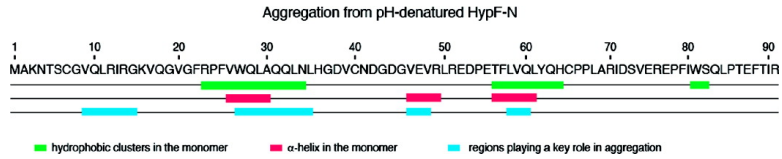
Article

Structure and Dynamics of a Partially Folded Protein Are Decoupled from Its Mechanism of Aggregation

Giulia Calloni, Christofer Lendel, Silvia Campioni, Silva Giannini, Alessandra Gliozzi, Annalisa Relini, Michele Vendruscolo, Christopher M. Dobson, Xavier Salvatella, and Fabrizio Chiti

J. Am. Chem. Soc., **2008**, 130 (39), 13040-13050 • DOI: 10.1021/ja8029224 • Publication Date (Web): 04 September 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES

Published on Web 09/04/2008

Structure and Dynamics of a Partially Folded Protein Are Decoupled from Its Mechanism of Aggregation

Giulia Calloni,[†,⊥] Christofer Lendel,[‡] Silvia Campioni,[†] Silva Giannini,[†,‡] Alessandra Gliozzi,[§] Annalisa Relini,[§] Michele Vendruscolo,[‡] Christopher M. Dobson,[‡] Xavier Salvatella,[*,‡,#] and Fabrizio Chiti[*,†,∥]

Dipartimento di Scienze Biochimiche, Università di Firenze, Viale Morgagni 50, 50134 Firenze, Italy, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K., Dipartimento di Fisica, Università di Genova and Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia (CNISM), via Dodecaneso 33, I-16146, Genova, Italy, and Consorzio interuniversitrio "Istituto Nazionale Biostrutture e Biosistemi" (INBB), Viale delle Medaglie d'Oro 305, 00136 Roma, Italy

Received April 28, 2008; E-mail: xs210@cam.ac.uk; fabrizio.chiti@unifi.it

Abstract: A common strategy to study the mechanism of amyloid formation is the characterization of the structure and dynamics of the precursor state, which is in most cases a partially folded protein. Here we investigated the highly dynamic conformational state formed by the protein domain HypF-N at low pH, before aggregation, using fluorescence, circular dichroism, and NMR spectroscopies. The NMR analysis allowed us, in particular, to identify the regions of the sequence that form hydrophobic interactions and adopt an α-helical secondary structure in the pH-denatured ensemble. To understand the role that this residual structure plays in the aggregation of this protein, we probed the mechanism of aggregation using protein engineering experiments and thus identified the regions of the sequence of HypF-N that play a critical role in the conversion of this dynamic state into thioflavin T-binding and β-sheet containing protofibrils. The combination of these two complementary approaches revealed that the aggregation of pH-denatured HypF-N is not structure-dependent, meaning that it is not driven by the regions of the protein that are either less or more protected in the initial partially folded state. It is, by contrast, promoted by discrete protein regions that have the highest intrinsic propensity to aggregate because of their physicochemical properties.

Introduction

Peptides and proteins have a generic tendency to convert from their soluble states into well-organized aggregates characterized by a fibrillar morphology and an extended cross-β structure, generally referred to as amyloid or amyloid-like fibrils.[1−3] This process is important for a number of reasons. From a physicochemical viewpoint, it represents an essential feature of the behavior of polypeptide chains that needs to be fully characterized to develop a detailed understanding of the nature and evolution of protein molecules.[1,4,5] From a biological and biomedical perspective, the aggregation process leads to the formation of cytotoxic oligomeric species that are damaging to living systems.[3,6,7] A detailed analysis of the initial steps of aggregation can facilitate our understanding of the crucial protective mechanisms that are in place in living organisms to prevent the proliferation of such detrimental species. The formation of amyloid fibrils or intracellular inclusions with amyloid-like characteristics is also associated with more than 40 pathological conditions in humans.[3] In spite of their link with pathology, fibrillar species with amyloid-like characteristics can serve a number of biological functions in living organisms, provided they assemble under controlled conditions.[3,8]

The soluble protein oligomers that precede the formation of mature fibrils are becoming increasingly important because they are intermediates in the mechanism of amyloid formation and may be the pathogenic species in a number of protein deposition diseases.[3,6,7] The formation of such soluble oligomers by normally folded proteins requires, in most cases, some degree of unfolding of the compact globular structure to generate an ensemble of partially unfolded conformations. Moreover, many systems that are associated with protein deposition diseases are unstructured peptides or intrinsically disordered proteins when studied in isolation. Examples of such proteins include the

[†] Università di Firenze.
[‡] University of Cambridge.
[§] Università di Genova and CNISM.
[∥] INBB.
[⊥] Present address: Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany.
[#] Present address: Institute for Research in Biomedicine, Parc Científic de Barcelona, Baldiri Reixac 10-12, 08028 Barcelona, Spain.

(1) Dobson, C. M. *Trends Biochem. Sci.* **1999**, *24*, 329–332.
(2) Uversky, V. N.; Fink, A. L. *Biochim. Biophys. Acta* **2004**, *1698*, 131–153.
(3) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(4) Jahn, T. R.; Radford, S. E. *FEBS J.* **2005**, *272*, 5962–5970.
(5) Monsellier, E.; Chiti, F. *EMBO Rep.* **2007**, *8*, 737–742.
(6) Stefani, M.; Dobson, C. M. *J. Mol. Med.* **2003**, *81*, 678–699.
(7) Walsh, D. M.; Selkoe, D. J. *Protein Pept. Lett.* **2004**, *11*, 213–228.
(8) Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W. *Trends Biochem. Sci.* **2007**, *32*, 217–224.

 10.1021/ja8029224 CCC: $40.75

amyloid β peptide (Aβ), the islet amyloid polypeptide, and proteins τ and α-synuclein.[3] The aggregation of these systems is a process of self-assembly that occurs directly from such states without the need for an initial partial unfolding step.

Despite its importance for the theoretical description of protein aggregation, the detailed mechanism by which partially folded states self-assemble to form initial oligomeric aggregates is not fully understood. To address this issue, it is important to study the relationship between the degree of structure present in the monomeric states populated by amyloidogenic proteins under conditions that promote aggregation and the mechanism of formation of the initial aggregates. In particular, it is crucial to identify the regions of the sequence that are most structured, and those that are most unfolded, in the initial monomeric state and determine the role that these regions play, if any, in the mechanism of aggregation.

The N-terminal domain of HypF from *Escherichia coli* (HypF-N) is particularly well suited for this purpose. In its native state, this α/β 91 residue protein has a ferredoxin-like fold with two α-helices packed against a five-stranded β-sheet.[9] At low pH or in the presence of trifluoroethanol (TFE), HypF-N unfolds across the major free energy barrier for unfolding into a partially structured state that self-assembles to form aggregates that are morphologically, structurally, and tinctorially indistinguishable from disease-associated amyloid fibrils.[10,11] Even under mildly destabilizing conditions that maintain the native state as the predominant species in solution, aggregation is promoted by a partially folded state that is scarcely populated (<2%) and in rapid equilibrium with the native state.[12] In addition, the extent to which HypF-N mutational variants are sequestered into inclusion bodies in *E. coli* cells following overexpression has been found to correlate inversely with the free energy change of unfolding ($\Delta G^{H_2O}_{U-F}$),[13] indicating that aggregation of this protein in its native cellular environment also requires partial unfolding of the native structure.[13] The formation of mature fibrils by HypF-N in the presence of moderate concentrations of TFE is preceded by the accumulation of soluble prefibrillar aggregates that were found to be cytotoxic when added to the extracellular medium of a number of cell lines,[14−16] to cause dysfunction of cholinergic neurons when microinjected into the brains of rats,[17] and to permeabilize liposomes.[11]

Here, we characterize the structure and dynamics of the partially folded state that HypF-N populates at low pH and low ionic strength and study the mechanism of its conversion into prefibrillar aggregates by using NMR spectroscopy and protein engineering, respectively. This highly favorable opportunity to study the relationship between the structure/dynamics of the initial precursor state and the process of aggregation is due to our ability to either stabilize the initial monomeric state of HypF-N or promote its aggregation into protofibrillar species by modifying the ionic strength of the solution.[18]

Methods

Cloning of the HypF-N Gene, Expression, and Protein Purification. Cloning, expression, and purification of wild-type HypF-N were carried out as previously described.[18] In brief, the gene for HypF-N was cloned in a modified pQE30-Xa plasmid (Qiagen S.p.A.), in which the DNA stretch coding for the factor Xa cleavage site was changed into a sequence coding for the thrombin cleavage site (pQE30-Th). Cultures of *E. coli* XL1 blue cells harboring the pQE30-Th/HypF-N vector were grown at 37 °C in LB medium with 100 μg/mL of ampicillin (Sigma-Aldrich) under vigorous shaking. At $OD_{600} \approx 0.6$, protein expression was induced for 4 h at 37 °C using 1 mM isopropyl β-D-thiogalactoside from Inalco. Cells were harvested and lysed, and the cell lysate was applied at 4 °C to an affinity chromatography column packed with the HIS-Select nickel affinity gel (Sigma-Aldrich). The column was washed and incubated overnight at 4 °C with 50 units of human thrombin (Sigma-Aldrich). Pure HypF-N separated from the His-tag was eluted with 50 mM phosphate, 50 mM NaCl, and 10 mM imidazole, pH 8.0. The protein sample was buffer-exchanged (5 mM acetate buffer, 2 mM dithiothreitol (DTT), pH 5.5) and concentrated at 4 °C. Mutations in the HypF-N genes were generated by using the QuikChange site-directed mutagenesis kit (Stratagene). Expression and purification of the mutants were carried out as described for the wild-type protein.

