Mass Spectroscopic Analysis of Sup35NM Prion Polymerization

Vladimir A. Goncharov
Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142

ABSTRACT Sup35NM, the prion determining domain of the protein responsible for the yeast prion phenomenon [Ψ], has become a powerful model for studying key processes in amyloid-related human diseases. One of these processes is a conformational conversion of soluble precursor protein into insoluble fibrillar structures. In this study, we created a set of Sup35NM mutants and used proteolytic digestion coupled with mass spectroscopy to monitor local structure of the protein during polymerization. Experimental data were compared to a network model and showed that during the conformational conversion residue Arg-28 became highly protected from cleavage, residue Arg-98 remained partially solvent exposed, and residues between 28 and 98 showed an intermediate degree of protection. In addition, we found that a distinct subset of proteolytic polypeptides spanning 28–98 residues segment spontaneously formed stable dimers. This finding suggests that the [29–98] region is the key interacting region of Sup35NM responsible for amyloid conversion.

INTRODUCTION

A range of human diseases that includes systemic amyloidosis, neurodegenerative diseases, and transmissible spongiform encephalopathies is accompanied by accumulation of highly ordered proteinaceous aggregates commonly called amyloids (1–5). The dominant component of these aggregates is a protein or a protein fragment that normally exists in a soluble, often well-folded globular conformation (6–8).

The spectrum of proteins capable of forming amyloids is ever increasing, leading to a suggestion that amyloid is an alternative misfolded state accessible for all proteins. This misfolded state is characterized by exceptionally high stability indicating that it populates significantly lower energy levels than its soluble counterpart. A distinct feature of the amyloids is that their structure consists of bundles of fibers made of protein molecules stabilized by a large number of hydrogen bonds (9–11). Although the structure of the soluble precursor protein could be arbitrary, the amyloid core formed by majority of proteins attains similar cross-β structures as evaluated by x-ray diffraction (9,12). This structural homomorphism between soluble precursor proteins and amyloid suggests that there may be a common mechanism governing the transition of a protein from its normal, usually soluble state into the amyloid assembly. Understanding the details of the amyloid transition may fuel the discovery of pharmacological compounds that would block and reverse the formation of amyloids. Such compounds could provide a therapy for the currently incurable, debilitating diseases associated with amyloids. Insight into amyloid transition also provides rational principles for the creation of new biologically inspired, protein-based nanomaterials and devices (13,14).

Sup35NM is an N-terminal fragment of prion protein Sup35p from *Saccharomyces cerevisiae*. It represents the intrinsically unstructured portion of otherwise well-folded cytosolic protein involved in translation termination activity (15). Secondary structure prediction algorithms applied to this fragment show the absence of α-helices and negligible β-sheet content that agree well with circular dichroism spectra for a freshly reconstituted purified fragment at physiological conditions (16). However, the lack of secondary structure does not necessarily mean that the protein is folded randomly because it was shown that natively unfolded proteins while lacking secondary structure may have residual structure (17–19).

The unfolded proteins have a number of properties placing them in a class distinct from the globular proteins, including exceptionally large hydrodynamic radii (20), high structural plasticity, and the ability of acquiring a fold upon binding to a cognate partner (21). High structural plasticity of natively unfolded proteins makes it easy for them to sample the conformational space, and because they cannot fold into stable globular conformation, they have a greater chance ''to find'' the amyloid state during their lifetime (21). Sup35NM is not an exception. It spontaneously forms homogeneous (as judged by electron microscopy (EM)) amyloid fibers under physiological conditions, thereby providing a model system for studying the general mechanism of amyloid conversion (22). The conversion of soluble Sup35NM into amyloid initiates spontaneously and resembles a nucleated polymerization reaction characterized by a measurable lag phase followed by rapid cooperative transition. When a fraction of preformed fibers is added to the soluble protein, the initial rate of reaction becomes a linear function of both the number of fiber ends and the concentration of soluble protein (23). Although a phenomenological model of this polymerization process has been developed (23), little is known about the structural transitions that accompany this conversion. The primary reason for the scarcity of structural information about amyloids is their physicochemical properties. For example, condensing polypeptides cause severe line broadening that precludes residue

Submitted April 11, 2005, and accepted for publication May 26, 2005.

Address reprint requests to Vladimir A. Goncharov, Tel.: 617-452-2632; Fax: 617-258-1966; E-mail: goncharov@wi.mit.edu.

0006-3495/05/12/4139/10 $2.00

doi: 10.1529/biophysj.105.063875

assignments in solution NMR. Additionally, because fibrils possess only one translational symmetry, x-ray crystallography cannot be applied because it requires that the sample has three-dimensional translational symmetry. Recent advances in solid-state NMR enable resolution of the final structural state of the fibrils (24–27). However, there are no high-resolution methods (1–10 Å) for the detection of conformational changes and transitional states during the conversion process. This leaves an open door for the application of the less efficient low-resolution methods (10–50 Å) including mass and optical spectroscopy (28–30), EPR (31,39), and reconstructive electron microscopy (32,33).

Methods to probe the structure and changes in the structure based on the proteolytic digestions coupled to mass spectroscopy were used in the past. The most recent and relevant to our work are studies of the structure of SH3 (34) and Ure2p (35) fibrils. Proteolytic mapping is based on the relative susceptibility of disordered and randomly folded regions of the protein to proteolytic cleavage, compared with the resistance of the regions that are tightly folded (36). A typical experiment involves a limited digestion with either a specific or nonspecific protease, followed by detection of the digestion products by mass spectrometry. The traditional methods of analysis employed in proteolytic mapping have two shortcomings. First, proteolytic peptides are not related quantitatively to their precursors. This makes it impossible to compare or to calculate the cleavage rates at different sites. In addition, the traditional methods assume that the release of proteolytic peptides in solution does not depend on the stage of amyloid formation. In reality, peptides located in the amyloidogenic regions could be dramatically less soluble if cleaved from the fiber than from the soluble protein.

We have developed a dynamic network model of proteolytic digestion to overcome the limitations of current methodology. Our model allows measuring the aggregation rates of proteolytic peptides and their changes, whereas this information is lost in the traditional approach. Additionally, it uses conservation laws existing between the proteolytic fragments to alleviate the experimental errors. These features of the model increase the robustness of information derived from the digests and allows the deduction of the degree of conformational protection occurring at chosen neighborhoods of the protein's local structure. We have used this model to interpret proteolytic digestion data and elucidate the structural changes occurring during the conformational conversion of Sup35NM from its soluble to amyloid state.

