Biochemical and Biophysical Research Communications 447 (2014) 255–262

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

The characterization and comparison of amyloidogenic segments and non-amyloidogenic segments shed light on amyloid formation

Shunmei Chen [a,b], Shan Gao [c], Dongqiang Cheng [a,b], Jingfei Huang [a,d,*]

[a] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32, Eastern Jiaochang Road, Kunming, Yunnan 650223, China
[b] Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China
[c] Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA
[d] Kunming Institute of Zoology – Chinese University of Hongkong Joint Research Center for Bio-Resources and Human Disease Mechanisms, Kunming 650223, China

ARTICLE INFO

Article history:
Received 20 March 2014
Available online 1 April 2014

Keywords:
Amyloid
Amyloidogenic segments
Non-amyloidogenic segments
Physico-chemical properties
Flexibility

ABSTRACT

Amyloid fibrillar aggregates of proteins or peptides are involved in the etiology of several neurodegenerative diseases and represent a major problem in healthcare. Short regions in the protein trigger this aggregation. It is important to understand the basis of such short regions aggregation and amyloidosis for therapeutic intervention. In this study, we describe specific physico-chemical properties of amyloidogenic segments and compare them with non-amyloidogenic segments. First, amyloidogenic segments are characterized by lower values for average net charge, electrostatic potential, solvent accessible surface area and *B*-factor when compared to the non-amyloidogenic segments of the same proteins. Second, they are enriched in hydrophobic residues and have a tendency to form hydrogen bonds. Thus, amyloidogenic segments have distinct physico-chemical properties that are different from those of non-amyloidogenic segments. Third, and quite unexpectedly, our dynamic simulation studies support the hypothesis that amyloidogenic segments have lower average flexibility than non-amyloidogenic segments. Furthermore, the presence of amyloidogenic segments in disordered proteins does not contradict the observation that amyloidogenic segments are less flexible.

1. Introduction

Amyloid refers to a specific peptide or protein that is normally soluble but is deposited as an insoluble aggregate under certain physiological conditions [1]. More than two dozen human diseases, including Alzheimer's, Parkinson's and Creutzfeldt-Jacob's neurodegenerative diseases, as well as type II diabetes and prion diseases, are characterized by the formation of amyloid aggregates. Additionally, several proteins and peptides not associated with diseases also have the ability to form amyloid [2]. Remarkably, amyloidogenic proteins having different primary sequences, structures, functions, and lengths can form structurally similar aggregates with similar physico-chemical properties [3,4].

The identification of factors that influence the protein aggregation of amyloids is one of the fundamental problems that must be addressed to understand the biogenesis of amyloid aggregates. One way to approach this problem is to correlate features of the primary sequence with its structure as determined by X-ray crystallography or NMR. However, investigation of factors that influence protein aggregation at the molecular level is not easy because it is difficult to crystallize amyloids and NMR limited to small proteins. Furthermore, the comparison of features of non-amyloidogenic segments with those of amyloidogenic segments could reveal factors that encode messages underlying the formation of amyloids aggregates. This information can potentially aid in the selection or development of therapeutic strategies to prevent or cure amyloid diseases. The aim of this work is to analyze and compare the primary amino acid sequences and structural features of amyloidogenic sequences with non-amyloidogenic sequences, especially with respect to their physico-chemical features.

Besides hydrogen bonds, the observation that van der Waals and hydrophobic interactions play a major role in aggregation formation [5]. Simulation studies showed that hydrophobic interactions are the principal driving force in these aggregations [6,7]. For amyloidogenic segments, specific side-chain and electrostatic interactions play an important role in the assembly of the β conformer [8]. The energetics of side chain conformation are now known with some accuracy, and they depend on many factors,

* Corresponding author at: State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32, Eastern Jiaochang Road, Kunming, Yunnan 650223, China. Fax: +86 871 5199200.
E-mail address: huangjf@mail.kiz.ac.cn (J. Huang).

http://dx.doi.org/10.1016/j.bbrc.2014.03.115
0006-291X/

including solvent exposure and local electrical potential [9]. It is the electrostatic interactions that determine the structure and flexibility of biopolymers [10].

Conformational flexibility is an inherent property of the protein structure [11]. Solvent accessibility is correlated with flexibility [9,12]. Flexible residues are more often found on the surface than in the core of proteins. Buried residues tend to be rigid, whereas exposed residues tend to be flexible [12]. The absolute value of solvent accessibility for an amino acid is its accessible surface area (ASA) in a protein structure. The relative solvent area (RSA) is obtained by normalizing the ASA value over the maximum value of the exposed surface area obtained for an extended tripeptide conformation of Ala-X-Ala or Gly-X-Gly [13].

As another measure for peptide chain flexibility, we chose the temperature factors, i.e., *B*-values (also referred to as *B*-factors), that are determined by X-ray crystallographic studies and provide information about the local mobility of atoms. Analysis of *B*-factors has provided novel insights into the flexibility of amino acids [14]. The values of *B*-factors is defined by $8\pi 2 <u2>$ to the unidirectional mean-square displacement, u^2, averaged over the lattice [15].

In addition to *B*-factors, which are directly related to atomic mobilities, the flexibility of a protein can also be derived from the trajectory of a molecular dynamics (MD) simulations by calculating the root mean-squared fluctuations (RMSFs) of individual atoms after removing the translational and rotational movements [16]. Vibrations around equilibrium are not random but dependent on local structure flexibility. Some regions of the protein may be flexible, some regions may be stiffer. RMSFs capture the fluctuation of each atom around average positions. Usually the analysis of the average atomic mobility of backbone atoms (N, Cα and C atoms) during MD simulation gives insight into the flexible and rigid regions of proteins.

2. Materials and methods

2.1. Database preparation

Amyloidogenic segments and non-amyloidogenic segments were found by surveying the literature. Only the segments validated by experimental evidence were included in this analysis. Specifically, amyloidogenic segments that have been shown to form amyloid fibrils experimentally *in vitro* were used.