Expression of ^{15}N−^{13}C HypF-N. Cultures of *E. coli* BL21 (DE3) cells harboring the pQE30-Th/HypF-N vector were grown in ^{15}N−^{13}C-enriched M9 minimal medium. ^{15}N−^{13}C-labeled HypF-N was expressed and purified as described for the unlabeled protein with the addition of a further purification step to remove any residual impurities. In brief, protein samples were applied to a size-exclusion chromatography column packed with 125 mL of Superdex G75 resin (Amersham-Pharmacia) equilibrated at 20 °C with two volumes of 50 mM ammonium acetate buffer. After elution, the buffer was exchanged with water using 5000 Da cutoff centriplus (Millipore), and samples were immediately used for NMR experiments.

Equilibrium Unfolding Measurements. Twenty-three buffer solutions were prepared at various pH values ranging from 7.4 to 1.3. For pH values of 7.4, 5.7−7.1, 3.9−5.1, and 2.1−3.7, solutions containing 50 mM Tris, 25 mM 3,3-dimethylglutarate, 50 mM acetate, and 50 mM citrate were used, respectively. pH values lower than 2.0 were obtained by adding different amounts of trifluoroacetic acid (TFA). Since a given compound can buffer the pH only in a limited range of pH ($pK_a - 1 <$ pH $< pK_a + 1$), it was necessary to use different compounds to cover the pH range under investigation (1.3−7.4). All solutions contained 2 mM mercaptoethanol to maintain the three cysteine residues in a reduced form, and a total ionic strength of 50 mM was achieved for all solutions by adding different amounts of NaCl. HypF-N was diluted to a final concentration of 19 μM in each solution, and the samples were incubated for 30 min at 28 °C. No aggregation occurred during this time period. The samples were then subjected to far-UV circular dichroism (CD) and intrinsic fluorescence analyses (see below), and plots of fluorescence emission at 320 nm and mean residue ellipticity at 222 nm were fitted to a sigmoidal function of the form:

$$y = [(q_1 + m_1 \times \mathrm{pH}) + (q_2 + m_2 \times \mathrm{pH}) \times \exp(-a + b \times \mathrm{pH})]/[1 + \exp(-a + b \times \mathrm{pH})] \quad (1)$$

where q_1 and m_1 represent the intercept and the slope, respectively, of the linear portion corresponding to the unfolded ensemble; q_2 and m_2 are the corresponding parameters for the native state; and

(9) Rosano, C.; Zuccotti, S.; Bucciantini, M.; Stefani, M.; Ramponi, G.; Bolognesi, M. *J. Mol. Biol.* **2002**, *321*, 785–796.
(10) Chiti, F.; Bucciantini, M.; Capanni, C.; Taddei, N.; Dobson, C. M.; Stefani, M. *Protein Sci.* **2001**, *10*, 2541–2547.
(11) Relini, A.; Torrassa, S.; Rolandi, R.; Gliozzi, A.; Rosano, C.; Canale, C.; Bolognesi, M.; Plakoutsi, G.; Bucciantini, M.; Chiti, F.; Stefani, M. *J. Mol. Biol.* **2004**, *338*, 943–957.
(12) Marcon, G.; Plakoutsi, G.; Canale, C.; Relini, A.; Taddei, N.; Dobson, C. M.; Ramponi, G.; Chiti, F. *J. Mol. Biol.* **2005**, *347*, 323–335.
(13) Calloni, G.; Zoffoli, S.; Stefani, M.; Dobson, C. M.; Chiti, F. *J. Biol. Chem.* **2005**, *280*, 10607–10613.
(14) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507–511.
(15) Bucciantini, M.; Calloni, G.; Chiti, F.; Formigli, L.; Nosi, D.; Dobson, C. M.; Stefani, M. *J. Biol. Chem.* **2004**, *279*, 31374–31382.
(16) Cecchi, C.; Baglioni, S.; Fiorillo, C.; Pensalfini, A.; Liguri, G.; Nosi, D.; Rigacci, S.; Bucciantini, M.; Stefani, M. *J. Cell Sci.* **2005**, *118*, 3459–3470.
(17) Baglioni, S.; Casamenti, F.; Bucciantini, M.; Luheshi, L. M.; Taddei, N.; Chiti, F.; Dobson, C. M.; Stefani, M. *J. Neurosci.* **2006**, *26*, 8160–8167.

a and *b* are parameters related to the position and width of the transition, respectively (the ratio between *a* and *b* represents the midpoint of the pH transition). The raw data were converted to give the fractions of folded versus unfolded protein as:

$$\text{fraction folded} = (y - y_U)/(y_N - y_U) \quad (2)$$

where *y* represents the observed spectroscopic signal at a given pH value, and y_U and y_N correspond to the signals of the unfolded and native states at the same pH value, respectively. The data were plotted against pH and again fitted by using eq 1.

In a different set of experiments, 30 samples containing different concentrations of urea, ranging from 0 to 8.7 M, were prepared. All contained 19 μM HypF-N, 20 mM TFA, 30 mM NaCl, and 2 mM DTT, pH 1.7, 25 °C. All samples were equilibrated for 30 min at 25 °C and subjected to far-UV CD and intrinsic fluorescence analyses (see the next two sections).

Far-UV CD. All spectra were acquired at 28 °C using 19 μM HypF-N and a 1-mm path length cell. A Jasco J-810 spectropolarimeter equipped with a thermostated cell holder attached to a Thermo Haake C25P water bath was used. For the experiments described in Figures 1a,c and 5b, each spectrum represents the average of four, five, and 10 scans, respectively; these were acquired at 100, 50, and 20 nm/min with a response time of 0.5, 1, and 4 s, respectively. All spectra were blank-subtracted and converted to molar ellipticity per residue. A time-course measurement was performed by acquiring the CD spectrum of a solution containing 48 μM HypF-N, in 20 mM TFA, 2 mM DTT, and 330 mM NaCl, pH 1.7, 25 °C at different times to follow the aggregation process. The spectra of the aggregates reported in Figure 5b and in its inset correspond to an individual scan performed at 20 nm/min with a response time of 2 s.

Intrinsic Fluorescence. All samples, initially containing 19 μM protein, were diluted 10-fold in the corresponding buffer solutions immediately before the measurements. Intrinsic fluorescence emission spectra were acquired at 28 °C from 300 to 450 nm with excitation at 280 nm. A Perkin-Elmer LS 55 spectrofluorimeter equipped with a thermostatted cell holder attached to a Haake F8 water bath was used. Each spectrum represents the average of four different scans acquired at 100 nm/min with an excitation slit of 4 nm and an emission slit of 2.5 nm. The maximum emission wavelength (λ_{max}) was obtained by a fitting procedure, applied to the fluorescence data around the peak in the spectrum, using a third-order polynomial function.

ThT Assay. Wild-type HypF-N and its variants were incubated at a concentration of 48 μM in 20 mM TFA and 2 mM DTT, pH 1.7, 25 °C in the presence of 330 mM NaCl, ionic strength 350 mM. At regular time intervals, 60-μL aliquots of each sample were added to 440 μL of a solution containing 25 μM ThT in 25 mM phosphate buffer to yield a solution of pH 6.0. The steady-state fluorescence of the resulting mixture was measured at 25 °C using the apparatus described above and a 2 × 10-mm path length quartz cell. The excitation and emission wavelengths were 440 and 485 nm, respectively. Plots of the ThT fluorescence intensity versus time were fitted to single-exponential functions of the form:

$$F(t) = F_{eq} + A \exp(-k_{app}t) \quad (3)$$

where F_{eq} is the value at the plateau, *A* is the amplitude of the fluorescence change upon aggregation, and k_{app} is the apparent rate constant.

Tapping Mode Atomic Force Microscopy (TM-AFM). Twenty-microliter aliquots of protein sample (48 μM HypF-N in 20 mM TFA, 2 mM DTT, and 330 mM NaCl, pH 1.7, 25 °C) were withdrawn at fixed times, diluted 330 times, deposited on a freshly cleaved mica substrate, and dried under a gentle nitrogen flow. Dilution of the sample was required to avoid salt crystallization.