Digestion model

The time course of a protein digestion was modeled using a graph, where nodes represent all possible fragments produced by specific protease cleavage joined by directed edges representing cleavage pathways. Fig. 1 shows a network model of the cleavage process for a peptide with two cleavage sites. This model places an explicit link between the observed cleavage product (the node's relative intensity) and the surface exposure of all residues that have to be cleaved to make that product (the subset of cleavage rates). The dynamics of this graph are described by a set of differential equations. In the initial stages of limited digestion, when enzyme concentration is significantly lower than the substrate concentration, the cleavage rates could be taken constant. Additionally, if we assume proteolytic peptide aggregation rates during the same period to be constant as well, then the model becomes a linear system with constant coefficients and can be solved algebraically. The number of nodes (differential equations) depends quadratically on the total number of cleavage sites. Because Sup35NM has two arginines, cleaving it with Arg-C protease yields five fragments. Therefore, the cleavage dynamics are described by the following system of differential equations (Eq. 1):

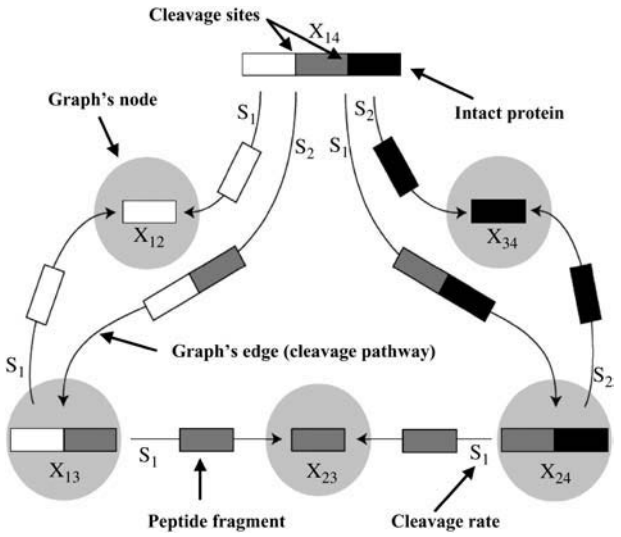

FIGURE 1 Digestion network model. Figure shows schematics of proteolytic digestion for a protein with two cleavage sites. Proteolysis of wild-type Sup35NM by Arg-C protease is described by this model. Rectangles represent different proteolytic peptides liberated during the digestion. Bended arrows represent the graph's edges. They signify cleavage pathways. Large light gray circles represent nodes of the graph. The intensities of the nodes correspond to the time-dependent normalized intensities of the proteolytic fragments X_{ij}. They are calculated as a solution to Eq. 1. The solution depends on S_1 and S_2—cleavage rates at proteolytic sites 1 and 2. The cleavage rates are determined by minimizing Eq. 2.

$$\frac{dx_{14}}{dt} = -(s_1 + s_2 + d_{14})x_{14}$$
$$\frac{dx_{13}}{dt} = s_2 x_{14} - (s_1 + d_{13})x_{13}$$
$$\frac{dx_{24}}{dt} = s_1 x_{14} - (s_2 + d_{24})x_{24}$$
$$\frac{dx_{12}}{dt} = s_1(x_{14} + x_{13}) - d_{12}x_{12}$$
$$\frac{dx_{23}}{dt} = s_1 x_{13} + s_2 x_{24} - d_{23}x_{23}$$

$$\frac{dx_{34}}{dt} = s_2(x_{14} + x_{24}) - d_{34}x_{34}, \quad (1)$$

where x_{ij} is concentration of a fragment ij, s_1, s_2 are cleavage rates at Arg-28 and Arg-98, and d_{ij} are the intrinsic decay (aggregation) rates. Each structural state of the protein is described by the set of coefficients s_1, s_2, and d_{ij}. Cleavage rates relate to the protein's topology whereas decay rates signify the peptide's propensity to aggregate. Fitting s_1, s_2, and d_{ij} to satisfy the observed data provides a structural snapshot of the protein. This is accomplished by a traditional least-squares method. Let $f_i^C = f_i^C(s_1, s_2, d_{ij}, t)$ be solutions of Eq. 1, $f_i^E(t)$ concentrations of observed fragments, and w_i statistical weights.

$$\psi(s_1, s_2, d_{ij}, t) = \sum_i w_i \left[f_i^E(t) - f_i^C(s_1, s_2, d_{ij}, t)\right]^2. \quad (2)$$

Then minimization of Ψ provides s_1, s_2, and d_{ij} corresponding to the duration of digestion time t, which is set to be constant for all samples in a set.

If the number of experimentally observed f_i^E is equal to or less than the total number of fitting parameters (d_{ij} and s_i), it becomes necessary to reduce the number of independent parameters. One possibility is to determine cleavage and decay rates in separate experiments. The first experiment is designed in a way that minimizes the aggregation effect by employing a disaggregating step before analysis. Then, the second experiment is designed to specifically determine the decay rates by performing peptides recovery from the digestion mixture under physiological conditions.

The number of independent solutions to Eq. 1 under certain conditions could be less than the number of equations. In aggregation-free experiments, we have a number of conservation laws relating the peptide concentrations. For example, concentration of fragment x_{12} in Sup35NM digestion is equal to the sum of x_{23} and x_{24}. Similar relation also exists for x_{34}. These conservation laws are easily generalized for an arbitrary number of cleavage sites (Eq. 3):

$$x_{12} = \sum_{i=3}^{N} x_{2i}$$
$$x_{N-1,N} = \sum_{i=2}^{N-1} x_{N-i,N-1}. \quad (3)$$

Equation 3 and other relations of that kind that exist between proteolytic fragments are instrumental for the detection and correction of systematic experimental errors. Such corrections enhance the robustness of data fits.

EXPERIMENTAL METHODS

Mutagenesis

Wild-type Sup35NM cDNA was subcloned into pET-22 vector (Novagen, Madison, WI) in tandem with C-terminal hexahistidine tag. Point mutations were introduced using QuikChange II site-directed mutagenesis kit (Stratagene, La Jolla, CA). The fidelity of all constructs was verified by sequencing.

Protein purification

Proteins were expressed in *Escherichia coli* and purified via metal chelation and reverse phase chromatography. Details are provided in Supplementary Material.

Preparation of fiber reactions

Proteins were rehydrated in 25 mM MES pH 4.5 and spin filtered through 100 KDa cutoff membrane (OMEGA100K, Pall, Port Washington, NY). Unseeded fibers were formed at 12.5 μM concentration in PBS in gently stirred reactions. Seeded reactions were initiated by addition of 5% (w/w) preformed seed. The seeded reactions were frozen on dry ice at indicated time (typically 1, 10, 30, 60 min after addition of seed).

