D grid on the protein and consider every grid point to be a potential binding site.
For comparison a predicted NMR spectrum is needed and the program SHIFTX2 is used to predict a ^{1}H-^{15}N spectrum of the protein. Every binding position needs to be assigned the corresponding values from the predicted NMR spectrum. To achieve that, all atomic coordinates that belong to a binding position are compared with the coordinates of the protein structure. For all protein coordinates that are found to be closer than 7 Å to any coordinate of a potential binding position, the corresponding chemical shift values are assigned to these binding positions and stored in a separate file. At the end of this step, a file with chemical shift values exists for all potential binding sites, here called "predicted spectra". They can be compared with the reduced spectrum. The distance threshold of 7 Å was chosen as it is assumed that receptor atoms effected by ligand binding are in a vicinity of around 7 Å from the ligand. For the different test cases, this proved to be a reasonable value. If the grid method is used to mimic binding sites, this value should be modified to match the ligand size.

After the reduced spectrum and the spectra of the potential binding sites have been generated, they can be compared to find the best matching position. For every peak in the reduced spectrum the closest peaks in the predicted spectra are determined. The spectra do not necessarily have the same number of peaks and double assignments are possible. A distance score is calculated according to

$$d\,score = \sum_{i=1}^{N} \frac{1}{1 + d_{min,i}} W_i, \tag{5.1}$$

with N being the number of peaks in the reduced spectrum, d_{min} the minimal distance found and W_i the weighting factor.
In common HSQC spectra, the range for ^{15}N chemical shift values is usually much larger than the range of ^{1}H values. To compensate for this in the distance score calculation, the chemical shift values are scaled by their standard deviation, which is calculated from all ^{1}H and ^{15}N values in the reduced and predicted spectrum.
As can be seen from formula 5.1, the distance score indirectly depends on the number of peaks in the predicted spectra because, in general, the more peaks are present in a spectrum the smaller distances can be found. To eliminate this effect, the distance score for 1000 spectra with random values for the chemical shift is calculated. The random values are uniformly distributed in the interval $[5.0 - 11.0)$ for proton and $[105.0 - 135.5)$ for nitrogen chemical shifts and the number of peaks is the same as

in the reduced spectrum. The Distance score is standardized with the mean distance score and its standard deviation of the random spectra:

$$Z_{score} = \frac{d\,score - \bar{d}\,score_{random}}{\sigma_{\bar{d}\,score_{random}}} \tag{5.2}$$

.

5.3 Results and Discussion

5.3.1 Simulated data

For a proof of principle the method was tested using simulated test systems. Five protein-ligand complexes with known structure and binding site were selected from the Protein Data Bank (PDB) and applied to the program. The test cases are glutamate mutase complexed with vitamin B12 nucleotide (*PDB code 1ID8*), influenza AM2-BM2 chimeric channel bound to rimatidine (*PDB code 2LCJ*), bacillus cereus metallo-Beta-lactamase BcII with R-thiomandelic acid (*PDB code 2M5D*) and monovalent lectin microvirin with the ligand man(Alpha)(1-2)man (*PDB code 2YHH*). The structures of all these complexes have been determined with solution NMR. The last simulated test case is the GABA(A)receptor-associated protein (GABARAP) (*PDB code 3D32*) with tryptophan. The reason this complex was chosen, is that for GABARAP and tryptophan, HSQC titration spectra also exist which can be used as a direct comparison between applicability of the method on simulated and experimental data.

Docking simulations were carried out with AutoDock4 and 1000 positions were generated to be tested for every system. The protein structures of the complexes were used to generate the position specific predicted spectra as described in section 5.2. The reduced spectrum was also derived from the predicted spectrum. To mimic experimental conditions as close as possible, the reduced spectrum was generated in the following way: Chemical shift values for the residues within a radius of 7 Å were selected. In a real experiment it can be assumed that not all of the residues in the reduced spectrum are in the vicinity of the binding site or that residues close to the binding site are not part of the reduced spectrum because their CSP is below the threshold. To account for this, for every test case 1– 4 peaks of the chemical shift values in the reduced spectrum are exchanged with chemical shift values that are not connected with the binding

site. As the predicted NMR spectra cannot be assumed to be perfectly identical with an experimental NMR spectrum, the reduced spectrum is slightly modified in another way.
The accuracy of SHIFTX2 is described by correlation coefficients between predicted and experimental spectra and the correlation coefficients for ^{15}N and ^{1}H are 0.9800 and 0.9714 with RMS errors of 1.1169 and 0.1711, respectively [41]. Thus, a random number was added to each chemical shift value. It is Gaussian distributed with 0 mean and standard deviation of 0.9714 for nitrogen chemical shifts and 0.1711 for hydrogen chemical shifts. The weighting factor introduced in equation 5.2 was set to one for the simulated test cases. Afterwards, Z-scores were calculated following the protocol described in section 5.2 for the AutoDock4 docking positions, the receptor atoms and 3D grids. The Z-score is plotted against the center-of-mass distance between the ligand and the tested binding positions. A decrease of the Z-score with larger distances to the ligand is found for all complexes, see figure 5.2. The results are also summarized in Table 5.3.1.
For better illustration of the results the grids were used to produce density maps that are scaled with the calculated Z-scores. This map can be looked at together with the receptor structure and the region with the highest Z-scores can be easily identified. This was done for all the test cases and is shown in figure 5.3. It can be seen that the correct ligand position lies in the region with highest Z-scores or in its direct vicinity.
In the Z-score plot in figure 5.2, one can see that for the AM2-BM2 channel using the receptor atoms as binding sites the resulting highest Z-scores are in a region that is slightly shifted from the correct position. But in the structure plot 5.3 the highlighted region is at the correct binding site. This supports the theoretical consideration that the grid method is better suited to find binding positions. The reason is that with this method it is also possible to test positions inside of channels or pockets on the receptor, which are common places for ligands to bind.
For all tested systems the highest Z-score gives a position that has maximal RMSD to the correct binding site of 5.2 Å.

The simulated test cases showed that the method is in general able to locate the binding site of a protein-ligand complex correctly and in the next section it is applied to experimental titration spectra.

