CHAPTER 7

Examination of the intrinsic disordered regions present in the $A\beta_{1-42}$ peptide

Examination of the intrinsic disordered regions present in the $A\beta_{1-42}$ peptide

7.1. Abstract:

Proteins are composed of amino acid residues that exhibit several structure levels. Many proteins or regions of proteins lack a stable 3-D structure, and are rather intrinsically disordered, for which the term has been coined as/for which they are termed as IDPs/IDRs, respectively. As the IDPs exist as an ensemble of rapidly interconverting conformations, their structural and functional characterization is a special challenge. Development of methods to predict the disordered regions in a protein in advance is getting a significant attention. Our protein of interest $A\beta_{1-42}$ peptide is an IDP that misfolds and aggregates to form senile plaques leading to the AD. In this Chapter, we have used disorder predictors to identify the disordered regions in $A\beta_{1-42}$ peptide. We found $A\beta_{1-42}$ peptide to have disordered regions in the CHC (17-22), and C-terminal regions (30-42) of $A\beta_{1-42}$ peptide.

7.2. Introduction:

Our traditional view of protein structure and function is the structure–function paradigm according to which a protein folds into a stable 3-D structure and impart its biological functions. However, almost 20 years ago it was suggested that many proteins or regions of proteins lack a stable 3-D structure, and are rather intrinsically disordered. These proteins are now termed as IDPs and the corresponding disordered regions in the protein are termed as IDRs [102, 103]. The word "intrinsically" indicates a sequence dependent characteristic [104]. The thermodynamic definition of disordered regions in a protein is the random coil structural state. The structural disorder which is prevalent in all organisms is found to play roles in cellular signaling [105] and regulation, due to which IDPs are implicated in diseases [106] and are sought as important drug targets [107].

As the IDPs exist as an ensemble of rapidly interconverting conformations, their structural and functional characterization is a special challenge. Although they cannot be directly characterized by X-ray crystallography, there are a variety of techniques that can report their highly dynamic structural state. NMR and X-ray crystallography provide site-specific information, whereas far-UV CD and size-exclusion chromatography provide qualitative and global information [102]. The current best

structural descriptions of IDPs/IDRs are solved by a combination of experimental and computational approaches.

Development of methods to predict in advance the disordered regions in a protein is getting a significant attention. The capability of a method to predict the aggregation propensity of a protein from its sequence will be useful to control the unwanted protein depositions through specific sequence targeted therapeutics. $A\beta_{1-42}$ peptide, that aggregates to form senile plaques, is one of the IDPs associated with neurodegenerative disease AD in particular. In this study, two reputable disorder predictors, AMYLPRED2 [108] and DisEMBL [109] were used to assess the disordered region in $A\beta_{1-42}$ peptide. AMYLPRED2 employs a consensus of different methods that have been specifically developed to predict features related to the formation of amyloid fibrils. Similarly, DisEMBL includes features from three predictors to detect the disorderness of a protein.

7.3. Materials & Methods:

The initial FASTA sequence for the examination of the intrinsic disordered regions of $A\beta_{1-42}$ peptide (PDB ID: 1IYT) [211] was used from the Protein Data Bank [212]. The disorder predictors AMYLPRED 2 [108] and DisEMBL [109] were used to investigate the disordered region.

AMYLPRED2 uses a consensus algorithm for the prediction of amyloidogenic regions from the sequence. It combines the result from 11 different softwares and algorithms: Aggrescan, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conformational Energy, NetCSSP, Pafig, SecStr (Possible Conformational Switches), Tango, and Waltz, thus its results tend to be more accurate than the individual predictors.

The consensus of these methods is defined as the hit overlap of at least n/2 (rounded down) out of n selected methods. This is an empirical threshold that was chosen based on many tests that were performed. Many individual methods provide several different settings. The values are provided on the basis of the best performances by each method. Additionally, tests were performed with multiple subsets of amyloidogenic proteins. For Amyloid Mutants, the default settings and the cross-beta pleat (serpentine) structural scheme were used. For Average Packing Density, values above 21.4 obtained from a five-residue long sliding window were considered as hits.

For Beta-strand contiguity, a threshold value of MbP = 1.2 was used and total y values above 20 as hits were considered. For Hexapeptide Conformational Energy, energy values below -27.00 were considered as hits. For NetCSSP, dual network architecture was used. The amyloidogenic hidden beta propensity (HbP) was calculated using the expression HbP = P(beta)/P(helix). Residues with values of HbP above 1 and of P (beta) above 6 were considered as hits. For Pafig, a threshold for the Reliability Index of 7 was used. For Tango, Tango 2.1 was used and scores above 5.00% for beta aggregation were considered as hits. For Waltz, pH= 7.0 and a threshold value of 79.0 was used.