Three-dimensional structures of the proteins used in this work were retrieved from the Protein Data Bank (PDB) [17]. The rules for selecting atom co-ordinates were as follows: First, UniProt IDs were used to search for a structure of the protein in PDB. Second, for structures with identical regions, the structure with the longest length of the protein was selected. Third, for structures with the same length, the structure with the highest resolution was selected, with a preference for single-protein structures over complexes. Fourth, if the resolution was same, the structure with the smallest *R*-value was selected. If all the structures retrieved were resolved by NMR, the structures with the smallest RMSD were selected; in the absence of RMSD, the structures submitted with the highest numbers of conformers were selected. Thirty protein structures met the above selection criteria and were thus included in this analysis. These structures contain 81 amyloidogenic segments, as the positive dataset and 99 non-amyloidogenic segments, as the negative dataset (in the Supplementary Tables 1 and 2).

2.2. Hydrogen bonds

Hydrogen bonds were calculated with the program hbplus [18] using the generally recommended [18] and default [19] parameters for angular and distance constraints between donor (D), hydrogen (H), and acceptor (A) atoms and the atom covalently bound to A (AA) [20]. The following default parameters were applied for the calculation: maximum distances for D–A, 3.9 Å and for H–A, 2.5 Å; minimum angles for D–H–A and D–A–AA, 90°. These values are based on the extensive analysis reported previously [18,19].

Hydrogen bonds were counted for the entire segment. This result was divided by the number of residues to yield the mean value of hydrogen bonds for every segment.

2.3. Hydrophobicity

The hydrophobicity of each amino acid sequence was calculated by the Kyte and Doolittle scale [21]. The residues included in the hydrophobic category were Ala, Met, Cys, Phe, Leu, Val, Ile, Pro, Tyr, Ser, Try, Thr, and Gly. The residues included in the hydrophilic category were Arg, Lys, Asp, Gln, Glu and His. The mean hydrophobicity is defined as the sum of the hydrophobicities of all residues divided by the number of residues in the peptide segment.

2.4. Charge

The mean net charge is defined as a net charge at pH 7.0 (total number of negatively charged Asp + Glu and positively charged Arg + Lys + His) divided by the total number of residues.

2.5. Electric potential

We used DS3.1 [22] to calculate the electrostatic potential of a molecular system, which has DelPhi built in, a program that solves the Poisson–Boltzmann equation on a cubical lattice using the finite-difference technique. Atomic charges and radii for DelPhi calculation were assigned using default charges and radii.

2.6. Surface accessibility

We used Naccess [23], which calculates the atomic and residue accessible area. The program uses the Lee & Richards method [13], whereby a probe of given radius is rolled around the surface of the molecule, and the path traced out by its center is the accessible surface. We also calculate the relative solvent area (RSA). This calculation gave accessible surface area per residue for the amyloidogenic and non-amyloidogenic segments. The mean RSA for each segment is defined as the sum of the RSA of all residues divided by the number of residues in the peptide segment.

2.7. Chain flexibility

The segments from NMR were excluded. We used the method by Yuan [24] to normalize *B*-factors. Because thermal factors can be affected by systematic errors such as the treatment of the solvent and the weights and types of refinement restraints, the thermal factors were normalized to a zero mean and unit variance, particularly for comparison amongst different structures. For each segment, the *B*-value was normalized by the following equation:

$$B\_\text{normalized} = (B - B\_\text{mean})/B\_\text{sigma}$$

Where *B*-mean and *B*-sigma are, respectively, the mean value and the standard deviation of the distribution of atomic thermal factors for a given protein structure and *B* is the atomic thermal factor as reported in the PDB file. Normalized *B*-factors derived from the unbiased structures were used as a measure of the flexibility of the residues in the segment. We compared the absolute value of mean normalized *B*-factors for amyloidogenic and non-amyloidogenic segments. The mean normalized *B*-factors for each segment is defined as the mean of the normalized *B*-factors of all residues divided by the number of residues in the peptide segment.

2.8. Molecular dynamics simulation

We randomly selected 54 segments from the positive and negative datasets for comparison. The flexibility of the different segments is revealed by looking at the root mean square fluctuation (RMSF) of each residue from its time-averaged position.

The Gromacs 4.5.5 software was used to perform molecular dynamics simulations [25]. The C and N termini of peptides were not capped according for the aggregation experiments. Under neutral pH, Lys and Arg residues are positively charged, and Glu and Asp are negatively charged. Charged systems were neutralized by adding 0.05 M NaCl, according to the standard protocol. All simulations were performed using the OPLS/AA force field [26], in combination with a cubic box containing the SPC water model [27] and using the periodic boundary condition. The distance of the box edge from the molecule's periphery was set to 0.7 nm. Initial energy minimization of the systems with the steepest-descent method until the maximum derivative was lower than 1000.0 $kJmol^{-1}nm^{-1}$. After the total energy of the system was nearly invariant upon further minimization, equilibration was performed for 1500–23000 ps in a constant-volume and constant-temperature conditions (NVT) ensemble and was followed by a 100–300 ps constant-pressure and constant-temperature conditions (NPT) ensemble. Finally, 11–295 ns production runs were performed. The simulations of datasets were run using the leapfrog Verlet algorithm with a 1-fs integration time step. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) summation with a grid size of 0.12 nm and a real-space cutoff of 0.9 nm [28]. The temperature was set to 310 K, and the pressure was set to 1 bar. No constraints were introduced for covalent bonds during simulation. A qualitative assessment of the degree of equilibrium for each peptide was made as follows. Cluster analysis was performed using g_cluster program of the GROMACS software package with a cutoff of 0.2 nm and the Gromos clustering algorithm proposed by Daura [29]. This method observed convergence of a simulation by considering the number of clusters as a function of time. The trajectory was then divided into two equal parts and the first part compared with the second. Convergence is deemed when samples from the same set of clusters in the first half of the trajectory as in the second half. For some segments, simulations were in the long run. If no more clusters appear in one-third part of trajectory and the curve plateaus, convergence is also deemed (in the Supplemental Figs. 1 and 2). The calculated root mean square fluctuation (RMSF) values were from the backbone (N, Cα and C) atoms of peptides during simulations using the g_rmsf command, which is a part of the GROMACS simulation package. For the final 5 ns of each simulation, RMSF values were calculated and compared the flexibility. The set up details of all MD simulations are listed (in the Supplementary Tables 3 and 4).