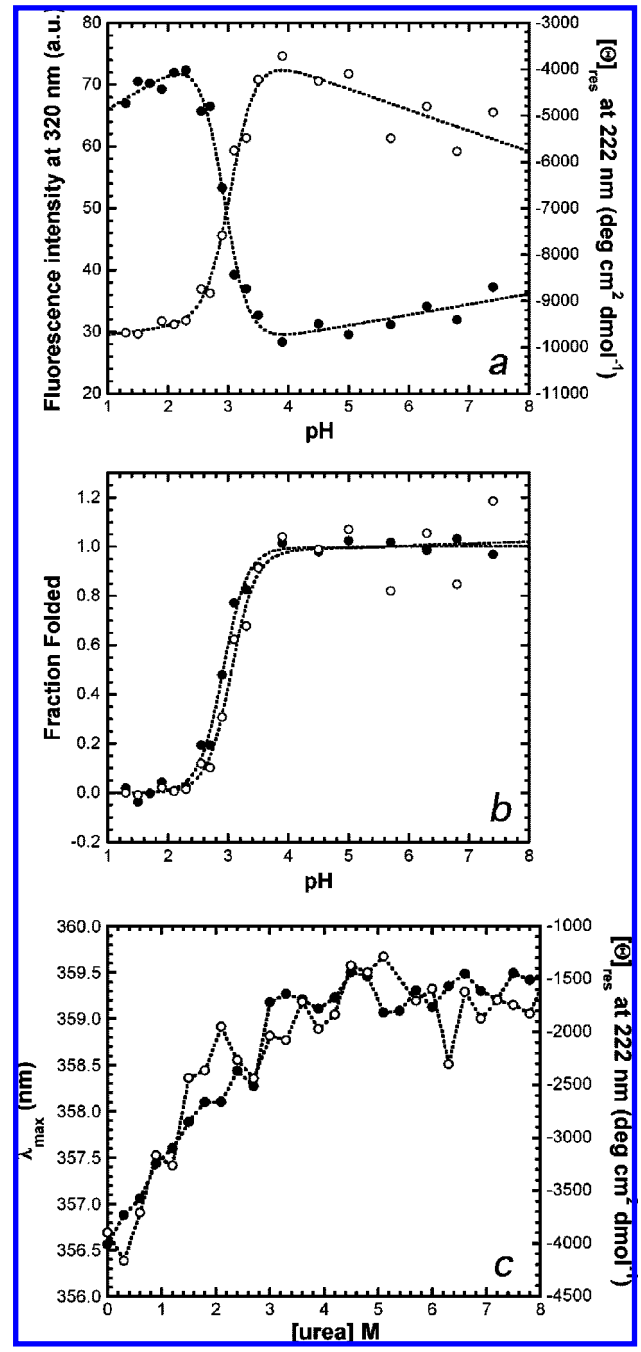


Figure 1. (a) pH-induced unfolding of HypF-N monitored by fluorescence intensity at 320 nm (○) and the mean residue ellipticity at 222 nm (●). The dashed lines represent the best fits to the sigmoidal function given in eq 1. (b) pH-induced unfolding curves obtained by converting the fluorescence (○) and far-UV CD (●) data shown in (a) into values of the fraction of folded molecules using eq 2. The dashed lines again result from a fitting procedure using eq 1. (c) Urea-induced unfolding curves of HypF-N, at pH 1.7, obtained by following the maximum emission wavelength, λ_{max} (○), and the mean residue ellipticity at 222 nm (●) as a function of urea concentration.

Alternatively, the protein aliquot was deposited on mica without dilution; after 1-min incubation, the sample was washed with Milli-Q water and then dried under nitrogen. TM-AFM images were acquired in air using a Dimension 3000 microscope (Digital Instruments-Veeco), equipped with a "G" scanning head (maximum scan size 100 μm) and driven by a Nanoscope IIIa controller. Single beam uncoated silicon cantilevers (type RTESPA, Veeco) were used. The drive frequency was between 280 and 320 kHz, and the scan rate was between 1.0 and 1.5 Hz. Vertical displacements

(18) Campioni, S.; Mossuto, M. F.; Torrassa, S.; Calloni, G.; de Laureto, P. P.; Relini, A.; Fontana, A.; Chiti, F. *J. Mol. Biol.* **2008**, *379*, 554–567.

were calibrated by measuring the depth of grating notches (180 nm) and the half-unit cell steps (1 nm) obtained by treating freshly cleaved mica with hydrofluoric acid. The horizontal displacements of the piezoelectric tube were calibrated using a 3-μm pitch diffraction grating. Aggregate sizes were obtained from the height in cross section of the topographic AFM images. The measured heights are reduced with respect to fully hydrated conditions due to the drying procedure. The aggregate sizes reported in the Results and Discussion were obtained by multiplying the measured height values by a shrinking factor of 1.4, determined by comparing the height values of native HypF-N measured under liquid and in air after drying under nitrogen. Standard errors are reported.

NMR Spectroscopy. All NMR experiments were carried out using 0.15−0.35 mM samples of ^{15}N- and ^{13}C-labeled HypF-N in 10% $^{2}H_2O$. The pH was adjusted to 1.9 by adding TFA, and 3-(trimethylsilyl)propionate was used as an internal chemical shift reference. All ^{1}H−^{15}N HSQC and triple resonance NMR spectra were acquired on a Bruker Avance 700 MHz spectrometer equipped with a TXI Cryoprobe or an Bruker Avance 500 MHz spectrometer equipped with a TCI-ATM cryoprobe at 25 °C. The spectra were processed using NMRPipe,[19] and referenced, visualized, and analyzed using Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco) or CCPNMR.[20] Complete sequential backbone assignments for the $^{1}H^{N}$, ^{15}N, ^{13}Cα, ^{13}C$'$, and ^{13}Cβ nuclei were obtained from triple resonance experiments HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB carried out on a sample in 6 M urea at pH 1.9; the ^{1}H and ^{15}N assignments at other concentrations of urea were measured by means of a urea titration carried out between 0 and 8 M; the Cα chemical shifts were measured at 0, 2, 6, and 8 M urea using HNCA experiments.

Results and Discussion

HypF-N Is Partially Folded at Low pH. Intrinsic fluorescence and far-UV CD spectra were acquired for HypF-N at various pH values, ranging from 1.3 to 7.4, while maintaining the ionic strength of the solutions at 50 mM in all samples. Both spectra show substantial changes as the pH is reduced from 3.5 to 2.5. The pH dependence of intrinsic fluorescence at 320 nm and mean residue ellipticity at 222 nm indicate a single sharp transition in the pH range 2.5−3.5 with well-defined pre- and post-transition regions (Figure 1a). When these plots are normalized to report on the fraction of folded protein, the transitions observed by these two techniques become superimposable within experimental error (Figure 1b). The presence of a single sharp transition and the superimposition of the plots obtained with two different spectroscopic probes reporting on different types of structure suggest that the pH-induced unfolding transition follows, to a good approximation, a two-state model with a midpoint at pH 2.95 ± 0.20. Moreover, the decreases of far-UV CD signal and intrinsic fluorescence indicate that the transition consists in an unfolding process as the pH is lowered.

To investigate whether nonrandom structure is present in pH-denatured HypF-N, the intrinsic fluorescence and far-UV CD spectra were measured at various urea concentrations (from 0 to 8 M) while maintaining a constant pH value of 1.7 and a constant ionic strength of 50 mM. The wavelength of maximum emission (λ_{max}) undergoes a nonlinear red-shift upon the addition of urea from ca. 356.5 to ca. 359.5 nm (Figure 1c). The mean residue ellipticity at 222 nm also changes nonlinearly from ca. −4000 to ca. −1500 deg cm^2 dmol^{-1} following the addition of urea (Figure 1c). These two spectroscopic probes indicate that the pH-denatured state of HypF-N presents nascent secondary structure and partial exclusion of water molecules from the vicinity of tryptophan side chains and that such transient structure fully unfolds, in a rather noncooperative manner, as the concentration of urea is increased (Figure 1c). Such noncooperative transitions are typically observed in the chemical denaturation of partially folded states devoid of persistent long-range tertiary contacts.[21,22] Overall, therefore, the experimental data indicate that at low pH HypF-N contains significant elements of nonrandom structure.

The observation that pH-denatured HypF-N is best described as an ensemble of fluctuating conformations with elements of nonrandom structure is consistent with previous studies that indicated that at low pH and in the absence of chemical denaturant this protein adopts a state with the characteristics of a premolten globule.[18] Indeed, from these studies it was found that pH-denatured HypF-N has values of secondary structure content, degree of solvent exposure of tryptophan and tyrosine residues, hydrodynamic diameter, and susceptibility to limited proteolysis that are intermediate between those of the fully folded protein and those of a fully unfolded polypeptide chain.[18] Importantly, experiments of dynamic light scattering and photoinduced cross-linking of unmodified proteins (PICUP) indicate that the pH-denatured state of HypF-N is a monomer under these conditions and that such a monomeric state is stable for at least 6 h (Figure S1 in Supporting Information). Thus, the stability of such a conformational state over time provides an opportunity to carry out a detailed structural characterization using NMR spectroscopy.

