Unbinding of Abeta peptides from amyloid fibrils: explicit solvent molecular dynamics study

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

By

Pamela Haradhan Mishra
Bachelor of Engineering
Mumbai University, 2004

Director: Dr.Dmitri Klimov, PhD.
Department of Bioinformatics and Computational Biology

Fall Semester 2008
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my family and friends who have been supportive all throughout my life.

ACKNOWLEDGEMENTS

I would like to thank especially Dr. Dmitri Klimov for his patience and guidance throughout this thesis. I am thankful to my committee members Dr.Iosiv Vaisman and Dr. Saleet Jafri for their support through out the course of this thesis. My gratitude goes out to Prabhu Raman for his invaluable assistance to teach me throughout this project.

Thanks to all my close friends especially Fayaz Seifuddin, Yogesh Wagle, and my brother Subrata Mishra who supported me during my educational journey at George Mason University.

TABLE OF CONTENTS

Page
List of Tables……v
List of Figures……..vi
List of Abbreviations/Symbols……vii
Abstract.....……viii
Introduction……..2
Background……..4
Protein folding……...……...4
Amyloid fibril formation……..5
Structure of Abeta amyloid fibrils……...6
Effects of external conditions on fibril formation……8
Materials and Methods……9
Molecular Dynamics……...9
Simulation of amyloid fibrils……..11
Clustering of molecular dynamics trajectory… ……12
Calculation of structural probes……..12
Results……..15
Hydrophobic peptide-fibril contacts ……...15
RMSD calculations……..16
Analysis of the contacts between residues……...…20
Analysis of solvent accessible surface area ……24
Analysis of unbinding of the edge peptides using clustering technique……...26
Discussion and Conclusion……...30
Appendix……..33
References……46

LIST OF TABLES

Table Page
Table 1: Characteristics of structural clusters of the edge peptides……..… 29

LIST OF FIGURES

Figure Page
Figure 1: Sequence of Abeta peptide……6
Figure 2: Aβ fibril hexamer fragment…….....7
Figure 3: Number of hydrophobic contacts formed by the edge peptides F5 and F6…..16
Figure 4: RMSD of Ca atoms in the peptides F5 and F6 as a function of time……...17
Figure 5: Difference between the RMSDs of Ca for F5 and F6 as a function of time ...18
Figure 6: RMSD values of F5 and F6 Ca atoms.....18
Figure 7: RMSD of the centers of mass of F5 and F6 side chains…….……..19
Figure 8: Probability distribution of the distance between the residues Phe19 in F5 and F3 (a) and in F6 and F4 (b) 22
Figure 9: Probability distribution of the distance between the residues Phe20 in F5 and F3 (a) and in F6 and F4(b)…….……...23
Figure 10: Probability distribution of the distance between the residues Lys28 and Asp23 in F5 and F3……..24
Figure 11: Total accessible surface area tASA of F5 and F6 as a function of time…… 25
Figure 12: AAS of the residue Lys28 as a function of time…… ...……...25
Figure 13: Distribution of cluster formation as a function of time for F5 (a) and F6(b).27

LIST OF ABBREVIATIONS OR SYMBOLS

Aβ - Abeta fibril

ASA - Accessible Surface Area

Ca – C-alpha atoms

F – Amyloid Fibril

HB – Hydrogen Bond

MD – Molecular Dynamics

RMSD - Root Mean Square Deviation

ABSTRACT

UNBINDING OF ABETA PEPTIDES FROM AMYLOID FIBRILS: EXPLICIT SOLVENT MOLECULAR DYNAMICS STUDY

Pamela Haradhan Mishra, M.S.

George Mason University, 2008

Thesis Director: Dr. Dmitri Klimov

We used all-atom molecular dynamics to investigate the unbinding of Alzheimer's Abeta peptides from amyloid fibrils. The peptide unbinding is driven by temperature-induced structural fluctuations and is thought of as an elementary step in the molecular recycling occurring between fibrillized and soluble Abeta species. Several conclusions can be drawn from our data. Early unbinding stages can be observed on the 100 ns timescale, although complete dissociation is likely to occur on much longer timescale. The unbinding pathway starts with fraying of β-strands from Abeta fibril. We predict that most peptide-fibril interactions will be initially lost by the N-terminal of Abeta peptide. This unbinding scenario can be modified by the specific location of dissociating peptide on the fibril edge. Our simulations provide new molecular details on the process of amyloidogenesis linked to Alzheimer's disease.

INTRODUCTION

Amyloidosis is a process of abnormal deposition of soluble proteins in the form of insoluble fibrils in organs and tissues. It has been associated with various diseases such as Alzheimer's, Type-2 diabetes, Parkinson's, and prion-related transmissible spongiform encephalopathies [1-4]. Although amyloidogenic proteins do not share any significant sequence or structural homology, the amyloid fibrils formed by these proteins are extremely similar [5, 6]. The fibrils appear as highly elongated and unbranched deposits with a diameter of approximately 80-150 Å [7, 8]. They all have characteristic dye-binding properties and a common core structure, which shows the cross-beta X-ray diffraction pattern. The amyloid cross-beta structure is unusually stable due to hydrogen bond stacks established between protein backbones. Recent experiments have also shown that a steric zipper formed by amino acids side chains plays important role in amyloid stability [9].

Many soluble proteins or peptides such as myoglobin, lysozyme or homopeptides (polythreonine or polylysine) are seen to form amyloid fibrils under certain circumstances [6, 10]. These proteins did not have any significant sequence similarity and are not known to aggregate under physiological conditions. It can be surmised that

amyloid formation is the inherent property of any polypeptide chain irrespective of its sequence. This makes the prediction of the occurrence of amyloidosis difficult.

The following sections discuss protein folding, structure of amyloids and temperature and pressure effects on amyloids and the importance of molecular dynamics in studying amyloids.

BACKGROUND

Protein folding

Native conformations of a protein is its most thermodynamically stable state [6]. Most of protein functions - binding of ligands, DNAs, RNAs, ions, transport etc - are performed only when protein is in native state. It is hypothesized that polypeptide sequence has to find its lowest energy (native) state through a stochastic search process guided by underlying free energy landscape. Many proteins fold quickly on micro- to millisecond time scale, while secondary structure elements form even faster. For example, helices can form in less than 100 nanoseconds and β–hairpins appear within few microseconds [6, 11].

How proteins fold to their native states has been investigated by a variety of experimental and theoretical methods. Mutation studies have been useful to probe the role of certain amino acids in the polypeptide sequence in protein folding. Many such studies indicate that the mechanism of folding involves the formation of a nucleus by a relatively small number of residues [12]. The remaining structure then condenses around this nucleus. For large proteins one or more intermediate folding stages usually occur [13]. Native

structure is attained, when all native interactions have been formed and water is expelled from the native core [14].

Amyloid fibril formation

Amyloid plaques are formed as a result of protein aggregation. The precise nature of conformational changes and the time evolution of polypeptides from their monomeric states to oligomers and to mature fibrils are still largely unknown. However, it is well established that the core structure of the fibril is stabilized by hydrogen bonds between protein backbone. Since the backbone is common for all polypeptides, fibrils can, in principle, be formed by any polypeptide. There is a significant similarity in the amyloid assembly for different proteins. The first phase is a formation of soluble oligomers that are later transformed into fibril protofilaments. Lateral association among these protofilaments and some structural rearrangements lead finally to the assembly of mature fibrils. It has been shown that the peptides as short as four residues long are capable of fibril formation [9]. Experimentally it has been shown that mature fibrils are able to elongate by adding monomers to the fibril edges [15, 16]. High stability and mechanical strength makes it difficult to degrade amyloid fibrils.

Determination of discreet molecular structure of amyloids has been difficult due to their non-crystalline and insoluble nature [17]. However, experiments such as NMR (Nuclear Magnetic Resonance), EPR (Electron Paramagnetic Resonance), proline scanning mutagenesis, Raman spectroscopy can be used to determine the secondary structure in

amyloids. It is now known that the secondary structure is almost exclusively based on β-sheets. The tertiary structure of fibrils involves the arrangement of these β-sheets into laminated stacks stabilized by hydrophobic interactions. The X-ray diffraction patterns of various amyloids show that the β-strands in sheets perpendicular to the axis [9]. Furthermore, NMR experiments indicate that polypeptides in β-sheets can form either parallel or anti-parallel registry.

Structure of Abeta amyloid fibril

Aggregation of 40-residue long Abeta peptides is linked to Alzheimer's disease.