Because the first and last residues usually fluctuate considerably and can introduce bias, the C and N termini of peptides were omitted from each structure. For each segment, the RMSF range was generated by the following equation:

$$\Delta RMSF = RMSFmax - RMSFmin$$

Where RMSFmax and RMSFmin are, respectively, the max value and the min value of the RMSF for a given segment. ΔRMSF reflects the range of fluctuation. We compared the median ΔRMSF for amyloidogenic and non-amyloidogenic segments. The mean ΔRMSF for each segment is defined as the mean of the ΔRMSF of a segment divided by the number of residues in the peptide segment.

2.9. Statistical analysis

The Mann–Whitney test and chi-square test were used for non-normally distributed data and were calculated using SPSS software (Version 17.0 for Windows; SPSS Inc., Chicago, Illinois, USA). Significance was defined at the $p < 0.05$ threshold.

3. Results and discussion

In this study, the characterization and comparison of amyloidogenic segments with non-amyloidogenic segments elucidated factors that encode messages underlying the formation of amyloids. A total of 81 amyloidogenic segments and 99 non-amyloidogenic segments were chosen for analysis, as validated by the experimental evidence shown (in the Supplementary Tables 1 and 2). Specific physico-chemical properties of these amyloidogenic and non-amyloidogenic segments, including H-bonds, hydrophobicity, charge, electrostatic potential, RSA, *B*-factor, and ΔRMSF, are listed in Tables 1 and 2.

3.1. Hydrogen bonds

Amyloidogenic segments have a higher median value of H-bond (0.44) than non-amyloidogenic segments (0.30), with p value 0.0431 (Table 3). Amyloidogenic proteins display a high density of hydrogen bonds that are solvent-exposed in the monomeric structure. Because of the larger number of H-bonds, amyloidogenic segments are more likely to be susceptible to attacks from water molecules than non-amyloidogenic segments in similar situations. During the formation of the cross-β, more hydrogen bonds will provide more possibilities to form the hydrogen-bond stacks that stabilize amyloid fibril formation.

3.2. Hydrophobicity

The median residue hydrophobicity for amyloidogenic segments is much greater (−0.18) than that for non-amyloidogenic segments (−6.00), with p value 2.6414e-05 (Table 3). This result shows that amyloidogenic segments prefer a hydrophobic environment, indicating the importance of hydrophobic interactions for amyloid structural determination.

3.3. Charge

Amyloidogenic segments showed little difference median values for positive charge, but lower median values for negative charge and total charge (0.10; 0.11; 0.20) compared to non-amyloidogenic segments (0.10; 0.14; 0.30), with p value 0.0991, 3.9412e-04, and 0.0013, respectively (Table 3). Increase net charge can interfere with the ability to form amyloid aggregates.

Therefore, we can conclude that a combination of a high overall hydrophobicity and low net charge represents a unique structural feature of amyloid segments.

3.4. Electric potential

Comparison of the electrostatic potential of both segments reveals a large difference between them. In general, amyloidogenic segments displayed much lower median electrostatic potential values (40.36) compared to non-amyloidogenic segments (60.32), with a p value of 1.0368e-05 (Table 3). This result suggests that lower electrostatic potentials facilitate the formation of cross-β conformations. Amyloids show not only low net charge but also lower electrostatic potential.

Increased H-bonds and reduced electrostatic potential most likely represent another unique structural feature of amyloid segments.

Table 1
Calculation hydrophobicity, hydrogen-bond, charge, electrostatic potential, *B*-factor, ΔRMSF frequency and accessible surface area in amyloidogenic segments.

