TASK QUARTERLY 18 No 3, 245–254

KEY FACTORS GOVERNING FIBRIL FORMATION OF PROTEINS: INSIGHTS FROM SIMULATIONS AND EXPERIMENTS

NGUYEN TRUONG CO[1], MAN HOANG VIET[2], PHAN MINH TRUONG[1], MAKSIM KOUZA[3] AND MAI SUAN LI[2]

[1]*Institute for Computational Science and Technology, SBI Building Quang Trung Software City, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam*

[2]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*

[3]*Faculty of Chemistry, Faculty of Chemistry University of Warsaw Pasteura 1, 02-093 Warsaw, Poland*

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Fibril formation of proteins and peptides is associated with a large group of major human diseases, including Alzheimer's disease, prion disorders, amyotrophic lateral sclerosis, type 2 diabetes, etc. Therefore, understanding the key factors that govern this process is of paramount importance. The fibrillogenesis of polypeptide chains depends on their intrinsic properties as well as on the external conditions. In this mini-review we discuss the relationship between fibril formation kinetics and the sequence, aromaticity, hydrophobicity, charge and population of the so called fibril-prone conformation in a monomer state. The higher the population, the faster is the fibril elongation and this dependence may be described by a single exponential function. This observation opens up a new way to understand the fibrillogenesis of bio-molecules at the monomer level. We will also discuss the influence of the environment with focus on the recently observed dual effect of crowders on the aggregation rates of polypeptide chains.

Keywords: fibril formation,binding free energy

1. Introduction

The protein folding and function take place in an environment crowded with biological macromolecules. As a result proteins are exposed to intermolecular interactions that may lead to aggregation [1]. In all, about 20 proteins and polypeptides such as polylysine or polyglutamic acid peptides, myoglobin, SH3 *et al.* are now implicated in amyloid formation *in vivo* [2]. In many cases protein

aggregates take the form of amyloid fibrils, which appear as unbranched rod-like nanostructures with the diameter of an order of 10 nm and varying length [3]. A large body of evidence suggests that amyloid fibrils and associated oligomeric intermediates are related to a number of diseases, including Alzheimer's, Parkinson's, Huntington's, and prion diseases [2]. For example, in the case of Alzheimer's disease the memory decline may result from the accumulation of the amyloid beta (Aβ) peptide present in two forms – 40 (Aβ_{40}) and 42 (Aβ_{42}) amino acids which are produced through endoproteolysis of the β-amyloid precursor transmembrane protein.

Although amyloid forming proteins and peptides exhibit no obvious sequence or structure homology, the common structural element shared by all amyloid fibrils is an extensive cross-β structure stabilized by backbone hydrogen bonds oriented parallel to the fibril axis. An important question that emerges then is: what are the general principles that govern the fibril formation process?

In this mini-review we consider the key factors that control the aggregation process. These factors are divided into two groups: intrinsic characteristics of proteins and external conditions. The first group involves properties of a polypeptide chain such as a sequence, aromaticity, charge, hydrophobicity *etc.* We highlight the role of an aggregation-prone ensemble of $\mathbf{N}^*$ structures [4] in the folding landscape of the monomer in determining the propensity of sequences to form fibrils. The external factors refer to the properties of the surrounding environment such as temperature, pH, salt concentration, crowding and others. Here, we focus on the crowding effects on aggregation kinetics.