Structure of HypF-N at Low pH Is Characterized by Transient Hydrophobic Interactions. Non-native states of proteins are challenging to characterize by NMR as they are ensembles of conformations that interconvert on the submillisecond time scale.[23−32] This great structural heterogeneity gives rise, through motional averaging, to the narrow chemical shift dispersion and consequent resonance overlap that is the hallmark of the spectra of non-native proteins. In addition, the exchange contribution to the line widths due to motions in the microsecond to millisecond time scale often gives rise to resonance broadening that can prevent observation of part of the NMR signals,[23,24] particularly those that stem from the most structured regions of the protein. Despite these inherent difficulties, information about the degree of structure present in partially folded states can be

(19) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(20) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 687–696.
(21) Schulman, B. A.; Kim, P. S.; Dobson, C. M.; Redfield, C. *Nat. Struct. Biol.* **1997**, *4*, 630–634.
(22) Greene, L. H.; Wijesinha-Bettoni, R.; Redfield, C. *Biochemistry* **2006**, *45*, 9475–9484.
(23) Baum, J.; Dobson, C. M.; Evans, P. A.; Hanley, C. *Biochemistry* **1989**, *28*, 7–13.
(24) Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063–1072.
(25) McParland, V. J.; Kalverda, A. P.; Homans, S. W.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 326–331.
(26) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.
(27) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607–3622.
(28) Jahn, T. R.; Parker, M. J.; Homans, S. W.; Radford, S. E. *Nat. Struct. Mol. Biol.* **2006**, *13*, 195–201.
(29) Mok, Y. K.; Kay, C. M.; Kay, L. E.; Forman-Kay, J. *J. Mol. Biol.* **1999**, *289*, 619–638.
(30) Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M.; Fersht, A. R. *Nature* **2005**, *437*, 1053–1056.
(31) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719–1722.
(32) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011–1032.

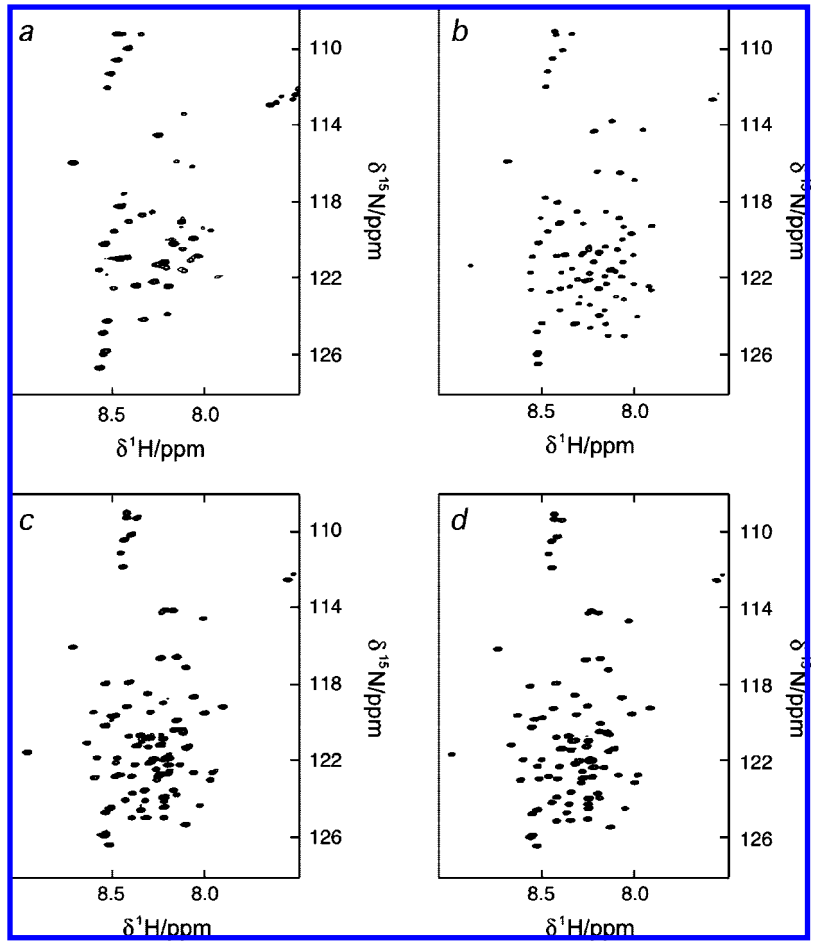


Figure 2. (a−d) $^{1}H-^{15}N$ HSQC spectra of $^{13}C-^{15}N$-labeled HypF-N acquired at pH 1.9 in 0, 2, 6, and 8 M urea respectively, at 700 MHz.

obtained in a variety of ways, for example, by monitoring the intensities of individual NMR resonances as a function of denaturant concentration.[21,22,25,33]

The $^{1}H-^{15}N$ heteronuclear single quantum correlation (HSQC) spectrum of HypF-N at pH 1.9 and low ionic strength (less than 50 mM) shows features typical of a partially structured protein (Figure 2a). Although many well-resolved amide cross-peaks are sharp and visible in the spectrum, a very significant number are absent or have very low intensity due to severe line broadening. This indicates the presence of structural fluctuations on the microsecond to millisecond time scale between conformations with different chemical shifts. In addition, the cross-peaks that can be observed have very low chemical shift dispersion, a situation typical of unstructured polypeptides. Taken together, these results indicate that in the absence of urea there are regions of the protein that are highly unfolded but others that are structured to a degree that leads to severe line broadening. By recording a series of $^{1}H-^{15}N$ HSQC spectra at pH 1.9 and low ionic strength at increasing concentrations of urea (from 0 to 8 M), we found that the line-broadening effects were greatly reduced (Figure 2a−d), allowing the assignment of most nonproline backbone amide resonances using standard triple resonance experiments.

In the $^{1}H-^{15}N$ HSQC spectra acquired in this way, a total of 48 amide cross-peaks were well-resolved, allowing their intensities to be accurately monitored at all urea concentrations and in the absence of chemical denaturant. We divided the residues giving rise to the peaks into four classes on the basis of their behavior as the concentration of denaturant was increased (Figure 3). The intensities of one group of resonances (assigned to residues 1−7 and corresponding to the N-terminus and the initial portion of strand β1 in the native structure) are already high in the absence of urea and do not change significantly with denaturant concentration (Figure 3a, type I). The intensities of the resonances of the other residues of HypF-N vary widely in the absence of denaturant and all increase with urea concentration, yielding quasi-sigmoidal curves with a midpoint at ca. 3.5 M urea (Figure 3b−d). In the absence of denaturant, the intensities of the peaks of residues 8−21 (corresponding to strand β1 and to the loop that connects this strand with helix α1) are 35−65% of the corresponding values in 6 M urea (Figure 3b, type II). The transition from type I to type II behavior is not completely distinct, as can be seen by comparing the behaviors of Thr5 and Ser6 with that of Gly8 and Val9 in Figure 3a,b, respectively. Three regions of the sequence (spanning approximately residues 36−55, 72−77, and 86−89 and corresponding in the native structure to the hairpin formed by strands β2−β3 and the following loop, the central portion of strand β4 and the central portion of strand β5, respectively) have an even more marked dependence on urea concentration (Figure 3c, type III). In the absence of urea, these peaks have only 10−30% of the intensities measured in 6 M urea. The intensity of the resonances of the remaining three regions of the sequence (approximately spanning residues 23−34, 56−64, and 81−82,

(33) Redfield, C. *Methods* **2004**, *34*, 121–132.