Protein	Region	Length	Hydrophobicity/ length	H-bond/ length	Charge/length			Electrostatic potential/ length	RSA/ length	*B*-factor/ length	ΔRMSF/ length
					Positive	Negative	Total				
Microtubule-associated protein tau	306–311	6	4.20	0.00	0.17	0.00	0.17	105.88	93.35		
Amyloid beta A4 protein	672–711	40	2.40	0.68	0.15	0.15	0.30	16.69	19.78		
	683–697	15	1.80	0.67	0.20	0.13	0.33	52.87	11.13		
	684–694	11	−5.39	0.64	0.27	0.18	0.45	77.61	10.60		0.0013
	683–695	13	2.99	0.62	0.23	0.15	0.38	57.15	11.52		0.0009
	682–696	15	−0.90	0.60	0.20	0.20	0.40	52.85	10.46		
	681–697	17	−3.06	0.59	0.18	0.18	0.35	46.31	11.50		
	680–698	19	−6.84	0.63	0.16	0.16	0.32	40.81	12.26		
	679–699	21	−11.55	0.62	0.19	0.14	0.33	36.81	13.42		
	678–700	23	−15.41	0.61	0.17	0.17	0.35	34.30	13.57		
	677–701	25	−17.00	0.68	0.20	0.16	0.36	29.09	13.34		
Alpha-synuclein	61–78	18	16.74	0.78	0.00	0.06	0.06	11.58	62.35		
	63–78	16	23.84	0.69	0.00	0.00	0.00	10.48	61.26		
	61–73	13	5.46	0.54	0.00	0.08	0.08	20.88	62.96		
	66–74	9	16.74	0.44	0.00	0.00	0.00	32.24	58.51		
	71–82	13	10.53	0.85	0.00	0.15	0.15	39.62	67.91		
Beta-2-microglobulin	79–91	13	−0.65	0.08	0.10	0.20	0.30	12.62	4.24		
	73–82	10	−4.60	0.10	0.10	0.10	0.20	78.45	7.96		0.0013
	78–87	10	−2.40	0.00	0.00	0.10	0.10	28.81	5.93		0.0019
	83–92	10	0.00	0.10	0.10	0.40	0.50	21.39	3.04		
	88–97	10	−18.80	0.60	0.11	0.16	0.26	84.52	8.02		
	103–108	6	−0.18	0.67	0.16	0.16	0.32	81.42	18.15		
	84–89	6	0.78	0.17	0.19	0.13	0.31	66.97	1.53		0.0009
	111–116	6	−3.30	0.17	0.13	0.13	0.25	143.57	6.97		
	78–83	6	−7.38	0.00	0.32	0.16	0.47	135.10	9.05		0.0012
	82–87	6	6.48	0.00	0.00	0.13	0.13	30.45	0.85		
	40–61	22	2.20	0.05	0.15	0.15	0.30	14.85	8.41		
	103–109	7	−3.71	0.57	0.08	0.08	0.17	46.30	19.10		
Myoglobin	2–20	19	8.17	0.84	0.00	0.22	0.22	24.92	37.67		0.0016
	5–35	31	3.41	0.84	0.00	0.13	0.13	21.61	25.87	0.000	
	21–36	16	1.76	0.88	0.10	0.10	0.20	51.82	18.06	0.001	
Myoglobin	1–29	29	−6.38	0.86	0.12	0.06	0.18	30.93	32.34		
	101–118	18	10.26	0.94	0.16	0.20	0.36	40.30	23.92	0.001	
Myohemerithrin	40–63	24	−5.04	1.08	0.06	0.25	0.31	14.42	31.99	0.001	
	69–87	19	−9.69	0.84	0.20	0.10	0.30	39.05	37.74	0.000	
Plastocyanin	1–15	15	20.85	0.40	0.40	0.10	0.50	32.31	1.49		
	24–43	20	−7.20	0.10	0.10	0.00	0.10	27.72	4.84		
	46–57	12	12.72	0.50	0.17	0.00	0.17	42.98	14.55		
	57–74	18	−0.36	0.22	0.00	0.17	0.17	25.78	21.06		0.0019
	67–74	8	5.60	0.00	0.33	0.17	0.50	31.66	24.55		
Glutathione S-transferase P	145–164	20	6.00	0.85	0.17	0.17	0.33	17.96	11.34	0.001	
Chemotaxis protein CheY	90–106	17	−1.36	0.59	0.00	0.00	0.00	51.32	37.66	0.001	
PL B1 protein	132–156	25	−30.50	0.60	0.00	0.17	0.17	35.99	10.34		
Immunoglobulin G-binding protein G	412–427	16	−15.04	0.63	0.00	0.17	0.17	41.78	8.28		
Human prion protein	176–185	10	5.10	0.80	0.00	0.17	0.17	86.02	36.70	0.001	0.0026
	178–191	14	−0.56	0.93	0.00	0.00	0.00	25.53	39.39	0.002	0.0013
Lysozyme C	23–32	10	−9.70	0.70	0.17	0.00	0.17	82.46	29.94	0.003	0.0016
	43–52	10	6.80	0.60	0.09	0.14	0.23	41.37	11.25	0.005	0.0014
Islet amyloid polypeptide	55–60	6	9.00	0.33	0.14	0.00	0.14	57.48	8.48		0.001
	48–53	6	6.00	0.17	0.00	0.00	0.00	59.05	8.13		0.0006
	55–62	8	7.44	0.38	0.00	0.00	0.00	53.75	13.34		0.0013
	63–70	8	−6.72	0.38	0.22	0.00	0.22	96.56	28.71		
	43–51	9	1.71	0.33	0.00	0.00	0.00	55.66	8.23		0.0026
	53–60	8	4.72	0.38	0.11	0.00	0.11	35.31	10.73		0.0014
	47–55	9	−4.50	0.33	0.08	0.00	0.08	45.09	9.56		0.0028
Transthyretin	30–40	11	17.38	0.18	0.09	0.09	0.18	34.73	23.64		
	125–135	11	9.24	0.18	0.00	0.00	0.00	10.21	20.17	0.001	
Gelsolin	209–219	11	1.21	0.18	0.00	0.18	0.18	34.25	7.95	0.002	0.0009
Lactotransferrin	556–563	8	7.44	0.00	0.00	0.13	0.13	43.18	15.14	0.004	
Eukaryotic peptide chain release factor GTP-binding subunit	7–13	7	−19.18	0.29	0.14	0.07	0.21	60.03	99.97		
Insulin	36–41	6	8.82	0.33	0.00	0.00	0.00	89.55	19.33	0.008	0.0009
	102–107	6	−4.20	0.33	0.10	0.10	0.20	76.93	48.48	0.005	
	35–40	6	8.82	0.33	0.10	0.10	0.20	79.93	8.15	0.007	0.0006
Ribonuclease pancreatic	20–25	6	−5.40	0.33	0.29	0.00	0.29	81.95	50.65	0.010	
	22–32	11	−6.71	0.18	0.00	0.09	0.09	13.71	55.30	0.002	
	75–80	6	−2.52	0.33	0.14	0.18	0.32	58.10	33.07	0.007	
Beta-lactoglobulin	11–20	10	−3.50	0.40	0.11	0.17	0.29	51.56	34.90	0.001	0.0027
	101–110	10	1.80	0.00	0.09	0.19	0.28	40.36	13.34	0.004	0.0016
	146–152	7	−0.91	0.00	0.00	0.00	0.00	57.96	53.77		
Apolipoprotein C-II	60–70	11	−2.42	0.55	0.00	0.06	0.06	50.50	8.67		