2. Intrinsic factors controlling aggregation rates of proteins

2.1. Sequences and mutations

The protein sequence determines its propensity not only to folding but also to misfolding. The role of sequence in aggregation may be understood by studying mutations which can alter aggregation rates and toxicity to cell and in some cases even change the morphology of fibril structures [1]. For illustration we consider the case of Aβ peptides. Since the turn region 21–23 of these peptides might play a crucial step in fibril formation, numerous experimental as well as theoretical studies have been performed for various mutations in this region including the Flemish (A21G) , Dutch (E22Q), Italian (E22K), Arctic (E22G), Iowa (D23N) and Osaka (ΔE22, deletion) variants (see Ref. [5, 6] and references therein). The G25L, G29L, G33L, G33A, G33I and G37L mutants of Aβ_{42} undergo β-sheet and fibril formation at an increased rate compared with wild-type (WT) Aβ_{42} [7]. On the other hand, as regions 1–8 of Aβ_{40} and 1–16 of Aβ_{42} were believed to be disordered in the fibril state [8, 9], the mutation in the N-terminal has attracted little attention of researchers. However, recent experiments [10] have suggested that residues at the N-terminal may be ordered and this terminal could carry some structural importance. The English (H6R) [11], Taiwanese (D7H) [12] and

Tottori (D7N) [13] mutations alter the fibril formation rate and the survival of cells without affecting the Aβ production [11]. The mutation A2V was found to enhance the Aβ40 aggregation kinetics, but the mixture of the Aβ40 WT and A2V peptides protects against AD [14]. All-atom simulations [15–18] were able to capture experimental findings on impacts of various mutations of Aβ peptides.

2.2. Charge

One of the best examples on the role of the charge in self-assembly is the 101-residue ribosomal protein S6 from *Thermus thermophilus* which does not aggregate but its chains form a tetramer, if charged residues are replaced by four neutral nonpolar amino acids via the quadrupole mutation E41A/E42I/R46M/R47V [19]. The removal of these so called structural gate-keepers promotes aggregation as the repulsion between chains due to electrostatic interaction is reduced. For the same reason the English (H6R) [11] and Tottori (D7N) [13] mutations enhance the fibril formation rate by about 10-fold because the net charge of Aβ peptides is reduced from $-3e$ to $-2e$ upon mutations. Overall, the reduction in an absolute value of a net charge promotes self-assembly of proteins and peptides [1].

2.3. Aromatic interaction and hydrophobicity

Having analyzed a variety of short functional fragments from unrelated amyloid-forming proteins, Gazit observed a remarkable occurrence of aromatic residues [20] pointing to the important role of π-stacking interactions in the formation of amyloid fibers. The stacking interactions provide energetic contribution as well as order and directionality in the self-assembly of amyloid structures. Diphenylalanine peptide FF shows remarkably high propensity to self-aggregation forming highly stable nanotubes at a physiological pH [21].

On the other hand, recent studies [22] have cast doubt on the suggestion that aromatic side-chain interactions play a central role in early self-assembly recognition events. Namely, it was found that in some amyloid peptides, aromatic residues could be mutated by other hydrophobic residues and these nonaromatic variant peptides still would retain competency to form amyloid fibril, although with attenuated kinetics. If two aromatic phenylalanine residues in the amphipathic peptide FKFE are replaced with nonaromatic natural residues with lower hydrophobicity (alanine, valine, and leucine), then the self-assembly does not occur. However, the propensity to fibril formation remains high if they are mutated with a nonnatural residue with greater hydrophobicity (cyclohexylalanine) [22]. Thus, aromatic interactions are not strictly required for amyloid formation and nonaromatic, but highly hydrophobic, cyclohexylalanine appears to promote self-assembly. Manipulating with different variants of fragment Aβ_{16-22} Senguen *et al.* [23] have shown that sequence hydrophobicity alone does not dictate self-assembly, but it is rather aromatic, hydrophobic, and steric considerations that collectively influence fibril formation.

2.4. Fibril-prone structure population

The fibril-prone structure $\mathbf{N}^*$ is defined as a structure of a monomer in the fibril state. The example of $\mathbf{N}^*$ for the fibril of truncated peptide Aβ_{17-42} (Figure 1 a) is given in Figure 1 (b), where the first 16 unstructured residues were omitted.

