



Revealing Well-Defined Soluble States during Amyloid Fibril Formation by Multilinear Analysis of NMR Diffusion Data

Kristine Steen Jensen,^{*,†} Sara Linse,[‡] Mathias Nilsson,[§] Mikael Akke,^{*,†} and Anders Malmendal^{*,‡,⊥}

[†]Biophysical Chemistry, Center for Molecular Protein Science, Department of Chemistry, Lund University, P.O. Box 124, SE-22100 Lund, Sweden

[‡]Biochemistry and Structural Biology, Center for Molecular Protein Science, Department of Chemistry, Lund University, P.O. Box 124, SE-22100 Lund, Sweden

[§]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, U.K.

[⊥]Department of Science and Environment, Roskilde University, P.O. Box 260, DK-4000 Roskilde, Denmark

Supporting Information

ABSTRACT: Amyloid fibril formation is a hallmark of neurodegenerative disease caused by protein aggregation. Oligomeric protein states that arise during the process of fibril formation often coexist with mature fibrils and are known to cause cell death in disease model systems. Progress in this field depends critically on development of analytical methods that can provide information about the mechanisms and species involved in oligomerization and fibril formation. Here, we demonstrate how the powerful combination of diffusion NMR and multilinear data analysis can efficiently disentangle the number of involved species, their kinetic rates of formation or disappearance, spectral contributions, and diffusion coefficients, even without prior knowledge of the time evolution of the process or chemical shift assignments of the various species. Using this method we identify oligomeric species that form transiently during aggregation of human superoxide dismutase 1 (SOD1), which is known to form misfolded aggregates in patients with amyotrophic lateral sclerosis. Specifically, over a time course of 42 days, during which SOD1 fibrils form, we detect the disappearance of the native monomeric species, formation of a partially unfolded intermediate in the dimer to tetramer size range, subsequent formation of a distinct similarly sized species that dominates the final spectrum detected by solution NMR, and concomitant appearance of small peptide fragments.

To obtain insight into the behavior of biological macromolecules it is imperative to establish efficient procedures for exploring their kinetic and equilibrium behavior. A particularly challenging case is the oligomerization and fibril formation of proteins associated with neurodegenerative diseases and other protein aggregation-related disorders. Oligomeric states appear during fibril formation and have been identified as cytotoxic components in many diseases.^{1,2} Oligomers are often present only transiently and in low amounts,^{3,4} thereby making their detection particularly challenging.

Nuclear magnetic resonance (NMR) spectroscopy can provide detailed chemical information from individual signals for virtually every atom in a molecule and, hence, makes it possible to follow the evolution of molecular species. For simple mixtures, the use of DOSY (diffusion-ordered spectroscopy) can resolve the individual NMR signals by distinguishing different species through their different rates of translational diffusion.^{5,6} By contrast, complex samples containing interconverting species pose a formidable challenge to the interpretation of NMR data by conventional means. However, the additional complication posed by an evolving mixture can be an advantage, because the time and diffusion dependences can be encoded as independent dimensions enabling the use of multilinear data analysis methods, such as parallel factor analysis, PARAFAC.^{7–10}

PARAFAC is an extension of the two-dimensional principal component analysis (PCA) to higher order arrays. While PCA suffers from rotational ambiguity, resulting in fitted components that often are linear combinations of different processes and describe deviations from the mean rather than actual amplitudes, a successful PARAFAC decomposition directly provides the user with the variation of each species in a form representing meaningful amplitudes, resulting in a tremendous improvement in interpretability.

Here, we show that the combination of DOSY and PARAFAC^{5,6,11} provides powerful information on an elusive oligomerization process underlying formation of amyloid fibrils. We employed three-dimensional ¹H–¹⁵N-DOSY-HSQC to acquire a series of spectra during fibril formation by SOD1 (153 amino acid residues), a protein associated with the neurodegenerative disease amyotrophic lateral sclerosis. The present investigation involves three independent dimensions: the NMR resonance frequencies, the signal decays caused by diffusion, and time. PARAFAC translates these dimensions into the sought information: the number of (NMR detectable) species and their time evolution profiles, sizes, and spectra. This is achieved without any constraints or

Received: July 31, 2019

Published: November 8, 2019

presumptions of kinetic models, hydrodynamic properties, or spectral characteristics of the species.