Cold shock protein CspB	1–22	22	−5.28	0.36	0.08	0.17	0.25	41.64	39.03		
	1–35	35	−3.50	0.54	0.07	0.17	0.24	28.20	39.02		
	36–67	32	−18.88	0.47	0.22	0.11	0.33	26.19	45.95		
Transforming growth factor-beta-induced protein ig-h3	515–525	11	17.49	0.82	0.20	0.13	0.33	16.85	6.63		0.0024
	515–532	18	16.38	0.78	0.20	0.07	0.27	17.89	5.34		0.001
Laminin subunit alpha-1	2919–2930	12	12.72	0.25	0.00	0.00	0.00	39.83	21.62	0.002	
Prolactin	7–21	15	−4.35	0.20	0.14	0.16	0.30	48.36	11.67		0.0015
	20–34	15	0.75	0.80	0.27	0.00	0.27	31.39	6.25		0.001
	71–85	15	−10.35	0.20	0.00	0.00	0.00	33.67	13.37		
Replication protein	26–34	9	25.38	0.67	0.00	0.00	0.00	40.58	18.19	0.002	
Transcription elongation Regulator 1	430–466	37	−49.21	0.38	0.00	0.00	0.00	30.83	13.15		
Median			−0.18	0.44	0.10	0.11	0.20	40.36	13.57	0.002	0.0013

3.5. Surface accessibility

The residues from the amyloidogenic segments showed a significantly lower median value of relative accessible surface area (13.57% RSA) compared to residues from the non-amyloidogenic segments (29.19% RSA), with a p value of 0.0374 (Table 3). Therefore, this result shows that amyloidogenic segments have a much lower average surface accessibility than non-amyloidogenic segments. It is likely that amyloidogenic segments minimize their solvent-accessible surface area. This tendency is consistent with the result demonstrating that median residue hydrophobicity for amyloidogenic segments is much greater than that for non-amyloidogenic segments. Amyloidogenic segments are thus more likely to occur in proteins with a hydrophobic core buried inside. This is most likely one of the most important characteristics of amyloidogenic segments. To prevent water-mediated attacks on the H-bond, residues of the amyloidogenic segments must be shielded from the solvent, and this necessity is reflected in a low RSA for amyloidogenic segments.

3.6. Chain flexibility

Very interestingly, amyloidogenic segments have a significantly lower median value for normalized B-factors (0.002) than non-amyloidogenic segments (0.024), with a p value of 2.9924e-10 (Table 3). This result suggests that the residues in amyloidogenic segments are less flexible than those in the non-amyloidogenic segments because flexible residues are more often found on the surface than in the core of proteins, while buried residues incline to be rigid. This result is consistent with the finding that amyloidogenic segments minimize solvent-accessible surface area and are buried in a protein's interior.

3.7. Molecular dynamics simulation

Quite unexpectedly, our dynamic simulation studies support that a model in which amyloidogenic segments have lower average flexibility relative to non-amyloidogenic segments, which seems to suggest that these regions are associated with the rigid core of the protein. This result tallies with the finding of a previous study by Tartaglia et al. [30] which reports that the regions of the amino acid sequence that are highly aggregation-prone should be protected in the folded state and buried in the native state before they can form stable intermolecular interactions.

MD simulation of datasets shows that amyloidogenic segments have a slightly lower median values for ΔRMSF (0.0013) compared to non-amyloidogenic segments (0.002), with a p value of 0.0029 (Table 3). This tendency supports the observation based on normalized B-factors that amyloidogenic segments are less flexible.

Taken together, it is worth noting that the B-factor suggests that amyloidogenic segments are less flexible, and the range of RMSF fluctuations during MD simulations also supports this observation. This result differs from tau aggregation in Alzheimer's disease, where tau protein has been shown to exist as a disorder protein. Through a deep analysis of our data, we arrive at the exciting finding that amyloidogenic segments are located in rigid regions of the protein and rarely in the disordered regions. In fact, loop coverage is a common strategy employed to avoid aggregation [31]. Additionally, using packing density as a parameter to predict amyloidogenic and disordered regions in protein chains, Galzitskaya and co-workers found that a stronger than expected packing density is responsible for amyloid formation. They further demonstrated that the regions with a weaker than expected packing density are responsible for the appearance of disordered regions [32]. To validate this exciting finding of our disorder prediction and its relationship with aggregation, we also calculated the packing density. Disorder prediction and packing density of amyloidogenic segments and non-amyloidogenic segments are listed (in the Supplementary Tables 5 and 6). The chi-square test results show that the overall difference between amyloidogenic and non-amyloidogenic segments in predicting segment disorder is significant; amyloidogenic segments are less disordered than non-amyloidogenic segments, with a p value of 0.0071 (in the Supplementary Table 7). In addition, amyloidogenic segments have a significantly higher median value for packing density (21.4) than non-amyloidogenic segments (20.86), with a p value 3.3470e-04, whereas disordered segments have a significantly lower packing density (21.1) than ordered segments (20.4), with a p value 7.2851e-04 (in the Supplementary Table 8). These results are consistent with the findings of Galzitskaya and co-workers' [32]. Therefore, from both prediction and digital analysis, we verified that amyloidogenic segments are preferentially located in rigid regions of the protein and rarely in the disordered regions. This finding represents a potential difference between amyloidgenic and disordered regions. We also investigated the parent proteins of these amyloidogenic and non-amyloidogenic segments in DisProt (Database of Protein Disorder), and we found that 8 in 30 of these proteins are disordered. Unsurprisingly, the amyloidogenic segments in these disordered proteins are located in rigid regions of the protein and only rarely in the disordered regions (in the Supplementary Tables 9 and 10). This result further validates our other findings. The presence of amyloidogenic segments in disordered proteins does not contradict the fact that amyloidogenic segments are less flexible. This is a very exciting finding and confirm that amyloidgenic segments are less flexible than non-amyloidgenic segments.

In summary, amyloidogenic segments are characterized by a lower average flexibility, lower average net charge, lower average electrostatic potential, and lower average RSA and B-factor than non-amyloidogenic segments. However, amyloidogenic segments are enriched in hydrophobic residues and show the presence of more hydrogen bonds.