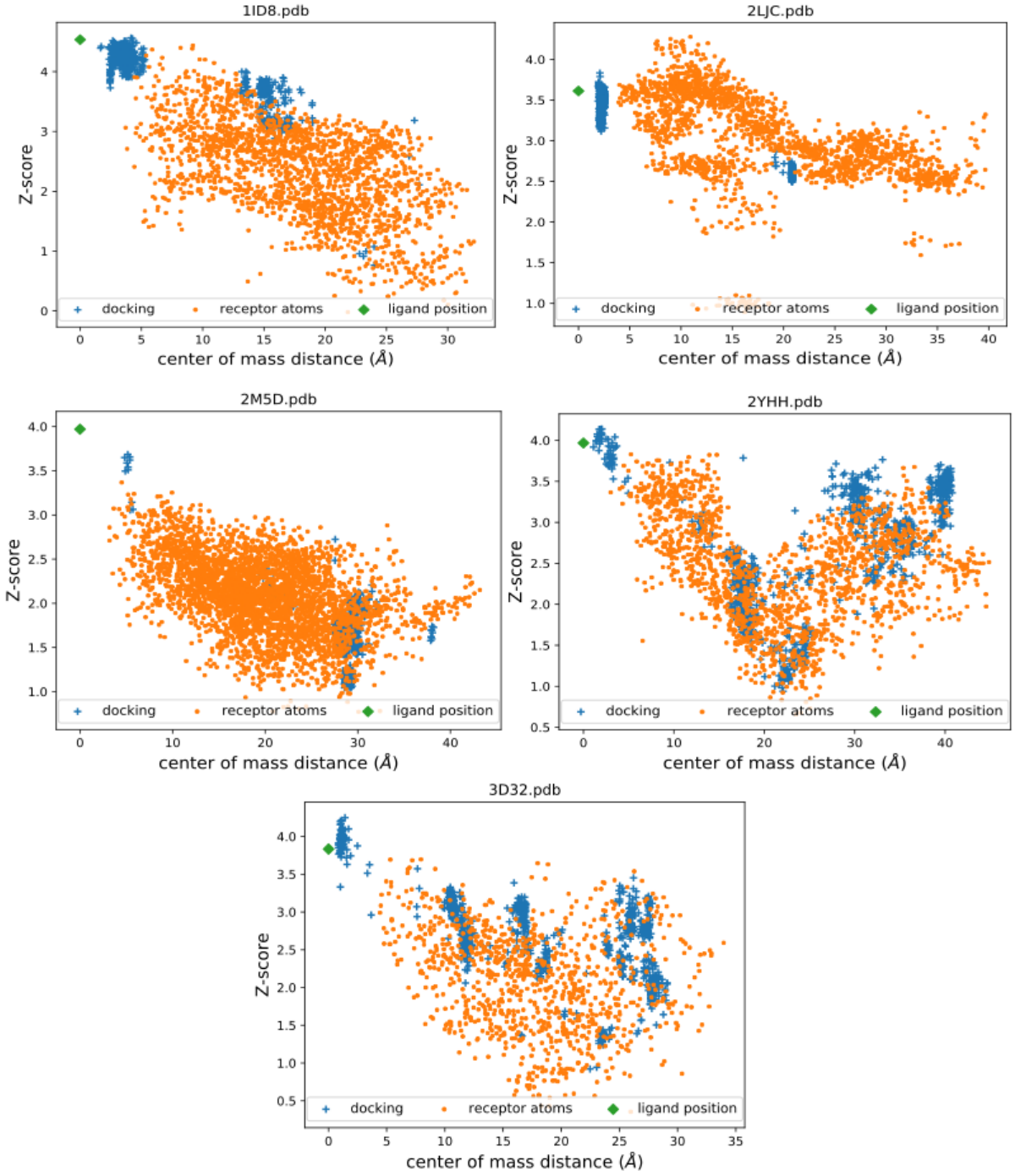

Figure 5.2: Simulated test cases. The Z-scores are plotted against the center of mass distance between all tested positions and the ligand for all receptors.

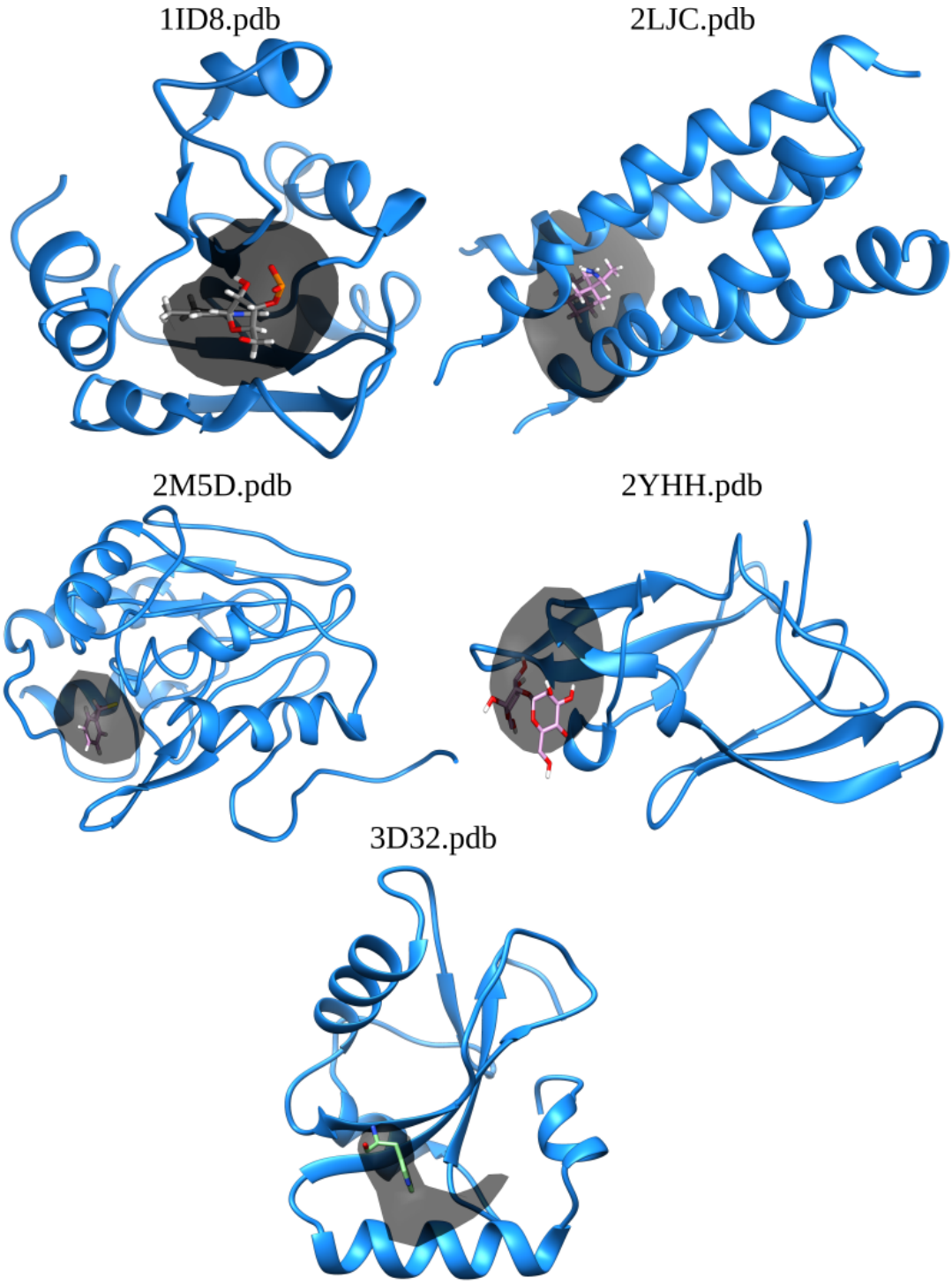

Figure 5.3: Simulated test cases. The receptors with their correct ligand positions are presented. The regions with the highest Z-scores are shown as dark surfaces.

PDB ID of receptor	Max Z-score	Distance(Å)	Max Z-score [Distance > 8 Å]
1ID8	4.564	4.153	4.00
2LJC	3.835	2.093	2.869
2M5D	3.685	5.160	2.724
2YHH	4.140	2.042	3.786
3D32	4.250	1.410	3.452

Table 5.1: Listed are the highest Z-score and the distance of the corresponding binding position to the ligand. In the last column the highest Z-scores of all positions with a distance greater than 8 Å to the ligand are shown.

