

Multidimensional Structure–Activity Relationship of a Protein in Its Aggregated States**

Lei Wang, David Schubert, Michael R. Sawaya, David Eisenberg, and Roland Riek*

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] Furthermore, aggregation into inclusion bodies during protein overexpression in *Escherichia coli* and during the production of protein pharmaceuticals are concerns in biotechnology,^[10] whereas deliberately induced aggregation is utilized as a biochemical method, as in trichloroacetic acid (TCA) precipitation.^[11] Astbury and Dickinson 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 and normal functions, as well as in bacterial inclusion bodies.^[4,6,13–21] Furthermore, protein aggregates induced by high temperature, high protein concentration, or extreme pH values show more extensive β -sheet secondary structure than their soluble counterparts,^[22–24] in contrast to pressure-induced aggregates.^[22] In this study, we explored the hypothesis that proteins aggregate into specific environment-dependent, ordered, structural states with many different properties. 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: amyloidlike 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 (Figure 1 a; see also the Supporting Information, including Figure S1). However, they all show a similar X-ray fiber diffraction pattern typically observed for the cross- β structure, with a major reflection at a resolution of about 4.7 Å interpreted as the spacing between strands in a β sheet and a diffuse reflection at approximately 10 Å interpreted as the spacing between β sheets (Figure 1 a). The circular profiles of these reflections, rather than the orthogonal positioning typical of cross- β -sheets, show that the cross- β -sheet structural entities in the aggregates are not strongly aligned. Fourier transform infrared (FTIR) spectroscopy further revealed that the cross- β structures in each of these aggregates are different (Figure 1 b; see also Figure S2 in the Supporting Information).^[21]

To localize the β sheets of the cross- β -sheet entities within the amino acid sequence, we used NMR spectroscopy to detect quenched hydrogen/deuterium exchange.^[16,28] This method enables the identification of solvent-protected backbone amide hydrogen atoms (see Figures S3–S8 in the Supporting Information). The H/D-exchange data in 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 extent residues 9–13 and 41–43 of TCA-precipitated aggregates display slow exchange rates of 10^{-2} – 10^{-4} h⁻¹. These residues are therefore considered to be involved in hydrogen bonds. The other residues are not or are only weakly protected and are hence considered to be disordered. For the minor population (see Figure S5 in the Supporting Information: $P < 1/2$), other amide residues (colored gray in Figure 2) also show 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 (see Figure S6 in the Supporting Information): each has different segments of the protein sequence involved in the cross- β -sheet structural motif. The presence of multiple aggregate structures of a protein is in line with the results of other studies.^[29–31]

The presence of distinct, but defined and reproducible conformations (see Figure S8 in the Supporting Information)

[*] L. Wang, R. Riek

Laboratorium für Physikalische Chemie
ETH Zürich, 8093 Zürich (Switzerland)
Fax: (+41) 44-633-1448
E-mail: roland.riek@phys.chem.ethz.ch