DisEMBL [109] is based on artificial neural networks trained for predicting several definitions of disorder. It predicts and displays the probability of disordered segments within a protein sequence. DisEMBL furthermore provides a pipeline interface for bulk predictions, essential for large- scale structural genomics.

In DisEMBL, the coils data set was constructed based on DSSP secondary structure assignments. Two different versions of this data set were used for training neural networks. One version simply consists of the "raw" labeling of residues, while the other version is filtered to only include regions of at least seven consecutive residues with the same labeling. A second data set was constructed for discriminating between ordered and disordered loops. B-factors from regions of regular secondary structure were used for normalization by establishing chain-specific cutoffs for discriminating between ordered and disordered regions. All loop regions with B-factors below the median for secondary structure elements were labeled as ordered loops, while only those above the 90% fractile were considered to be disordered loops. Loops were identified in a similar manner as in coil data set. Finally, a data set was constructed for the prediction of missing coordinates. Artificial neural networks were trained on symmetric sequence windows centered at the position to be predicted. All neural networks were trained with a learning rate of 0.005. The size of these windows was systematically varied from 3 to 51 residues. The coil and hot loops neural network ensembles, the score distributions of positive and negative test examples were estimated using Gaussian kernel density estimation. Based on these distributions, a calibration curve for converting neural network output scores to probabilities was constructed.

7.4. Results & Discussions:

Figure 7.1 illustrates the residues in $A\beta_{1-42}$ peptide which are amyloidogenic in nature. Residues: 16-21 and 32-26 marked in blue dots are above the threshold value (**Figure 7.1**). According to the AMYLPRED2 algorithm, these regions are amyloidogenic. This result is in good agreement with the unrestrained folding simulation study, wherein we have found appearance of β-strands at the middle region as well as at the C-terminal region.

The probability of disorder for $A\beta_{1-42}$ peptide obtained from DisEMBL is shown graphically, as illustrated in **7.2.** The random expectation levels for the different predictors: Loops or coil; Hot-loops; Remark-465 and Tango are shown on the graph as horizontal lines. The green curve is the predictions for missing coordinates, red for the hot loop network, blue for coil and orange to predict cross-β aggregating segments by Tango. The horizontal lines correspond to the random expectation level for each predictor. Loops/coils are defined by DSSP. Residues are assigned as belonging to one of several secondary structure types: α-helix, 3_{10}-helix or β-strand. Hot loops are those with a high degree of mobility as determined from C-α temperature B-factors. It follows that highly dynamic loops should be considered protein disorder. Missing coordinates are defined by REMARK-465. From **Figure 7.2,** we can see that residues: 1-13 were defined as disordered by Loops/coils and hot-loops definition; residues: 17-21 and 30-42 were defined as β-strand aggregation region by Tango.

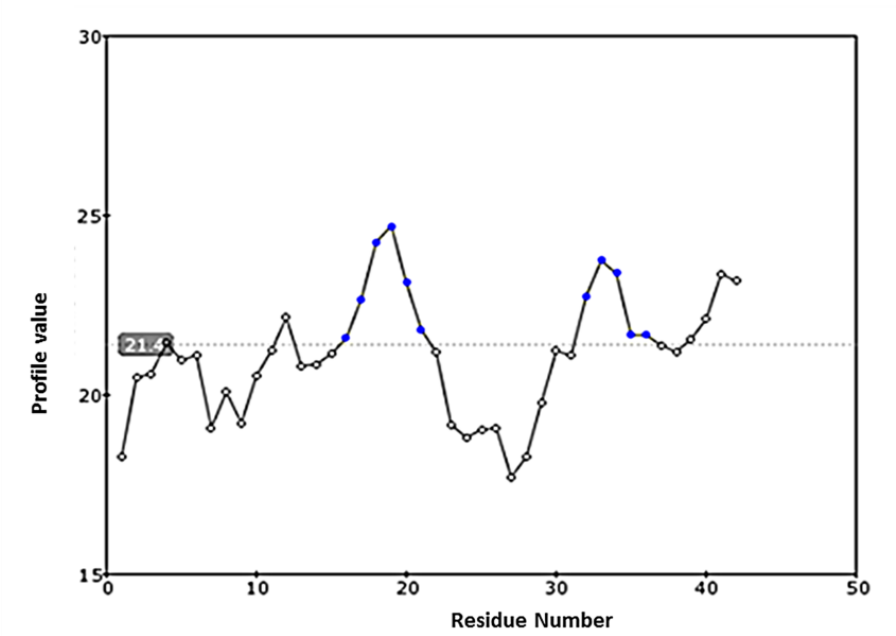

Figure 7.1. *Amyloidogenic residues of Aβ1-42 peptide as predicted by AMYLPRED2 software.*