5.3.2 Experimental data

The first system tested is the $GABA_R$ receptor associated protein (GABARAP) in complex with tryptophan. This complex is well suited to be used with the method because molecular docking simulations and mapping of CSP from HSQC titration spectra have been used to identify two binding sites on GABARAP, namely HP1 and HP2 [93]. HP1 has clearly proven to be the favored binding site [93] and can be used as a reference to analyze the results. From the NMR experiments, HSQC titration spectra are available which are used as experimental input in the program. Form these spectra the 10 peaks with the largest CSPs are extracted to become the reduced spectrum. The Z-score is calculated and plotted in a the same way as the simulated data, only this time also weighting factors are used. Starting from the highest Z-score the location of the binding position of HP1 could be determined within a radius of 1 Å. The Z-score density map together with the structure of GABARAP also shows that the correct binding site was identified, see figure 5.4.

Additional complexes tested with the method are SPF45 and Pex14 for which data was kindly provided by Michael Sattler TU Munich.
Two titration spectra were provided for splicing factor SPF45, one of them is a cyclic peptide with the sequence KSRWDE, the other one is a small compound called TOK98. Both molecules were designed to bind to the C-terminal U2AF homology (UHM) domain of SPF45. Two crystal structures for the UHM domain of SPF45 bound to (cyclic) peptides exist (*PDB codes: 2PEH and 5LSO*) [48][12] and the results of this method are compared with this binding position. From the titration spectra, seven and eight peaks are extracted to become the reduced spectrum for the cyclic peptide and TOK98, respectively. As receptor, a crystal structure of the free form of SPF45 (*PDB code 2PE8*) is used. Z-scores are calculated in the same way as before using a 3D grid

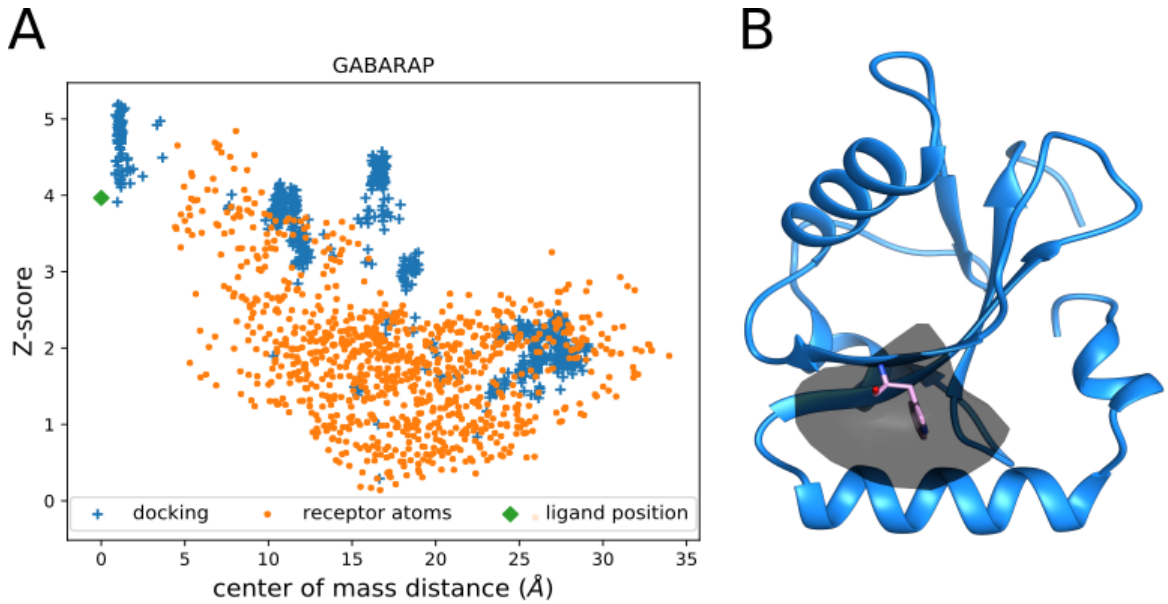

Figure 5.4: Z-scores for GABARAP. In (**A**) Z-scores are plotted against the center-of-mass RMSD. (**B**) Shows the structure of GABARAP with the ligand at binding position HP1 and the region with highest Z-scores as a dark surface.

as potential binding poses. For illustration, maps are calculated from the Z-scores and they are shown together with a crystal structure of SPF45 binding a cyclic peptide (*PDB code 5LSO*). The maps representing the regions with the highest Z-scores for the cyclic peptide and TOK98 are directly next to each other and the peptide in the crystal structure, see figure 5.5 .

The next complex tested was Pex14 together with a peptide that has the amino acid sequence GPMEDWAQHFAAHQQH and is called W1. This peptide was designed to bind to Pex14 in the same way as Pex5 which is for example described in [64]. 10 peaks form the Pex14-W1 titration spectrum were selected for the reduced spectrum and a crystal structure of the N-terminal domain of Pex14 (*PDB code 5AON*) was used as receptor. Z-scores were again calculated using the 3D grid method. A map of the highest Z-scores is shown together with Pex14 binding a small molecule at the same position that Pex5 binds (*PDB code 5L87*) [17]. As W1 was designed to bind at this binding pose, its predicted position should be same. As can be seen in figure 5.5, the predicted binding site of W1 and the peptide binding position are at the same location.

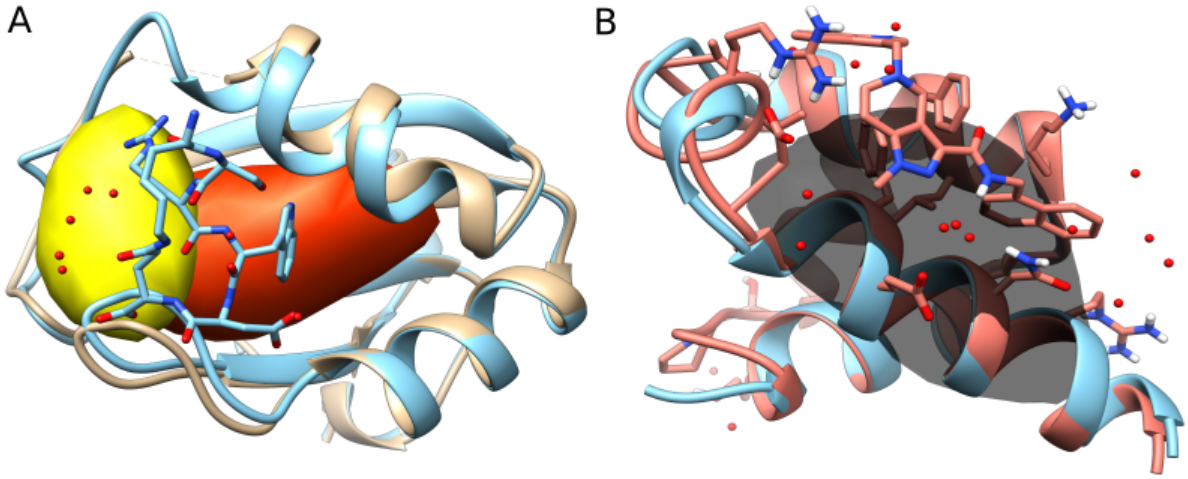

Figure 5.5: Z-scores for SPF45 and Pex14. In (**A**) a crystal structure of free SPF45 is shown (beige) together with a crystal structure of SPF45 binding a peptide (blue). The yellow and red surface indicate the predicted binding position for the cyclic peptide and TOK98, respectively, according to the highest Z-scores. (**B**) shows the free structure of Pex14 (blue) and Pex14 together with a ligand (red). The predicted binding site, according to the highest Z-scores, is shown as a dark surface.