D. Schubert
Cellular Neurobiology Laboratory
The Salk Institute for Biological Studies
La Jolla, CA 92037 (USA)

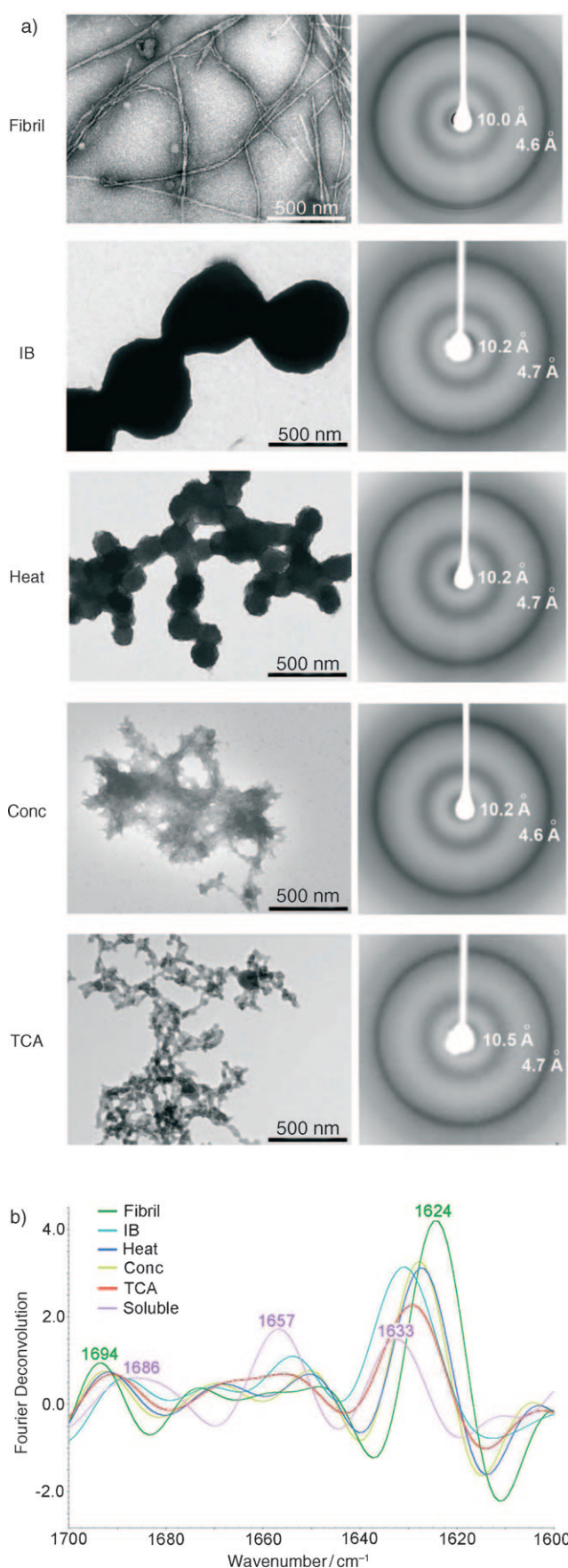
M. R. Sawaya, D. Eisenberg
Howard Hughes Medical Institute
University of California, Los Angeles, CA 90095 (USA)

M. R. Sawaya, D. Eisenberg
UCLA-DOE Institute for Genomics and Proteomics
University of California, Los Angeles

R. Riek
Structural Biology Laboratory
The Salk Institute for Biological Studies

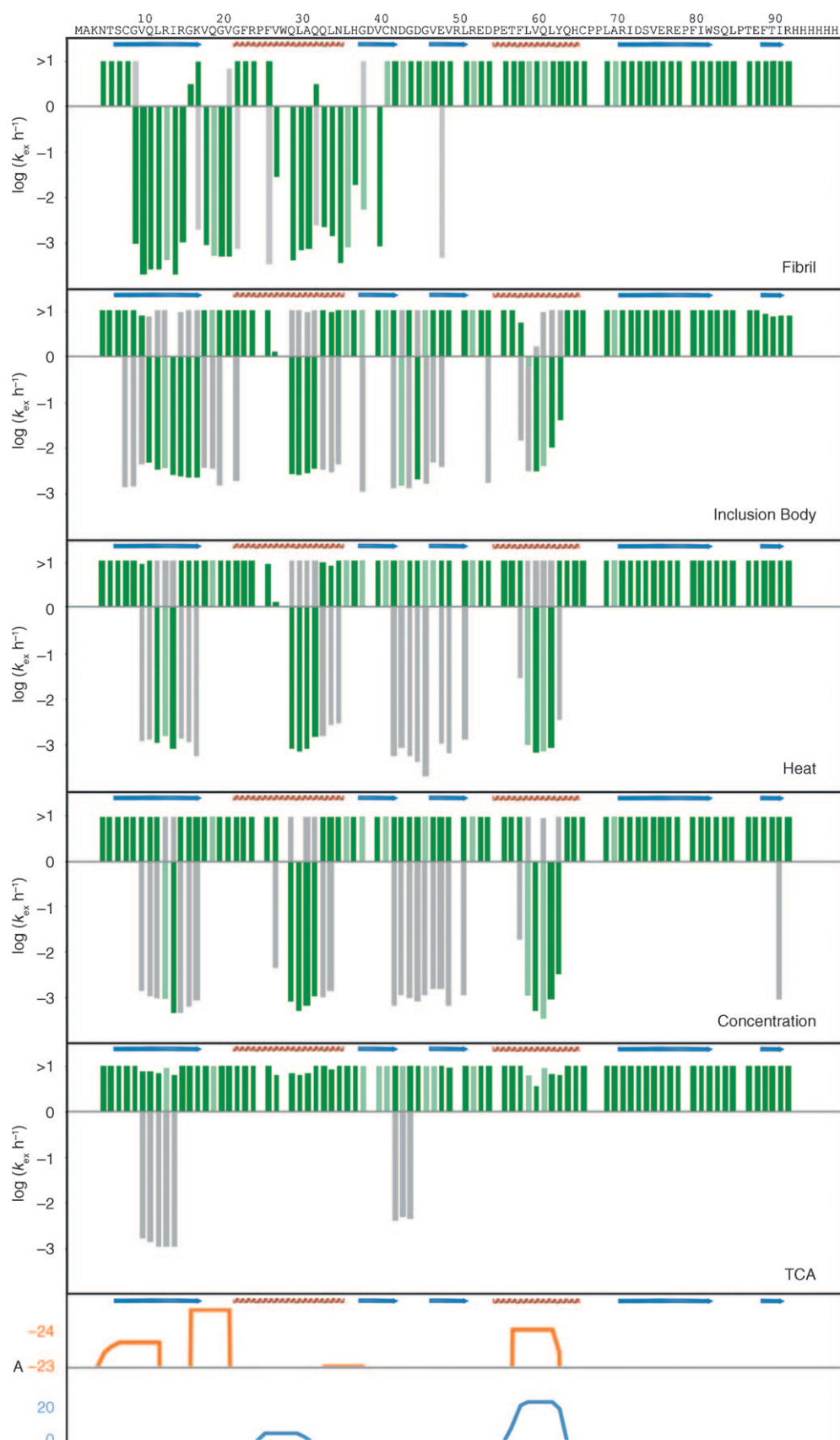
[**] We thank Dr. S. Choe, Dr. L. Goldschmidt, Dr. E. Evangelisti, Dr. C. Cecchi, Dr. P. Arosio, and Dr. M. Morbidelli for experimental assistance and discussions. This research was supported by the Swiss National Science Foundation.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/ange.201000068>.