Proteolytic digestions

In general, enzyme manufacturer's recommendations were followed. Arg-C enzyme [E.C. 3.4.22.8] was purchased from Roche. Digestion time was set at 20 min. To stop the digestions at desired time points, double volume of 6 M GdmCl was added and samples were desalted.

Desalting

POROS R2 reverse-phase resin was added to the samples and mixed on a vortex at 30°C for 30 min. Next, resin with bound peptides was separated by spin filtration in a 0.45 μM NANOSEP MF GP device (Pall) and washed three times with 0.1% TFA in water. Peptides were eluted with 50% MeOH 0.1% TFA.

Mass spectrometry

Mass spectrometry was performed on Applied BioSystems Voyager-DE STR MALDI-TOF. Sinapinic acid was used as crystallization matrix. Mass/charge values for proteolytic fragments were calculated using ProteinProspector (University of California, San Francisco). Spectra were analyzed by M/Z software (ProteoMetrics, New York, NY).

Normalizing the MALDI mass spectra

The raw intensity of each fragment I_{ij} was calculated by adding intensity of the singly charged species i_{ij}^{1+} plus 1/2 of intensity of the doubly charged species i_{ij}^{2+} plus double intensity of the corresponding homodimer d_{ij}^{homo} plus intensity of the corresponding heterodimer d_{ij}^{hetero}. The normalized intensity of a fragment x_{ij} was obtained by multiplying raw intensity I_{ij} by 100% and dividing it by the combined raw intensities of all fragments present in the spectrum.

$$I_{ij} = i_{ij}^{1+} + \frac{i_{ij}^{2+}}{2} + 2d_{ij}^{homo} + d_{ij}^{hetero}$$
$$x_{ij} = 100\% \frac{I_{ij}}{\sum I_{ij}}.$$

Thioflavin-T binding

TfT binding was performed as previously described (23). Details are provided in Supplementary Material.

Electron microscopy

Seeded fibril samples were negatively stained with uranyl acetate and imaged on JEOL JEM-2200FS microscope.

Data fits

All calculations and data fits were performed using CAS *Mathematica* (Wolfram, Champaign, IL).

RESULTS

Arg-28 and Arg-98 gain protection during polymerization at different rates

The main aim of this study was to elucidate the conformational changes occurring at the N-terminal region of Sup35NM during the fiber assembly. Although the proposed model accounts for peptide aggregation, we left the experimental determination of the aggregation rates outside of the scope of this study, and performed the experiments under conditions that minimize the effects of aggregation. We prepared a set of five samples representing five time points of the polymerization reaction by freezing aliquots of seeded fiber formation reaction on dry ice. The first sample had no seed and thus represented the soluble, intrinsically unstructured state. The second sample was withdrawn 1 min after addition of the seed and had ~5–10% of fibrous material. The third and fourth samples had ~50% and ~75% of polymerized protein. The last sample was taken two hours after addition of seed and contained >90% fibrous material. Evaluation of the ratio between soluble and fibrous material was made by TfT binding (Fig. 4 *B*). Upon completion of the reaction, the samples were simultaneously thawed out, digested with a protease that hydrolyzes peptide bond on the C-terminal side of arginine residue (Arg-C protease) and disaggregated with GdmHCl.

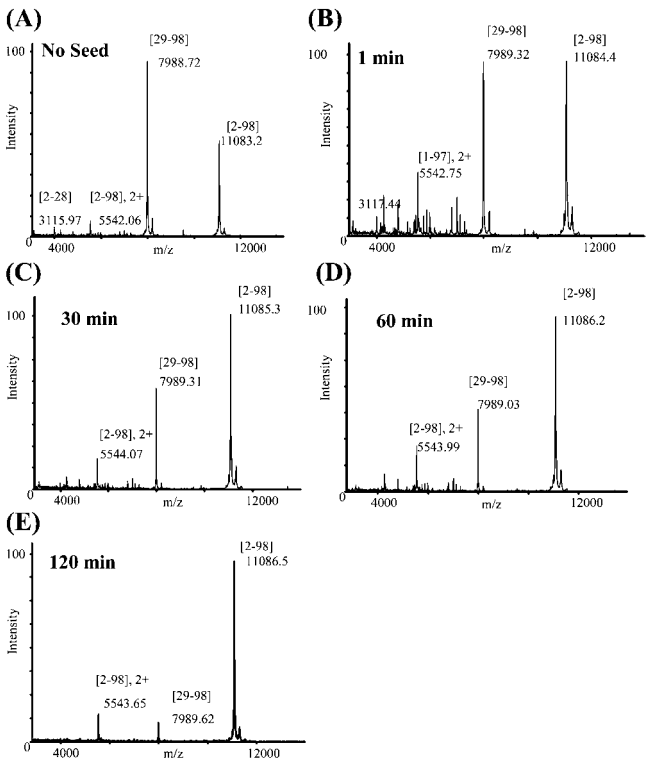

FIGURE 2 Sup35NM Arg-C digestion spectra. MALDI spectra of Sup35NM digested with Arg-C protease. The fiber formation reactions were initiated by addition of seed and then frozen at various stages of polymerization. (*A*) No seed added; (*B*) 1 min after addition of seed; (*C*) 30 min; (*D*) 60 min; (*E*) 120 min. All samples were digested simultaneously, desalted and analyzed by MALDI.

Digestion of Sup35NM with Arg-C protease produces up to five proteolytic fragments, because wild-type Sup35NM has two arginines. Visual examination of mass spectra in Fig. 2 shows that relative fragment intensities are changing dramatically between the samples collected at initial and terminal stages of fiber formation. In the initial, unseeded sample, the [29–98] fragment is the most intense (Fig. 2 *A*). This fragment is produced when Sup35NM is cleaved at both sites, Arg-28 and Arg-98. The second highest peak is the [2–98] fragment that contains one missed cleavage at Arg-28. The observed digestion pattern suggests that both cleavage sites are accessible (although to different degrees) on a soluble protein before the ''seed'' is added to initiate fiber formation. The sample taken immediately after addition of the ''seed'' shows that the peak corresponding to [29–98] drops in intensity and becomes nearly equal to the [2–98] fragment (Fig. 2 B). This result is only possible if the cleavage rate at the Arg-28 site decreases. The next samples in sequence show that the [29–98] peak progressively declines whereas the [2–98] remains the highest (base) peak of the spectrum (Fig. 2, *C* and *D*). Two hours after addition of the seed, the [29–98] loses 87% of its intensity (Fig. 2 E).