Limitations of the methods

The method was tested with another three examples: The HCN2 channel with cAMP, the SH3 domain of Bin1 binding cMyc and MdmX. For all examples the Z-score can not detect the correct binding site and the reasons for that are discussed below. As described for the other complexes, the structures (*PDB codes 1Q43, 1MUZ and 3DAB*) are used to predict NMR spectra and from the titration spectra the reduced spectra are calculated. Z-scores are calculated for docking positions generated with AutoDock4 and all receptor atoms. For MdmX only the grid method is used to generate bindings poses and Z-scores are calculated for them.

In figure 5.6 the Z-scores are plotted against the center of mass distance from the tested binding position to the correct ligand position. No connection can be seen between the Z-score and the distance to the ligand. For the complex of HCN2 and cAMP this is because the channel undergoes a large conformational change upon binding of cAMP and this changes chemical shift values also far away from the binding site. Thus, many peaks in the titration spectrum have a large CSP, but many of them are not related to the binding site.
The situation for Bin1 is different and in this case the method fails because the ligand

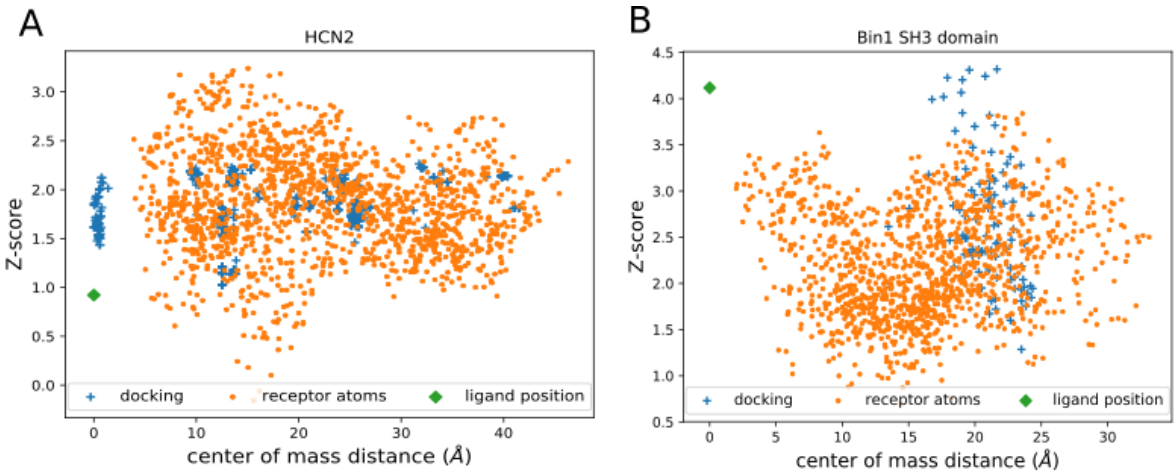

Figure 5.6: In (**A**) Z-scores calculated for HCN2 channel are plotted against the ligand center-of-mass distance. In (**B**) the same plot is shown for the SH3 domain of Bin1.

cMyc is relatively large compared to the receptor, meaning that is has approximately 20 % of the number of atoms of the SH3 domain. In the GABARAP-tryptophan complex the ligand has only around 1 percent of the receptor atoms. This means that many receptor atoms are found within 7 Å of the ligand and the predicted spectra contain many chemical shift pairs. This is inopportune for peak assignment between each predicted and reduced spectrum as wrong assignments are more likely to happen. Predicted spectra that are unrelated to the correct binding site seem to match well with the reduced spectrum and distort the result.

For MdmX, three titration spectra with three different compounds, called Cmp13 Cmp22 and Cmp24, were provided. The exact 3D structure of the tested compounds is not known and therefore only the grid method is used to produce binding sites from which Z-scores are calculated afterwards. However, it is known that Cmp13, Cmp22 and Cmp24 are all indole-based and so it is assumed that they would bind to the same binding site of MdmX as the tumor suppressor transactivation domain p53. For all three compounds the regions with the highest Z-scores are presented as surfaces and displayed together with the crystal structure of MdmX (*PDB ID 3DAB*) binding p53. In this figure (figure 5.7) one can see that the three compounds are predicted to bind at slightly different positions and also that the binding site of p53 is different from all predicted positions. The origin of the poor prediction is probably the inaccurate prediction of chemical shift values by SHIFTX2. To illustrate this the experimental spectrum of MdmX and the predicted one are plotted in the same graph. It can be

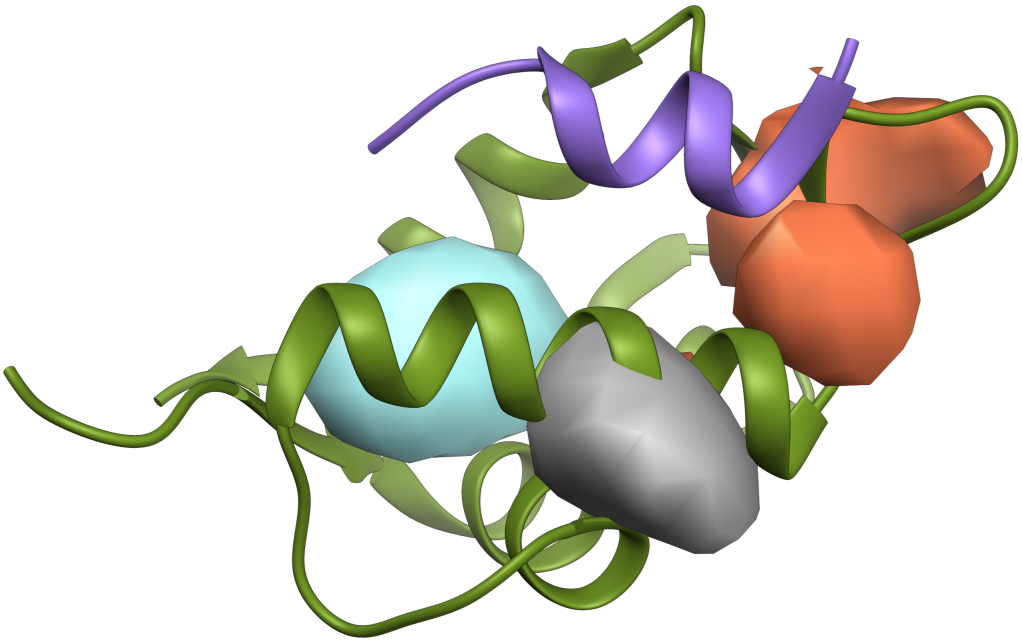

Figure 5.7: MdmX with predicted binding sites. MdmX and its original ligand (purple) are shown together with the predicted binding sites. The blue, grey and red surfaces mark the regions with highest Z-scores and originate from Cmp13, Cmp22 and Cmp24 titration spectra.

seen very clearly that their agreement is poor, see figure 5.8 .

As a reference a similar plot is shown for GABARAP where the experimental protein-only and the predicted spectrum are shown simultaneously. The agreement between the two spectra is not perfect either but clearly better than for MdmX and it has been shown before that this extent of agreement between experimental and predicted spectra is sufficient for the correct prediction of a ligand binding site.