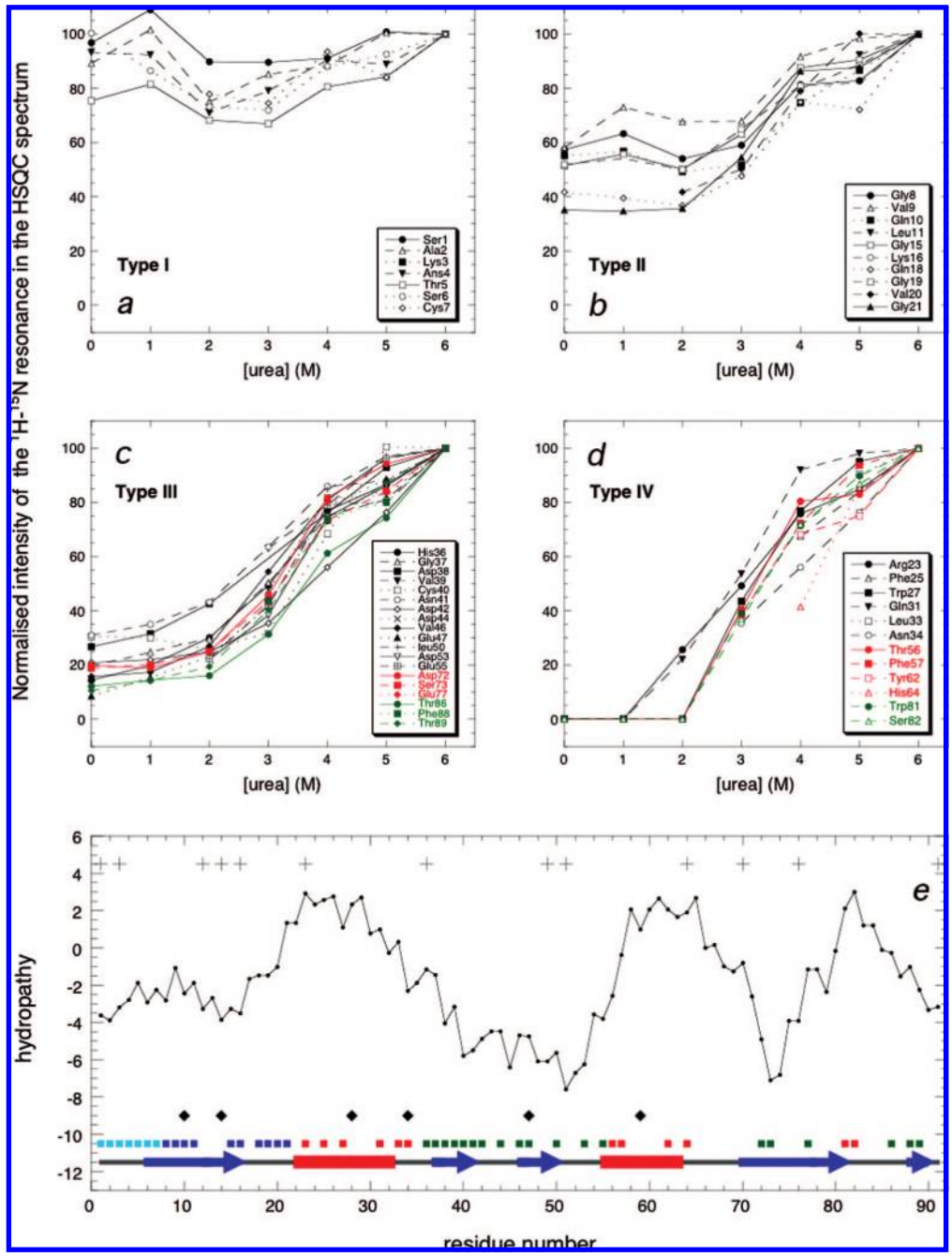


Figure 3. Urea concentration dependence of the intensities of amide cross-peaks from the ^{1}H−^{15}N HSQC spectrum of acid-denatured HypF-N. (a) Type I resonances having intensities with at most a small dependence on urea concentration. (b) Type II resonances having intensity values in 0 M urea that are 35−65% of those measured in 6 M urea. (c) Type III resonances having intensity values in 0 M urea that are 10−30% of those measured in 6 M urea. (d) Type IV resonances having no detectable intensity in 0 and 1 M urea. (e) Distribution of positive charges at pH 2.0 along the sequence (+) and hydropathy profile (●) of HypF-N calculated using the reported hydrophobicity scale of the 20 residues.[34] A sliding window of nine residues was used, except at the C- and N-termini where shorter windows were used. The diamonds below the profile indicate the positions of the mutations that are found to perturb the aggregation rate (Figure 6a). The filled colored squares below the profile show the urea-titration behavior of the residues indicated: pale blue, type I; blue, type II; green, type III; red, type IV. The sequence positions of α-helices (red) and β-strands (blue) in the native structure (PDB entry 1GXU) are also indicated as determined by MOLMOL.[35]

and corresponding, in the native structure, largely to helix α1 and the following loop, helix α2, and the loop that connects strands β4 and β5, respectively) are not detectable at urea concentrations below 2 M, but become clearly visible at higher concentrations (Figure 3d, type IV).

Thus, these three regions of the sequence (residues 23−34, 56−64, and 81−82) appear to fluctuate, at low pH and in the absence of urea, between structurally dissimilar conformations on the microsecond−millisecond time scale (i.e., significantly more slowly than the picosecond−nanosecond motions of very

(34) Pawar, A. P.; Dubay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J. Mol. Biol.* **2005**, *350*, 379–392.

(35) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graphics* **1996**, *14*, 51–55.

unstructured states). Such segments correspond to regions of the sequence that have the highest hydrophobicity and lowest density of positive charges (Figure 3e). In addition, they correspond to the part of the native state that has the lowest density of salt bridges, suggesting that protonation of acidic residues at low pH does not contribute to the destabilization of these regions through loss of specific electrostatic interactions. The presence of hydrophobic clusters stabilized by the inter-residue interactions that lead to severe line-broadening is supported by the observation that HypF-N binds 8-anilino-1-naphthalene sulfonic acid at acidic pH and low ionic strength.[18]

The segments adjacent to these hydrophobic regions (residues 36−55, 72−77, and 86−89) are less structured at low pH and no urea, although their flexibility may be limited by the restraints imposed by the adjacent hydrophobic stretches. These regions have low hydrophobicity and some dispersed positive charges (Figure 3e), in agreement with their lower tendency to become structured in the pH-denatured ensemble. A significant number of salt bridges also appear to stabilize these regions in the native structure (collectively forming most of the native β-sheet), suggesting that loss of such bridges as a consequence of the protonation of acidic residues may contribute to the destabilization of the native structure at low pH. Finally, the N-terminus (residues 1−21) is the most dynamic part of the protein with a degree of flexibility that increases with the distance from the first hydrophobic segment (residues 23−34). This is in agreement with the low hydrophobicity and high density of positive charges of this segment in the sequence at low pH (Figure 3e) and with the very probable loss, in the conditions of this study, of all three salt bridges present in strand β1 of the native state. Within the overall pH-denatured ensemble of HypF-N, there is no evidence for the presence of subdomains with cooperatively folded and persistent structure. Indeed, the low intensity of amide cross-peaks from the most structured regions in the $^{1}H-^{15}N$ HSQC spectrum in the absence of urea indicates that conformational exchange occurs even for these regions.

HypF-N at Low pH Contains Native and Non-Native α-Helical Structure. The urea concentration dependence of the $^{13}C\alpha$ chemical shifts of acid-denatured HypF-N was monitored using HNCA experiments, which allow the identification of sequence regions that present significant secondary structure, particularly α-helical, in the monomeric precursor state. The results of this analysis are presented in Figure 4a, where the differences [$\Delta\delta$(Cα)] between the $^{13}C\alpha$ chemical shifts of HypF-N in 0, 2, and 6 M urea and those in 8 M urea, where the protein is expected to be devoid of secondary structure, are presented. Regions of the sequence with increased $\Delta\delta$(Cα) values correspond to sequence stretches that have a tendency to populate α-helical or turnlike conformations. The results of this analysis show that residues 26−30 and 56−61, corresponding in the native structure largely to helix α1 and α2, respectively, have the highest probability of being in such a conformation ($\Delta\delta > 0.7$). A significant tendency to form α-helical or turn structure is also shown by residues 46−49, corresponding to strand β3 in the native state ($\Delta\delta > 0.4$). However, the lower values of $\Delta\delta$(Cα) in this region suggest that such structure is either more loosely formed or adopted by a smaller fraction of protein molecules in the ensemble.

A comparison between the $\Delta\delta$(Cα) values (Figure 4a) and the intrinsic helical propensities predicted by the AGADIR algorithm[37] (Figure 4b) shows a very good agreement. This suggests that the non-native nascent α-helical structure of residues 46−49 in the pH-denatured state, which forms strand β4 in native HypF-N, arises from the intrinsic helical propensity of this segment of the sequence. The network of medium- and long-range interactions that stabilize the native state plays an important role in defining the elements of secondary structure that are observed in protein structures and causes; in this specific case, formation of a β-strand conformation. In conditions such as those used in this work, where the native fold is destabilized at low pH, this network of interactions is severely compromised, allowing the intrinsic helical propensity of the sequence, which is dominated by short-range interactions,[37] to result in an α-helical conformation in this segment. The differences [$\Delta\delta$(Cβ)] between the $^{13}C\beta$ chemical shifts of HypF-N in 0 M urea and those in 8 M urea are close to or below zero (Figure 4c). Since positive $\Delta\delta$(Cβ) values can be used to identify β-sheet structure,