of the various aggregates studied suggests that each aggregate type should display distinct properties. To gain insight into the suggested multidimensional structure–activity relationship, we measured a few diverse activities of biological interest: 1) The aggregates have different conformational stabilities in urea (Figure 3a), with the following decreasing order of stability: amyloid fibrils > soluble HypF-N > inclusion bodies > concentration-induced aggregates > heat-precipitated aggregates > TCA-precipitated aggregates. 2) Because soluble HypF-N is an acylphosphatase-like protein that can bind a phosphate anion,^[27] we tested the affinities of soluble HypF-N and the aggregates for adenosine-5'-triphosphate (ATP; Figure 3b). All structural entities bind ATP less strongly than soluble HypF-N, with the following order of binding strength: soluble HypF-N > fibrils > heat-precipitated aggregates ≈ concentration-induced aggregates ≈ TCA-precipitated aggregates > inclusion bodies. 3) The affinity of the various aggregates for the amyloid dye thioflavin T (ThT; Figure 3c) follows the same trend as the stability measurements, with the exception of soluble HypF-N, which does not bind to ThT. 4) In contrast to soluble HypF-N, most aggregate entities are able to bind DNA (Figure 3d), with the following order of binding strength: fibrils > heat-precipitated aggregates > concentration-induced aggregates ≈ TCA-precipitated aggregates > inclusion bodies ≈ soluble HypF-N. 5) The capacity for binding to micelles composed of the zwitterionic phospholipid 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC) follows the order (Figure 3e): fibrils > TCA-precipitated aggregates > concentration-induced aggregates ≈ soluble HypF-N aggregates > heat-precipitated aggregates ≈ inclusion bodies. 6) Finally, we tested whether the various aggregates and monomeric HypF-N are able to disturb cell function in the reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). This assay measures amyloid toxicity.^[32] The extent to which the

Figure 1. a) Electron micrographs (left) and X-ray fiber diffraction patterns (right) of the five aggregate types of HypF-N, as indicated. Fibrils have a width of 6 ± 2 nm; inclusion bodies have an irregular spherulike shape with a radius of 280 ± 40 nm; heat-precipitated aggregates have an irregular spherulike shape with a radius of 80 ± 20 nm; concentration-induced and TCA-precipitated aggregates have an amorphous morphology with a wide range of sizes, whereby the latter has irregular spherulike substructures with a radius of 20 ± 4 nm (\pm indicates the standard error). The sharp reflection of the fiber diffraction pattern observed at approximately 4.7 Å is interpreted as the spacing between strands in a β sheet, and the diffuse reflection at approximately 10 Å is interpreted as the spacing between β sheets. b) FTIR spectra of HypF-N aggregates (fibrils, green; inclusion bodies, cyan; heat-precipitated aggregates, blue; concentration-induced aggregates, yellow; TCA-precipitated aggregates, red) and soluble HypF-N (purple). In the amide I region, soluble HypF-N shows two major peaks around 1633 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 a newly formed β -sheet structure in the aggregate.^[21] Whereas for TCA-precipitated aggregates 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.