These results show that the assumption, that peaks with large CSP originate from ligand binding, is fundamental for this method to work. In general, one can say that not more than 30 % of the total peaks in a titration spectrum should have large CSP. It was also demonstrated that too large ligands can not be used in this method. As a rule of thumb the predicted spectra should have a similar amount of peaks as the predicted binding site spectra.

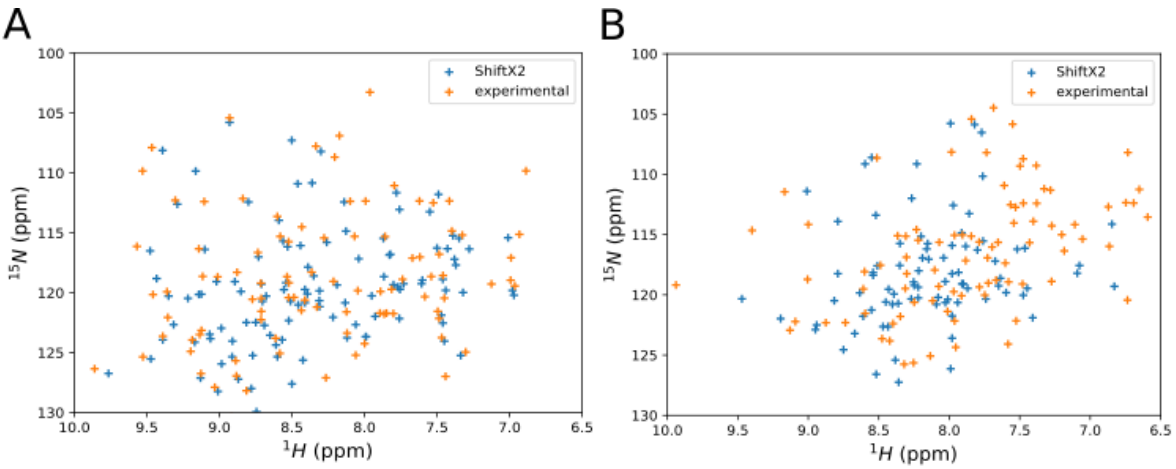

Figure 5.8: An illustration of the agreement between experimental and predicted (by SHIFTX2) NMR spectra for GABARAP (**A**) and MdmX (**B**).

5.3.3 Conclusion

In summary, the presented method is well suited for ligand binding site prediction. All simulated data and the GABARAP complex could precisely predict the correct binding site. Also, for SPF45 and Pex14 the tested ligands were predicted to be at the same location as similar molecules, which binding sites have been determined before with X-ray crystalloghraphy. The HCN2 channel and the Bin1/cMyc complex are exceptions as the method does not work on their specific characteristics. The HCN2 channel undergoes large conformational changes and Bin1 binds to a large ligand which is not taken into account by the method. The example of MdmX show that for some proteins the SHIFTX2 prediction is not accurate enough. The agreement between experimental and predicted spectra is crucial as the Z-score calculation is based on that. To conclude, the method gives the most reliable results for small ligands and proteins without conformational change.

6 Conclusion and Outlook

In conclusion, this thesis presents the first high-resolution cryo-EM structure of an amyloid-β fibril, which reveals new insight into the fibril architecture. In particular, a completely new subunit arrangement and dimer interface was found. In general, samples of Aβ fibrils are heterogeneous, making structure determination difficult. In this work, unique sample preparation, latest electron detector hardware, and extensive image processing made structure determination possible. In addition to this high-resolution fibril structure, two other much less populated polymorphs were observed in the sample that could be determined only to medium resolution because the amount of data was not sufficient to reach higher resolution.

The high-resolution structure reported in this thesis is an excellent starting point for further investigation by molecular dynamics simulations to study the fibrillization process. Specifically, since the fibril ends are different from each other the suspected polarity of growth could be tested. The structure might also be important for drug development, where compounds can be designed to interfere with fibril growth or formation. The structure could further be used to rationalize familial mutations (for example A2T and A2V). Solving the structures of fibrils from mutated Aβ or other Aβ variants could in a similar way help understanding putative differences in structure and interaction of fibril subunits and the dimer interface.

Together with the recently published structures of tau filaments (from the MRC Cambridge) Aβ-fibril is a breakthrough in the structural biology of amyloids in general and those are the first high-resolution cryo-EM structures of amyloid fibrils. It can be expected that this breakthrough stimulates more studies of amyloids by cryo-EM: structures of other disease associated fibrils are going to be solved in the near future, from both recombinant and brain-derived material. The techniques used in this study for reconstructing high-resolution density maps should help future studies on resolving other amyloid fibril structures.

Progress has also been made in studying fibrils inside cells by cryo-electron tomography

and although the obtained structures do not reach high resolution yet, this might be achieved in the future. In regards of neurodegenerative diseases, this would mean that fibrils could be studied in their native environment, revealing more information about the fibrillization process and interactions with other molecules in the brain.

Bibliography

[1] Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010.

[2] Pavel V. Afonine, Ralf W. Grosse-Kunstleve, Nathaniel Echols, Jeffrey J. Headd, Nigel W. Moriarty, Marat Mustyakimov, Thomas C. Terwilliger, Alexandre Urzhumtsev, Peter H. Zwart, and Paul D. Adams. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, 68(4):352–367, 2012.

[3] Mahiuddin Ahmed, Judianne Davis, Darryl Aucoin, Takeshi Sato, Shivani Ahuja, Saburo Aimoto, James I. Elliott, William E. Van Nostrand, and Steven O. Smith. Structural conversion of neurotoxic amyloid-β 1-42 oligomers to fibrils. *Nature Structural and Molecular Biology*, 17(5):561–567, 2010.

[4] Benjamin A. Barad, Nathaniel Echols, Ray Yu Ruei Wang, Yifan Cheng, Frank Dimaio, Paul D. Adams, and James S. Fraser. EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods*, 12(10):943–946, 2015.

[5] Elmar Behrmann, Guozhi Tao, David L. Stokes, Edward H. Egelman, Stefan Raunser, and Pawel A. Penczek. Real-space processing of helical filaments in SPARX. *Journal of Structural Biology*, 177(2):302–313, 2012.

[6] Oleksandr Brener, Tina Dunkelmann, Lothar Gremer, Thomas Van Groen, Ewa A. Mirecka, Inga Kadish, Antje Willuweit, Janine Kutzsche, Dagmar Jürgens, Stephan Rudolph, Markus Tusche, Patrick Bongen, Jörg Pietruszka, Filipp

Oesterhelt, Karl Josef Langen, Hans Ulrich Demuth, Arnold Janssen, Wolfgang Hoyer, Susanne A. Funke, Luitgard Nagel-Steger, and Dieter Willbold. QIAD assay for quantitating a compounds efficacy in elimination of toxic Aβ oligomers. *Scientific Reports*, 5(July):1–13, 2015.

[7] Marta Carroni and Helen R. Saibil. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods*, 95:78–85, 2016.

[8] Vincent B. Chen, W. Bryan Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21, 2010.

[9] Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):439–449, 2015.

[10] V.W. Chow, M.P. Mattson, P.C. Wong, and M. Gleichmann. Neuronal Activity and the Expression of Clathrin Assembly Protein AP180. *Biochem Biophys Res Commun*, 12(1):1–12, 2011.