Arg-28 gains protection faster than Arg-98, but remains partially exposed in fiber

To confirm visual observations and to put them into the semiquantitative perspective, we fitted the digestion model to the experimental data. The values of cleavage rates obtained in these fits are strictly semiquantitative, and could only be used to illustrate the changes in cleavage rate ratios between samples at different time points. The results are summarized in Table 1. ''No seed'' sample gives ~3.1 average ratio between Arg-28 and Arg-98 rates, that grows rapidly to over 18 for the ''120-min'' sample. This tells us that Arg-28, while it is partially exposed in soluble protein, becomes rapidly sequestered inside of the protein during the fiber assembly. Arg-28 is buried significantly, if not completely inside the fiber core, because intensity of the [2–28] fragment quickly reduces to zero and intensity of the [29–98]

fragment reduces from 65 to 8%. In addition, the fitted model confirms that Arg-98 becomes progressively less exposed to the solvent during fiber formation, because the value of the rate declines during the conversion process. However, a significant portion of Arg-98 remains at the surface of the fiber core, as judged by the presence of the [2–98] fragment in the fibrous sample. This indicates that Arg-98 could be in at least two topologically nonequivalent positions with respect to the fiber surface.

Insertion of additional arginines expands the method's coverage

Initial success in detecting the conformational changes in Sup35NM using Arg-C encouraged us to further experiment by making a set of arginine mutants. We decided to use arginine point mutations as engineered cleavage sites to probe the conformational changes at chosen points of protein fold. Therefore, we created six mutants of Sup35NM: G7R, N21R, G51R, G58R, G68R, and G96R; we expressed them in *E. coli*, purified, and verified identities of the resulting proteins by LC-MS. We collectively refer to that set as Sup35RM mutants.

RM behave similarly to NM

Because all of the mutants lead to the placement of a charged residue in the region sensitive to the charge mutations (37), we tested whether these mutations abolish the fiber formation, or alter the fiber formation properties. First, we formed fibrils by dissolving lyophilized proteins in phosphate-buffered saline and examined them by electron microscopy. EM images (Fig. 3) show that all mutants share morphology similar to the wild-type Sup35NM fibrils, characterized by uniform nonbranched filaments of 10–30-nm diameter up to several micrometers long (22).

Next, we examined kinetics of unseeded and seeded fiber formation (Fig. 4) using Thioflavin-T binding as an indicator of amyloid conversion (23). In unseeded reactions, all except G7R mutants showed kinetics (chiefly characterized by the duration of lag phase) comparable to the wild-type Sup35NM (Fig. 4 *A*). G7R had a lag phase 45% longer than the rest. In seeded reactions, all but N21R had assembly kinetics similar to wild type (Fig. 4 *B*). N21R had ~50% decrease in apparent rates of assembly. Thus, we concluded that seeded fibers formed by five out of six arginine mutants in vitro were comparable to wild type and thus could provide the structural information that is possible to integrate with that of wild-type NM.

TABLE 1 Fitted model of Sup35NM/Arg-C digestion

Time (min)	S_1^* (min^{-1})	$S_2^\dagger$ (min^{-1})	S_2/S_1	$\Xi^{2\ddagger}$
NS[§]	0.0555	0.1697	3.1	4.9
1	0.0338	0.1589	4.7	3.6
30	0.0219	0.1258	5.8	5.5
60	0.0159	0.1395	8.8	5.7
120	0.0054	0.1001	18.6	2.1

*Cleavage rate at Arg-28.
†Cleavage rate at Arg-98.
‡Residual sum of squares.
§No-seed sample.

Protection gain is faster for sites closer to N-terminus

The set of six Sup35RM mutants was subjected to the same Arg-C/MS analysis as wild type described above. Fig. 5 shows the MALDI spectra for the digests of the initial and the final fiber formation samples. The figure shows that the distribution of intensities changes significantly between initial and terminal samples for all but N21R proteins. Table 2 summarizes the results of data fits. With the notable exceptions of N21R and G96R, cleavage rates at Arg-28 and Arg-98 sites of Sup35RM mutants followed the course similar to the wild type. First, Arg-28 was gaining protection faster than Arg-98. Second, all cleavage sites gained some protection during the fiber formation. The gain in protection was highest for the sites close to the N-terminus. In particular, Arg-7, Arg-21, and Arg-28 gained protection faster than the other sites, and Arg-21 appears to be protected more than Arg-28 at all stages of polymerization. Cleavages

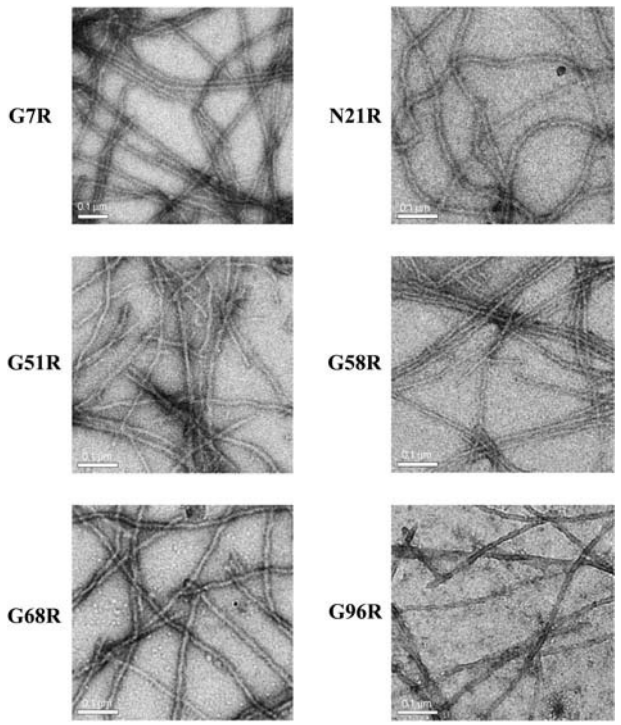

FIGURE 3 Sup35RM EM images. Electron micrographs of the negatively stained seeded fiber samples: Sup35NM-G7R, Sup35NM-N21R, Sup35NM-G51R, Sup35NM-G58R, Sup35NM-G68R, and Sup35NM-G96R. All proteins assemble into fibers with similar morphology. Scale bar, 100 nm.

at Arg-58 and Arg-68 were moderately frequent throughout all stages of polymerization, and gained less protection than N-terminal sites, indicating that these sites could be in several topologically nonequivalent positions in respect to the fiber surface, similarly to Arg-98 (see Sup35NM digestion results). G96R was different from the rest. A concern about fidelity of G96R data fits forced us to rely on visual interpretation of the digestion. Thus, we provide here only the visual interpretation of the digest (Fig. 5). At the initial stage the [29–96] fragment is the most intense, [2–96] is second, [29–98] is third, and [2–28] is present. This tells us that initially all three sites are exposed, although to different degrees. Arg-98 appears to be twice less exposed than Arg-96, and Arg-28 is the least exposed. Two hours later [2–28] and [28–98] disappear, [2–96] becomes the dominant peak, [29–96] loses about half of its intensity. This shows that Arg-28 becomes sequestered inside of the fiber core, whereas Arg-96 and Arg-98 remain partially exposed. Because peak ratios of [29–96]:[29–98] in the ''no-seed'' sample and [2–96]:[2–98] in ''120-min'' samples are ~2:1, it is possible