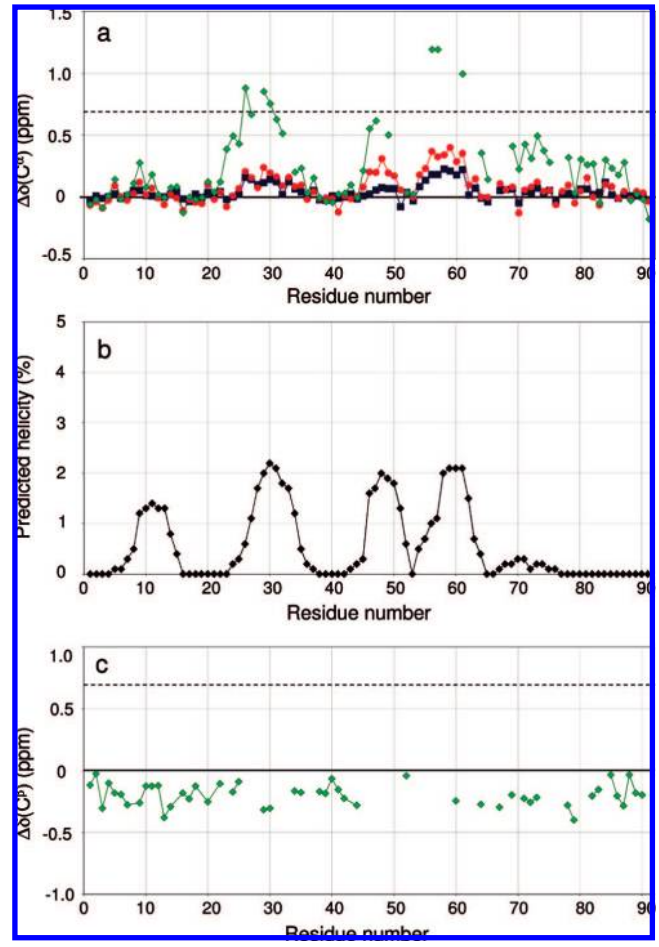


Figure 4. (a) Differences in the $^{13}C\alpha$ chemical shifts ($\Delta\delta$), as a function of residue number, between a HypF-N sample in 8 M urea and equivalent samples at 0 M (green ◆), 2 M (red ●), and 6 M (blue ■) urea. In all cases, the $\Delta\delta$ value was calculated by subtracting the chemical shift measured in 8 M urea from that measured at the indicated urea concentration. The expected difference between the $^{13}C\alpha$ chemical shift of a residue in a persistent α-helix conformation and that of the same residue in a random coil conformation[36] is shown as a dashed horizontal line. (b) Intrinsic helix propensities predicted by the AGADIR algorithm[37] for the sequence of HypF-N at pH 2.0 and 25 °C. (c) Differences in the $^{13}C\beta$ chemical shifts ($\Delta\delta$) between a HypF-N sample in 8 M urea and an equivalent sample at 0 M urea. The expected difference between the $^{13}C\beta$ chemical shift of a residue in a persistent β-sheet conformation and that of the same residue in a random coil conformation[36] is shown as a dashed horizontal line.

(36) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363–392.
(37) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173–191.

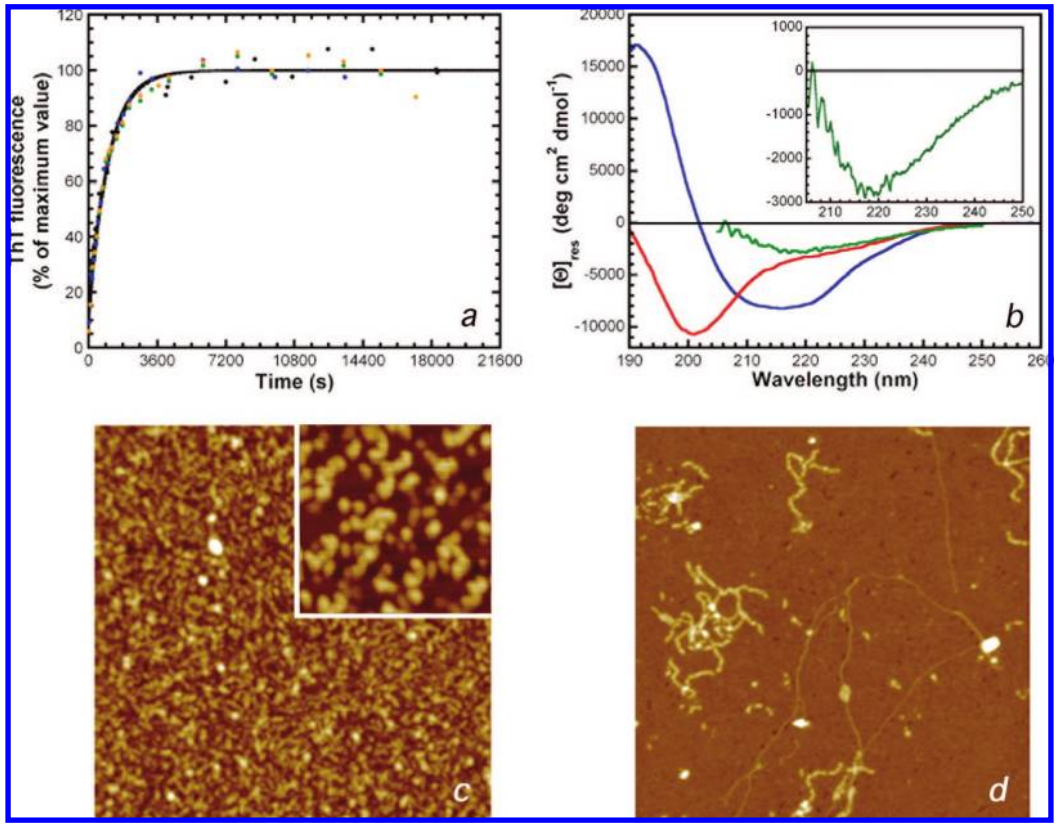


Figure 5. Aggregation behavior of HypF-N at acidic pH and high ionic strength. (a) Time courses of ThT fluorescence obtained with protein samples containing 48 μM HypF-N, pH 1.7, ionic strength 350 mM, 25 °C. The figure shows data points from four independent experiments (indicated by black, green, blue, and orange colors); the corresponding lines represent the best fits of the data to a single-exponential function (eq 3). All data points were blank-subtracted and normalized such that the 100% value corresponds to the maximum fluorescence observed at the apparent plateau. (b) Far-UV CD spectra of native HypF-N in 10 mM phosphate buffer, pH 7.0, 28 °C (blue), acid-denatured HypF-N in 10 M HCl, pH 2.0, 28 °C (red), aggregated HypF-N in 20 mM TFA, 2 mM DTT, and 330 mM NaCl (350 mM overall ionic strength), pH 1.7, 25 °C, after 4 h (green). The inset shows the latter spectrum on an expanded *y* axis. (c and d) TM-AFM images (height data) from HypF-N samples after 4 h (c) or 56 days (d) of incubation at a protein concentration of 48 μM in 20 mM TFA, 2 mM DTT, and 330 mM NaCl (350 mM overall ionic strength), pH 1.7, 25 °C. Scan size 1.3 μm; Z range 10 nm. The inset in (c) shows a detail of the main panel at higher magnification (scan size 700 nm; Z range 5 nm).

we conclude that acid-denatured HypF-N does not contain this type of secondary structure.

The majority of residues whose resonances experience severe line broadening effects, notably residues 27−34 and 56−62, is located in stretches of the sequence that show a very significant degree of helicity according to the $^{13}C\alpha$ chemical shift analysis. It is interesting to note, however, that the stretch of sequence in which non-native helical structure has been found does not have severely broadened resonances (residues 46−49). In analogy, one of the regions with severely broadened resonances (residues 81−82) does not present any significant secondary structure. These two observations suggest that the partial formation of helical structure and the establishment of hydrophobic interactions are not necessarily coupled but are on the contrary primarily determined by the respective intrinsic propensities to form either type of structure. Overall, the NMR analysis confirmed the partially folded nature of acid-denatured HypF-N and allowed the regions participating to the formation of hydrophobic clusters and α-helical structure to be identified accurately.

At Low pH and High Ionic Strength, HypF-N Aggregates into Amyloid-like Protofibrils. Although HypF-N is relatively stable in its monomeric state at low pH, increasing the ionic strength of the solution by addition of NaCl was found to induce rapid aggregation of the protein.[18] In particular, after incubation for about 2 h at low pH under high ionic strength conditions (350 mM) the protein sample induces a 4.5-fold increase in ThT fluorescence (Figure 5a). Such a fluorescence increase follows a highly reproducible, single exponential time course, with an observed rate constant (k_{wt}) of 1.0 (±0.2) 10^{-3} s^{-1} under these conditions (Figure 5a). The far-UV CD spectrum of the protein sample after a 4-h incubation has a single broad minimum centered at ca. 217−219 nm and crosses the baseline ($[\theta] = 0$) at about 205 nm (Figure 5b, green spectrum). Both features are typical of a highly β-sheet-rich protein. The mean residue ellipticity of this spectrum is low in absolute value because of a phenomenon termed differential absorption flattening.[38] This phenomenon can be observed in the CD spectra of samples containing suspensions of solid-phase material and is known to cause both a decrease in intensity and red-shift of all CD bands.[38] In contrast to the CD spectrum acquired for this sample, the spectrum acquired at low pH and low ionic strength is characteristic of a largely, although not completely, unstructured species (Figure 5b, red spectrum).

The protein sample at low pH and high ionic strength was then analyzed using TM-AFM. After incubation for 4 h, a large number of approximately spherical beadlike aggregates were found to be present, along with very short protofibrils that appear to result from the assembly of a variable number of beads

(38) Castiglioni, E.; Abbate, S.; Longhi, G.; Gangemi, R.; Lauceri, R.; Purrello, R. *Chirality* **2007**, *19*, 642–646.

