Published as: *Angew Chem Int Ed Engl.* 2010 May 25; 49(23): 3904–3908.

A multi-dimensional Structure-Activity Relationship of a protein in its aggregated states

Lei Wang¹, David Schubert², Michael R. Sawaya^{3,4}, David Eisenberg^{3,4}, and Roland Riek^{1,5,*}

¹ Laboratory of Physical Chemistry, Swiss Federal Institute of Technology (ETH), Zurich, 8093, Switzerland ² Cellular Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, 92037, United States of America ³ Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California, 90095, United States of America ⁴ University of California, Los Angeles–Department of Energy (UCLA-DOE) Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California, 90095, United States of America ⁵ Structural Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, 92037, United States of America

Abstract

Protein aggregates are both associated with disease and function. Because a variety of factors induce protein aggregation, a given protein can aggregate into different states. Here, we compare the structures and activities of five distinct protein aggregates of a single protein. Despite the diverse chemical, physical and biological treatments used to induce aggregation, all aggregate types contain the cross- β -sheet motif. However, they are structurally distinct, having different segments of the protein sequence involved in secondary structure formation. Because of these structural differences each aggregate has a unique set of properties. These include affinity to ATP, Thioflavin T, DNA, and membrane mimics, and interference with cell viability. The key to their multiple properties may be that the repetitive nature of the cross- β -sheet motif guarantees for many potent activities through cooperativity. The observed multidimensional structure-activity relationship of protein aggregates may be important for amyloid diseases but may also be advantageous in nanotechnology.

Keywords

Protein Aggregation; Amyloid Fibrils; Inclusion Bodies; Cross- β -Sheet; EM; X-ray Diffraction; FT-IR; NMR; Hydrogen/Deuterium Exchange; Structure-Activity

Protein aggregation has been associated with pathological conditions,[1-4] is an ever-present challenge for the cell machinery,[5] but is also involved in biological functions.[6-9] In addition, aggregation into inclusion bodies during protein over-expression in *E. coli* and during production of protein pharmaceuticals are concerns in biotechnology,[10] while deliberately induced aggregation is utilized as a biochemical method, such as trichloroacetic acid (TCA)-precipitation.[11] Astbury suggested from studies of poached egg white that when proteins aggregate, they go into alternative, fibrous conformational states composed of β -sheets, that run across the fibril.[12] This so-called cross- β sheet structural motif is also present in amyloid fibrils associated with disease, normal functions, as well as in bacterial

* To whom correspondence should be addressed. Telephone: +41 44 632 6139; Fax: + 41 44 633 1448; roland.riek@phys.chem.ethz.ch.

inclusion bodies.[4, 6, 13-21] Furthermore, protein aggregates induced by high temperature, high protein concentration or extreme pH show more extensive β -sheet secondary structure than their soluble counterparts, [22-24] which contrasts to pressure-induced aggregates.[22] We explored here the hypothesis that proteins aggregate into specific environment-dependent, ordered, structural states that yield many properties. To do so, we studied the structural landscape and the structure-activity relationships of aggregates of the N-terminal domain of the hydrogenase maturation factor HypF (HypF-N), which is an established model system of protein aggregation.[25-27] Five distinct protein aggregates of HypF-N were studied: amyloid-like fibrils, bacterial inclusion bodies, heat-precipitated aggregates, concentration-induced aggregates, and TCA-precipitated aggregates. These aggregates appear to be a representative set because diverse chemical, physical and biological treatments were used to induce aggregation.