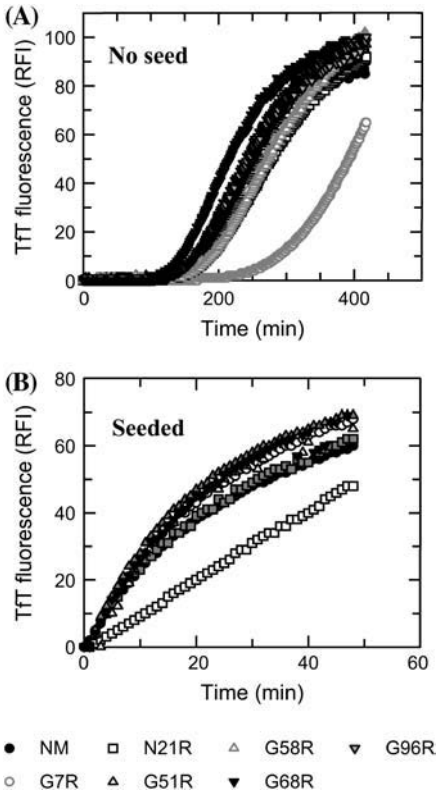

FIGURE 4 Sup35RM thioflavin-T kinetics. (*A*) Unseeded kinetics. Soluble protein samples were mixed with TfT and gently stirred during the course of reaction. The changes in TfT fluorescence occurs when TfT binds to the fibers. All proteins except Sup35NM-G7R show unseeded kinetics similar to the wild-type Sup35NM. Sup35NM-G7R has lag phase ~40% larger than wild type. (*B*) Seeded kinetics. Seeded polymerization reactions were monitored by TfT binding. All proteins except Sup35NM-N21R show kinetics similar to wild-type Sup35NM. N21R mutant has slower apparent assembly rates than the wild type.

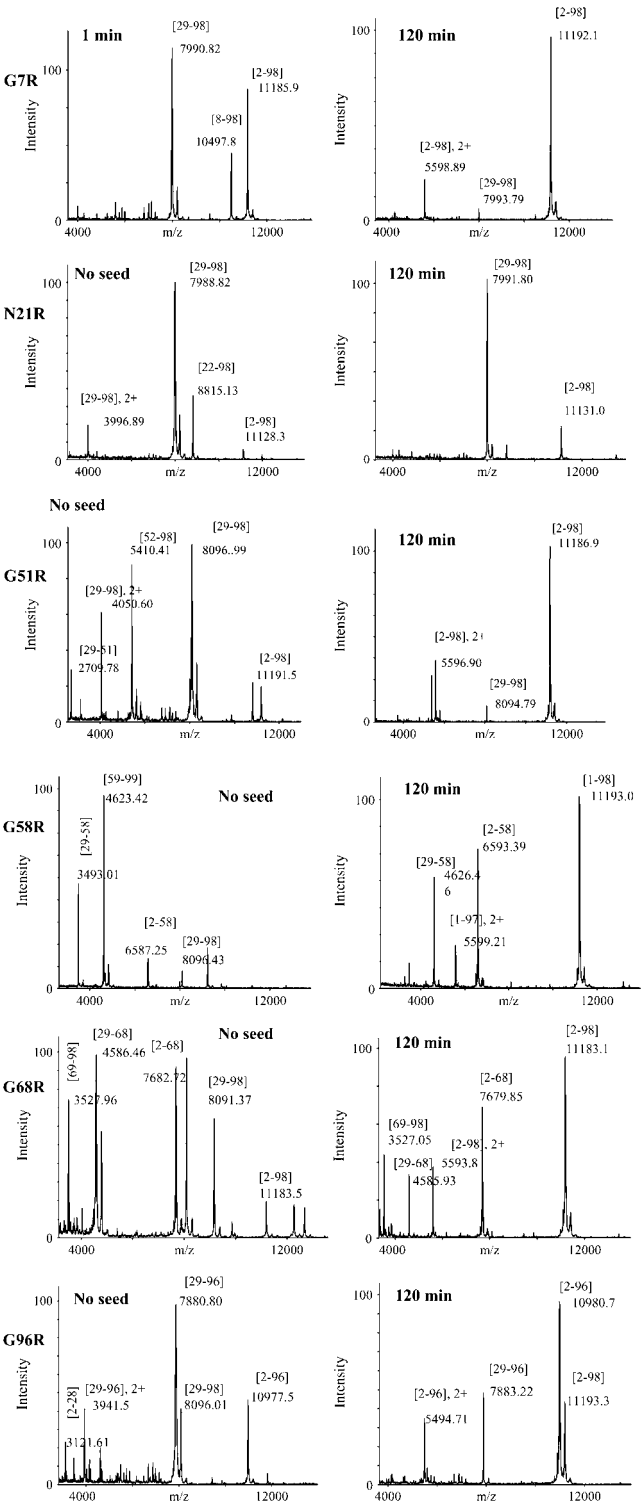

FIGURE 5 Sup35RM digestion spectra MALDI spectra of initial no seed (*left*) and fibrous (*right*) samples digested with Arg-C protease. Relative intensities of the fragments change significantly during the course of fiber formation. Spectra of initial samples is dominated by completely cut end-fragments, whereas ''120 min'' fiber samples show significant presence of undigested fragments containing one or two cleavage sites.

that cleavage rates at Arg-96 is about twice that of Arg-98 site, and is constant during fiber formation.

Proteolytic peptides form highly stable dimers

Initial examination of the MALDI spectra showed the presence of peptides with masses that differed from the masses of the expected proteolytic fragments and their multiple charged states. Further analysis showed that these peaks correspond to the homo- and heterodimers formed by the proteolytic peptides liberated during the Arg-C cleavage (Fig. 6). The detection of the peptide dimers in the physical vacuum of the mass spectrometer is perplexing; and because our mass spectrometer had insufficient resolution we cannot say at this point if they are stabilized by noncovalent interactions or are joined covalently. Nevertheless, the fact that only selected peptides formed dimers indicate that there must be specific interaction between them.