The sequence of Abeta peptide is given in Figure 1. The residues numbered 10 through 40 are shown in this figure. Residues forming two beta strands β1 and β2 are in red.

Y_{10}EVHHQKLVFFAEDVGSNKGAIIGLMVGGVV_{40}

Figure 1: Sequence of Abeta peptide.

The fibril structure of Abeta peptides was resolved in the NMR experiments of Petkova et. al (Figure 2) [21]. These experiments have shown that the N-terminal residues 1-9 are structurally disordered and have minor impact on Abeta amyloid assembly [21]. Consequently, these residues are not considered. The two β-strands in Figure 2 are formed by the residues 10-22 and 30-40. These β-strands form in-register parallel β-sheets. Residues 23-29 form a loop. The fibril structure consists of four layered β-sheets.

The fragment of fibril shown in Figure 2 consists of six peptides denoted as F1, F2, F3, F4, F5, and F6.

In the fibril structure an interpeptide salt bridge exists between the residues Asp23 and Lys28 [21, 22]. This buried salt bridge is thought to be important for the stability of the fibril [22]. Another interesting feature of the fibril structure is its distinct edges. In Figure 2 Aβ fibril has concave and convex edges, of which the former demonstrates higher affinity toward incoming peptides [23].

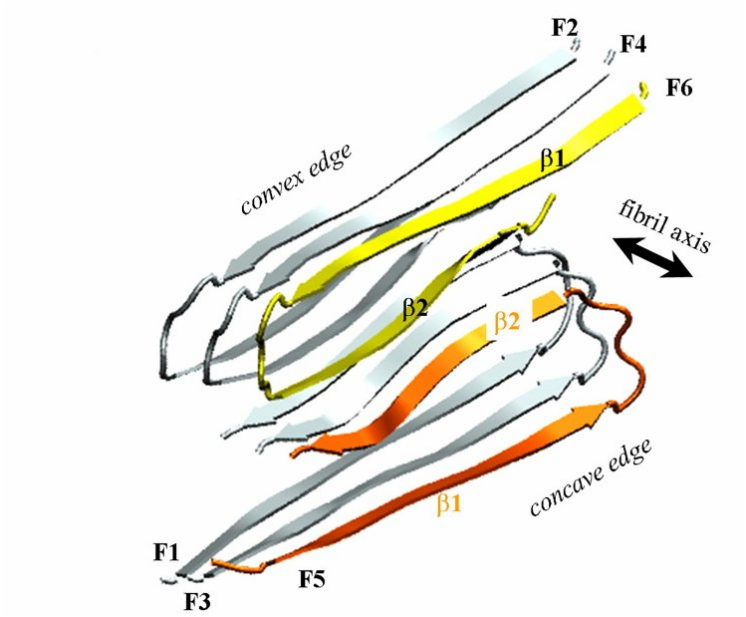

Figure 2: Aβ fibril hexamer fragment. β1 and β2 denote the β-strands. The fibril axis is perpendicular to β-strands (the figure is taken from [23]).

Effects of external conditions on fibril formation

The formation of amyloid fibrils is hierarchical, involving the initial assembly of oligomers, their conversion into protofibrils, and final association of protofibrils into mature fibrils. Studies have shown the importance of non-covalent interactions in the stability of amyloid fibrils [18]. Experiments probing early fibril formation [19, 20] suggest that the assembly is mostly driven by hydrophobic and electrostatic interactions. Consequently, amyloid aggregates can be dissociated by changing external conditions, for instance by applying high hydrostatic pressure or raising temperature.

In this thesis, we have studied the unbinding of Abeta peptides from the fibril caused by temperature induced structural fluctuations. The fibril fragment was taken from the experimental structure resolved by Petkova et al (Figure 2). The aim of this thesis is to study using molecular dynamics the unbinding of the peptides F5 and F6 from the fibril core. To this end, we analyzed the interactions between F5 and F6 and the fibril, which include hydrogen bonds, electrostatic and hydrophobic contacts as well as solvent accessible surface area and RMSD of backbone atoms and side-chains. Taken together this information allowed us to reconstruct the initial stages in unbinding process.

MATERIALS AND METHODS

In this section we describe the steps taken to produce the results, graphs and give also the background for the analysis techniques.

Molecular dynamics

Typically molecular level information about structural transitions in biomolecules can only be obtained through molecular simulations. Molecular dynamics (MD) is the most widely used technique to simulate proteins in all-atom detail. In most cases the starting configuration in those simulations is an experimentally determined structure, for which Newton's equations of motions are applied

$$F = ma, \quad \textbf{(1)}$$

where m is the mass of atom, a is acceleration, and F is the force applied to an atom. Force is calculated using

$$F = -dEp / dx \quad \textbf{(2)}$$

where E_p is the potential energy and x is the atomic position. The potential energy in MD is decomposed into bonded and non-bonded contributions.

$$E_p = E_{bonded} + E_{nonbonded} \quad \textbf{(3)}$$

The first is represented by bond length, bond angle, and dihedral potentials

$$E_{bonded} = E_{bl} + E_{ba} + E_{di} \quad \textbf{(4)}$$

These potentials act on the atoms which are located in close proximity along the sequence. The specific form of three bonded potentials is give by

$$E_{bonded} = \sum \mathrm{k}^{r}{}_{i}(r_i - r_{eq})^2 + \sum k^{\theta}{}_{i}(\theta_i - \theta_{eq})^2 + \sum k^{\phi}{}_{i}(1 + \cos(\phi_i - \gamma_m)) \quad \textbf{(5)}$$

where r_i and r_{eq} are the distance between bonded atoms *i* and *i+1* in a given conformation and its equilibrium value, θ_i and θ_{eq} are the bond angle between the bonds *(i-1, i)* and *(i, i+1)* and its equilibrium value, and ϕ_i is the dihedral angle describing the rotation of the atom *i+4* with respect to the plane *(i, i+1, i+2)*. The coefficients $k^{r}{}_{i}$, $k^{\theta}{}_{i}$, and k^{ϕ} are the constants determining the strength of respective potentials.

The non-bonded interactions correspond to Van-der-Waals and electrostatic forces, which are assumed to be additive

$$E_{nonbonded} = E_{el} + E_{vdw} \quad \textbf{(6)}$$

Electrostatic interactions are given by Coulombic potential

$$E_{el} = \sum q_i q_j / 4\varepsilon_o r_{ij} \quad \textbf{(7)}$$

where q_i and q_j are charges on the atoms *i* and *j*, ε_0 is the permittivity of free space, r_{ij} is the distance between the atoms. Van-der-Waals interactions are described by Lennard-Jones 12-6 potentials

$$E_{vdw} = \sum 4\varepsilon_h[(\sigma / r_{ij})^{12} - (\sigma / r_{ij})^{6}] \quad \textbf{(8)}$$

where σ is an equilibrium distance between the atoms *i* and *j* and ε_h is the strength of atomic contact. To reduce computational load cut-off techniques are used, which exclude

the non-bonded interactions between distant atoms. In addition, particle–mesh Ewald (PME) summations are applied to electrostatic interactions to accelerate computations. In our study, we used NAMD molecular dynamics program coupled with CHARMM22 all-atom force field, which supplied all the parameters for the potentials described above.

Simulation of amyloid fibrils

Simulations of temperature-induced unbinding of Abeta peptides were performed starting with the experimental Abeta fibril structure (Figure 2). Specifically, we used the hexamer fragment of the fibril consisting of Abeta peptides F1-F6. The structure was solvated in the 81Å x 73Å x 52Å water box with the density of 1.00 g/cm^3. In all, the system contains 29,676 atoms. To eliminate confinement effects we applied periodic boundary conditions. Using NAMD program the system was heated to 330K and equilibrated for 300 ps. The aim of the simulations was to investigate the temperature-induced unbinding of the edge Abeta peptides, F5 and F6 (Figure 2), from the “large” fibril sample represented by the peptides F1-F4. Consequently, to mimic the stability of the ``large'' fibril sample, the backbone heavy atoms of the peptides F1-F4 were constrained to their experimental fibril positions using soft harmonic potentials at all simulation stages. During heating and equilibration the peptides F5 and F6 were also constrained to fibril positions. These constraints were released once equilibration was completed. The 100 ns production simulations were performed using NVE ensemble after equilibration. During production simulations the unbinding of F5 and F6 was monitored. The structural snapshots were saved every 10 ps.