aggregates perturbed MTT reduction followed the order (Figure 3 f): 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, whereby the fibrils show the highest activities overall. Although the inclusion bodies do not interact with membrane mimetics, ATP, or DNA, they interfere with the cell viability. In contrast, little activity in membrane binding and the MTT assay was observed for the concentration- and temperature-

Figure 2. Sequence-resolved amide exchange rates, $k_{\text{ex}} \text{h}^{-1}$, indicative of secondary structure for all five types of HypF-N aggregate studied, as indicated. The measured H/D exchange is mostly of a biexponential nature, which suggests the presence of two populations. The exchange rates of the major population are indicated in green. If the minor population is present as more than 1/3 of the sample, the corresponding exchange rates are shown in gray. Because of spectral overlap, the exchange rates for some residues can be difficult to determine. However, most of these overlap problems could be resolved on the basis of the assumption that neighboring residues in the sequence show a similar extent of exchange. The exchange rates that were extracted in this way are indicated in light green. The panel labeled “A” shows aggregation-prone segments of HypF-N, as predicted by using two distinct algorithms (see the Supporting Information for details): 3D profile (orange) and TANGO (blue). The secondary structures of soluble HypF-N are highlighted in red for α helices and blue for β sheets.

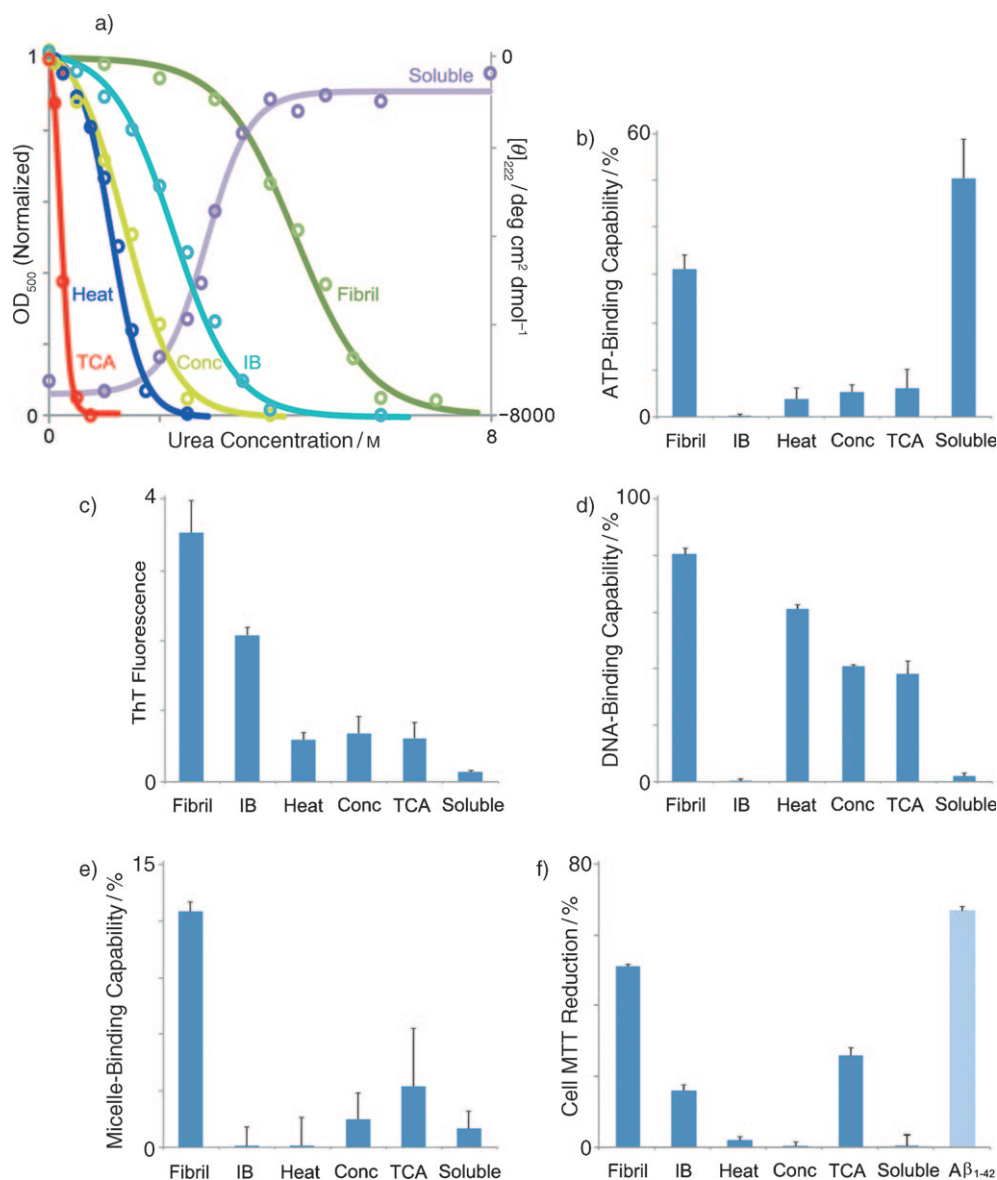


Figure 3. Activities of HypF-N aggregates. a) Disaggregation of HypF-N aggregates and unfolding of soluble HypF-N, as labeled, in the presence of urea. b) ATP-binding capability of the aggregates. ATP and the aggregates were mixed in a 1:5 monomer ratio; the heights of the columns represent the percentage of bound ATP versus the total amount of ATP. c) Enhancement of ThT fluorescence upon binding to the aggregates. The heights of the columns represent the fluorescence intensity in arbitrary units. d) DNA-binding capability of the aggregates. DNA and the aggregates were mixed in a 1:20 monomer ratio; the heights of the columns represent the percentage of bound DNA versus the total amount of DNA. e) DHPC-micelle-binding capability of the aggregates. DHPC micelles and the aggregates were mixed in a 1:1 concentration ratio; the heights of the columns represent the percentage of bound DHPC micelles versus the total amount of DHPC micelles. f) MTT reduction by cells upon the addition of different aggregates. The heights of the columns represent the percentage