Table 3 summarizes the dimers observed during the cleavage of Sup35RM with Arg-C. The most ubiquitous homodimers are formed by the [2–98] and [29–98] peptides. In the case of G96R these are replaced by [2–96] and [28–96], apparently because the cleavage at Arg-96 is more efficient than at Arg-98. Homodimers are typically formed at initial stages of polymerization, when the cleavage at Arg-28 is still frequent. At later stages, abundance of the [29–98] fragment drops and heterodimers become more prevalent. The heterodimers are typically formed by the [2–98] and [29–98] peptides. In addition, [8–98], [22–98], [52–98], [58–98], [2–68], and [29–68] participate in the dimerization. The [2–28], [2–51], [29–51], [29–58], [69–98], [69–275], [97–275], and [99–275] fragments apparently do not participate in the dimerization, because all homodimers or heterodimers were formed without their participation.

Superimposing the sequences of the interacting peptides and subtracting the sequences of the noninteracting peptides reveals the region of Sup35NM that is most likely to be involved in the formation of fiber core (Fig. 7). The broad segment [29–98] certainly encompasses the entire fiber-forming region. But the narrower fragment [29–68] probably encompasses the most strongly interacting part of the fiber core, because all dimerizing peptides either contain this region or share at least a 10-amino-acid-long stretch of it. Notably, [29–68] is the shortest of all dimerizing peptides.

DISCUSSION

What are the conformational changes that Sup35p purportedly self-perpetuate?

At the heart of [Ψ] prion phenomena lies the propagation of self-perpetuating conformational changes in Sup35p protein (38). Although Sup35p is actively studied in several laboratories around the world and more than two dozen articles were published about it in the last eight years, no systematic study of the Sup35p conformational changes on

TABLE 2 Fitted model of Sup35RM/Arg-C digestion

Protein	Time (min)	S_1^* (min^{-1})	$S_2^\dagger$ (min^{-1})	$S_3^\ddagger$ (min^{-1})	S_3/S_1	S_3/S_2	$\Xi^{2\S}$
G7R	1 min	0.0200	0.0365	0.1907	9.5	5.2	1.9
	30 min	0.0046	0.0137	0.1799	39.3	13.1	8.9
	60 min	0.0049	0.0128	0.1348	27.8	10.6	10.6
	120min	0.0010	0.0031	0.1316	135.7	42.4	9.2
N21R	NS¶	0.0721	0.0580	0.1724	2.4	3.0	5.8
	1 min	0.0760	0.0721	0.1905	2.5	2.6	4.2
	30 min	0.0233	0.0675	0.1299	5.6	1.9	39.8
	60 min	0.0217	0.0673	0.1401	6.5	2.1	32.9
G51R	NS¶	0.1412	0.0258	0.1797	1.3	7.0	4.5
	1 min	0.0248	0.0205	0.1742	7.0	8.5	2.1
	30 min	0.0126	0.0114	0.1473	11.7	12.9	12.9
	60 min	0.0060	0.0121	0.2166	36.2	17.9	2.1
	120min	0.0038	0.0084	0.1728	45.7	20.6	0.8
G58R	NS¶	0.0467	0.0500	0.1038	2.2	2.1	96.9
	1 min	0.0219	0.0263	0.0631	2.9	2.4	31.3
	30 min	0.0189	0.0209	0.0499	2.6	2.4	26.6
	60 min	0.0138	0.0292	0.0825	6.0	2.8	38.1
	120min	0.0107	0.0138	0.0752	7.0	5.4	52.4
G68R	NS¶	0.0815	0.0111	0.0299	0.4	2.7	821.1
	1 min	0.0161	0.0079	0.0428	2.7	5.4	556.9
	30 min	0.0046	0.0102	0.0464	10.0	4.5	286.8
	60 min	0.0044	0.0107	0.0525	12.0	4.9	206.0
	120min	0.0034	0.0123	0.0521	15.5	4.2	173.2

*†‡Cleavage rates at first, second, and third cleavage sites counted from the N-terminus.
§Residual sum of squares.
¶No-seed sample.

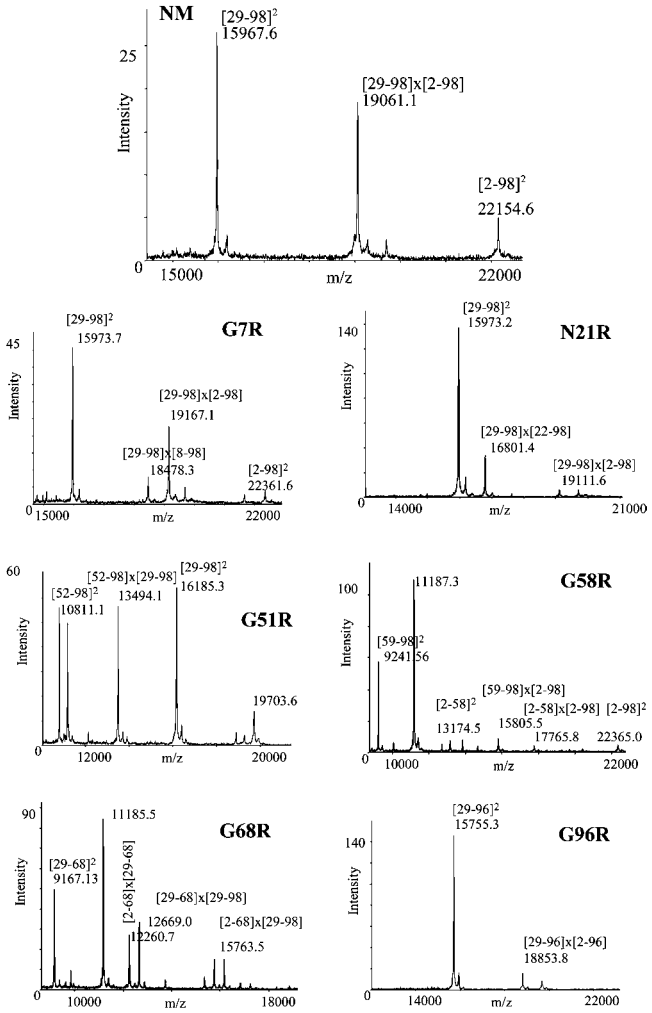

FIGURE 6 Sup35 peptide dimers MALDI spectra show the presence of stable dimers formed from the proteolytic peptides. Dimers are formed only if both peptides span the residue subset from [29–98]. Thus, homodimers $[2–98]^2$, $[29–98]^2$ and heterodimers [29–98] × [2–98] are formed, but not [2–28] × [2–98] or $[2–28]^2$.

a molecular level were ever published. The first structural study of Sup35NM fibers on the molecular level came in 2004 (39). It shows mobility profiles of paramagnetic probes attached at different positions along the Sup35 sequence detected by EPR. The study showed that the minimum mobility in the fiber segment [26–116] corresponds to the [36–76] region. That study determined our choice of residues for the conformational experiments. Before using Arg-C protease we experimented with trypsin, chymotrypsin, proteinase-K and Asp-N. The pilot study showed that indeed the [141–253] part of Sup35 is highly flexible, although not totally random at both soluble and fibrous states, while N-terminus undergoes some ''protection'' from cleavage (V. A. Goncharov, unpublished data). The deviation of the cleavage pattern from randomness (characterized as equal cleavage at all sites) for the monomeric state of Sup35NM shows that it has some residual structure.