The five types of HypF-N aggregates display distinct morphologies under the electron microscope (EM) (Figures 1a and S1, see also Supporting Information). However, they all show a similar X-ray fiber diffraction pattern typically observed for the cross- β structure, with a major reflection at 4.7 Å resolution interpreted as the spacing between strands in a β -sheet and a diffuse reflection at ~ 10 Å interpreted as the spacing between β -sheets (Figure 1a). The circular profiles of these reflections, rather than the cross- β -typical orthogonal positions, show that the cross- β -sheet structural entities in the aggregates are not strongly aligned. Fourier transform infrared red (FT-IR) spectroscopy further revealed that the cross- β structures in these aggregates are different (Figures 1b & S2).[21]

To localize the β -sheets of the cross- β -sheet entities within the amino acid sequence, quenched hydrogen/deuterium (H/D) exchange nuclear magnetic resonance (NMR)[16, 28] was measured, which allows the identification of solvent-protected backbone amide protons (Figures S3 – S8). The H/D exchange data of Figure 2 indicate that residues 8-14, 17-20, 26-30 and 32-39 of fibrils, residues 10-16, 28-31, 42-44 and 58-62 of inclusion bodies, residues 11-13, 28-31 and 58-61 of heat-precipitated aggregates, residues 12-13, 28-31 and 58-62 of concentration-induced aggregates, and to a small percentage residues 9-13 and 41-43 of TCA-precipitated aggregates display slow exchange rates of 10^{-2} to 10^{-4} h $^{-1}$ and are considered to be involved in hydrogen bonds, while the other residues are not or only weakly protected and hence considered to be disordered. For the minor population (Figure S5: $P < 1/2$) other amide residues (colored grey in Figure 2) show also slow exchange rates indicative of structural heterogeneity within a given type of aggregate. In combination with the other biophysical data, the H/D exchange data indicate that each aggregate entity occupies a distinct conformational state, different from that of soluble HypF-N (Figure S6), having different segments of the protein sequence involved in the cross- β -sheet structural motif. The presence of multiple aggregate structures per protein is in line with other studies. [29-31]

The presence of distinct, but defined and reproducible (Figure S8) conformations of the various aggregates studied suggests that each aggregate type displays distinct properties. To get insights into such a suggested multi-dimensional structure-activity relationship we measured a few diverse (biologically interesting) activities: (i) The aggregates have a different conformational stabilities in urea (Figure 3a) with the ranking amyloid fibrils > soluble HypF-N > inclusion bodies > concentration-induced aggregates > heat-precipitated aggregates > TCA-precipitated aggregates. (ii) Because soluble HypF-N is an acylphosphatase-like protein that can bind a phosphate anion,[27] we tested the affinities for adenosine-5'-triphosphate (ATP) of soluble HypF-N and the aggregates (Figure 3b). All structural entities bind ATP less than soluble HypF-N with the following ranking: soluble HypF-N > fibrils > heat-precipitated aggregates $\approx$ concentration-induced aggregates $\approx$ TCA-precipitated aggregates > inclusion bodies. (iii) The affinity of the various aggregates for the

amyloid dye Thioflavin T (ThT) (Figure 3c) follows the same trend as the stability measurements with exception of the soluble HypF-N which does not bind to Thio T. (iv) In contrast to soluble HypF-N, most aggregate entities are able to bind the DNA oligonucleotide (Figure 3d) with the following ranking: fibrils > heat-precipitated aggregates > concentration-induced aggregates $\approx$ TCA-precipitated aggregates > inclusion bodies $\approx$ soluble HypF-N. (v) The capacity for binding to micelles composed of the zwitter-ionic phospholipid 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC) follows the ranking (Figure 3e): fibrils > TCA-precipitated aggregates > concentration-induced aggregates $\approx$ soluble HypF-N aggregates > heat-precipitated $\approx$ inclusion bodies. (vi) Finally, we tested whether the various aggregates and the monomeric HypF-N are able to disturb the cell function in reducing 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT), which is an assay to measure amyloid toxicity.[32] The MTT reduction is perturbed in the order (Figure 3f): fibrils > TCA-precipitated aggregates > inclusion bodies > heat-precipitated aggregates $\approx$ concentration-induced aggregates $\approx$ soluble HypF-N. In summary, each of the five distinct aggregated structures displays distinct properties and activities with the fibrils having the highest activities overall. While the inclusion bodies do not interact with membrane-mimetics, ATP and DNA, they interfere with the cell viability. In contrast, little activity in membrane binding and the MTT assay is observed for the concentration and temperature-induced aggregates, while they bind considerably DNA. TCA-precipitated aggregates bind DNA and show a toxic response in the MTT assay.