Clustering of MD trajectory

MD trajectory generates thousands of structures. To reduce random structural fluctuations and identify common structural properties in the large pool of conformations, we apply clustering technique [24-26]. Specifically, the peptide structures were clustered using pattern recognition algorithm [27]. To avoid implicit introduction of progress variable, conformations were clustered using the positions of peptide Ca atoms.

To cluster conformations we used the cut-off radius, R_c. R_c represents the maximum Euclidian distance between the Ca coordinates of a cluster and a structure. After the initial stage of "learning" clusters were "refined" to obtain optimal distribution of structures in clusters. As long as peptides remain bound to the fibril, this approach results in meaningful distribution of clusters with distinct structural and energetic characteristics.

The peptide structural clusters were defined with the cut-off radii R_c=29 Å (F5) and 20 Å (F6). To select R_c we scan the values in the range from 10 to 50 Å. Small R_c lead to the appearance of numerous structurally and energetically similar clusters, whereas large R_c result in merging structurally distinct clusters.

Calculation of structural probes

To characterize the interactions between the peptide F5 (F6) with the fibril (F1-F4), we computed the number of side chain contacts. A side chain contact between amino acids *i* and *j* is assumed to be formed, if the distance between the centers of mass of their side chains is less than 6.5 Å. Backbone hydrogen bonds (HB) between NH and CO groups

were assigned according to Kabsch and Sander [28]. This definition postulates that HB is formed when the total electrostatic energy of interactions between C, O, N, and H atoms is less than 0.5 kcal/mol. In all, we considered two classes of HBs. The first includes any backbone HB formed between the edge peptide and the fibril. The second class is restricted to those HBs between the edge peptide and fibril, which simultaneously occur in a given structure and in the intact fibril-like conformation. We termed such HBs fibril ones. This class of HBs is formed by the edge peptides F5 and F6 in Figure 2.

Solvent accessible surface area (ASA) of a protein is defined as the protein surface exposed to solvent. The exposure of polar residues is energetically favorable, whereas large ASA of hydrophobic residues typically increases the free energy of the system. Therefore, ASA is a useful quantity which can distinguish stable protein conformations. We used program STRIDE [29] to calculate the ASA of individual residues.

To analyze fluctuations in local structure we computed the distances between selected pairs of atoms and obtained their probability distributions. In particular, we considered the distances between hydrophobic residues Phe19 in the peptides F5 (F6) and F3 (F4). Also selected were Phe20 and the pair of charged amino acids Asp23 and Lys28. In all these cases one amino acid was located in the edge peptide (F5 or F6), whereas the other was taken from the fibril peptide (F3 or F4, respectively).

Throughout the thesis averaging is performed over time or over the structures assigned to a given cluster.

RESULTS

We have analyzed the 100 ns MD trajectory using the probes described in Materials and Methods. In this section we present the results of our analysis.

Hydrophobic peptide-fibril contacts

In the course of simulations we computed the number of side chain contacts formed by individual residues in the edge peptides F5 or F6. To characterize hydrophobic interactions we restricted our computations to the contacts formed between hydrophobic residues. Figure 3 shows the average number of contacts formed by individual residues in F5 and F6.

In general, the distribution of contacts for both peptides is similar. Most of the contacts occur near the centers of either β1 or β2 strands (Figure 1). The absence of contacts in the turn region between β1 and β2 (residues 24-29) is due to the lack of hydrophobic residues in that sequence region. Though the patterns of contacts made by both peptides are similar, on an average F6 appears to form more contacts than F5. This observation reflects the fact that F6 forms larger number of hydrophobic contacts in the intact fibril conformation than F5.

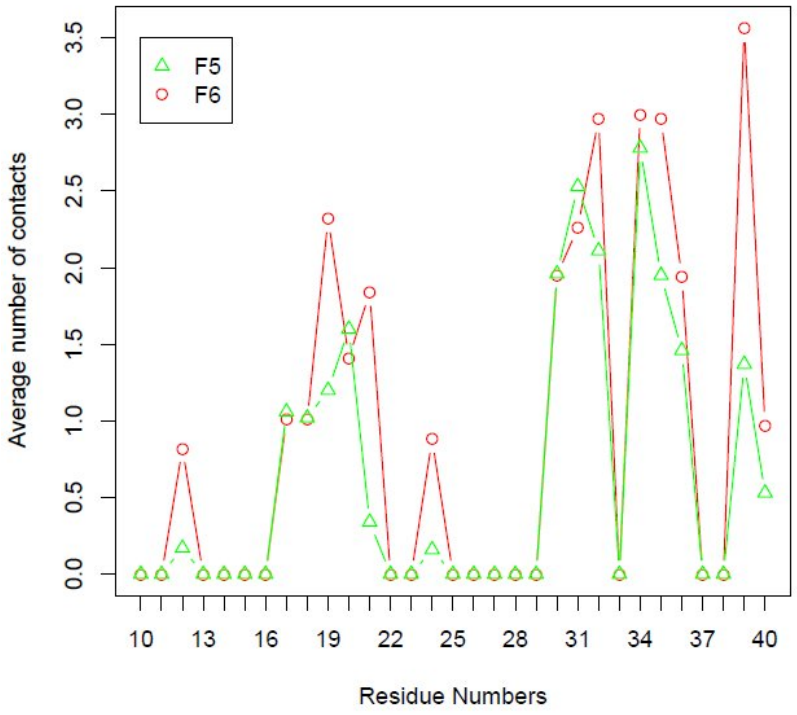

Figure 3: Number of hydrophobic contacts formed by the edge peptides F5 and F6. The data are obtained by averaging over 100ns trajectory.

RMSD calculations

RMSD (root mean square deviation) is a probe, which measures the fluctuations of protein atoms with respect to the reference structure. As a reference structure we used the first conformation from the production simulations obtained at 0 ns. In this structure the edge peptides F5 and F6 adopt unperturbed fibril-like conformations. The computations of RMSD were performed using VMD [30] (see Appendix). Because the

fibril fragment consisting of the peptides F1-F4 is constrained, there is no need to remove translational and rotational motions in the computations of RMSD. To verify this point, we computed RMSD with and without the removal of these motions. The resulting RMSD plots were almost identical.

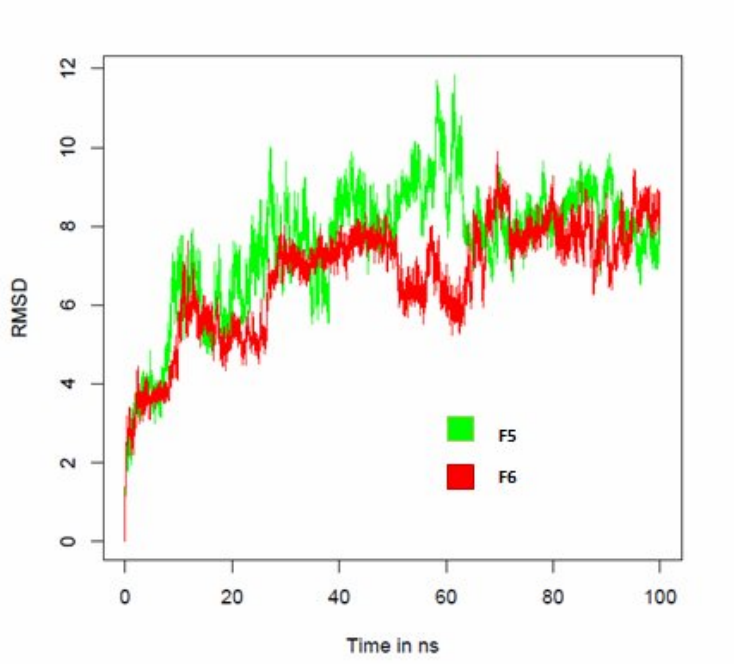

Figure 4: RMSD of Ca atoms in the peptides F5 and F6 as a function of time.