We used the pseudo wild type SOD1 variant C6A/F50E/G51E/C57S/C111A/C146S (pwtSOD1^{ΔC}) that mimics the monomeric metal-free and disulfide-reduced form of SOD1,^{3,12,13} which is known to be a key precursor in the formation of aberrant oligomers and fibrils.^{14–17}

We performed ¹H–¹⁵N-DOSY-HSQC experiments based on the X-STE pulse sequence¹⁸ with a diffusion time of 500 ms and 10 different gradient strengths, yielding a total experimental time of 18.6 h. This experiment was repeated 10 times on the same sample over a period of 6 weeks, during which time the sample temperature was maintained at 37 °C, to cover the full fibril formation of pwtSOD1^{ΔC} (Supporting Information (SI) methods). Over this time period the ¹H–¹⁵N-HSQC spectrum of pwtSOD1^{ΔC} becomes crowded as peaks from partially unfolded species build up (Figure S1). For the present analysis we selected a total of 223 peaks, which may well have contributions from overlapping resonances.

The data were analyzed using PARAFAC to identify the minimum number of species (factors) needed to describe the variance of the data set, see SI methods for details. We computed PARAFAC models including up to 5 factors. Anomalous behavior was identified for three weak signals from histidines affected by exchange processes at this pH, and these were removed from further analysis. The analysis identified four factors based on various quality assessment tools, such as residuals and core consistency diagnostics¹⁹ (Table S1). Factor 4 comprises only three signals that we excluded to generate a separate set of PARAFAC models with improved descriptors for factors 1–3.

The resulting factors are characterized by unique combinations of time evolution, size, and spectral contribution (Figure 1) that make physical and chemical sense: the amplitudes of individual factors change smoothly with time (Figure 1A–D) and decrease exponentially with the square of the gradient strength, indicating that each factor represents a unique, monodisperse species (Figure 1E–H). Due to extensive overlap, individual peaks may represent more than one molecular species, but this is resolved by the PARAFAC decomposition. Most peaks include contributions from more than one factor (Figure 1I–L), indicating that species I, II, and III contain different levels of similarly folded and unfolded segments.

At the starting point of the reaction the dominant species (state I, factor 1) is the folded, monomeric apo form of pwtSOD1^{ΔC}, as identified by the chemical shifts.³ The diffusion coefficient, $D = (1.40 \pm 0.02) \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$, for this species corresponds to a Stokes' radius, $r_s = 23 \text{ \AA}$ (eq S2), which corresponds very well with the previously determined value of $r_s = 22.5 \text{ \AA}$ for monomeric apo disulfide-reduced SOD1.²⁰

Factor 2 describes an intermediate (state II) that builds up rapidly from the start, displays the same short lag phase as the monomer curve, peaks around day 8–15, and declines over days 15–42 (Figure 1B). The time evolution of factor 2 is similar to that of prefibrillar intermediates of A β 42 and transthyretin, associated with Alzheimer's disease and transthyretin amyloidosis, respectively.^{21,22} State II has a lower diffusion coefficient than state I, $D = (0.86 \pm 0.01) \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$, corresponding to $r_s = 38 \text{ \AA}$. If states I and II behave like globular species (despite the presence of disordered segments), the ratio of their r_s values suggests that state II is a tetramer.