The data presented shows that when HypF-N aggregates it generally forms a cross- β -sheet entity comprising a specific and reproducible structure that is determined by the aggregation conditions. Each of the aggregation states is characterized by several short protein segments participating in cross- β structure while the remaining residues are less structured or disordered. Each aggregate has different segments of the protein sequence involved in this β -sheet structure formation. Because these protein aggregates are structured, they show activities that are gained upon aggregation. HypF-N aggregates show binding capacities for a diverse set of molecules with estimated affinities per HypF-N molecule down to the μ M range per HypF-N molecule of the aggregate (note: only an upper value is given since the size of the aggregates and the number of binding sites therein is unknown). It is evident that the aggregate-dependent activities arise from the different structures. Interestingly, the aggregates appear to have a plurality of activities in the μ M range. This property is attributed here by the repetitive nature of the cross- β -sheet motif, that is able to enhance on a multiplicative level a weak activity into a strong one by cooperativity and therefore may generate many potent activities. Some of these activities may be toxic to the host yielding a multifactorial toxicity typically observed in amyloid diseases.[1-4]

Together with the multiple conformations of aggregates the cooperativity effect of the repetitive cross- β sheet structure may provide a basis for the presence of prion strains:[4, 33] The aggregation conditions activate different sequence segments into the cross- β -sheet motif resulting in different structures with different properties that can account for a variety of phenotypic behaviors. Alternatively, slight structural alterations of the aggregates may result in distinct activities because the repetitive nature of the cross- β -sheet entity may be a multiplier of a weak activity into a strong one by cooperativity.

The observed plurality of μ M activities of protein aggregates indicates that protein aggregation is rather a primitive folding event than misfolding. This also supports the idea that protein aggregation has played a key role in the early evolution of proteins.[7, 34] In contrast to the unlikeliness that in the early phase of life, amino acid sequences were established that resulted in well-folded soluble proteins, it can be envisioned that polypeptides of primitive sequence composition were aggregating reproducibly into an aggregate entity with decent functionalities.

In conclusion, the variety of structures and activities we find for HypF-N aggregates highlight the presence of specific and defined conformations for each aggregate entity. Hence, in contrast to its soluble protein state with a single conformation, HypF-N aggregates have a complex structural landscape accompanied with multiple, aggregate-specific activities resulting in a multidimensional structure-activity relationship. It is likely that such complex relationships may play important roles both in amyloid diseases and functional amyloids, and may have been key in the early evolution.[34]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. S. Choe, L. Goldschmidt, E. Evangelisti, C. Cecchi, P. Arosio, M. Morbidelli for experimental assistance and discussions. This research was supported by the Swiss National Science.