Here, we have investigated and characterized some of the features of amyloidogenic sequences. The phenomenon of protein aggregation is a ubiquitous problem in biomedical and

Table 2
Calculation hydrophobicity, hydrogen-bond, charge, electrostatic potential, *B*-factor, ΔRMSF frequency and accessible surface area in non-amyloidogenic segments.

Protein	Region	Length	Hydrophobiciy/ length	H-bonds/ length	Charge/length			Electrostatic potential/lengh	RSA/ length	*B*-factor/ length	ΔRMSF/ length
					Positive	Negative	Total				
Amyloid beta A4 protein	687–691	5	9.70	0.20	0.00	0.20	0.20	223.88	12.84		
	686–692	7	7.98	0.43	0.00	0.14	0.14	104.83	10.46		0.0019
	685–693	9	1.26	0.67	0.11	0.22	0.33	63.62	9.27		
Beta-2-microglobulin	21–36	16	−17.60	0.06	0.06	0.25	0.31	13.57	5.64		0.0019
	37–49	13	−4.81	0.08	0.00	0.08	0.08	10.64	3.82		
	50–61	12	−0.96	0.08	0.25	0.17	0.42	34.13	12.71		0.0022
	62–79	18	−23.94	0.17	0.28	0.22	0.50	40.19	15.98		
	107–119	13	−11.44	0.08	0.15	0.23	0.38	68.61	18.99		
	23–32	10	−12.10	0.10	0.00	0.30	0.30	48.31	6.03		0.0012
	28–37	10	−15.90	0.00	0.10	0.20	0.30	30.66	5.80		
	33–42	10	−15.80	0.20	0.10	0.20	0.30	78.14	3.93		
	43–52	10	2.50	0.00	0.00	0.10	0.10	13.64	5.19		0.0021
	48–57	10	−2.30	0.10	0.20	0.10	0.30	40.96	15.48		
	53–62	10	−2.40	0.10	0.30	0.10	0.40	43.23	14.31		
	58–67	10	−10.70	0.30	0.30	0.20	0.50	112.86	19.29		
	63–72	10	−14.60	0.10	0.30	0.30	0.60	93.82	20.42		
	68–77	10	−5.70	0.10	0.20	0.20	0.40	39.79	9.02		
	93–102	10	−12.40	0.40	0.30	0.20	0.50	61.81	5.94		0.0092
	98–107	10	3.30	0.40	0.00	0.20	0.20	40.82	4.75		
	103–112	10	−4.70	0.40	0.00	0.20	0.20	63.25	18.17		
Regulatory protein cro	51–70	20	−6.40	0.50	0.05	0.15	0.20	44.86	13.55		
Myoglobin	35–40	6	−13.68	0.50	0.17	0.33	0.50	167.55	36.87	0.087	
	37–51	15	−22.20	0.60	0.20	0.40	0.60	98.21	39.94	0.021	
	52–58	7	−6.09	0.71	0.29	0.14	0.43	183.13	45.80	0.027	0.0028
	59–78	20	6.80	1.00	0.10	0.20	0.30	48.66	29.97	0.001	
	79–86	8	−19.84	0.38	0.25	0.50	0.75	196.98	46.08	0.029	
	86–95	10	−5.30	0.90	0.10	0.20	0.30	55.64	27.76	0.022	
	95–101	7	−7.00	0.14	0.00	0.43	0.43	141.33	37.61	0.020	
Myohemerithrin	1–18	18	−8.46	0.33	0.22	0.06	0.28	31.53	44.82		0.0025
	109–118	10	−11.40	0.40	0.10	0.30	0.40	143.76	34.47		
	18–38	21	−20.16	1.10	0.29	0.24	0.52	39.47	28.12	0.008	
	37–41	5	−10.50	0.20	0.20	0.20	0.40	136.74	48.64	0.050	
	63–70	8	−5.20	0.25	0.25	0.13	0.38	111.30	49.30	0.006	
	86–92	7	3.50	0.00	0.14	0.00	0.14	113.43	47.99	0.044	0.0018
	93–108	16	−9.92	0.75	0.13	0.25	0.38	53.14	24.63	0.014	
Plastocyanin	1–8	8	10.48	0.00	0.13	0.00	0.13	41.46	2.38		0.0013
	7–12	6	−2.10	0.33	0.17	0.00	0.17	93.32	1.47		
	11–20	10	10.30	0.10	0.10	0.00	0.10	41.22	2.18		
	17–21	5	1.90	0.00	0.20	0.00	0.20	117.34	3.82		
	17–26	10	−8.30	0.20	0.20	0.10	0.30	66.47	7.54		
	26–33	8	−1.52	0.00	0.00	0.25	0.25	81.03	5.36		
	26–37	12	−3.96	0.08	0.00	0.25	0.25	43.40	4.01		0.0028
	30–39	10	−10.80	0.10	0.00	0.20	0.20	55.49	1.34		0.0014
	36–47	12	−8.16	0.25	0.33	0.08	0.42	34.78	4.19		0.0013
	45–50	6	4.98	0.17	0.17	0.00	0.17	68.18	8.78		0.001
	51–56	6	2.28	0.83	0.17	0.17	0.33	197.03	18.18		0.0045
	57–63	7	−2.59	0.14	0.43	0.00	0.43	109.00	19.10		
	61–70	10	−5.10	0.30	0.20	0.00	0.20	46.94	19.18		0.0028
	64–70	7	−9.17	0.29	0.14	0.00	0.14	59.17	20.96		0.0036
	65–69	5	−4.40	0.20	0.20	0.00	0.20	129.98	25.14		
	68–73	6	2.22	0.00	0.17	0.00	0.17	47.58	29.60		0.0014
	72–80	9	−3.24	0.22	0.11	0.11	0.22	74.68	13.01		0.0021
	79–84	6	1.20	0.00	0.00	0.00	0.00	30.08	1.07		
	92–99	8	6.00	0.13	0.00	0.13	0.13	45.00	1.98		
Pancreatic trypsin inhibitor	51–63	13	−2.21	0.31	0.00	0.23	0.23	40.23	39.12	0.018	
Glutathione S-transferase P	80–107	28	−6.44	1.04	0.21	0.11	0.32	27.86	20.99	0.010	
	109–133	25	−23.50	0.92	0.20	0.20	0.40	44.86	35.09	0.013	
	172–183	12	12.00	0.83	0.00	0.08	0.08	30.78	28.14	0.012	0.0015
	185–192	8	5.28	0.63	0.00	0.25	0.25	117.43	39.31	0.019	
Human prion protein	121–130	10	15.00	0.10	0.00	0.00	0.00	21.24	50.26		0.0027
	126–135	10	5.30	0.00	0.00	0.00	0.00	25.55	54.98	0.035	
	131–140	10	1.40	0.00	0.00	0.20	0.20	34.78	58.62	0.054	
	141–150	10	−17.30	0.60	0.30	0.10	0.40	81.35	53.16	0.013	
	146–155	10	−26.90	0.80	0.30	0.30	0.60	98.20	35.77	0.005	
	151–160	10	−27.20	0.30	0.10	0.30	0.40	48.79	34.25	0.025	0.0015
	156–165	10	−18.90	0.00	0.00	0.20	0.20	27.92	30.66	0.012	
	161–170	10	−11.70	0.30	0.20	0.10	0.30	60.32	48.14	0.012	
	166–175	10	−18.40	0.20	0.20	0.00	0.20	50.70	62.15	0.007	
	171–180	10	−7.00	0.60	0.10	0.10	0.20	48.54	46.26	0.062	
	181–190	10	−3.00	0.90	0.00	0.20	0.20	69.00	39.34	0.039	0.0015
	186–195	10	−10.30	0.30	0.00	0.20	0.20	46.68	54.43	0.029	
	191–200	10	−14.80	0.30	0.20	0.10	0.30	67.32	58.28	0.045	
	196–205	10	−10.40	0.60	0.30	0.10	0.40	70.25	46.52	0.025	