of reduced cells in the presence of the aggregates versus the amount of reduced cells without the aggregates. A β_{1-42} amyloid fibrils were used as a positive control. Activities of soluble HypF-N were also measured as a reference. When we sent HypF-N fibrils to the Chiti research group for testing, they did not show an influence on MTT reduction by the human neurotypic SH-SY5Y cell line used in Ref. [35]. At least two independent experiments were carried out for each system (\pm indicates the standard error of the experimental measurements for individual preparations).

induced aggregates, but they showed significant binding to DNA. TCA-precipitated aggregates were found to bind DNA and showed a toxic response in the MTT assay.

These results show that when HypF-N aggregates, it generally forms a cross- β -sheet entity with a specific and reproducible structure that is determined by the aggregation conditions. Each of the aggregation states is characterized by several short protein segments that participate in the cross- β structure, whereas the remaining residues are less structured or disordered. Each aggregate has different segments of the protein sequence involved in the formation of the β -sheet structure. 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 of the aggregate down to the micromolar range (only an upper limit 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 micromolar range. We attribute this property to the repetitive nature of the cross- β -sheet motif, which is able to enhance a weak activity on a multiplicative level by cooperativity and may therefore generate many potent activities. Some of these activities may be toxic to the host; in this way, a multifactorial toxicity typically observed in amyloid diseases may result.^[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 to form part of the cross- β -sheet motif,

and the resulting structures have different properties that can account for a variety of phenotypic behaviors. Alternatively, slight structural variations of the aggregates may result in distinct activities because the repetitive nature of the cross- β -sheet entity may increase a weak activity by cooperativity so that a strong activity is observed.

The observed plurality of the micromolar activity of protein aggregates indicates that protein aggregation is rather a primitive folding event than misfolding and supports the idea that protein aggregation played a key role in the early evolution of proteins.^[7,34] It is unlikely 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 aggregated reproducibly into aggregate entities with decent functionality.