Figure 4 shows the time dependence of the RMSD for the edge peptides F5 and F6. The plot suggests that after about 50 ns the edge peptides deviate from the fibril conformation by roughly 8 Å. The RMSD baseline does not increase until the end of trajectory suggesting that F5 and F6 sample partially destabilized bound conformations. The difference in the RMSD of F5 and F6 ΔRMSD=RMSD(F6)-RMSD(F5) is plotted in Figure 5. In this plot the largest difference between the two is ~6Å. Thus, RMSD plots

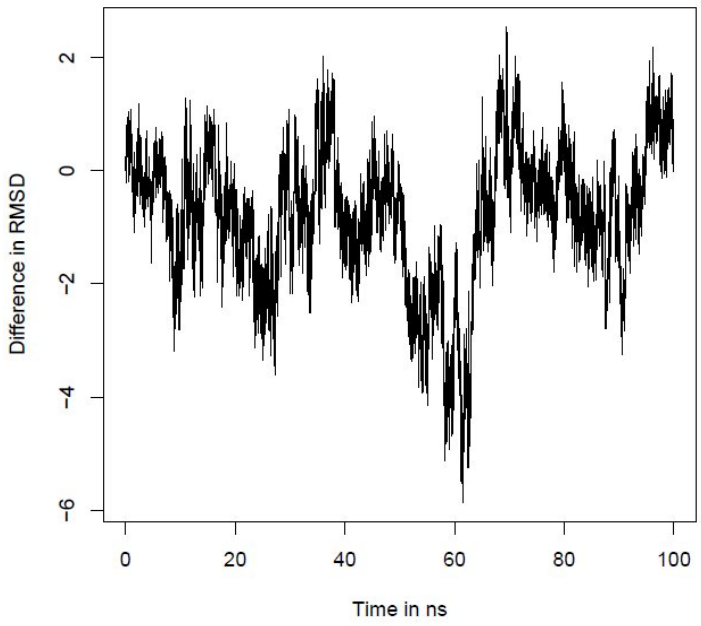

Figure 5: Difference RMSD for Ca atoms for F5 and F6 as a function of time.

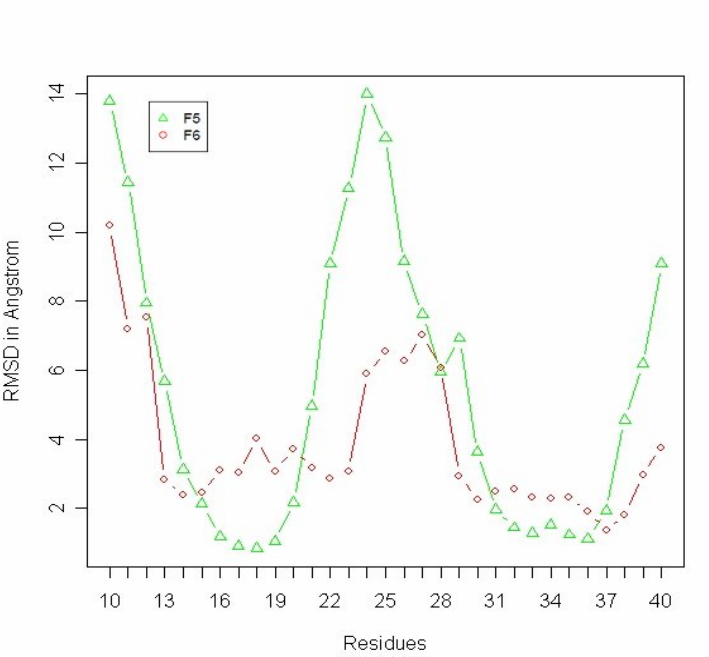

Figure 6: RMSD values of F5 and F6 Ca atoms. The RMSD values are obtained by averaging over the 100 ns trajectory.

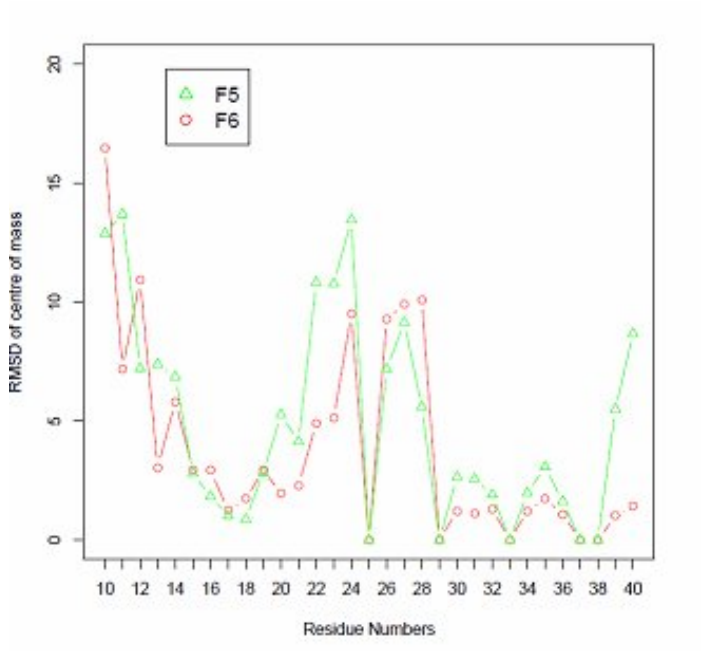

Figure 7: RMSD of the centers of mass of F5 and F6 side chains. The RMSD values are obtained by averaging over the 100 ns trajectory. The RMSD values for Gly are set to zero.

indicate that the unbinding patterns on ABF5 and ABF6 are similar, but F5 experiences somewhat larger fluctuations while being bound to the fibril edge. These findings are consistent with the computations of hydrophobic contacts above.

To further investigate peptide fluctuations, we computed the RMSDs of individual Ca atoms of the residue from F5 and F6. Figure 6 demonstrates that the midpoints of β1 and β2 strands are rigid experiencing the fluctuations of ~2Å, but the C- and N-terminals and the turn region are highly mobile (RMSD is up to 20 Å). The distribution of fluctuations

in both peptides is very similar, although as already demonstrated above F5 reveals larger structural motions.

In-depth knowledge about unbinding process can be obtained by comparing the RMSD fluctuations of the side-chains in F5 and F6. These computations are important, because in amyloid fibrils side chains play important role in packing [31]. We calculated the centers of mass of each residue and the corresponding RMSDs are plotted in Figure 7. Similar to the patterns seen in Figure 6, the turn and the terminals have the highest fluctuations. Although the RMSDs of side chains are generally similar to those of Ca, Figure 7 shows that the C-terminal is relatively rigid compared to the N-terminal. We explain this result by the observation that in Figure 2 the C-terminal of Abeta peptide is buried in the laminated fibril structure.

Analysis of the contacts between residues

As in native proteins hydrophobic effect plays an important role in the formation of amyloid fibrils. We investigated the formation of contacts between pairs of hydrophobic residues (see Methods). The distribution of distances between the side chains of Phe19 *P(r)* are shown in Figure 8 for the peptides pairs F5-F3 and F6-F4. It is clear that the typical distance between Phe19 is about 5Å and no appreciable differences are observed between F5 and F6. In both cases, the distribution *P(r)* is unimodal and given that *P(r)* reaches maximum at $r \approx 5Å < 6.5Å$ the Phe19-Phe19 contact is mostly formed.

In contrast, the distributions of distances between Phe20 residues in the peptide pairs F5-F3 and F6-F4 are qualitatively different. In Figure 9a *P(r)* for the peptides F5-F3 has bimodal distribution, whereas there is still only one peak for F6 and F4 in Figure 9b. Figure 9a suggests that the contact between Phe20 residues toggles between "on" and "off" states. The "off" state corresponds to the second peak at ~7Å. The "on" state occurs when *r*~5Å. Therefore, unlike other hydrophobic contacts considered so far the interpeptide hydrophobic contact Phe20-Phe20 between the edge peptide F5 and the fibril is relatively unstable. This observation is consistent with lower binding affinity of F5 compared to F6

The probability distribution of the distances between the residues Asp23 (from F3 or F4) and Lys28 (from F5 or F6) shows that this electrostatic contact is frequently disrupted. Indeed, the maximum in *P(r)* is reached at *r*~7Å suggesting that the salt bridge between oppositely charged Lys28 and Asp23 is less stable than the aromatic Phe-Phe hydrophobic tethers (Figures 8 and 9). The stability of Lys-Asp contact, which is believed to be important for fibril stability, is further analyzed below.

.