	201–210	10	0.30	0.70	0.20	0.20	0.40	99.00	47.48	0.015	
	206–215	10	4.20	0.70	0.20	0.10	0.30	65.51	54.94	0.009	
	211–220	10	−11.60	0.80	0.20	0.10	0.30	70.39	56.40	0.024	
	216–225	10	−20.80	0.90	0.20	0.10	0.30	71.79	62.43		
Lysozyme C	28–37	10	−8.40	0.40	0.10	0.30	0.40	90.09	40.13	0.047	
	33–42	10	−1.10	0.40	0.10	0.10	0.20	57.47	40.87	0.030	
	38–47	10	0.80	0.60	0.00	0.10	0.10	47.30	25.80	0.037	
	48–57	10	−6.00	0.60	0.10	0.10	0.20	62.52	20.20	0.021	
	53–62	10	−17.10	0.30	0.10	0.10	0.20	42.54	33.13	0.051	
	58–67	10	−14.50	0.00	0.10	0.10	0.20	29.66	51.51	0.005	
	63–72	10	−17.70	0.60	0.20	0.10	0.30	69.93	40.54	0.023	
	68–77	10	−2.90	0.40	0.10	0.10	0.20	38.03	10.81	0.030	
	73–82	10	−2.90	0.40	0.00	0.10	0.10	27.63	11.95	0.038	
	78–87	10	−19.60	0.40	0.10	0.20	0.30	81.16	32.96	0.050	
	83–92	10	−5.50	0.30	0.10	0.10	0.20	94.94	46.43	0.008	0.0025
	88–97	10	4.70	0.20	0.00	0.10	0.10	22.48	40.54	0.002	
	93–102	10	7.90	0.70	0.00	0.10	0.10	32.24	29.78	0.030	
	103–112	10	3.90	0.30	0.20	0.00	0.20	65.28	35.95	0.042	
	108–117	10	6.20	0.70	0.10	0.20	0.30	110.70	19.34	0.028	
	113–122	10	−8.80	0.60	0.10	0.30	0.40	86.28	29.19	0.042	
	118–127	10	−8.20	0.40	0.10	0.20	0.30	77.02	33.98	0.028	
	123–132	10	−2.00	0.70	0.00	0.20	0.20	59.15	29.19	0.023	
	128–137	10	−16.20	0.70	0.00	0.30	0.30	72.43	43.59	0.001	0.002
	133–142	10	−22.10	0.30	0.10	0.30	0.40	76.66	46.04	0.003	0.0013
	138–147	10	−6.20	0.40	0.10	0.10	0.20	45.46	50.17		0.002
Ribonuclease pancreatic	33–38	5	−10.25	0.00	0.17	0.33	0.5	121.75	43.78	0.057	
Median			−6.00	0.30	0.10	0.14	0.30	60.32	29.19	0.026	0.002

Table 3
Comparation hydrogen-bond, hydrophobicity, charge, electrostatic potential, *B*-factor, ΔRMSF frequency and accessible surface area in amyloidogenic segments with non-amyloidogenic segments.

	Amyloidogenic segments Median (QR)[a]	Non-amyloidogenic segments Median (QR)	Mann–Whitney test *p*-value[b]	
H-bond/length	0.4 (0.5)	0.3 (0.5)	0.0431	*
Hydrophobicity/length	−0.18 (11.87)	−6.00 (13.10)	2.6414e−05	***
Positive charge/length	0.10 (0.17)	0.10 (0.20)	0.0991	
Negative charge/length	0.11 (0.17)	0.14 (0.10)	3.9412e−04	***
Total charge/length	0.20 (0.21)	0.30 (0.20)	0.001	**
Electrostatic potential/length	40.36 (28.58)	60.32 (48.11)	1.0368e−05	***
RSA/length	13.57 (24.91)	29.19 (31.59)	0.0374	*
B-factor/length	0.002 (0.0033)	0.024 (0.0257)	2.9924e−10	***
ΔRMSF/length	0.0013 (0.0008)	0.002 (0.0012)	0.0029	**

[a] QR, Quartile range.
[b] Asterisks representative significance (*p*), *<0.05; **<0.01; ***<0.001.

biotechnological research [33]. Our studies provide and identify fundamental differences in the characteristics of amyloidogenic and non-amyloidogenic segments, which could be useful when designing a therapeutic regimen for the treatment of diseases involving protein aggregation.