Recently we have developed a simple lattice model to study the kinetics of fibril formation [24]. The 8-bead sequence in this model is +HHPPHH-, where H and P refer to hydrophobic and polar residues while two oppositely charged residues are assigned at the ends to speed up aggregation. Information about the interaction energies on a cubic lattice is available in our previous works [24, 4]. The fibril-like structure of the lattice model with 10 chains is given in Figure 1 (c) and the corresponding structure of $\mathbf{N}^*$ is shown in Figure 1 (d).

The population of $\mathbf{N}^*$ in a monomer state is $P_{N^*} = \exp(-\Delta E/k_BT)/Z$, where ΔE is the gap separating it from the ground state (GS) (Figure 1 e) and Z refers to the partition function. For conventional proteins, GS is the native state, whereas for intrinsically disordered proteins like Aβ peptides, GS refers to the state with the lowest energy. The high population of this state with a narrow gap would facilitate escape from GS enhancing the propensity to aggregation [25].

In order to support this hypothesis one can use the lattice model [24]. Since all possible conformations of a 8-bead chain on a simple cubic lattice can be enumerated, the population P_{N^*} was computed exactly. We have shown that the fibril formation time τ_{fib} depends on P_{N^*} exponentially implying that P_{N^*} is a key factor that governs aggregation of polypeptide chains [4].

The important role of the population of $\mathbf{N}^*$ was also revealed by all-atom simulation of fibril formation of short peptides using different force fields [26, 27]. Enhancement of P_{N^*} either by mutation or chemical cross linking should increase fibril formation rates. Indeed, a recent experiment [28] showed that the aggregation rate of Aβ_{1-40}-lactam[D23–K28], in which the residues D23 and K28 are chemically constrained by a lactam bridge, is nearly 1000 times greater than in the wild-type. Since the salt bridge constraint increases the population of the $\mathbf{N}^*$ conformation in the monomeric state [27], this observation is consistent with our hypothesis about the important role of $\mathbf{N}^*$. Another example supporting this hypothesis is that Aβ_{42} peptides aggregate into β-sheet fibril much faster than Aβ_{40} ones [29] because the former has higher β-content in the monomer state [30, 31].

The relationship between P_{N^*} in a monomer state and the aggregation rate has been invoked to understand a number of experiments on mutations of Aβ peptides [16–18]. Using all-atom molecular dynamics simulations in explicit solvent one can show [18] that upon D7H mutation, the helix content of Aβ_{42} remains almost zero, while main variations in a coil occur in the disordered N-terminal (Figure 2). The substantial reduction in the β-structure in fibril-prone regions expanded over residues 19–23, 31–35 and 39–41, and the increase in turn at the segments 19–21 and 30–41 may be interpreted as reduction of the population of $\mathbf{N}^*$ in the monomeric state. Thus, such a reduction is in agreement with the