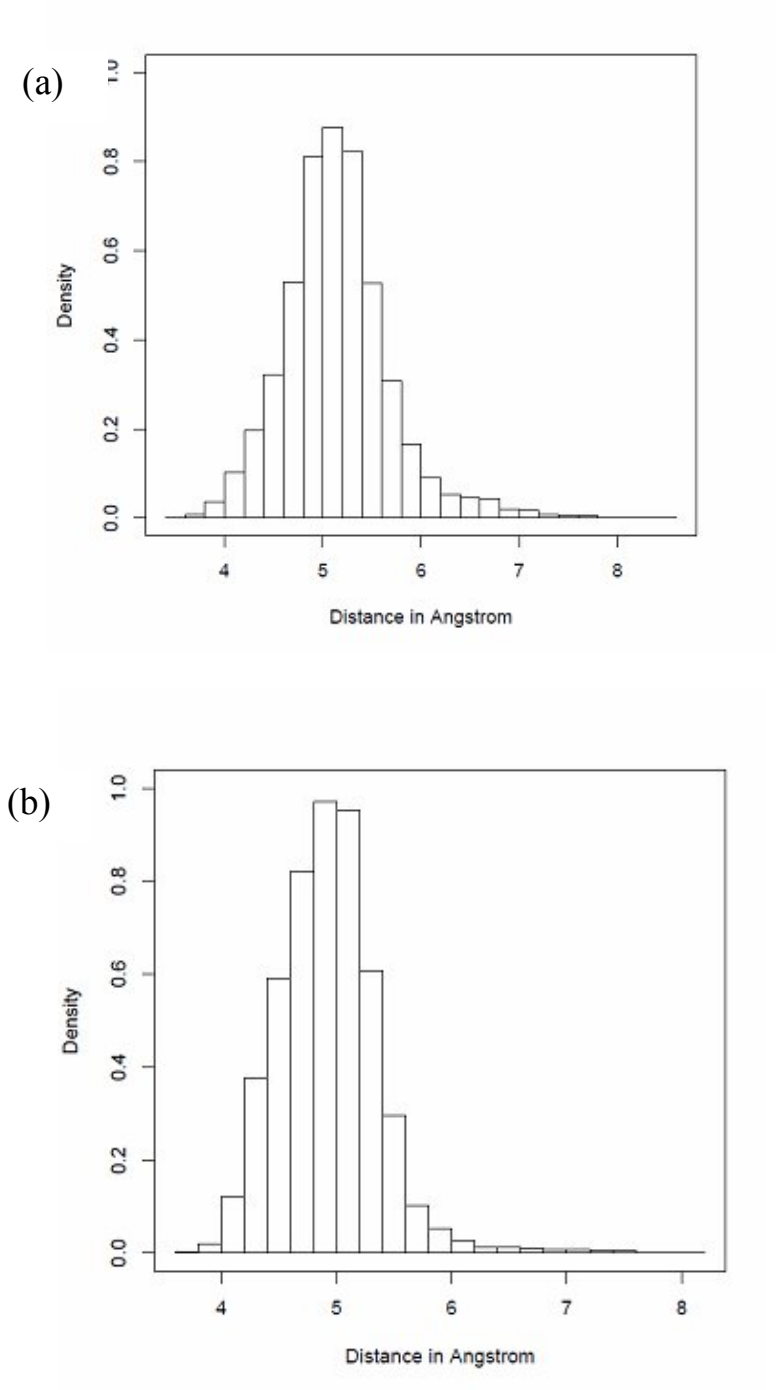

Figure 8: Probability distribution of the distance between the residues Phe19 in F5 and F3 (a) and in F6 and F4 (b).

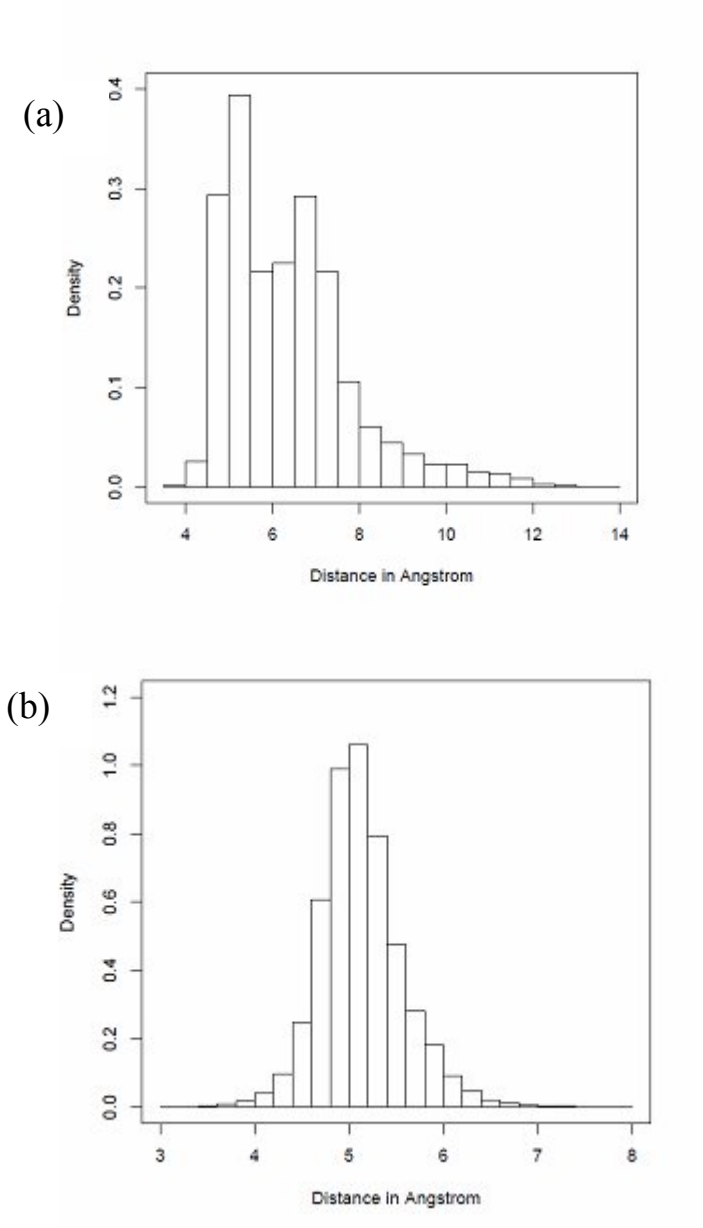

Figure 9: Probability distribution of the distance between the residues Phe20 in F5 and F3 (a) and in F6 and F4 (b).

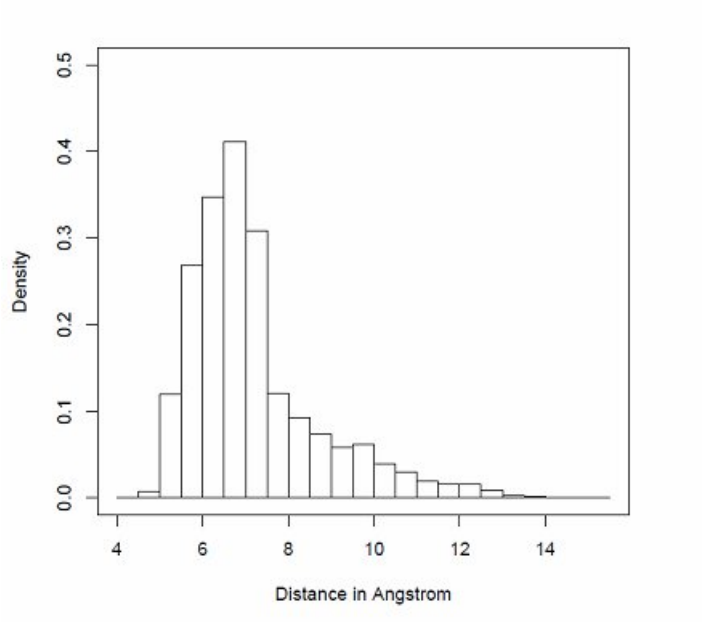

Figure 10: Probability distribution of the distance between the residues Asp23 and Lys28 in F5 and F3

Analysis of solvent accessible surface area

Using the program STRIDE we computed the total solvent accessible surface area (tASA) of the edge peptides F5 and F6. tASA values represent the sum of ASA of individual residues in polypeptide. The time dependence of tASA is shown in Figure 11. The plot suggests both edge peptides undergo structural transitions, which reduce their exposure to solvent. However, while F6 reveals roughly monotonically declining tASA, there are considerable fluctuations in tASA of F5, which changes by up to 20%. As we demonstrated above the edge peptide F5 forms relatively weak interactions with the fibril compared to F6. Computation of ASA for selected individual residues also indicates that considerable fluctuations in solvent exposure of amino acids occur. For example, ASA of

Lys28 in the edge peptide F5 is shown in Figure 12. The exposure of this charged residue fluctuates by more than 50% in the course of MD trajectory. This observation is in agreement with the flickering of Lys28-Asp23 salt bridge observed above.