The advantages of a networking model

The method used in this study relies on the detection of changes in degree of protection from cleavage as a manifestation of conformational changes in the protein's structure. The novelty of our approach lies in the usage of a network model that integrates the pieces of proteolytic data into one system. It improves reliability and allows probing the structural changes of the amyloid fibers and other protein complexes that are difficult to study.

Structural transitions in Sup35NM

Initially, soluble Sup35NM is predominantly monomeric (23). Sup35NM at this stage is in an extended structural state. This state is not a random coil, but probably a collection of small (3–7 amino acids) segments with stable intraresidue contacts joined by segments of random coil. Threefold differences in cleavage rates at Arg-28 and Arg-98 show that Arg-28 is partially protected in one such small folding element. Partial protection at Arg-7, as well as nonrandom pattern of chymotrypsin digest (data not shown) also supports this view. The very N-terminus of Sup35NM [2–12] could be involved in nucleation, because G7R is the only mutant that had shown ~45% extension of the lag phase in unseeded reaction (Fig. 4 *A*). However, it is unlikely to participate in assembly, because: a), this mutant had similar rates of assembly to the wild type; and b), Arg-C was cleaving at Arg-7 position even in the fibrous state. Another abnormal mutant was also found at the N-terminus. N21R had slower apparent assembly rate in seeded reaction, indicating that the charged residue at this position is interfering with assembly process. Addition of the ''seed'' had a strong depressing effect on the accessibility of N-terminal residues, Arg-28 in particular. Immediately after addition, the cleavage rate at Arg-28 dropped, while little fibrous material was produced at this stage. These results suggest that presence of fibrous template was somehow ''communicated'' to the rest of the protein molecules, triggering the beginning of conformational changes. The ''communication'' effect could explain the less-than-linear dependence of assembly rates on the protein concentration at high monomer concentrations. Polymerization that follows addition of the seed leads to the gain in protection from cleavage for the most of the mutations.

Interacting region

The proliferation of the dimers formed from proteolytic peptides points to the strong interaction that exists in the [29–68] segment of the Sup35NM. This finding is in agreement with previous work that showed that minimum mobility of side chain linked paramagnetic probes correspond to roughly the same region in the fibrous structure (39). If the [29–68] fragment alone can form amyloid fibers comparable to Sup35NM, it could greatly simplify the fiber

TABLE 3 Summary of observed dimers

Protein	Homodimers	Heterodimers
NM	$[29–98]^2$; $[2–98]^2$	[2–98] × [29–98]
G7R	$[29–98]^2$; $[2–98]^2$	[2–98] × [29–98]; [8–98] × [29–98]
N21R	$[29–98]^2$	[2–98] × [29–98]; [2–98] × [22–98]
G51R	$[29–98]^2$; $[2–98]^2$; $[52–98]^2$	[52–98] × [29–98]
G58R	$[2–58]^2$; $[2–98]^2$; $[59–98]^2$	[59–98] × [2–98]; [2–58] × [2–98]
G68R	$[29–98]^2$; $[2–98]^2$; $[29–68]^2$; $[2–68]^2$	[2–98] × [29–98]; [2–68] × [29–98]; [2–98] × [2–68];[29–68] × [2–68]; [29–98] × [29–68]
G96R	$[29–96]^2$; $[2–96]^2$	[2–96] × [29–96]

The left number in the square brackets represents the leftmost (N-terminal) amino acid and the right number represents the rightmost (C-terminal) amino acid of the proteolytic fragment.

polymerization model and make it tractable for molecular dynamics simulations as well as more amendable for biochemical studies.

The ensemble averaging problem

We are fully aware of one fundamental problem that is inherent in all low-dimension, low-resolution methods including fluorescence, mass, or Fourier transform infrared spectroscopy. That is the uncontrolled averaging over multiple structural states present at each sampling point. The detected signal represents all conformations present and could be interpreted as a state that in fact is never present in the pure form. The typical example is ''partial'' solvent accessibility. If partial number reaches above the error limit of experiment, let us say ~10–15%, it becomes quite probable that in fact more than one structural configuration contributes to the signal. In other words, it is most likely that 10% of these residues are fully exposed, while the rest are buried, rather than assuming that the residue is close to the surface and ''pops out'' 10% of the time. In case of Sup35NM, Arg-98 is such a residue. It gains protection during fiber assembly, but even in the mature fiber a significant portion of Arg-98 is solvent exposed. We interpret this as Arg-98 being in more than one topologically nonequivalent position. Additional experiments and enhancements of the technique are necessary to untangle these structural ensembles. One approach that could be fruitful is the use of a very low number of molecules per experiment, ideally just one at a time.

The results of this work present only a tiny piece of information, a short glimpse at the processes happening during the Sup35NM fiber assembly. The experimental part of our method is traditional, but we used a novel method of analysis that looks at proteolytic disintegration as a dynamic process, and accounts for all parts of the multivariate system. This ''network'' approach could offer dramatically more information about protein structure than the traditional ''piecewise'' method. This new method should be applicable to a variety of proteins and protein complexes that are difficult to study, including other amyloids, glycoproteins, and membrane proteins.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The author thanks Kazuoshi Murata for help with electron microscopy, Oliver King, Korie Handwerger, and Leah Cowen for comments on the manuscript, Tom DiCesare for help with figures, and Olga Goncharova for proofreading the manuscript. In addition, the author thanks Sue Lindquist, whose support made this publication possible.

This work was supported by DuPont-Massachusetts Institute of Technology Alliance grant ''Determining the Assembly Mechanism of Peptide Nanofibers to Facilitate the Fabrication of Nanoscale Devices''.

REFERENCES

1. Tan, S. Y., and M. B. Pepys. 1994. Amyloidosis. *Histopathology*. 25:403–414.
2. Thomas, P. J., B. H. Qu, and P. L. Pedersen. 1995. Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* 20: 456–459.
3. Prusiner, S. B. 1998. Prions. *Proc. Natl. Acad. Sci. USA*. 95:13363–13383.
4. Koo, E. H., P. T. Lansbury Jr., and J. W. Kelly. 1999. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. USA*. 96:9989–9990.
5. Dobson, C. M. 2001. The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356:133–145.
6. Fandrich, M., M. A. Fletcher, and C. M. Dobson. 2001. Amyloid fibrils from muscle myoglobin. *Nature*. 410:165–166.