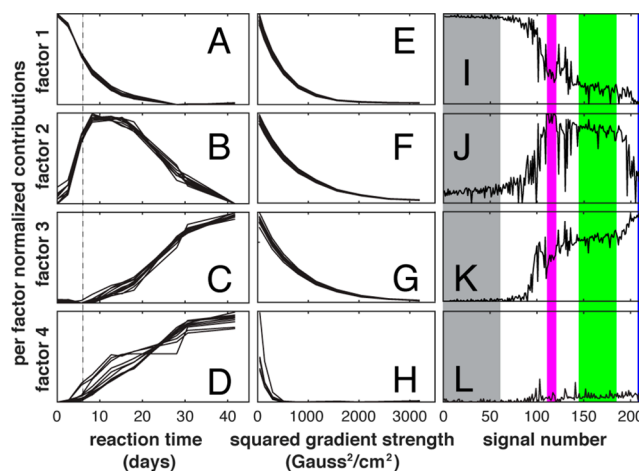


Figure 1. PARAFAC model of pwtSOD1^{ΔC} ¹⁵N-HSQC-DOSY experiments. (A–D) Factor contributions to the NMR spectrum as a function of reaction time. (E–H) Intensity profiles versus gradient strength for each factor. (I–L) Factor contributions to the signal intensity of each peak number (1–223); the contributions were scaled individually for each factor. The contributions to each factor cannot be directly interpreted as the relative population of each species, because differences in relaxation between residues and states affect the signal intensities, but they do show how each factor varies with time and signal number. Factor 1 (A, E, I) describes a disappearing monomeric state (I), factor 2 (B, F, J) an intermediate state (II), and factor 3 (C, G, K) a state (III) that builds up during the process. Factor 4 (D, H, L) describes a state (IV) that involves only three peaks. The dashed line in panel A–D indicates the time point at which the population of state I is reduced to 50% of its starting value. The signals in (I–L) are ordered such that the contribution from factor 1 decreases with increasing signal number, and the contribution from factor 3 increases. The colored areas indicate signals dominated by factor 1 in gray, factor 2 in pink, factors 2 + 3 in green, factor 3 in blue, and factor 4 in red. The individual curves in panels A–H represent 10 models created using 50% randomly chosen signals.

For comparison, the native Cu,Zn-SOD1 dimer has $r_s = 30.3 \text{ \AA}$,²⁰ further supporting the conclusion that state II is tri- or tetrameric (SI). A trimeric SOD1 species with partially unfolded regions has previously been implicated.²³

Factor 3 describes the final visible state III, which has essentially the same diffusion coefficient as state II, $D = (0.89 \pm 0.01) \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$, corresponding to $r_s = 37 \text{ \AA}$. The similar diffusion coefficients and the large number of peaks with similar contributions from factors 2 and 3 (green in Figure 1J,K) imply that states II and III have similar structures.

Factor 4 describes state IV, which has a much higher diffusion coefficient than the other three states, $D = (5.90 \pm 0.2) \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$, yielding $r_s \approx 5 \text{ \AA}$. State IV involves only three signals and appears to build up simultaneously with state III (Figure 1C,D), suggesting that the two events are linked; most likely, state IV represents a short peptide that is cleaved off from the protein.

Variations in intensity and diffusion rate for the different types of signals are shown in Figure 2. Note that the data in Figure 2A describe individual peaks, many of which contain contributions from more than one PARAFAC factor. The evolution profiles and diffusion coefficients of the three main states suggest the following linear reaction model: state I (monomer) → state II (partly unfolded oligomer) → state III (partly unfolded and cleaved oligomer) + state IV (peptide). The observed buildup of state III to a plateau value (Figure