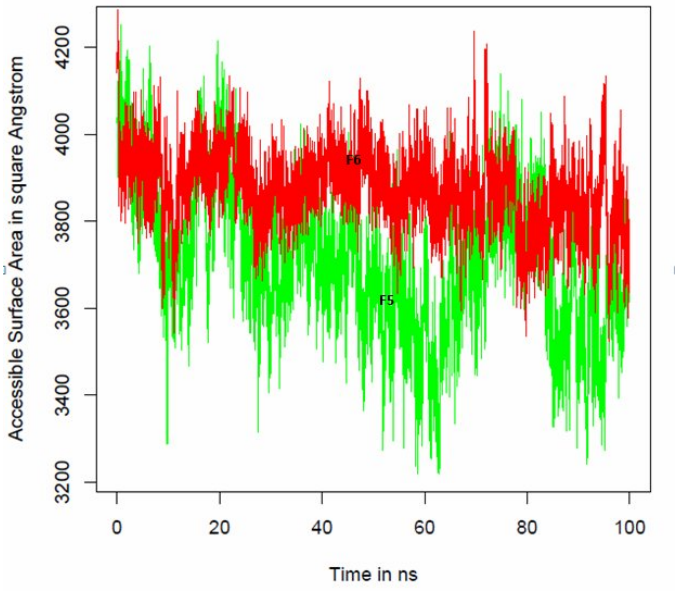

Figure 11: Total accessible surface area tASA of F5 (in green) and F6 (in red) as a function of time.

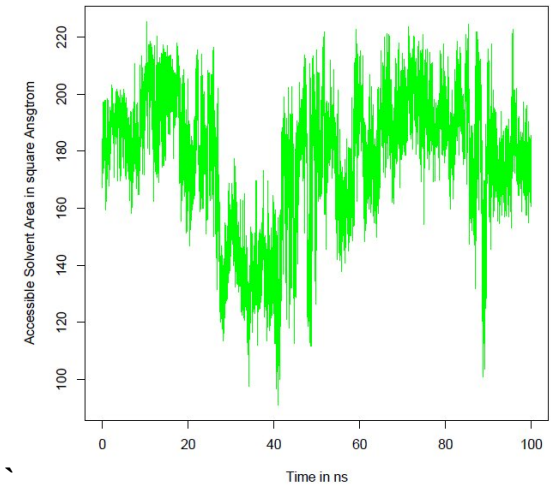

Figure 12: ASA of residue Lys28 is plotted as a function of time.

Analysis of unbinding of the edge peptides using clustering technique

As described in the Methods we clustered the conformations of the edge peptides F5 and F6 sampled in 100 ns MD trajectory. Clustering algorithm groups conformations into small number of distinct structural clusters. Appearance of clusters along the MD trajectory formed by the peptides F5 and F6 is shown in Figure 13. As a rule, cluster number correlates with MD time. For example, CL1 appears at the beginning of the trajectory, while "high" number clusters, such as CL4 or CL5, occur later. The cluster characteristics are given in Table 1.

It is seen that all clusters except CL5 (F6) contain large number of structures (more or about 100 structures). Because the total number of saved structures is 10,000, their probability of occurrence is in excess of 0.1. From the analysis of structural characteristics of F5 clusters the following conclusions can be made. First, there is little change in the number of hydrophobic contacts C_{HH} across all clusters. The total number of HBs N_{hb} also experience relatively small variations (from 14 to 16). Interestingly, N_{hb} differs from the number of fibril HBs by no more than one bond. Therefore, in general F5 does not form new HBs, which are not already present in the fibril structure. Second, significant differences between the clusters can be seen if one considers the number of HBs formed by the strands β1 and β2. For example, in the cluster CL1 the ratio $N_{hb}(\beta 1)/N_{hb}(\beta 2)$ is about 0.95, but it is reduced to 0.6 in the clusters CL2 and CL3.

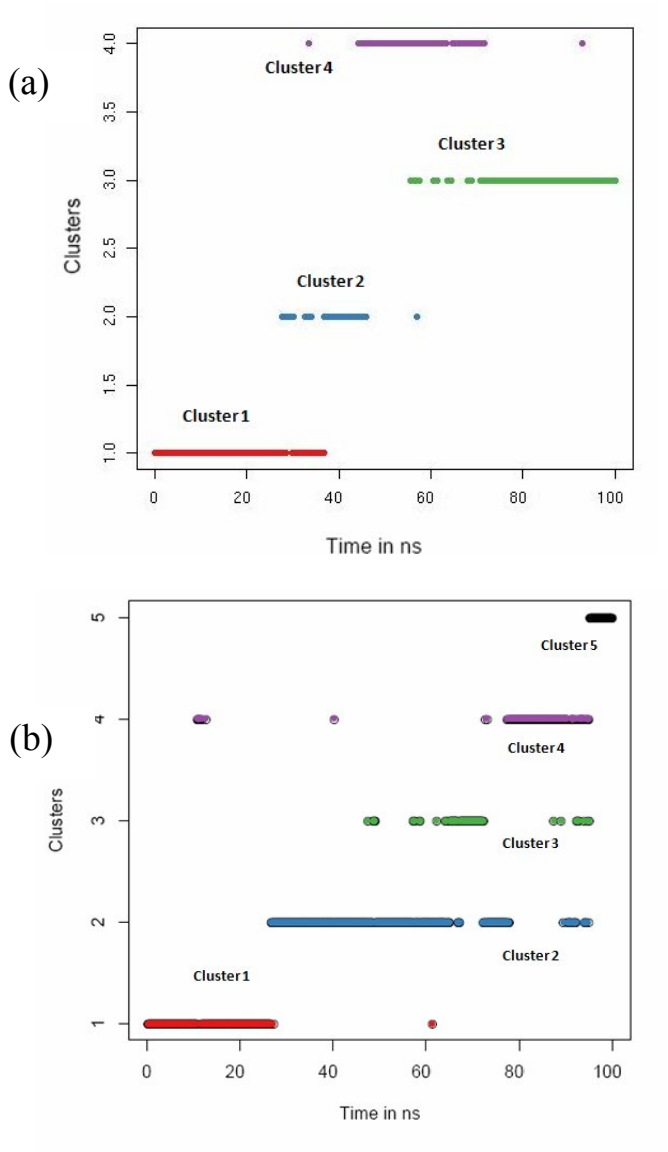

Figure 13: Distribution of cluster formation as a function of time for F5 (a) and F6 (b)

Similar trend is observed in the number of hydrophobic contacts formed by the strands β1 and β2. The ratio $C_{HH}(\beta 1)/C_{HH}(\beta 2)$ is approximately 0.5 in CL1, but it becomes only 0.3 in CL2-CL4. Finally, there are significant variations in the number of electrostatic contacts C_{EL}. In CL1 and CL2 the average number of C_{EL} is 0.1, but in CL3 it is ~0.7.

Consequently, the electrostatic contacts are formed with the average probability of about 0.3.

The analysis of clusters formed by F6 reveals that they are structurally similar to those of F5. As in F5 the total numbers of hydrophobic contacts C_{HH} and hydrogen bonds N_{hb} between the edge peptide F6 and the fibril show relatively small variations. For example, excluding rare cluster CL5 N_{hb} changes by no more than five that constitutes 22% decrease compared to N_{hb} in the cluster CL1. More importantly, as in F5 a significant loss of HBs is observed in the strand β1. In CL1-CL3 $N_{hb}(\beta 1) \approx N_{hb}(\beta 2)$, whereas in CL4 $N_{hb}(\beta 1)/N_{hb}(\beta 2) < 0.7$. Similarly, hydrophobic contacts are predominantly lost within the strand β1. For instance, in CL1-CL3 $C_{HH}(\beta 1)/C_{HH}(\beta 2) \approx 0.5$, but this ratio is less than 0.4 in CL4. In all clusters except CL1 about one electrostatic contact is present. Consequently, the average probability for forming salt bridge between F6 and the fibril is 0.75 that is much higher than for F5.