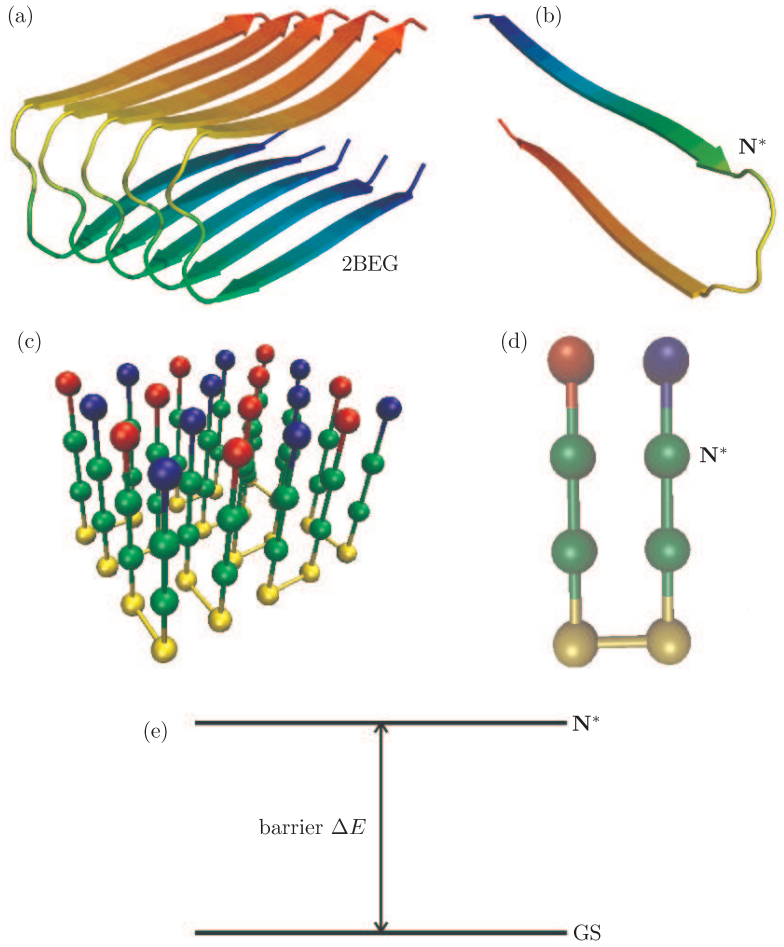

Figure 1. (a) the protofibril structure of truncated amyloid beta peptides 5Aβ_{17-42} (PDB ID: 2BEG); (b) the fibril-prone structure $\mathbf{N}^*$ of 5Aβ_{17-42}; (c) fibril-like structure of the 10 chains in the lattice model. Charged residues are in blue and red, while green and golden stand for hydrophobic and polar residues, respectively; (d) fibril-prone structure $\mathbf{N}^*$ in a lattice model; (e) schematic cartoon for the energy barrier between $\mathbf{N}^*$ and the ground state of the monomer

experiment [12] that the Taiwanese mutation slows down aggregation of the Aβ_{42} peptide. It is presumably useful to gather together the knowledge about the population of $\mathbf{N}^*$ for elucidating the fibrillogenesis at a single-monomer level. This is of paramount importance because the fibril formation is an extremely slow process which is difficult for numerical study of many chains.

3. Impact of environment on aggregation process

Such external factors as temperature, pH, salt concentration *etc.* may alter the propensity to aggregation of polypeptide chains. Myoglobin can form fibrillar structures at pH 9.0 and 65°C conditions under which the native fold