[11] Michael T Colvin, Robert Silvers, Qing Zhe Ni, Thach V Can, Ivan Sergeyev, Melanie Rosay, Kevin J Donovan, Brian Michael, Joseph Wall, Sara Linse, G Griffin, Francis Bitter, United States, Structural Biology, United States, and United States. Atomic-resolution structure of a disease-relevant Abeta(1-42) amyloid fibril. *Journal of American Chemical Society*, 138(30):9663–9674, 2017.

[12] Lorenzo Corsini, Sophie Bonnal, Jerome Basquin, Michael Hothorn, Klaus Scheffzek, Juan Valcárcel, and Michael Sattler. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nature Structural and Molecular Biology*, 14(7):620–629, 2007.

[13] Joseph M. Courtney, Qing Ye, Anna E. Nesbitt, Ming Tang, MD Tuttle, Eric D. Watt Watt, Kristin M. Nuzzio, Lindsay J. Sperling, Gemma Comellas, Joseph R. Peterson, James H. Morrissey, and Chad M. Rienstra. Experimental Protein Structure Verification by Scoring with a Single, Unassigned NMR Spectrum. *Structure*, 26:390–394, 2015.

[14] Anthony Crowther. Straight and paired helical filaments in Alzheimer disease

have a common structural unit. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6):2288–2292, 1991.

[15] Mariusz Czarnocki-Cieciura and Marcin Nowotny. Introduction to high-resolution cryo-electron microscopy. *Postepy biochemii*, 62(3):383–394, 2016.

[16] Charlotta S. Damberg, Vladislav Y. Orekhov, and Martin Billeter. Automated analysis of large sets of heteronuclear correlation spectra in NMR-based drug discovery. *Journal of Medicinal Chemistry*, 45(26):5649–5654, 2002.

[17] M. Dawidowski, L. Emmanouilidis, V. C. Kalel, K. Tripsianes, K. Schorpp, K. Hadian, M. Kaiser, P. Mäser, M. Kolonko, S. Tanghe, A. Rodriguez, W. Schliebs, R. Erdmann, M. Sattler, and G. M. Popowicz. Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. *Science*, 355(6332):1416–1420, 2017.

[18] Angela H. DePace and Jonathan S. Weissman. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Structural Biology*, 9(5):389–396, 2002.

[19] Ruben Diaz, William J. Rice, and David L. Stokes. *Fourier-bessel reconstruction of helical assemblies*, volume 482. Elsevier Inc., 1 edition, 2010.

[20] Dennis W. Dickson, Howard A. Crystal, Caroline Bevona, William Honer, Inez Vincent, and Peter Davies. Correlations of synaptic and pathological markers with cognition of the elderly. *Neurobiology of Aging*, 16(3):285–298, 1995.

[21] Kelly T. Dineley, Xuefeng Xia, Duy Bui, J. David Sweatt, and Hui Zheng. Accelerated plaque accumulation, associative learning deficits, and up-regulation of alpha7 nicotinic receptor protein in transgenic mice co-expressing mutant human presenilin 1 and amyloid precursor proteins. *Journal of Biological Chemistry*, 277(25):22768–22780, 2002.

[22] M.-P. Dubuisson and A.K. Jain. A modified Hausdorff distance for object matching. *Proceedings of 12th International Conference on Pattern Recognition*, 1(1):566–568, 1994.

[23] Edward H. Egelman. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy*, 85(4):225–234, 2000.

[24] Edward H. Egelman. *Reconstruction of helical filaments and tubes*, volume 482. Elsevier Inc., 1 edition, 2010.

[25] David S. Eisenberg and Michael R. Sawaya. Structural Studies of Amyloid Proteins at the Molecular Level. *Annual Review of Biochemistry*, 86(1):69–95, 2017.

[26] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501, 2010.

[27] Thomas Evangelidis, Santrupti Nerli, Jiří Nováček, Andrew E. Brereton, P. Andrew Karplus, Rochelle R. Dotas, Vincenzo Venditti, Nikolaos G. Sgourakis, and Konstantinos Tripsianes. Automated NMR resonance assignments and structure determination using a minimal set of 4D spectra. *Nature Communications*, 9(1):384, 2018.

[28] Marcus Fändrich. Oligomeric intermediates in amyloid formation: Structure determination and mechanisms of toxicity. *Journal of Molecular Biology*, 421(4-5):427–440, 2012.

[29] Marcus Fändrich, Jessica Meinhardt, and Nikolaus Grigorieff. Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion*, 3(June):89–93, 2009.

[30] A. R. Faruqi and G. McMullan. *Electronic detectors for electron microscopy*, volume 44. 2011.

[31] Verena H. Finder and Rudi Glockshuber. Amyloid-β aggregation. *Neurodegenerative Diseases*, 4(1):13–27, 2007.

[32] Anthony W.P. Fitzpatrick, Benjamin Falcon, Shaoda He, Alexey G. Murzin, Garib Murshodov, Holly J. Garringer, Anthony R. Crowther, Bernadino Ghetti, Michel Goedert, and Sjors H.W. Scheres. Cryo-EM structures of Tau filaments from Alzheimer's disease brain. *Nature*, 547:185–190, 2017.

[33] Joachim Frank. 2 . Electron Microscopy of Macromolecular Assemblies. In *Three Dimensional Electron Microscopy of Molecular Assemblies*. 1996.

[34] Joachim Frank. 4. Multivariate Statistical Analysis and Classification of Images. In *Three Dimensional Electron Microscopy of Molecular Assemblies*, pages 126–181. 1996.

[35] Joachim Frank. 5. Three-Dimensional Reconstruction. In *Three Dimensional Electron Microscopy of Molecular Assemblies*. 1996.

[36] S. A. Fromm and C. Sachse. *Cryo-EM Structure Determination Using Segmented Helical Image Reconstruction*. Elsevier Inc., 1 edition, 2016.

[37] Lisa Fukui and Yuan Chen. NvMap: automated analysis of NMR chemical shift perturbation data. *Bioinformatics*, 23(3):378–380, 2007.

[38] Benoit I. Giasson, Ian V J Murray, John Q. Trojanowski, and Virginia M Y Lee. A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of alpha-Synuclein Is Essential for Filament Assembly. *Journal of Biological Chemistry*, 276(4):2380–2386, 2001.

[39] Lothar Gremer, Daniel Schölzel, Carla Schenk, Elke Reinartz, Jörg Labahn, Raimond B G Ravelli, Markus Tusche, Carmen Lopez-iglesias, Wolfgang Hoyer, Henrike Heise, Dieter Willbold, and Gunnar F Schröder. Fibril structure of amyloid-b (1 – 42) by cryo – electron microscopy. *Science*, 358(October):116–119, 2017.

[40] N. Grigorieff. *Frealign: An Exploratory Tool for Single-Particle Cryo-EM*, volume 579. Elsevier Inc., 1 edition, 2016.

[41] Beomsoo Han, Yifeng Liu, Simon W. Ginzinger, and David S. Wishart. SHIFTX2: Significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR*, 50(1):43–57, 2011.

[42] John A. Hardy and Gerald A. Higgins. Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science*, 256, 1992.

[43] Shaoda He and Sjors H.W. Scheres. Helical reconstruction in RELION. *Journal of Structural Biology*, 198(3):163–176, 2017.

[44] Elizabeth Head, Jennifer Tomic, Theresa M Mcintire, Rakez Kayed, Elizabeth Head, Jennifer L Thompson, Theresa M Mcintire, Saskia C Milton, and Carl W Cotman. Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis . Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science*, 486(MAY):486–490, 2003.