Table 1: Characteristics of structural clusters of the edge peptides

	Total number of clusters	Cluster and number of structures	Total number of electrostatic interactions (C_{EL})	Hydrophobic contacts with the fibril	Total number of hydrophobic contacts (C_{HH})	Hydrogen bond with fibril (N_{hb})
F5	4	Cluster 1:343	0.1	β1:6.4	19.7	β1:7.6
				β2:12.4		β2:8.0
		Cluster 2:113	0.1	β1:4.9	21.1	β1:5.2
				β2:14.8		β2:7.9
		Cluster 3:322	0.7	β1:4.9	21.2	β1:5.0
				β2:16.1		β2:8.1
		Cluster 4:222	0.4	β1:4.7	19.3	β1:5.0
				β2:13.7		β2:6.9
F6	5	Cluster 1:258	0.1	β1:7.8	27.6	β1:10
				β2:18.1		β2:9.0
		Cluster 2:442	1.0	β1:9.1	29.9	β1:8.5
				β2:18.6		β2:8.8
		Cluster 3:92	0.9	β1:9.0	28.8	β1:7.8
				β2:17.8		β2:7.9
		Cluster 4:159	1.0	β1:8.3	29.6	β1:6.5
				β2:20.1		β2:8.9
		Cluster 5:49	1.1	β1:5.1	25.2	β1:4.3
				β2:18.3		β2:8.9

DISCUSSION AND CONCLUSION

In this thesis we used molecular dynamics to investigate the early stages of unbinding of edge Abeta peptides from amyloid fibril. The unbinding process is caused by temperature-induced structural fluctuations and can be thought as an elementary step in the molecular recycling occurring between fibrillized and soluble Abeta species [6]. Taken together our analysis has led us to the following four conclusions:

1. Early steps in unbinding can be observed on 100 ns timescale, but complete unbinding is likely to take place on much longer timescale. Previous MD simulations using short seven-residue fragments of Abeta peptide demonstrated that the complete unbinding timescale is at least 1 μs [32]. Abeta peptide considered here is four times longer, so it is conceivable that the unbinding timescale may approach millisecond timescale as it follows from the recent experiments on Abeta fibril growth [33-35].
2. The unbinding pathway sampled by the edge peptides F5 and F6 is qualitatively similar. RMSD computations indicate that the residues located in the middle of the strands β1 and β2 are rigid, whereas the turn region and the N- and C-terminals of Abeta peptides are highly mobile. In agreement with this unbinding pathway the interpeptide salt bridge Lys28-Asp23 located in the turn region is flickering and appears to be less stable than the Phe-Phe hydrophobic tethers in

the middle of the strand β1. Therefore, unbinding process begins with fraying of β-strand ends and their gradual peeling off from the fibril.

3. Most of peptide-fibril interactions lost upon unbinding are localized within the N-terminal of Abeta peptide, which includes the strand β1. Clustering analysis reveals that β1 loses from 30 to 40% of the hydrogen bonds linking it to the fibril, whereas almost all HBs formed by β2 remain intact. These findings are supported by the analysis of hydrophobic contacts formed by both β-strands and by the computation of RMSD of side chains. Therefore, unbinding process may not only involve fraying of the β-strands, but is also likely to proceed with initial dissociation of the strand β1. This suggestion is consistent with the burial of the C-terminal in the Abeta fibril structure and exposure of the N-terminal [22].
4. Despite the overall similarity of F5 and F6 unbinding RMSD computations and clustering analysis suggest that the edge peptide F6 is more tightly coupled with the fibril compared to F5. Indeed, the difference between Ca RMSD values computed for F6 and F5, ΔRMSD, is mostly negative indicating that F5 experiences larger fluctuations than F6. It appears that this finding can be explained by different initial positions of the peptides F5 and F6 on the fibril edge. In the intact fibril conformations the peptide F6 forms slightly larger number of HBs and hydrophobic interactions with the fibril. Therefore, details of unbinding may depend on the original location of the peptide on the fibril edge.

Our simulations provide new molecular details on the process of amyloidogenesis linked to Alzheimer's disease. Future simulations will probe unbinding process on longer timescales and by using more advanced sampling algorithms such as replica exchange simulations.

APPENDIX

To load files into VMD

```
 set runmin 0
set runmax 99
for {set i $runmin} {$i <= $runmax} {incr i} {
  if {$i<10 } { set filename "abf_quench010$i.pdb" }
  if {$i>=10} {set filename "abf_quench01$i.pdb" }
  animate read pdb $filename waitfor all
}
```

To remove translation of the frames in the trajectory

```
##############################################################################
# Prints the RMSD of the protein atoms between each timestep and the first \timestep for
# the given molecule id (default: top)
proc print_rmsd_through_time {{mol top}}
set outfile [open rmsd_66_CA.dat w]
# use frame 0 for the reference
set reference [atomselect $mol "segname ABF6" frame 0]
# set subreference [atomselect $mol "segname ABF5" frame 0] the frame being
#  compared
  set compare [atomselect $mol "segname ABF6"]
 # set subcompare [atomselect $mol "segname ABF"]
```

```
  set num_steps [molinfo $mol get numframes]
for {set frame 0} {$frame < $num_steps} {incr frame} {
# get the correct frame
$compare frame $frame
#$subcompare frame $frame
# compute the transformation
set trans_mat [measure fit $compare $reference]
#puts  $outfile $trans_mat
# do the alignment
 $compare move $trans_mat
# compute the RMSD
set rmsd [measure rmsd $compare $reference]
#set rmsd [measure rmsd $subcompare $subreference]
# print the RMSD
puts $outfile $rmsd
}
close $outfile
}
#########################################################################
#THIS PROGRAM CALCULATES THE CONTACTS BETWEEN EVERY RESIDUE
#IN
# THE TRAJECTORY.2 FILES "All_CA" AND "Contacts.TXT" IS
```

```
# GENERATED.
        ##########################################################################
use strict;
my $file;
our %Hash_contacts;
our $fin=0;
open(OUT, ">abf5_ca_trans.pdb");
for (my $i=0;$i<100;$i++)
        {
        if ($i<10)
        {
                $file="abf_quench010".$i.".pdb";
                print "$file\n";
                read_file($file);
        }
        else
        {
                $file="abf_quench01".$i.".pdb";
                print "$file\n";
                read_file($file);
        }
        }#End of for loop
```

```
print_contacts(\%Hash_contacts);
#####################################################################
#SUBROUTINE:READ_FILE
#FUNCTION:PARSES ALL THE TRAJECTORY FILES, EXTRACTS THE 'CA'
#POSITION ATOMS AND WRITES TO FILE 'All_CA', SENDS EACH
#TRAJECTORY TO SUBROUTINE DISTANCE
#####################################################################
sub read_file
{
my ($file)=shift;
my @traj;
my @line_pos;
my @all_CA;

open(IN, $file);
#Open this file to write all the CA values in all the trajectories
#open(OUT, ">All_CA");
my $count=0;
while(<IN>)
{

        chomp;
```

```
        my @val=split(/\s+/,$_);
        foreach my $val_1(@val)
        {
                if (($val_1 eq "CA") || ($val_1 eq "END"))
                {
                        print OUT "$_\n";
                }
        }
}
#close(OUT);
close(IN);
open(IN1, "All_CA");
while(<IN1>)
{
        chomp;
        push(@all_CA,$_);
}
my $j=0;
#Trying to find out the posiions of END
for (my $i=0;$i<=$#all_CA;$i++)
        {
        if ($all_CA[$i] eq "END")
```

```
                {
                push(@line_pos,$i);
                }
                else
                {
                }
        }
#Send to subroutine distancE() to calculate distance
my @array_traj="";
my $num=0;
my $ii=0;
my $m=0;
my @dist_fin;
my $g;
for (my $i=0;$i<=$#line_pos;$i++)
        {
        push(@array_traj,@all_CA[$m..$line_pos[$i]]);
        @dist_fin=distance(\@array_traj);
        $m=$line_pos[$i]+1;
        @array_traj=" ";
        $Hash_contacts{$fin}=[@dist_fin];
        $fin++;
```

```
        $g=$i;
        }
#JUST TO CHECK THE VALUES IN THE HASH!!
my @f1;
for my $family ( keys %Hash_contacts )
        {
         @f1=@{ $Hash_contacts{$family}};
        #print $#f1;
#       print "$family: @{ $Hash_contacts{$family} }\n";
        }

}#END OF SUB
#########################################################################
#SUBROUTINE:DISTANCE                                    #
#FUNCTION:EXTRACTS ALL THE X,Y,Z CO-ORDINATES, CALCULATES THE
#DISTANCE BETWEEN THESE CO-ORDINATES,MAKES THE DECISION OF A
#CONTACT OR NOT BY THE CUT-OFF, RETURNS A ARRAY @FINAL          #
#########################################################################
sub distance
{
my ($xdist,$ydist,$zdist, $diff);
my @final="";
```

```
my ($posCA)=@_;
pop(@$posCA);
my $num=0;
my @pos_CA_cood;

#@arr contains one trajectory at a time
my @arr=@$posCA;
my $i=0;
foreach my $line(@arr)
        {
        my @pro=split(/\s+/,$line);
        $pos_CA_cood[$i][0]=$pro[6];
        $pos_CA_cood[$i][1]=$pro[7];
        $pos_CA_cood[$i][2]=$pro[8];
        $i++;
        }
for(my $i=1;$i<$#pos_CA_cood;$i++)
{
 for(my $j=$i+1;$j<=$#pos_CA_cood;$j++)
        {
        $xdist=($pos_CA_cood[$i][0]-$pos_CA_cood[$j][0])**2;
        $ydist=($pos_CA_cood[$i][1]-$pos_CA_cood[$j][1])**2;
```

```
        $zdist=($pos_CA_cood[$i][2]-$pos_CA_cood[$j][2])**2;
        $diff=sqrt($xdist+$ydist+$zdist);
        #$final[$num]=$diff;
        #$num++;
        if ($diff <7.00)
                {
        #               push(@final,1)
                        $final[$num]=1;
                        $num++;
                }
                else
                {
        #               push(@final,0);
                        $final[$num]=0;
                        $num++;
                }
        }
}#End of i loop
return @final;
}#END OF SUB
#########################################################################
#SUBROUTINE:PRINT_CONTACTS
```

```
#FUNCTION:PRINTS THE %HASH_CONTACTS WHICH HAS ALL THE
#CONTACTS  #
#INTO THE FILE CALLED 'Contacts.TXT'
#SOME LOOPS ARE HARD-CODED, MAY NEED TO BE CHANGED
#########################################################################
sub print_contacts
{
my ($hash)=@_;
my %Hash_con=%$hash;
my $hash_size;
for my $keys(keys %Hash_con)
{
        $hash_size= $#{$Hash_con{$keys}};
}
my @final_contacts;
my @arr_keys =keys( %Hash_con);
my $sum=0;
for(my $i=0;$i<=$hash_size;$i++)
{
        foreach my $k(@arr_keys)
        {
                $sum=$sum+$Hash_con{$k}[$i];
```

```
        }
$final_contacts[$i]=$sum;
$sum=0;
}
#print"FINAL: $#final_contacts";
open(FIN, ">Contacts.TXT");
my @fi;
foreach( @final_contacts)
{
my $n=$_/10000;
push(@fi,$n);
#my @fi= @final_contacts/10000;
}
my $start=0;
for (my $i=0;$i<185;$i++)
{
        for(my $j=$i+1;$j<=185;$j++)
        {
                my $ii=$i+1;
                my $jj=$j+1;
                print FIN "$ii\t$jj\t$fi[$start]\n";
                $start++;
```

```
            }
}
close(FIN);
}#END OF SUB
```