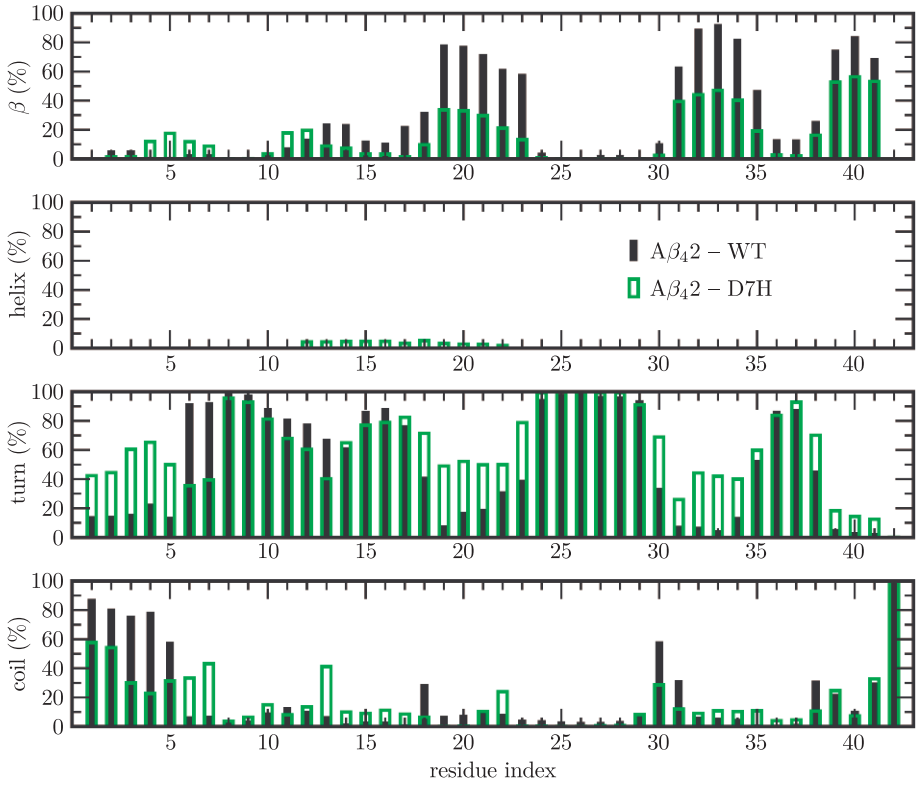

Figure 2. Per-residue distributions of secondary structures of the Aβ_{42} monomer in the presence and absence of D7N mutation. The results were obtained at $T = 311.8$ K using the replica exchange molecular dynamics simulation with the OPLS force field [32] and implicit solvent [33]. Taken from Truong *et al.* [18]

is substantially destabilized [34]. The insulin molecule is, however, indicated to convert to β-sheet and assemble to a fibrillar structure from the native state in vitro at much lower pH ≈ 2.0 [35]. Denaturant, urea, decreased the lag time of insulin aggregation, whereas the stabilizers, trimethylamine N-oxide dihydrate (TMAO) and sucrose, increased the lag times [36]. The presence of a membrane can substantially modulate not only self-assembly rates but also the polymorphism of fibril. Disordered peptides like Aβ can penetrate the membrane forming ion-conducting channels as indicated by electrophysiological data [37].

Recently some progress has been made in apprehending the effects of macromolecular crowding on folding, conformation and function of proteins [38]. The aggregation of proteins is a more complex process because it depends not only on the monomer sequence but also on intermolecular interactions, nucleation rates, diffusive properties *etc.* [39]. Despite this apparent complexity one can delineate general features for fibril growth in crowded environment showing that the propensity to self-assembly does not have to depend on some details of the studied systems. For example, having used a quartz crystal microbalance

(QCM) assay with a high level of accuracy in measuring fibril growth rates, Dobson et al. have shown that cosolutes accelerate the fibril elongation [40]. This finding is supported by numerical simulations [41, 42] and in line with the depletion theory [38, 43]. Similar dependence of fibril formation rates on crowder concentration has been found by Linse et al. [44] who have shown that copolymeric nanoparticles accelerate β_2-microglobulin. The opposing effect has been observed for amyloid β peptides the fibril growth of which is retarded by nanoparticles [45]. It has been interpreted as a result of depletion of solution concentration leading to a block of the growing ends. The most striking experimental observation has been made by Cabaleiro-Lago et al. who reported on the dual effect of amino modified polystyrene nanoparticles on Aβ_{40} and Aβ_{42} peptide fibrillation [46]. At a constant peptide concentration, the fibril formation is accelerated by crowders at low crowder concentration, while at high concentration this process is slowed down. Thus, contrary to other groups which have seen either acceleration [44, 47] or retardation [45, 48], they have observed both effects by one nanoparticle type. This interesting dual effect occurs presumably due to the competition between two different mechanisms of interaction between protein and crowders which is tuned by the total crowder surface area A_{total} [46] (Figure 3 a).

Since simulation of fibril formation kinetics of full-length Aβ peptides is beyond present computational facilities, in order to explain the dual effect of crowders on fibril formation rates we have extended our lattice model [24] to include crowding agents which are modeled as impenetrable cubes of various sizes (Figure 3 b) [49]. Although the 8-bead sequence is different from Aβ_{40} and Aβ_{42} our model is still capable of capturing the crowding effect due to its universal nature. Namely, as seen below, the dual effect occurs as a switch between entropy-driven and energy-driven regimes independent of details of involved systems.