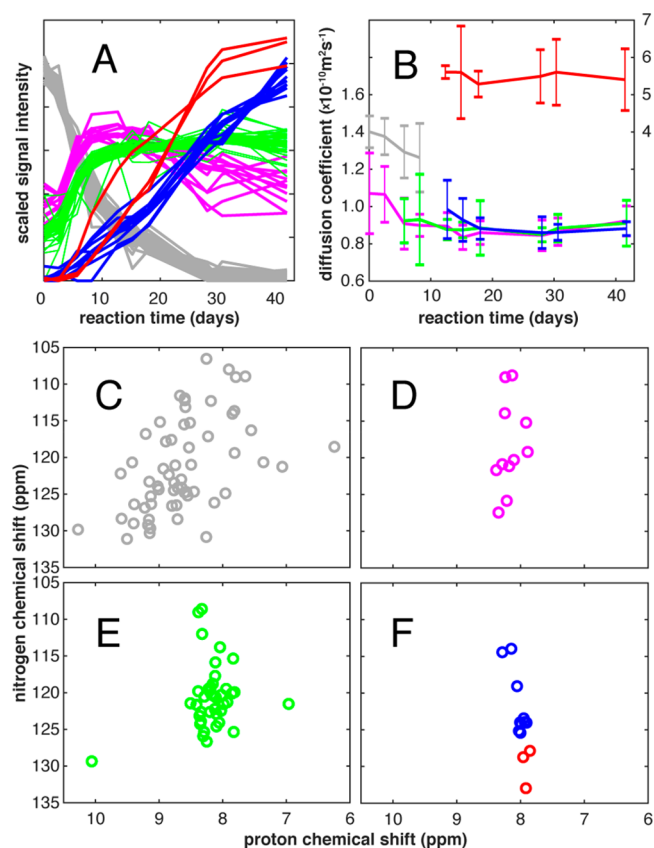


Figure 2. Intensity variation with time (A), average diffusion rates (B), and peak positions (C–F) of signals dominated by factor 1 (gray; $n = 60$), factor 2 (magenta; $n = 11$), factor 2 and 3 (green; $n = 40$), factor 3 (blue; $n = 10$), and factor 4 (red; $n = 3$). The signals are marked in panels I–L of Figure 1. Intensities in (A) are scaled to unit variance for each residue in the same way as for the PARAFAC analysis. In panel (B), the left-hand vertical axis reports the diffusion rate for factors 1–3, while the right-hand axis reports on factor 4. Error bars correspond to one standard deviation.

1C) could result from an equilibrium with fibrils, reflect depletion of monomeric pwtSOD1^{ΔC}, which might be necessary for the turnover of state III (I + III → fibrils), or indicate that state III is a dead end product of proteolysis. Note that the end-stage fibrils are too large to be detected and studied by liquid state NMR, although flexible parts of fibrils could potentially be observable and mistakenly identified as unbound objects.²⁴ However, the diffusion coefficients of state II and III are not low enough to correspond to flexible parts of a large fibrillar state.

The ¹H chemical shifts of the resonances contributing to factor 2 or 3 cluster around 7.0–8.5 ppm, as typical for disordered protein segments. The absence of signals with chemical shifts typical of folded proteins is expected, because signals from structured parts of large oligomers are often broadened beyond detection. We note that an alternative interpretation of factors 2 and 3 could be that states II and III are fully unfolded monomers (which are also expected to have increased r_s), rather than partly unfolded oligomers. However, a number of observations argue against this interpretation. First, the unfolding/folding equilibrium is shifted toward the folded state under the present experimental conditions and the unfolding/folding dynamics are much faster than the observed buildup of state II.²⁵ Second, we validated our interpretation of

the results by small-angle X-ray scattering, which unequivocally shows that pwtSOD1^{ΔC} is monomeric at day 0, but polydisperse with larger oligomeric species by day 19 (SI Figure S2). Both samples are dominated by globular species with disordered regions, but the disordered contribution is only slightly increased at the later time point (SI Figure S2B), implying that states II and III are far from fully unfolded. The buildup of oligomers was further confirmed by chemical cross-linking of lysine residues, followed by MALDI TOF/TOF mass spectrometry of cross-linked, intact pwtSOD1^{ΔC} samples obtained at day 0 and day 19 (SI Figure S3). We also used cryo-transmission electron microscopy to verify that fibrils formed during the course of the 6-week experiment (SI Figure S4). It should be noted that our results do not exclude the possibility that the formation of the oligomeric state II takes place via the unfolded monomer, even though this state is weakly populated and the equilibrium unfolding dynamics is not rate limiting. Future experiments will reveal whether the identified states are on-path precursors to fibril formation.

We have shown that ¹H–¹⁵N-DOSY-HSQC spectroscopy in combination with PARAFAC analysis allows straightforward identification of four NMR-detectable species of pwtSOD1^{ΔC} present during fibril formation. PARAFAC yields the time evolution profiles and diffusion coefficients of these species, and identifies their chemical shifts. Thus, the approach enables powerful characterization of molecular processes that occur alongside fibril formation, or other macromolecular transitions occurring on similar time scales.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b07952.