[45] Michael Hohn, Grant Tang, Grant Goodyear, P. R. Baldwin, Zhong Huang, Pawel A. Penczek, Chao Yang, Robert M. Glaeser, Paul D. Adams, and Steven J.

Ludtke. SPARX, a new environment for Cryo-EM image processing. *Journal of Structural Biology*, 157(1):47–55, 2007.

[46] S. Hoyer. Brain glucose and energy metabolism abnormalities in sporadic Alzheimer disease. Causes and consequences: An update. *Experimental Gerontology*, 35(9-10):1363–1372, 2000.

[47] Yuji Inoue, Aiko Kishimoto, Jun Hirao, Masasuke Yoshida, and Hideki Taguchi. Strong Growth Polarity of Yeast Prion Fiber Revealed by Single Fiber Imaging. *Journal of Biological Chemistry*, 276(38):35227–35230, 2001.

[48] Pravin Kumar Ankush Jagtap, Divita Garg, Tobias G. Kapp, Cindy L. Will, Oliver Demmer, Reinhard Lührmann, Horst Kessler, and Michael Sattler. Rational Design of Cyclic Peptide Inhibitors of U2AF Homology Motif (UHM) Domains to Modulate Pre-mRNA Splicing. *Journal of Medicinal Chemistry*, 59(22):10190–10197, 2016.

[49] Mathias Jucker and Lary C. Walker. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, 501(7465):45–51, 2013.

[50] Tuomas P.J. Knowles, Michele Vendruscolo, and Christopher M. Dobson. The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecular Cell Biology*, 15(6):384–396, 2014.

[51] M. P. Lambert, A. K. Barlow, B. A. Chromy, C. Edwards, R. Freed, M. Liosatos, T. E. Morgan, I. Rozovsky, B. Trommer, K. L. Viola, P. Wals, C. Zhang, C. E. Finch, G. A. Krafft, and W. L. Klein. Diffusible, nonfibrillar ligands derived from Abeta 1-42 are potent central nervous system neurotoxins. *Proceedings of the National Academy of Sciences*, 95(11):6448–6453, 1998.

[52] Claire J Lansdall. Review An effective treatment for Alzheimer's disease must consider both amyloid and tau. *Bioscience Horizons*, 7(December):1–11, 2014.

[53] Hilal A. Lashuel, Dean Hartley, Benjamin M. Petre, Thomas Walz, and Peter T Lansbury Jr. Amyloid pores from pathogenic mutations. *Nature*, 418(July), 2002.

[54] Da Wei Li, Dan Meng, and Rafael Brüschweiler. Reliable resonance assignments of selected residues of proteins with known structure based on empirical NMR chemical shift prediction. *Journal of Magnetic Resonance*, 254:93–97, 2015.

[55] H Lin, R Bhatia, and R Lal. Amyloid beta protein forms ion channels: impli-

cations for Alzheimer's disease pathophysiology. *Faseb J.*, 15(1530-6860 SB - IM):2433–2444, 2001.

[56] Thorsten Lührs, Christiane Ritter, Marc Adrian, Dominique Riek-loher, Bernd Bohrmann, Heinz Döbeli, David Schubert, and Roland Riek. 3D structure of Alzheimer's amyloid-β(1 – 42) fibrils. *Proceedings of the National Academy of Sciences*, 2005.

[57] Sergey Lyskov and Jeffrey J. Gray. The RosettaDock server for local protein-protein docking. *Nucleic acids research*, 36(Web Server issue):233–238, 2008.

[58] Dominique Marion. An Introduction to Biological NMR Spectroscopy. *Molecular & Cellular Proteomics*, 12(11):3006–3025, 2013.

[59] Iris A. Mastrangelo, Mahiuddin Ahmed, Takeshi Sato, Wei Liu, Chengpu Wang, Paul Hough, and Steven O. Smith. High-resolution atomic force microscopy of soluble Abeta42 oligomers. *Journal of Molecular Biology*, 358(1):106–119, 2006.

[60] Beat H. Meier, Roland Riek, and Anja Böckmann. Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. *Trends in Biochemical Sciences*, 42(10):777–787, 2017.

[61] Jacqueline L.S. Milne, Mario J. Borgnia, Alberto Bartesaghi, Erin E.H. Tran, Lesley A. Earl, David M. Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo-electron microscopy - A primer for the non-microscopist. *FEBS Journal*, 280(1):28–45, 2013.

[62] Gm Morris and Ruth Huey. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of ...*, 30(16):2785–2791, 2009.

[63] Kyle L Morris and Louise C Serpell. X-Ray Fibre Diffraction Studies of Amyloid Fibrils. In *Amyloid Proteins*, volume 849, pages 121–135. Springer Science+Business Media, 2012.

[64] Christian Neufeld, Fabian V. Filipp, Bernd Simon, Alexander Neuhaus, Nicole Schüller, Christine David, Hamed Kooshapur, Tobias Madl, Ralf Erdmann, Wolfgang Schliebs, Matthias Wilmanns, and Michael Sattler. Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO Journal*, 28(6):745–754, 2009.

[65] Christoph Nitsche and Gottfried Otting. NMR studies of ligand binding. *Current Opinion in Structural Biology*, 48:16–22, 2018.

[66] Kenjiro Ono. Alzheimer's disease as oligomeropathy. *Neurochemistry International*, 2017.

[67] Anant K. Paravastu, Richard D. Leapman, Wai-Ming Yau, and Robert Tycko. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proceedings of the National Academy of Sciences*, 105(47):18349–18354, 2008.

[68] Hugh A. Pearson and Chris Peers. Physiological roles for amyloid β peptides. *Journal of Physiology*, 575(1):5–10, 2006.

[69] Pawel A. Penczek. *Fundamentals of Three-Dimensional reconstruction from projections*, volume 482. Elsevier Inc., 1 edition, 2010.

[70] Pawel A. Penczek. *Image restoration in cryo-electron microscopy*, volume 482. Elsevier Inc., 1 edition, 2010.

[71] Pawel A. Penczek. *Resolution measures in molecular electron microscopy*, volume 482. Elsevier Inc., 1 edition, 2010.

[72] Pawel A Penczek, Robert Renka, and Hermann Schomberg. Gridding-based direct Fourier inversion of the three-dimensional ray transform. *Journal of the Optical Society of America A*, 21(4):499, 2004.

[73] Chen Peng, Stephen W. Unger, Fabian V. Filipp, Michael Sattler, and Sándor Szalma. Automated evaluation of chemical shift perturbation spectra: New approaches to quantitative analysis of receptor-ligand interaction NMR spectra. *Journal of Biomolecular NMR*, 29(4):491–504, 2004.

[74] Aneta Petkova, Wai-Ming Yau, and Robert Tycko. Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry*, 45(2):498–512, 2006.

[75] Johnny D. Pham, Borries Demeler, and James S. Nowick. Polymorphism of oligomers of a peptide from β-amyloid. *Journal of the American Chemical Society*, 136(14):5432–5442, 2014.

[76] Iva Pritišanac, Matteo T. Degiacomi, T. Reid Alderson, Marta G. Carneiro, Eiso Ab, Gregg Siegal, and Andrew J. Baldwin. Automatic Assignment of Methyl-

NMR Spectra of Supramolecular Machines Using Graph Theory. *Journal of the American Chemical Society*, 139(28):9523–9533, 2017.