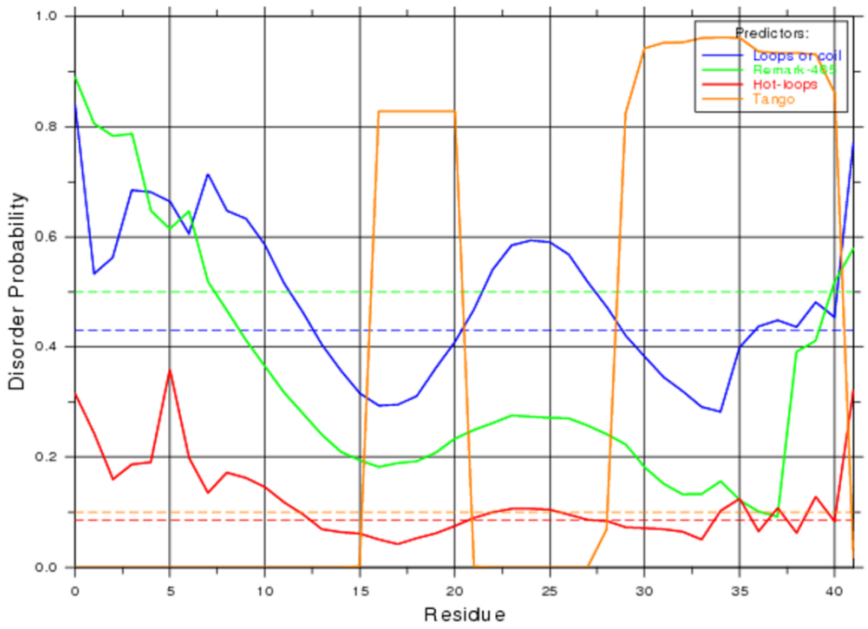

Figure 7.2. *Disorder Probability Graph for Aβ1-42 peptide as predicted by DisEMBL.*

In addition, we have also shown the hot spot regions in $A\beta_{1-42}$ peptide that were determined by AGGRESCAN [216]. **Table 7.1** shows the list of hot spot residues on the $A\beta_{1-42}$ peptide that may lead to aggregation. Hot spot residues are marked in blue. Residues: 17-22 and 30-42 of $A\beta_{1-42}$ peptide are marked to be hot-spots for aggregation by AGGRESCAN. The two regions that are marked to be the hot-spots belong to the CHC region and C-terminal region, respectively. This result obtained using AGGRECAN is in good agreement with the result obtained using AMYLPRED2 and DisEMBL.

The disordered regions found in the $A\beta_{1-42}$ peptide that were obtained using the disorder predictors are consistent with the probable seed structure of $A\beta_{1-42}$ peptide responsible for aggregation that we have determined in Chapter 4. **Figure 7.3.A** illustrates the structure of $A\beta_{1-42}$ peptide with the disordered regions highlighted in red color. **Figure 7.3.B** illustrates the seed structure of $A\beta_{1-42}$ peptide obtained in Chapter 3. From **Figure 7.3.B** we can notice appearance of β-strands in the CHC region and C-terminal region which are also marked as disordered regions by the disorder predictors. **Figure 7.3.C** shows the specific sequence of $A\beta_{1-42}$ peptide with the disordered regions in red color.