REFERENCES

REFERENCES

1. Falk RH, Comenzo RL, Skinner M: The systemic amyloidoses. *The New England journal of medicine* 1997, 337(13):898-909.

2. Masters CL, Simms G, Weinman NA, Multhaup G, McDonald BL, Beyreuther K: Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 1985, 82(12):4245-4249.

3. Hardy J, Selkoe DJ: The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science (New York, NY* 2002, 297(5580):353-356.

4. Martin JB: Molecular basis of the neurodegenerative disorders. *The New England journal of medicine* 1999, 340(25):1970-1980.

5. Wetzel R: Ideas of order for amyloid fibril structure. *Structure* 2002, 10(8):1031-1036.

6. Dobson CM: Protein folding and misfolding. *Nature* 2003, 426(6968):884-890.

7. Selkoe DJ: Folding proteins in fatal ways. *Nature* 2003, 426(6968):900-904.

8. Eisenberg D: Structure from Micro-crystals of the Cross-Beta spine of amyloid fibril. *The European Synchroton Radiation Facility* 2005.

9. Nelson Rea: Structure of the cross-bold beta spine of amyloid-like fibrils. *Nature* 2005, 435.

10. Dobson CM: The structural basis of protein folding and its links with human disease. *Philosophical transactions of the Royal Society of London* 2001, 356(1406):133-145.

11. W. A. Eaton. ea: Kinetics and dynamics of loops, -helices, -hairpins, and fast-folding proteins. *Acc Chem Res 31* 1998.

12. Fersht AR: Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus

mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 2000, 97(4):1525-1529.

13. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M: Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in biochemical sciences* 2000, 25(7):331-339.

14. Cheung MS, Garcia AE, Onuchic JN: Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences of the United States of America* 2002, 99(2):685-690.

15. Kusumoto Y, Lomakin A, Teplow DB, Benedek GB: Temperature dependence of amyloid beta-protein fibrillization. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(21):12277-12282.

16. Tseng BP, Esler WP, Clish CB, Stimson ER, Ghilardi JR, Vinters HV, Mantyh PW, Lee JP, Maggio JE: Deposition of monomeric, not oligomeric, Abeta mediates growth of Alzheimer's disease amyloid plaques in human brain preparations. *Biochemistry* 1999, 38(32):10424-10431.

17. Benzinger TL, Gregory DM, Burkoth TS, Miller-Auer H, Lynn DG, Botto RE, Meredith SC: Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with residues in exact register. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95(23):13407-13412.

18. Meersman F, Dobson CM: Probing the pressure-temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochimica et biophysica acta* 2006, 1764(3):452-460.

19. Dirix C, Meersman F, MacPhee CE, Dobson CM, Heremans K: High hydrostatic pressure dissociates early aggregates of TTR105-115, but not the mature amyloid fibrils. *Journal of molecular biology* 2005, 347(5):903-909.

20. Niraula TN, Konno T, Li H, Yamada H, Akasaka K, Tachibana H: Pressure-dissociable reversible assembly of intrinsically denatured lysozyme is a precursor for amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101(12):4089-4093.

21. Petkova AT, Yau WM, Tycko R: Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry* 2006, 45(2):498-512.

22. Sciarretta KL, Gordon DJ, Petkova AT, Tycko R, Meredith SC: Abeta40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* 2005, 44(16):6003-6014.

23. Takeda, T. and Klimov D. K. : Temperature induced dissociation of Abeta monomers from amyloid fibrils. *Biophysical J* 2008, 95.

24. Torda, AE, Gunsteren WFv.: Algorithms for clustering molecular dynamics configuration. *Journ Comp Chem* 1994, 15.

25. Unger R, Harel D, Wherland S, Sussman JL: A 3D building blocks approach to analyzing and predicting structure of proteins. *Proteins* 1989, 5(4):355-373.

26. Rooman MJ, Rodriguez J, Wodak SJ: Automatic definition of recurrent local structure motifs in proteins. *Journal of molecular biology* 1990, 213(2):327-336.

27. Klimov DK and Thirumalai, D.: Progressing from folding trajectories to transition state ensemble in proteins. *Chem Phys* 2004, 307:251-258.

28. Kabsch W, Sander C: Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983, 22(12):2577-2637.

29. Frishman D, Argos P: Knowledge-based protein secondary structure assignment. *Proteins* 1995, 23(4):566-579.

30. Humphrey W, Dalke A, Schulten K: VMD: visual molecular dynamics. *Journal of molecular graphics* 1996, 14(1):33-38, 27-38.

31. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC: Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(2):315-320.

32. Takeda T, Klimov DK: Dissociation of Abeta(16-22) amyloid fibrils probed by molecular dynamics. *Journal of molecular biology* 2007, 368(4):1202-1213.

33. Chatani E, Kato M, Kawai T, Naiki H, Goto Y: Main-chain dominated amyloid structures demonstrated by the effect of high pressure. *Journal of molecular biology* 2005, 352(4):941-951.

34. Sasahara K, Naiki H, Goto Y: Kinetically controlled thermal response of beta2-microglobulin amyloid fibrils. *Journal of molecular biology* 2005, 352(3):700-711.

35. Yamaguchi K, Takahashi S, Kawai T, Naiki H, Goto Y: Seeding-dependent propagation and maturation of amyloid fibril conformation. *Journal of molecular biology* 2005, 352(4):952-960.

CURRICULUM VITAE

Pamela H. Mishra graduated with a degree in Electronics Engineering from Mumbai University in 2004. She did her diploma in Bioinformatics from Pune University in 2005 after which she graduated as Masters in Bioinformatics and Computational Biology in 2008 from George Mason University, VA. She is keen to pursue her Phd in a few years.