[77] Wei Qiang, Wai-Ming Yau, Jun-Xia Lu, John Collinge, and Robert Tycko. Structural Variation in Amyloid-β Fibrils from Alzheimer's Disease Clinical Subtypes. *Nature*, 541(2):217–221, 2017.

[78] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *Journal of Microscopy*, 146(2):113–136, 1987.

[79] Jay Rasmussen, Jasmin Mahler, Natalie Beschorner, Stephan A. Kaeser, Lisa M. Häsler, Frank Baumann, Sofie Nyström, Erik Portelius, Kaj Blennow, Tammaryn Lashley, Nick C. Fox, Diego Sepulveda-Falla, Markus Glatzel, Adrian L. Oblak, Bernardino Ghetti, K. Peter R. Nilsson, Per Hammarström, Matthias Staufenbiel, Lary C. Walker, and Mathias Jucker. Amyloid polymorphisms constitute distinct clouds of conformational variants in different etiological subtypes of Alzheimer's disease. *Proceedings of the National Academy of Sciences*, 114(49):201713215, 2017.

[80] Michael L. Reese and Volker Dötsch. Fast Mapping of Protein-Protein Interfaces by NMR Spectroscopy. *Journal of the American Chemical Society*, 125(47):14250–14251, 2003.

[81] Peter B. Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003.

[82] Carsten Sachse, Marcus Fändrich, and Nikolaus Grigorieff. Paired beta-sheet structure of an Abeta(1-40) amyloid fibril revealed by electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21):7462–7466, 2008.

[83] A. Sandberg, L. M. Luheshi, S. Sollvander, T. Pereira de Barros, B. Macao, T. P. J. Knowles, H. Biverstal, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. M. Dobson, and T. Hard. Stabilization of neurotoxic Alzheimer amyloid- oligomers by protein engineering. *Proceedings of the National Academy of Sciences*, 107(35):15595–15600, 2010.

[84] Paul Schanda and Bernhard Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22):8014–8015, 2005.

[85] Matthias Schmidt, Alexis Rohou, Keren Lasker, Jay K. Yadav, Cordelia Schiene-Fischer, Marcus Fändrich, and Nikolaus Grigorieff. Peptide dimer structure in an Aβ(1–42) fibril visualized with cryo-EM. *Proceedings of the National Academy of Sciences*, 112(38):11858–11863, 2015.

[86] Gunnar F. Schröder, Axel T. Brunger, and Michael Levitt. Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution. *Structure*, 15(12):1630–1641, 2007.

[87] Anne K. Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert, Joseph Wall, Peter Güntert, Anja Bcöckmann, Rudi Glockshuber, and Beat H. Meier. Atomic-resolution three-dimensional structure of amyloid b fibrils bearing the osaka mutation. *Angewandte Chemie - International Edition*, 54(1):331–335, 2015.

[88] Dennis J Selkoe and John Hardy. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine*, 8(6):595–608, 2016.

[89] Suzanne B. Shuker, Philip J. Hajduk, Robert P. Meadows, and Stephen W. Fesik. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science*, 274(November):1531–1534, 1996.

[90] S. Singh, A. S. Kushwah, R. Singh, M. Farswan, and R. Kaur. Current therapeutic strategy in Alzheimer's disease. *European review for medical and pharmacological sciences*, 16(12):1651–1664, 2012.

[91] J. C. Stroud, C. Liu, P. K. Teng, and D. Eisenberg. Toxic fibrillar oligomers of amyloid- have cross- structure. *Proceedings of the National Academy of Sciences*, 109(20):7717–7722, 2012.

[92] Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007.

[93] Yvonne Thielmann, Jeannine Mohrlüder, Bernd W. Koenig, Thomas Stangler, Rudolf Hartmann, Karin Becker, Hans Dieter Höltje, and Dieter Willbold. An

indole-binding site is a major determinant of the ligand specificity of the GABA type A receptor-associated protein GABARAP. *ChemBioChem*, 9(11):1767–1775, 2008.

[94] Wouter G. Touw, Coos Baakman, Jon Black, Tim A.H. Te Beek, E. Krieger, Robbie P. Joosten, and Gert Vriend. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1):D364–D368, 2015.

[95] Yuki Toyama, Yoko Mase, Hanaho Kano, Mariko Yokogawa, Masanori Osawa, and Ichio Shimada. Nuclear Magnetic Resonance Approaches for Characterizing Protein-Protein Interactions. In *Potassium Channels: Methods and Protocols*, volume 1684, chapter 10, pages 115–128. Springer Science+Business Media, 2018.

[96] Matthias Trautwein, Kai Fredriksson, Heiko M. Möller, and Thomas E. Exner. Automated assignment of NMR chemical shifts based on a known structure and 4D spectra. *Journal of Biomolecular NMR*, 65(3-4):217–236, 2016.

[97] MD Tuttle, G Comellas, AJ Nieuwkoop, DJ Covell, and DA Berthold. Solid-state NMR structure of a pathogenic fibril of full-length human alpha synuclein. *Nat. Struct. Mol. Biol*, 23(5):409–415, 2016.

[98] Marin Van Heel. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987.

[99] Marin van Heel and Joachim Frank. USE OF MULTIVARIATE STATISTICS IN ANALYSING THE IMAGES OF BIOLOGICAL MACROMOLECULES. *Ultramicroscopy*, 6:187–194, 1981.

[100] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005.

[101] Niels Volkmann, HongJun Liu, Larnele Hazelwood, Elena B. Krementsova, Susan Lowey, Kathleen M. Trybus, and Dorit Hanein. The structural basis of myosin V processive movement as revealed by electron cryomicroscopy. *Molecular Cell*, 19(5):595–605, 2005.

[102] Michael J. Volles and Peter T. Lansbury. Vesicle permeabilization by protofibrillar α-synuclein is sensitive to Parkinson's disease-linked mutations and occurs by a pore-like mechanism. *Biochemistry*, 41(14):4595–4602, 2002.

[103] Dominic M. Walsh and Dennis J. Selkoe. Aβ oligomers - A decade of discovery. *Journal of Neurochemistry*, 101(5):1172–1184, 2007.

[104] Marielle Aulikki Wälti, Francesco Ravotti, Hiromi Arai, Charles G. Glabe, Joseph S. Wall, Anja Böckmann, Peter Güntert, Beat H. Meier, and Roland Riek. Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proceedings of the National Academy of Sciences*, 113(34):E4976–E4984, 2016.

[105] Rui Zhang, Xiaoyan Hu, Htet Khant, Steven J Ludtke, Wah Chiu, Michael F Schmid, Carl Frieden, and Jin-Moo Lee. Interprotofilament interactions between Alzheimer's Abeta1-42 peptides in amyloid fibrils revealed by cryoEM. *Proceedings of the National Academy of Sciences of the United States of America*, 106(12):4653–4658, 2009.

[106] Qin Zhiao, Ronnie Frederick, Kory Seder, Sandy Thao, Hassan Sreenath, Francis Peterson, Brian F. Volkman, John L. Markley, and Brian G. Fox. Production in two-liter beverage bottles of proteins for NMR strucutre determination labeled with either 15N- or 13C-15N. *Journal of Structural and Functional Genomics*, 5:87–93, 2004.

Selbstständigkeitserklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die vorliegende Arbeit wurde nicht an anderen Fakultäten vorgelegt. Es wurden zuvor keine Promotionsversuche unternommen.

Jülich, Carla Schenk