As follows from Figure 3 (c), we obtained good agreement with the experimental result on the Aβ fibril formation in the presence of a modified polystyrene nanoparticle [46] that τ_{fib} first decreases and then increases with the coverage of nanoparticles (Figure 3 a). As explained above, at small crowder coverages the entropy of aggregating agents is largely resulting in slow self-assembly. This situation is similar to slow protein folding at high temperatures [50]. The retardation at large surface areas comes from the reduction of entropy of proteins in a very crowded environment. The fast fibril formation occurs in the region where the energetics and entropic contributions compromise.

We have also developed a lattice model to study the impact of confinement on aggregation kinetics [49]. It was shown that the fibril elongation was fast at some optimal size of confining cavity but it became slow in small (energy driven) and large (entropy driven) encapsuled space.

4. Conclusion

The manner in which chains of amino acids fold into fiber depends both on the intrinsic properties of the amino-acid sequence and on multiple contributing

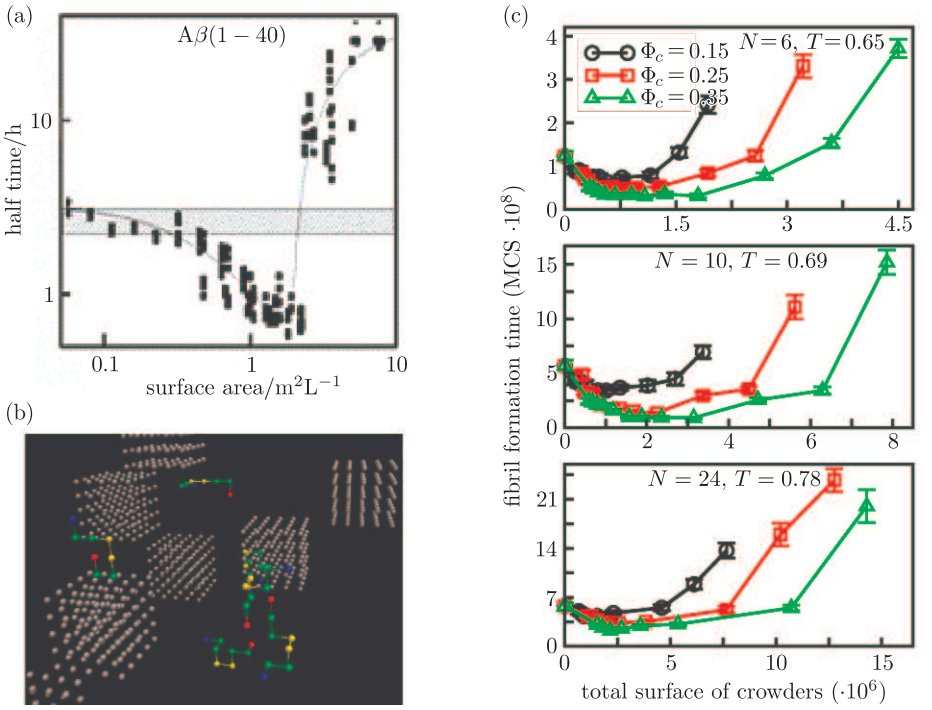

Figure 3. (a) dependence of the half time of lag phase of $A\beta_{1-40}$ peptides on the total surface area of amine-modified polystyrene nanoparticles. These experimental data was taken from Cabaleiro *et al.* [46]; (b) snapshot of crowders (grey cubes) and polypeptide chains (color); (c) dependence of $\tau_{\rm fib}$ on the total surface area of crowders at three values of crowder concentration $\Phi_c = 0.15$, 0.25, and 0.35; $T = 0.65$, 0.69 and 0.78 for the number of chains $N = 6$, 10, and 24, respectively; taken from Co *et al.* [49]

influences from the crowded cellular environment. We have reviewed the key factors such as charge, hydrophobicity, aromatic rings and population of the fibril-prone state in controlling aggregation of polypeptide chains. However, these factors are not independent from each other depending on protein sequences. External conditions, like temperature, pH, salt concentration, milieu crowding *etc.* also considerably affect aggregation.