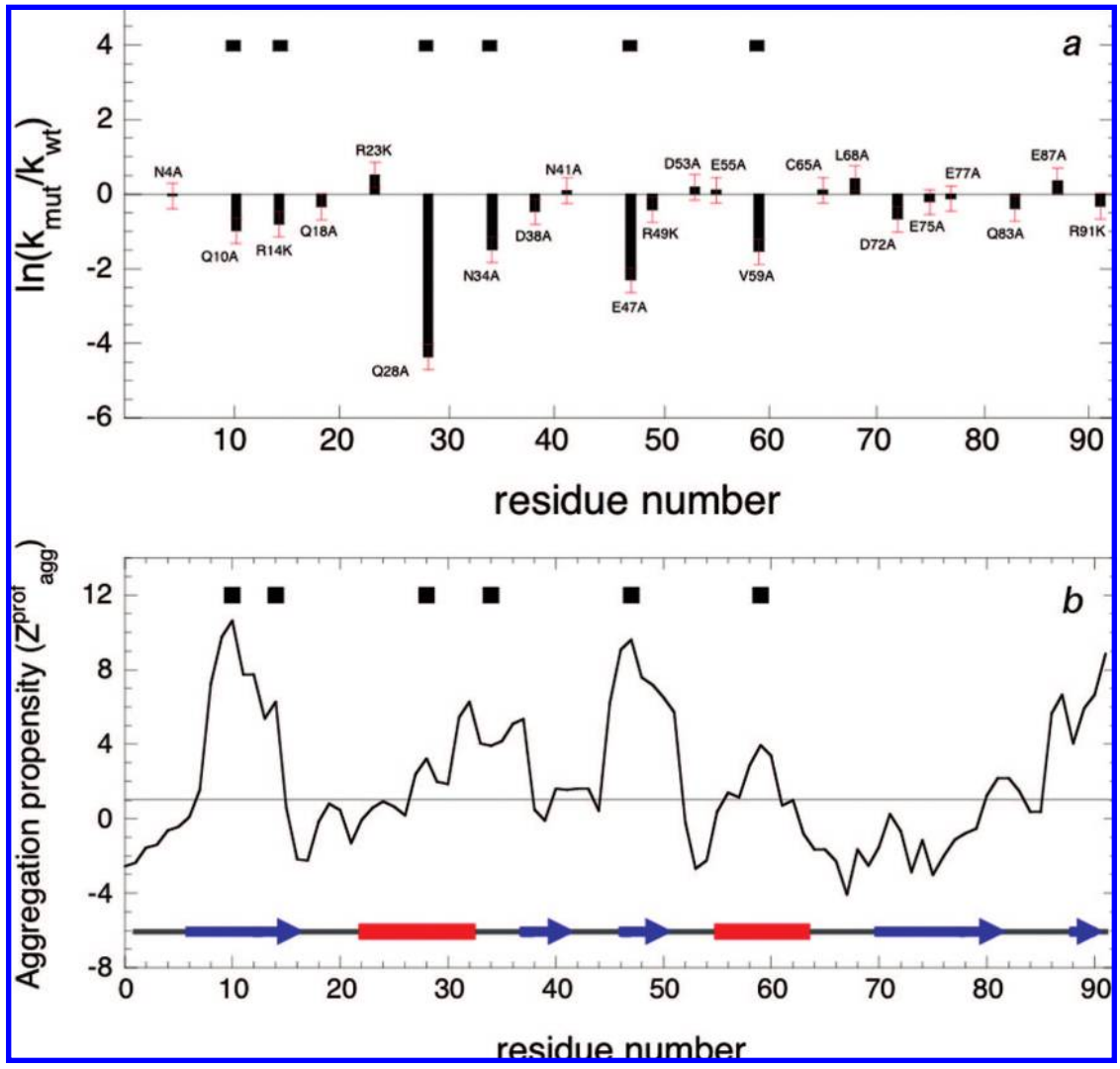


Figure 6. (a) Effect of single-point mutations on the rate of conversion of HypF-N from the acid-denatured state to β-structured and ThT binding oligomers (spherical protofibrils). Each bar refers to a single-point mutant having a mutation at the position indicated on the *x*-axis; error bars are given as standard deviations (SD). The mutants with aggregation rates that differ from that of the wild-type protein (i.e., with $\ln(k_{mut}/k_{wt})$ more than 2 SD from zero) are reported as ■. (b) Intrinsic aggregation propensity profile (Z^{prof}_{agg}) of acid-denatured HypF-N calculated at pH 2.0 as described previously.[34] The horizontal line at $Z^{prof}_{agg} = 1$ indicates a threshold of high aggregation propensity.[34] The mutations found to perturb the rate of protofibril formation are reported (■). The sequence positions of α-helices (red) and β-strands (blue) in the native structure are also indicated as determined by MOLMOL[35] (PDB entry 1GXU).

(Figure 5c). The longest assemblies usually display a tendency toward a curvilinear rather than straight conformation and often exhibit kinks. The heights of both types of aggregates were determined from cross sections of the AFM images to be 3.0 ± 0.2 nm after correction for the shrinking effect resulting from sample drying (see Methods). After 56 days of incubation under the same conditions, these aggregates converted into longer protofibrillar structures (Figure 5d), many of which appear to be curvilinear, or wormlike chains of beads, with a height of 2.6 ± 0.2 nm (Figure 5d). Long, thin filaments with a height of 0.9 ± 0.1 nm were also evident (Figure 5d). Both the globular and curvilinear species observed here for HypF-N were highly reminiscent of spherical and chainlike amyloid protofibrils observed with disease-associated systems, including the Aβ peptide, α-synuclein, transthyretin, and islet amyloid polypeptide.[39−43] Overall, the ThT fluorescence, far-UV CD, and TM-AFM analyses indicate that incubation of HypF-N under these conditions leads to the formation of amyloid-like protofibrils that bind ThT and have a largely β-sheet secondary structure.

Aggregation-Promoting Regions Coincide with Segments with Intrinsically High Aggregation Propensity. To gain insight into which regions of the sequence of HypF-N promote formation of protofibrillar aggregates, 22 single mutants of HypF-N were produced and their aggregation rates measured under conditions of low pH and high ionic strength. The mutations, which are distributed throughout the sequence on average every four residues, were designed to maintain the net

(39) Harper, J. D.; Wong, S. S.; Lieber, C. M.; Lansbury, P. T. *Chem. Biol.* **1997**, *4*, 119–125.
(40) Walsh, D. M.; Lomakin, A.; Benedek, G. B.; Condron, M. M.; Teplow, D. B. *J. Biol. Chem.* **1997**, *272*, 22364–22372.
(41) Conway, K. A.; Harper, J. D.; Lansbury, P. T., Jr *Biochemistry* **2000**, *39*, 2552–2563.
(42) Quintas, A.; Vaz, D. C.; Cardoso, I.; Saraiva, M. J.; Brito, R. M. *J. Biol. Chem.* **2001**, *276*, 27207–27213.
(43) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. *Science* **2003**, *300*, 486–489.

charge of the protein unchanged. Indeed, single or multiple mutations that decrease or increase the net charge of the protein are expected to alter the aggregation rate to an extent proportional to the change of charge, regardless of the site of the mutations.[44] The aggregation time course of each mutant protein was monitored using the ThT assay. The results for all 22 mutants are summarized in Figure 6a. Some of the mutants were found to aggregate with rates similar to that of the wild-type protein [$\ln(k_{mut}/k_{wt}) \approx 0$ within experimental error]. By contrast, other mutants aggregate with rates significantly different from that of the wild-type protein [$\ln(k_{mut}/ k_{wt}) \neq 0$], indicating that the regions of the sequence in which the mutated residues are located play an important role in aggregation.

The positions of the mutations that affect the aggregation rate were examined in the light of the intrinsic aggregation propensity profile of HypF-N (Figure 6b). This profile was generated using the approach described previously that takes into account the hydrophobicity, α-helical and β-sheet propensity, and charge of the various residues along the sequence.[34] Importantly, the hydrophobicity, charge, and secondary structure propensities of a given residue are estimated independently of whether that residue forms hydrophobic clusters, α-helical or β-sheet structure, or salt bridges. The algorithm therefore defines, for a given residue, a value of aggregation propensity that is intrinsic and independent of the structure present initially in the pH-denatured state.[34] According to the generated profile, the regions of the sequence with a value of $Z^{prof} > 1$ are considered "peaks", that is, regions with a high intrinsic aggregation propensity.[34]

The mutations that affect the aggregation rate are located at the peaks in the predicted profile (Figure 6b). The Q10A and R14K mutations are found within the first peak in the profile, the Q28A and N34A mutants colocalize with the second peak, whereas the E47A and V59A mutations coincide with the third and fourth peaks in the profile, respectively (Figure 6b). The only variation to this general trend is that mutations located within the peak at the C-terminus have little or no effect on the aggregation rate of the protein. A possible explanation for this observation is that this region of the sequence is separated by ca. 25 residues from the nearest of the other four regions of high aggregation propensity; such large separations were predicted to reduce significantly the probability of a given region to contribute to the rate-determining step in the aggregation reaction.[45]

Overall, a correlation between mutation-sensitive regions and regions with predicted high aggregation propensity is observed. On the contrary, there is no detectable correlation between mutation-sensitive regions and the degree of structure of the precursor state. The first region, which includes Gln10 and Arg14, corresponds to a segment of the sequence that experiments indicate is one of the least structured regions of the protein (compare Figures 3e and 6b). By contrast, the second region including Gln28 and Asn34, and the fourth region including Val59, overlap with two of the most highly structured regions in the low pH ensemble (compare Figures 3e and 6b). The third region with high aggregation propensity, which includes Glu47, has a level of structure intermediate between these two cases. It is also evident that the most highly structured regions identified by the NMR experiments do not always coincide with the mutation-sensitive segments identified through our protein engineering and computational analysis. Nor do the most unstructured regions overlap with such segments.