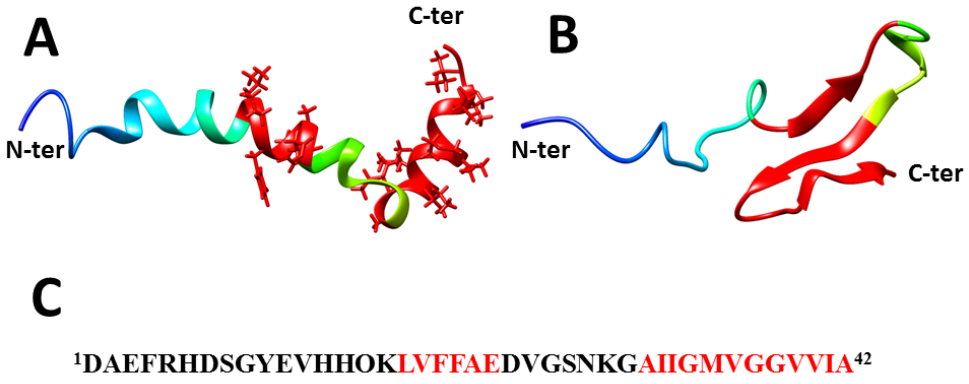

Figure 7.3.*A) $A\beta_{1-42}$ peptide showing the disordered regions in red color; B) seed structure of $A\beta_{1-42}$ peptide obtained in Chapter 3; C) sequence of $A\beta_{1-42}$ peptide.*

Table 7.1: *Hot spots of $A\beta_{1-42}$ peptide as predicted by AGGRESCAN.AA: amino acid; a^4v: a^3v average; HSA: hot spot area; NHSA: normalized HSA; a^4vAHS: a^4v average in the Hot Spot.*

	AA	a^4v	HSA	NHSA	a^4vAHS
1	D	-0.631	0.000	0.000	0.000
2	A	-0.631	0.000	0.000	0.000
3	E	-0.554	0.000	0.000	0.000
4	F	-0.393	0.000	0.000	0.000
5	R	-0.753	0.000	0.000	0.000
6	H	-0.530	0.000	0.000	0.000
7	D	-0.988	0.000	0.000	0.000
8	S	-0.508	0.000	0.000	0.000
9	G	-0.584	0.000	0.000	0.000
10	Y	0.102	0.000	0.000	0.000
11	E	-0.045	0.000	0.000	0.000
12	V	-0.145	0.000	0.000	0.000
13	H	-0.623	0.000	0.000	0.000
14	H	-0.527	0.000	0.000	0.000
15	Q	-0.570	0.000	0.000	0.000
16	K	-0.044	0.000	0.000	0.000
17	**L**	**0.513**	**3.821**	**0.637**	**0.617**
18	**V**	**1.110**	**3.821**	**0.637**	**0.617**
19	**F**	**1.289**	**3.821**	**0.637**	**0.617**
20	**F**	**0.731**	**3.821**	**0.637**	**0.617**
21	**A**	**0.045**	**3.821**	**0.637**	**0.617**
22	**E**	**0.013**	**3.821**	**0.637**	**0.617**
23	D	-0.445	0.000	0.000	0.000
24	V	-0.497	0.000	0.000	0.000
25	G	-0.475	0.000	0.000	0.000
26	S	0.294	0.000	0.000	0.000
27	N	-0.719	0.000	0.000	0.000
28	K	-0.620	0.000	0.000	0.000
29	G	-0.196	0.000	0.000	0.000
30	**A**	**0.428**	**9.906**	**0.762**	**0.742**
31	**I**	**0.508**	**9.906**	**0.762**	**0.742**
32	**I**	**0.891**	**9.906**	**0.762**	**0.742**
33	**G**	**1.080**	**9.906**	**0.762**	**0.742**
34	**L**	**1.034**	**9.906**	**0.762**	**0.742**
35	**M**	**0.891**	**9.906**	**0.762**	**0.742**
36	**V**	**0.563**	**9.906**	**0.762**	**0.742**
37	**G**	**0.606**	**9.906**	**0.762**	**0.742**
38	**G**	**0.742**	**9.906**	**0.762**	**0.742**
39	**V**	**0.788**	**9.906**	**0.762**	**0.742**
40	**V**	**0.888**	**9.906**	**0.762**	**0.742**
41	**I**	**0.778**	**9.906**	**0.762**	**0.742**
42	**A**	**0.778**	**9.906**	**0.762**	**0.742**

7.5. Conclusions:

In this study, we have identified the disordered regions in the $A\beta_{1-42}$ peptide. We have used two disorder predictors: AMYLPRED2, and DisEMBL to determine the disordered regions in $A\beta_{1-42}$ peptide. Our findings show the region 17-21 that falls under CHC region, and the region 30-42 belonging to the C-terminal region are disordered. Upon the comparison of our result with the probable seed structure of $A\beta_{1-42}$ peptide obtained from the unrestrained folding simulation in Chapter 4, we found the regions that formed β-strands in the seed structure to be the disordered regions in the $A\beta_{1-42}$ peptide.