REFERENCES

1. Selkoe DJ. Nature. 2003; 426:900. [PubMed: 14685251]
2. Dobson CM. Nature. 2003; 426:884. [PubMed: 14685248]
3. Kelly JW. N. Engl. J. Med. 2005; 352:722. [PubMed: 15716569]
4. Tanaka M, Collins SR, Toyama BH, Weissman JS. Nature. 2006; 442:585. [PubMed: 16810177]
5. Kopito RR. Trends Cell Biol. 2000; 10:524. [PubMed: 11121744]
6. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. PLoS Biol. 2008; 6:e195. [PubMed: 18684013]
7. True HL, Lindquist SL. Nature. 2000; 407:477. [PubMed: 11028992]
8. Fowler DM, Koulov AV, Alory-Jost C, Marks MS, Balch WE, Kelly JW. PLoS Biol. 2006; 4:e6. [PubMed: 16300414]
9. Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, Singru PS, Nilsson KP, Simon R, Schubert D, Eisenberg D, Rivier J, Sawchenko P, Vale W, Riek R. Science. 2009
10. Ventura S, Villaverde A. Trends Biotechnol. 2006; 24:179. [PubMed: 16503059]
11. Isaacson T, Damasceno CM, Saravanan RS, He Y, Catala C, Saladie M, Rose JK. Nat. Protoc. 2006; 1:769. [PubMed: 17406306]
12. Astbury WT, Dickinson S. Biochem. J. 1935; 29:2351. [PubMed: 16745914]
13. Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CC. J. Mol. Biol. 1997; 273:729. [PubMed: 9356260]
14. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D. Nature. 2005; 435:773. [PubMed: 15944695]
15. Ritter C, Maddelein ML, Siemer AB, Luhrs T, Ernst M, Meier BH, Saupe SJ, Riek R. Nature. 2005; 435:844. [PubMed: 15944710]
16. Luhrs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Dobeli H, Schubert D, Riek R. Proc. Natl. Acad. Sci. USA. 2005; 102:17342. [PubMed: 16293696]
17. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen AO, Riekel C, Eisenberg D. Nature. 2007; 447:453. [PubMed: 17468747]
18. Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH. Science. 2008; 319:1523. [PubMed: 18339938]
19. Ignatova Z, Gierasch LM. Biochemistry. 2005; 44:7266. [PubMed: 15882065]
20. Wang L. Prion. 2009; 3:139. [PubMed: 19806034]
21. Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S. J. Mol. Biol. 2005; 347:1025. [PubMed: 15784261]
22. Okuno A, Kato M, Taniguchi Y. Biochim. Biophys. Acta. 2006; 1764:1407. [PubMed: 16920040]

23. Chang ES, Liao TY, Lim TS, Fann W, Chen RP. *J. Mol. Biol.* 2009; 385:1257. [PubMed: 19041877]
24. Kendrick BS, Cleland JL, Lam X, Nguyen T, Randolph TW, Manning MC, Carpenter JF. *J. Pharm. Sci.* 1998; 87:1069. [PubMed: 9724556]
25. Chiti F, Bucciantini M, Capanni C, Taddei N, Dobson CM, Stefani M. *Protein Sci.* 2001; 10:2541. [PubMed: 11714922]
26. Calloni G, Lendel C, Campioni S, Giannini S, Gliozzi A, Relini A, Vendruscolo M, Dobson CM, Salvatella X, Chiti F. *J. Am. Chem. Soc.* 2008; 130:13040. [PubMed: 18767849]
27. Rosano C, Zuccotti S, Bucciantini M, Stefani M, Ramponi G, Bolognesi M. *J. Mol. Biol.* 2002; 321:785. [PubMed: 12206761]
28. Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H, Goto Y. *Nat. Struct. Biol.* 2002; 9:332. [PubMed: 11967567]
29. Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, Tycko R. *Science.* 2005; 307:262. [PubMed: 15653506]
30. Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D, Baldus M. *Proc. Natl. Acad. Sci. USA.* 2005; 102:15871. [PubMed: 16247008]
31. van der Wel PC, Lewandowski JR, Griffin RG. *J. Am. Chem. Soc.* 2007; 129:5117. [PubMed: 17397156]
32. Liu Y, Schubert D. *J. Neurochem.* 1997; 69:2285. [PubMed: 9375659]
33. Toyama BH, Kelly MJ, Gross JD, Weissman JS. *Nature.* 2007; 449:233. [PubMed: 17767153]
34. Maury CP. *Orig. Life Evol. Biosph.* 2009; 39:141. [PubMed: 19301141]
35. Cecchi C, Pensalfini A, Liguri G, Baglioni S, Fiorillo C, Guadagna S, Zampagni M, Formigli L, Nosi D, Stefani M. *Neurochem. Res.* 2008; 33:2516. [PubMed: 18307032]

Figure 1.