The increased ionic strength is expected to enhance the stability of the hydrophobic clusters formed by the most structured regions.[18,46] This rules out the possibility that regions appearing structured at low ionic strength may become irrelevant in the aggregation process at high ionic strength because of an increase in their dynamical character. We can therefore conclude that the most unfolded regions of the sequence do not appear to play a key role under the conditions investigated here. On the contrary, regions with an inherent propensity to aggregate promote the aggregation process independently of the extent and stability of the nonpersistent structure present in the monomeric precursor state.

These findings are not in any way inconsistent with the well-established notion that aggregation requires local or global unfolding of the native state of a globular protein.[1−3,47] Indeed, the cooperative and persistent nature of the structure present in globular proteins inhibits aggregation substantially and prevents segments of the protein with an inherently high aggregation propensity from promoting aggregation. The structure present in the pH-denatured state of HypF-N is not cooperatively folded and is not persistent. It is, on the contrary, inherently dynamic and fluctuates between structurally dissimilar conformations. Within such an ensemble, the differences in degree of structure across the sequence appear not to dominate the aggregation process, as all are sufficiently flexible and unstable to initiate aggregation. The intrinsic aggregation propensity is thus the dominant factor in determining which regions of the sequence of HypF-N promote aggregation.

Precursor State Structure and Aggregation Mechanisms Are Decoupled: A Comparison with Other Systems. The data presented thus far on acid-denatured HypF-N indicate that while hydrophobicity and intrinsic propensity to form α-helical structure appear to determine the presence and location of, respectively, hydrophobic clusters and α-helices in the initial monomeric state, the identity of the regions playing a key role in aggregation is determined mainly by the intrinsic aggregation propensity. The question arises as to whether these main conclusions are also applicable to other systems associated with protein deposition diseases. Similarly to acid-denatured HypF-N, aggregation of largely unstructured peptides and intrinsically disordered proteins, such as the Aβ peptide and α-synuclein, appears to be promoted by regions of the sequence with high intrinsic aggregation propensities[34,48] independently of the degree of structure present in various portions of the monomeric species. Unlike Aβ and α-synuclein, which are intrinsically disordered proteins and thus unfolded at neutral pH, HypF-N loses its compact fold only at low pH or in other denaturing media. For this reason, a proper comparison between intrinsically disordered systems and proteins that are by contrast folded at neutral pH can only be made at different pH values.

In α-synuclein, the residues located within the core of the resulting fibrils are all within a long stretch of sequence encompassing approximately residues 30−100.[49−52] Within this

(44) Chiti, F.; Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16419–16426.
(45) Hall, D.; Hirota, N.; Dobson, C. M. *J. Mol. Biol.* **2005**, *351*, 195–205.
(46) Raman, B.; Chatani, E.; Kihara, M.; Ban, T.; Sakay, M.; Hasegawa, K.; Naiki, H.; Rao, C. M.; Goto, Y. *Biochemistry* **2005**, *44*, 1288–1299.
(47) Kelly, J. W. *Curr. Opin. Struct. Biol.* **1998**, *8*, 101–106.
(48) Bemporad, F.; Calloni, G.; Campioni, S.; Plakoutsi, G.; Taddei, N.; Chiti, F. *Acc. Chem. Res.* **2006**, *39*, 620–627.
(49) Miake, H.; Mizusawa, H.; Iwatsubo, T.; Hasegawa, M. *J. Biol. Chem.* **2002**, *277*, 19213–19219.

region, segments adopting β-sheet conformation in the fibrils were identified[52] and found to overlap with those regions that have a high aggregation propensity.[48] In monomeric α-synuclein, such regions were found from NMR studies to form significant interactions with the C-terminal portion of the protein.[26,53] However, they still promote aggregation because the lack of persistence of these interactions allows them to play a key role in the formation of the cross-β core of the resulting fibrils. By contrast, the C-terminal region involved in the same interactions is not involved in the cross-β structural core of the fibril, thus showing that the ability to form transient interactions may or may not be associated with a subsequent role in the aggregation process.

A similar scenario can be described for the Aβ peptide. The structures of monomeric Aβ_{1-40} and Aβ_{1-42} were investigated using solution NMR and limited proteolysis.[54,55] Using these techniques, we found that residues 7−11 and 20−26 (or 21−30) formed structured turn- or bendlike motifs. Chemical shift analysis also suggests that nascent α-helical or β-strand conformations may be present in other regions.[54] Such putative secondary structure is not, however, persistent as clearly indicated by very rapid hydrogen/deuterium exchange rates and the lack of medium- and long-range (more than four residue apart) NOEs.[54] Structural investigation of the fibrils formed by Aβ_{1-40} and Aβ_{1-42} using solid-state NMR has led to the identification of two regions of the sequence, spanning approximately residues 12−24 and 30−40, as the strands forming the β-sheet core of the fibrils.[56−59] These regions appear to adopt an unfolded conformation in the monomer, as do residues 1−6, which do not participate in the formation of the fibril β-sheet core. By contrast, regions 12−24 and 30−40 correspond well to the two regions of the sequence with the highest intrinsic aggregation propensity in Aβ.[34,48]

Conclusions

At low pH, HypF-N is denatured but presents a significant degree of nonpersistent, fluctuating structure. Using NMR spectroscopy, we identified the regions of the sequence that form hydrophobic clusters (residues 23−34, 56−64, and 81−82) and α-helical structure (residues 26−30, 56−61, and 46−49). These correspond, to a very good approximation, to the regions of the sequence that have the highest hydrophobicity and propensity to form α-helical structure. Using a protein engineering approach, we also described the regions of the sequence that play key roles in the conversion of the pH-denatured state of HypF-N into ThT-binding and β-sheet containing protofibrillar species (approximately residues 9−15, 27−35, 46−48, and 58−60). These groups of residues correspond to the regions of the sequence of HypF-N that have the highest intrinsic aggregation propensity and do not coincide with either the most structured or the most unfolded regions of the precursor state.

These results, in the context of those reported for several intrinsically disordered systems, provide support to the view that the aggregation process appears to be only marginally related to the structure and dynamics of the monomeric precursor state, since all regions of the protein are sufficiently flexible and unstable to initiate aggregation. Therefore, segments of a polypeptide chain with an intrinsically high aggregation propensity will prompt aggregation independently of the extent and stability of the nonpersistent structure present in the initial monomeric state, on the condition that such segments are either fully unfolded or, if partially folded, that the residual structure does not form cooperatively and is not persistent. By contrast, the stability and cooperativity of the native or quasi-native states of globular proteins protect against aggregation because aggregation-prone regions are effectively shielded by their burial in the protein interior. In this case, the small amplitude of the fluctuations within the native structure prevents the formation of specific intermolecular interactions that could otherwise initiate the aggregation process. These findings indicate that, given the complexity of the aggregation process, strategies solely based on the characterization of the partially unfolded states populated before aggregation are unlikely to provide a detailed description of the mechanism of aggregation.

Acknowledgment. We thank Silvia Torrassa for help with the AFM measurements. This work was partially supported by grants from the Italian MIUR (PRIN 2005053998, FIRB RBNE03PX83), Fondazione CARIGE, the Leverhulme Trust, the Wellcome Trust, the EMBO Young Investigator Programme, the Swedish Research Council, and the Royal Society.

Supporting Information Available: Experimental procedures, dynamic light scattering data, and PICUP data. This material is available free of charge via the Internet at http://pubs.acs.org.

JA8029224

(50) Der-Sarkissian, A.; Jao, C. C.; Chen, J.; Langen, R. *J. Biol. Chem.* **2003**, *278*, 37530–37535.
(51) Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L., Jr *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15477–15482.
(52) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15871–15876.
(53) Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430–1435.
(54) Hou, L.; Shao, H.; Zhang, Y.; Li, H.; Menon, N. K.; Neuhaus, E. B.; Brewer, J. M.; Byeon, I. J.; Ray, D. G.; Vitek, M. P.; Iwashita, T.; Makula, R. A.; Przybyla, A. B.; Zagorski, M. G. *J. Am. Chem. Soc.* **2004**, *126*, 1992–2005.
(55) Lazo, N. D.; Grant, M. A.; Condron, M. C.; Rigby, A. C.; Teplow, D. B. *Protein Sci.* **2005**, *14*, 1581–1596.
(56) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzukin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16742–16747.
(57) Torok, M.; Milton, S.; Kayed, R.; Wu, P.; McIntire, T.; Glabe, C. G.; Langen, R. *J. Biol. Chem.* **2002**, *277*, 40810–40815.
(58) Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17342–17347.
(59) Petkova, A. T.; Yau, W. M.; Tycko, R. *Biochemistry* **2006**, *45*, 498–512.