Acknowledgements

This work was supported by the National Basic Research Program of China (grant number 2013CB835100), and the National Natural Science Foundation of China (grant number 31123005 to J.F.H.). Simulation work supported by HPC Platform, Large-scale Instrument Regional Center of Biodiversity, Kunming Institute of Zoology, CAS, China. We thank Xu Shao Bin of HPC Platform for technology support, we also thank Liang Hao of MolSimu Technology Ltd for suggestion and assistance in molecular simulation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.03.115.

References

[1] Jeffery W. Kelly, The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, Curr. Opin. Struct. Biol. 8 (1998) 101–106.
[2] C.M. Dobson, Protein misfolding, evolution and disease, Trends Biochem. Sci. 24 (1999) 329–332.
[3] E.H. Koo, P.T. Lansbury, J.W. Kelly, Amyloid diseases: abnormal protein aggregation in neurodegeneration, Proc. Nat. Acad. Sci. 96 (1999) 9989.
[4] J.L. Jiménez, E.J. Nettleton, M. Bouchard, C.V. Robinson, C.M. Dobson, H.R. Saibil, The protofilament structure of insulin amyloid fibrils, Proc. Nat. Acad. Sci. 99 (2002) 9196.
[5] J. Park, B. Kahng, W. Hwang, Thermodynamic selection of steric zipper patterns in the amyloid cross-β spine, PLoS Comput. Biol. 5 (2009) e1000492.
[6] L. Chang, J. Zhao, H. Liu, J. Wu, C. Chuang, K. Liu, J. Chen, W. Tsai, Y. Ho, The importance of steric zipper on the aggregation of the MVGGVV peptide derived from the amyloid beta peptide, J. Biomol. Struct. Dyn. 28 (2010) 39.
[7] W.M. Berhanu, A.E. Masunov, Molecular dynamic simulation of wild type and mutants of the polymorphic amyloid NNQNTF segments of elk prion: structural stability and thermodynamic of association, Biopolymers 95 (2011) 573–590.
[8] B. Ciani, E.G. Hutchinson, R.B. Sessions, D.N. Woolfson, A designed system for assessing how sequence affects α to β conformational transitions in proteins, J. Biol. Chem. 277 (2002) 10150–10155.
[9] O. Carugo, P. Argos, Correlation between side chain mobility and conformation in protein structures, Protein Eng. 10 (1997) 777.
[10] N.A. Baker, J.A. McCammon, Electrostatic Interactions, Structural Bioinformatics, John Wiley & Sons, Inc., 2005. pp. 427–440.
[11] I.B. Kuznetsov, Simplified computational methods for the analysis of protein flexibility, Curr. Protein Pept. Sci. 10 (2010) 607.
[12] A. Schlessinger, B. Rost, Protein flexibility and rigidity predicted from sequence, Proteins Struct. Funct. Bioinf. 61 (2005) 115–126.
[13] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400 (IN3–IN4).
[14] S. Parthasarathy, M.R.N. Murthy, Protein thermal stability: insights from atomic displacement parameters *B* values, Protein Eng. 13 (2000) 9–13.
[15] D.M. Blow, Outline of Crystallography for Biologists, Oxford University Press, USA, 2002.

[16] K. Bidmon, G. Reina, F. Bos, J. Pleiss, T. Ertl, Time-based haptic analysis of protein dynamics, IEEE (2007) 537–542.
[17] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The protein data bank: a computer-based archival file for macromolecular structures*, J. Mol. Biol. 112 (1977) 535–542.
[18] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[19] E. Baker, R. Hubbard, Hydrogen bonding in globular proteins, Prog. Biophys. Mol. Biol. 44 (1984) 97.
[20] G. Vogt, P. Argos, Protein thermal stability: hydrogen bonds or internal packing?, Fold Des 2 (1997) S40–S46.
[21] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[22] T.M. Doran, E.A. Anderson, S.E. Latchney, L.A. Opanashuk, B.L. Nilsson, Turn nucleation perturbs amyloid β self-assembly and cytotoxicity, J. Mol. Biol. 421 (2012) 315–328.
[23] S.J. Hubbard, J.M. Thornton, NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College London 2 (1993).
[24] Z. Yuan, J. Zhao, Z.X. Wang, Flexibility analysis of enzyme active sites by crystallographic temperature factors, Protein Eng. 16 (2003) 109.
[25] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[26] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. B 105 (2001) 6474–6487.
[27] H. Berendsen, J. Grigera, T. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271.
[28] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577.
[29] X. Daura, W.F. van Gunsteren, A.E. Mark, Folding–unfolding thermodynamics of a β-heptapeptide from equilibrium simulations, Proteins Struct. Funct. Bioinf. 34 (1999) 269–280.
[30] G.G. Tartaglia, M. Vendruscolo, Proteome-level interplay between folding and aggregation propensities of proteins, J. Mol. Biol. 402 (2010) 919–928.
[31] J.S. Richardson, D.C. Richardson, Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation, Proc. Nat. Acad. Sci. 99 (2002) 2754.
[32] O.V. Galzitskaya, S.O. Garbuzynskiy, M.Y. Lobanov, Prediction of amyloidogenic and disordered regions in protein chains, PLoS Comput. Biol. 2 (2006) e177.
[33] M.R.H. Krebs, D.K. Wilkins, E.W. Chung, M.C. Pitkeathly, A.K. Chamberlain, J. Zurdo, C.V. Robinson, C.M. Dobson, Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the [beta]-domain 1, J. Mol. Biol. 300 (2000) 541–549.