(a) Electron micrographs (EM) (left) and X-ray fiber diffraction patterns (right) of the five aggregate types of HypF-N as indicated. Scale bars indicate 500 nm. Fibrils have a width of 6 ± 2 nm, inclusion bodies have an irregular sphere-like shape with a radius of 280 ± 40 nm, heat-precipitated aggregates also have an irregular sphere-like shape with a radius of 80 ± 20 nm, concentration-induced and TCA-precipitated aggregates have amorphous morphology with widely ranging sizes, with the latter having irregular sphere-like substructures with a radius of 20 ± 4 nm. ($\pm$: standard error). The sharp reflection of the fiber diffraction pattern observed at approximately $4.7 \text{ \AA}$ is interpreted as the spacing between strands in a β -sheet, and the diffuse reflection at approximately $10 \text{ \AA}$ is interpreted as the spacing between β -sheets. (b) FT-IR spectra of HypF-N aggregates and soluble HypF-N. Fibrils (green); inclusion bodies (cyan); heat aggregates (blue), concentration aggregates (yellow); TCA aggregates (red); soluble HypF-N (purple). In the Amide I region, soluble HypF-N shows two major peaks around 1633 cm^{-1} and 1657 cm^{-1} that are usually assigned to intramolecular β -sheet and α -helix secondary structures.[21] For fibrils, there is a sharp peak around 1624 cm^{-1} and a peak around 1694 cm^{-1} that are indicative of newly formed β -sheet structure in the aggregate.[21] While for TCA-precipitated aggregates only little newly formed β -sheet structure is observed, the other three aggregate entities contain increased amounts of β -sheet structure. Spectra were normalized at the tyrosine band around 1513 cm^{-1} to account for differences in the total protein content.

Figure 2.

Sequence-resolved amide exchange rates k_{ex}/h indicative of secondary structures for all five types of HypF-N aggregates studied as indicated. The measured H/D exchange is mostly of bi-exponential nature suggesting the presence of two populations. The exchange rates of the major population are colored green. If the minor population is present more than 1/3, the corresponding exchange rates are shown in grey. Because of spectral overlap, the analysis of the exchange rates for some residues may be difficult to determine. However, most of these overlap problems could be resolved by the assumption that sequential neighboring residues show a similar extent of exchange. The exchange rates that have been extracted following this procedure are colored in light green. In the panel labeled with “A”, predicted aggregation-prone segments of HypF-N are shown using two distinct algorithms: 3DPROFILE in orange and TANGO in blue. The secondary structures of the soluble HypF-N are highlighted in red for α -helix and blue for β -sheet, respectively. For details of prediction algorithms, please see Supporting Information.

Figure 3.

Activities of HypF-N aggregates. (a) Urea disaggregation of HypF-N aggregates and unfolding of soluble HypF-N as labeled. (b) ATP binding capability of different aggregates. ATP and aggregates are mixed at a 1:5 monomer ratio, and the column values represent the percentage of bound ATP vs. total amount of ATP. (c) ThT fluorescence enhancement upon binding to different aggregates. The column values represent the fluorescence intensity at arbitrary unit (AU). (d) DNA binding capability of different aggregates. DNA and aggregates are mixed at a 1:20 monomer ratio, and the column values represent the percentage of bound DNA vs. total amount of DNA. (e) DHPC micelle binding capability of different aggregates. DHPC micelle and aggregates are mixed at a 1:1 concentration ratio and the column values represent the percentage of bound DHPC micelles vs. total amount of DHPC micelles. (f) MTT cell reduction upon addition of different aggregates. The column values represent the percentage of reduced cells with addition of aggregates vs. reduced cells without aggregates. A β_{1-42} amyloid fibrils were measured as a positive control. Activities of soluble HypF-N were also measured as reference. It should be noted that HypF-N fibrils do not show an influence in MTT reduction on the human neurotypic SH-SY5Y cells line used by ref [35] when sent by us to the Chiti group. At least two independent experiments were carried out for each system. ($\pm$: standard error of the experimental measurements for individual preparations)