Acknowledgements

This work was supported by Narodowe Centrum Nauki in Poland (Grant No 2011/01/B/NZ1/01622), Polish Ministry of Science and Higher Education Grant No. IP2012 016872, and the Department of Science and Technology of the Ho Chi Minh city, Vietnam.

References

[1] Dobson C M 2003 *Nature* **426** 884
[2] Chiti F, Dobson C M 2006 *Annual Rev. Biochemistry* **75** 333
[3] Tycko R 2004 *Curr. Opin. Struct. Biol.* **14** 96
[4] Li M S, Co N T, Hu C K, Straub J E, Thirumalai D 2010 *Phys. Rev. Lett.* **105** 218101

[5] Tomiyama T, Nagata T, Shimada H, Teraoka R, Fukushima A, Kanemitsu H, Takuma H, Kuwano R, Imagawa M, Ataka S *et al.* 2008 *Ann. Nerol.* **63** 377
[6] Coskuner O, Wise-Scira O, Perry G, Kitahara T 2013 *ACS Chemical Neuroscience* **4** 310
[7] Lu Y, Wei G H, Derreumaux P 2011 *J. Phys. Chem. B* **115** 1282
[8] Paravastu A K, Leapman R D, Yau W M, Tycko R 2008 *PNAS* **105** 18349
[9] Luhrs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Doeli H, Schubert D, Riek R 2005 *Proc. Natl. Acad. Sci. USA* **102** 17342
[10] Lu J X, Qiang W, Yau W M, Schwieters C D, Meredith S C, Tycko R 2013 *Cell* **154** 1257
[11] Hori Y, Hashimoto T, Wakutani Y, Urakami K, Nakashima K, Condron M M, Tsubuki S, Saido T C, Teplow D B, Iwatsubo T 2007 *J. Bio. Chem.* **282** 4916
[12] Chen W T, Hong C J, Lin Y T, Chang W H, Huang H T, Liao J Y, Chang Y J, Hsieh Y F, Cheng C Y, Liu H C, Chen Y R, Cheng I H 2012 *Plos One* **7**, e35807
[13] Ono K, Condron M M, Teplow D B 2010 *J. Bio. Chem.* **285** 23186
[14] Fede G D, Catania M, Morbin M, Rossi G, Suardi S, Mazzoleni G, Merlin M, Giovagnoli A R, Prioni S, Erbetta A *et al.* 2009 *Science* **323** 1473
[15] Lin Y-S, Pande V S 2012 *Biophys. J.* **103**, L47
[16] Viet M H, Nguyen P H, Ngo S T, Li M S, Derreumaux P 2013 *ACS Chem. Neurosci.* **4** 1446
[17] Viet M H, Nguyen P H, Derreumaux P, Li M S 2014 *ACS Chem. Neurosci.* DOI: 10.1021/cn500007j
[18] Truong P M, Viet M H, Nguyen P H, Hu C K, Li M S 2014 *J. Phys. Chem. B* **118** 8972
[19] Otzen D E, Kristensen O, Oliveberg M 2000 *Proc. Natl. Acad. Sci. (USA)* **97** 9907
[20] Gazit E 2002 *FASEB Journal* **16** 77
[21] Adler-Abramovich L, Gazit E 2014 *Chem. Soc. Rev.* **43** 6881
[22] Bowerman C J, Ryan D M, Nissan D A, Nilsson B L 2009 *Mol Biosyst* **5** 1058
[23] Senguen F T, Lee N R, Gu X, Ryan D M, Doran T M, Anderson E A, Nilsson B L 2011 *Mol Biosyst* **7** 486
[24] Li M S, Klimov D K, Straub J E, Thirumalai D 2008 *J. Chem. Phys.* **129** 175101
[25] Straub J E, Thirumalai D 2010 *Curr. Opin. Struct. Biol.* **20** 187
[26] Nam H B, Kouza M, Zung H, Li M S 2010 *J. Chem. Phys.* **132** 165104
[27] Reddy G, Straub J E, Thirumalai D 2009 *J. Phys. Chem. B* **113** 1162
[28] Sciarretta K, Gordon D, Petkova A, Tycko R, Meredith S 2005 *Biochemistry* **44** 6003
[29] Snyder S W, Ladror U S, Wade W S, Wang G T, Barrett L W, Matayoshi E D, Huffaker H J, Krafft G A, Holzman T F 1994 *Biophys. J.* **67** 1216
[30] Sgourakis N G, Yan Y L, McCallum S A, Wang C Y, Garcia A E 2007 *J. Mol. Biol* **368** 1448
[31] Yang M, Teplow D B 2008 *J. Mol. Biol.* **384** 450
[32] Kaminski G A, Friesner R A 2001 *J. Phys. Chem. B* **105** 6474
[33] Jorgensen J W, Chandrasekhar J, Madura J D, Imprey R W, Klein M L 1983 *J. Chem. Phys.* **79** 926
[34] Dobson C M 2001 *Nature* **410** 165
[35] Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, Vyas S, Uversky V N, Fink A L 2001 *Biochemistry* **40** 6036 PMID: 11352739
[36] Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, Vyas S, Uversky V N, Fink A L 2001 *Biochemistry* **40** 6036
[37] Jang H, Arce F T, Ramachandran S, Kagan B L, Lal R, Nussinov R 2014 *Chem Soc Rev* **43** 6750
[38] Zhou H X, Rivas G N, Minton A P 2008 *Annual Rev. Biophys.* **37** 375
[39] Kuriyan J, Eisenberg D 2007 *Nature* **450** 983
[40] White D A, Buell A K, Knowles T P J, Welland M E, Dobson C M 2010 *J. Am. Chem. Soc.* **132** 5170