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

Received: January 6, 2010

Revised: March 10, 2010

Published online: April 15, 2010

Keywords: aggregation · amyloid fibrils · biological activity · IR spectroscopy · protein structures

- [1] D. J. Selkoe, *Nature* **2003**, 426, 900.
- [2] C. M. Dobson, *Nature* **2003**, 426, 884.
- [3] J. W. Kelly, *N. Engl. J. Med.* **2005**, 352, 722.
- [4] M. Tanaka, S. R. Collins, B. H. Toyama, J. S. Weissman, *Nature* **2006**, 442, 585.
- [5] R. R. Kopito, *Trends Cell Biol.* **2000**, 10, 524.
- [6] L. Wang, S. K. Maji, M. R. Sawaya, D. Eisenberg, R. Riek, *PLoS Biol.* **2008**, 6, e195.
- [7] H. L. True, S. L. Lindquist, *Nature* **2000**, 407, 477.
- [8] D. M. Fowler, A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch, J. W. Kelly, *PLoS Biol.* **2006**, 4, e6.
- [9] S. K. Maji, M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. R. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale, R. Riek, *Science* **2009**, 325, 328.
- [10] S. Ventura, A. Villaverde, *Trends Biotechnol.* **2006**, 24, 179.
- [11] T. Isaacson, C. M. B. Damasceno, R. S. Saravanan, Y. He, C. Catalá, M. Saladié, J. K. C. Rose, *Nat. Protoc.* **2006**, 1, 769.
- [12] W. T. Astbury, S. Dickinson, *Biochem. J.* **1935**, 29, 2351.
- [13] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, C. C. Blake, *J. Mol. Biol.* **1997**, 273, 729.
- [14] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riek, R. Grothe, D. Eisenberg, *Nature* **2005**, 435, 773.
- [15] C. Ritter, M. L. Maddelein, A. B. Siemer, T. Luhrs, M. Ernst, B. H. Meier, S. J. Saupe, R. Riek, *Nature* **2005**, 435, 844.
- [16] T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert, R. Riek, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 17342.
- [17] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riek, D. Eisenberg, *Nature* **2007**, 447, 453.
- [18] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, *Science* **2008**, 319, 1523.
- [19] Z. Ignatova, L. M. Gierasch, *Biochemistry* **2005**, 44, 7266.
- [20] L. Wang, *Prion* **2009**, 3, 139.
- [21] M. Carrió, N. González-Montalbán, A. Vera, A. Villaverde, S. Ventura, *J. Mol. Biol.* **2005**, 347, 1025.
- [22] A. Okuno, M. Kato, Y. Taniguchi, *Biochim. Biophys. Acta Proteins Proteomics* **2006**, 1764, 1407.
- [23] E. S. Chang, T. Y. Liao, T. S. Lim, W. Fann, R. P. Chen, *J. Mol. Biol.* **2009**, 385, 1257.
- [24] B. S. Kendrick, J. L. Cleland, X. Lam, T. Nguyen, T. W. Randolph, M. C. Manning, J. F. Carpenter, *J. Pharm. Sci.* **1998**, 87, 1069.
- [25] F. Chiti, M. Bucciantini, C. Capanni, N. Taddei, C. M. Dobson, M. Stefani, *Protein Sci.* **2001**, 10, 2541.
- [26] G. Calloni, C. Lendel, S. Campioni, S. Giannini, A. Gliozzi, A. Relini, M. Vendruscolo, C. M. Dobson, X. Salvatella, F. Chiti, *J. Am. Chem. Soc.* **2008**, 130, 13040.
- [27] C. Rosano, S. Zuccotti, M. Bucciantini, M. Stefani, G. Ramponi, M. Bolognesi, *J. Mol. Biol.* **2002**, 321, 785.
- [28] M. Hoshino, H. Katou, Y. Hagihara, K. Hasegawa, H. Naiki, Y. Goto, *Nat. Struct. Biol.* **2002**, 9, 332.
- [29] A. T. Petkova, R. D. Leapman, Z. Guo, W. M. Yau, M. P. Mattson, R. Tycko, *Science* **2005**, 307, 262.
- [30] H. Heise, W. Hoyer, S. Becker, O. C. Andronesi, D. Riedel, M. Baldus, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 15871.
- [31] P. C. van der Wel, J. R. Lewandowski, R. G. Griffin, *J. Am. Chem. Soc.* **2007**, 129, 5117.
- [32] Y. Liu, D. Schubert, *J. Neurochem.* **1997**, 69, 2285.
- [33] B. H. Toyama, M. J. Kelly, J. D. Gross, J. S. Weissman, *Nature* **2007**, 449, 233.
- [34] C. P. Maury, *Origins Life Evol. Biosphere* **2009**, 39, 141.
- [35] C. Cecchi, A. Pensalfini, G. Liguri, S. Baglioni, C. Fiorillo, S. Guadagna, M. Zampagni, L. Formigli, D. Nosi, M. Stefani, *Neurochem. Res.* **2008**, 33, 2516.