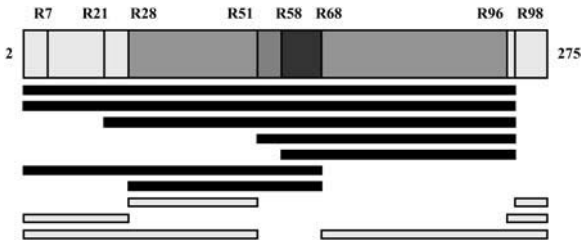

FIGURE 7 Interacting regions of Sup35NM. At the top of the figure, intensity of black color corresponds to the region's propensity to dimerize. (*Bottom*) Solid black rectangles at the bottom of the figure signify the peptides that formed dimers; light gray rectangles signify the peptides that did not form dimers.

7. Tahiri-Alaoui, A., and W. James. 2004. Rapid formation of amyloid from monomeric recombinant human PrP in vitro. *Prot. Sci.* 14:942–947.
8. Vendruscolo, M., and C. M. Dobson. 2005. Towards complete descriptions of the free energy landscapes of proteins. *Phil. Trans. R. Soc. A.* 363:433–452.
9. Sunde, M., and C. C. F. Blake. 1997. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Prot. Chem.* 50:123–159.
10. Serpell, L. C., M. Sunde, M. D. Benson, G. A. Tennent, M. B. Pepys, and P. E. Fraser. 2000. The protofilament substructure of amyloid fibrils. *J. Mol. Biol.* 300:1033–1039.
11. Jimenez, J. L., E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, and H. R. Saibil. 2002. The protofilament structure of insulin amyloid fibrils. *Proc. Natl. Acad. Sci. USA.* 99:9196–9201.
12. Kishimoto, A., K. Hasegawa, H. Suzuki, H. Taguchi, K. Namba, and M. Yoshida. 2004. beta-Helix is a likely core structure of yeast prion Sup35 amyloid fibers. *Biochem. Biophys. Res. Commun.* 315:739–745.
13. Scheibel, T., R. Parthasarathy, G. Sawicki, X. M. Lin, H. Jaeger, and S. L. Lindquist. 2003. Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proc. Natl. Acad. Sci. USA.* 100:4527–4532.
14. MacPhee, C. E., and D. N. Woolfson. 2004. Engineered and designed peptide-based fibrous biomaterials. *Curr. Opin. in Solid State and Mat. Sci.* 8:141–149.
15. Paushkin, S., V. Kushnirov, V. Smirnov, and M. Ter-Avanesyan. 1997. In vitro propagation of the prion-like state of yeast Sup35 protein. *Science.* 277:381–383.
16. Scheibel, T., and S. L. Lindquist. 2001. The role of conformational flexibility in prion propagation and maintenance for Sup35p. *Nat. Struct. Biol.* 8:958–962.
17. Bussell Jr., R., and D. Eliezer. 2001. Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein. *J. Biol. Chem.* 276:45996–46003.
18. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338:1015–1026.
19. Csizmok, V., M. Bokor, P. Banki, E. Klement, K. F. Medzihradszky, P. Friedrich, K. Tompa, and P. Tompa. 2005. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry.* 44:3955–3964.
20. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
21. Uversky, V. N., and A. L. Fink. 2004. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta.* 1698:131–153.
22. Glover, J. R., A. S. Kowal, E. C. Schirmer, M. M. Patino, J. J. Liu, and S. Lindquist. 1997. Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of S. cerevisiae. *Cell.* 89:811–819.
23. Collins, S. R., A. Douglass, R. D. Vale, and J. S. Weissman. 2004. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol.* 2:e321.
24. Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. 2002. A structural model for Alzheimer's -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. USA.* 99:16742–16747.
25. Tycko, R. 2003. Insights into the amyloid folding problem from solid-state NMR. *Biochemistry.* 42:3151–3159.
26. Petkova, A. T., G. Buntkowsky, F. Dyda, R. D. Leapman, W. M. Yau, and R. Tycko. 2004. Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J. Mol. Biol.* 335:247–260.
27. Jaroniec, C. P., C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin. 2004. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA.* 101: 711–716.
28. Bouchard, M., J. Zurdo, E. J. Nettleton, C. M. Dobson, and C. V. Robinson. 2000. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci.* 9:1960–1967.
29. Kheterpal, I., A. Williams, C. Murphy, B. Bledsoe, and R. Wetzel. 2001. Structural features of the Abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry.* 40:11757–11767.
30. Kheterpal, I., H. A. Lashuel, D. M. Hartley, T. Walz, P. T. Lansbury Jr., and R. Wetzel. 2003. Abeta protofibrils possess a stable core structure resistant to hydrogen exchange. *Biochemistry.* 42:14092–14098.
31. Jao, C. C., A. Der-Sarkissian, J. Chen, and R. Langen. 2004. Structure of membrane bound alpha-synuclein studied by site-directed spin labeling. *Proc. Natl. Acad. Sci. USA.* 101:8331–8336.
32. Wille, H., M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard, and S. B. Prusiner. 2002. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. USA.* 99:3563–3568.
33. Jimenez, J. L., J. I. Guijarro, E. Orlova, J. Zurdo, C. M. Dobson, M. Sunde, and H. R. Saibil. 1999. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* 18:815–821.
34. Polverino de Laureto, P., N. Taddei, E. Frare, C. Capanni, S. Constantini, J. Zurdo, F. Chiti, C. M. Dobson, and A. Fontana. 2003. Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J. Mol. Biol.* 334:129–141.
35. Bousset, L., V. Redeker, P. Decottignies, S. Dubois, P. Le Marechal, and R. Melki. 2004. Structural characterization of the fibrillar form of the yeast Saccharomyces cerevisiae prion Ure2p. *Biochemistry.* 43: 5022–5032.
36. Fontana, A., G. Fassina, C. Vita, D. Dalzoppo, M. Zamai, and M. Zambonin. 1986. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry.* 25:1847–1851.
37. DePace, A. H., A. Santoso, P. Hillner, and J. S. Weissman. 1998. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell.* 93:1241–1252.
38. Serio, T. R., A. G. Cashikar, A. S. Kowal, G. J. Sawicki, and S. L. Lindquist. 2001. Self-perpetuating changes in Sup35 protein conformation as a mechanism of heredity in yeast. *Biochem. Soc. Symp.* 68:35–43.
39. Tanaka, M., P. Chien, N. Naber, R. Cooke, and J. S. Weissman. 2004. Conformational variations in an infectious protein determine prion strain differences. *Nature.* 428:323–328.