[41] Magno A, Caflisch A, Pellarin R 2010 *J. Phys. Chem. Lett.* **1** 3027
[42] O'Brien E P, Straub J E, Brooks B R, Thirumalai D 2011 *J. Phys. Chem. Lett.* **2** 1171
[43] Asakura S, Oosawa F 1954 *J. Chem. Phys.* **22** 1255
[44] Linse S, Cabaleiro-Lago C, Xue W F, Lynch I, Lindman S, Thulin E, Radford S E, Dawson K A 2007 *Proc. Natl. Acad. Sci. (USA)* **104** 8691
[45] Cabaleiro-Lago C, Quinlan-Pluck F, Lynch I, Lindman S, Minogue A M, Thulin E, Walsh D M, Dawson K A, Linse S 2008 *J. Am. Chem. Soc.* **130** 15437
[46] Cabaleiro-Lago C, Quinlan-Pluck F, Lynch I, Dawson K A, Linse S 2010 *ACS Chem. NeuroSci.* **1** 279
[47] Wu W H, Sun X, Yu Y P, Hu J, Zhao L, Liu Q, Zhao Y F, Li Y M 2008 *Biochem. Biophys. Res. Commun.* **373** 315
[48] Klajnert B, Cortijo-Arellano M, Bryszewska M, Cladera J 2006 *Biochem. Biophys. Res. Commun.* **339** 577
[49] Co N T, Hu C K, Li M S 2013 *J. Chem Phys* **138** 185101
[50] Cieplak M, Hoang T X, Li M S 1999 *Phys. Rev. Lett.* **83** 1684