Experimental procedures and supporting figures with additional data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: kristine.steen_jensen@bpc.lu.se (K.S.J.).

*E-mail: mikael.akke@bpc.lu.se (M.A.).

*E-mail: amalm@ruc.dk (A.M.).

ORCID

Sara Linse: 0000-0001-9629-7109

Mathias Nilsson: 0000-0003-3301-7899

Mikael Akke: 0000-0002-2395-825X

Anders Malmendal: 0000-0002-8413-9717

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Ulrich Weininger for help with HSQC-based DOSY experiments; Annette Langkilde, Saskia Bucciarelli, Bente Vestergaard, and the P12 EMBL beamline staff (DESY, Hamburg) for assistance with SAXS experiments; Katja Bernfur for assistance with mass-spectrometry; and D. Flemming Hansen for inspiring discussions. This work was supported by the Swedish Research Council (2014-5815 to M.A., 2015-00143 to S.L.), Danish Council for Independent Research, Sapere Aude: DFF Research Talent (DFF – 4002-00258 to K.S.J.), the European Research Council (340890 to

S.L.), and the Engineering Physical Sciences Research Council (EP/E05899X/1 to M.N.).

REFERENCES

- (1) Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, M. J.; Selkoe, D. J. Naturally Secreted Oligomers of Amyloid β Protein Potently Inhibit Hippocampal Long-Term Potentiation in Vivo. *Nature* **2002**, *416*, 535–539.
- (2) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. Inherent Toxicity of Aggregates Implies a Common Mechanism for Protein Misfolding Diseases. *Nature* **2002**, *416*, 507–511.
- (3) Teilum, K.; Smith, M. H.; Schulz, E.; Christensen, L. C.; Solomentsev, G.; Oliveberg, M.; Akke, M. Transient Structural Distortion of Metal-Free Cu/Zn Superoxide Dismutase Triggers Aberrant Oligomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18273–18278.
- (4) Sekhar, A.; Rumfeldt, J. A. O.; Broom, H. R.; Doyle, C. M.; Bouvignies, G.; Meiering, E. M.; Kay, L. E. Thermal Fluctuations of Immature SOD1 Lead to Separate Folding and Misfolding Pathways. *eLife* **2015**, *4*, No. e07296.
- (5) Khajeh, M.; Botana, A.; Bernstein, M. A.; Nilsson, M.; Morris, G. A. Reaction Kinetics Studied Using Diffusion-Ordered Spectroscopy and Multiway Chemometrics. *Anal. Chem.* **2010**, *82* (5), 2102–2108.
- (6) Nilsson, M.; Khajeh, M.; Botana, A.; Bernstein, M. A.; Morris, G. A. Diffusion NMR and Trilinear Analysis in the Study of Reaction Kinetics. *Chem. Commun.* **2009**, No. 10, 1252–1254.
- (7) Bro, R. PARAFAC. Tutorial and Applications. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149–171.
- (8) Dal Poggetto, G.; Castañar, L.; Adams, R. W.; Morris, G. A.; Nilsson, M. Relaxation-Encoded NMR Experiments for Mixture Analysis: REST and Beer. *Chem. Commun.* **2017**, *53*, 7461–7464.
- (9) Colbourne, A. A.; Meier, S.; Morris, G. A.; Nilsson, M. Unmixing the NMR Spectra of Similar Species – Vive La Différence. *Chem. Commun.* **2013**, *49*, 10510–10512.
- (10) Colbourne, A. A.; Morris, G. A.; Nilsson, M. Local Covariance Order Diffusion-Ordered Spectroscopy: A Powerful Tool for Mixture Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 7640–7643.
- (11) Björnerås, J.; Botana, A.; Morris, G. A.; Nilsson, M. Resolving Complex Mixtures: Trilinear Diffusion Data. *J. Biomol. NMR* **2014**, *58*, 251–257.
- (12) Bertini, I.; Piccioli, M.; Viezzoli, M. S.; Chiu, C. Y.; Mullenbach, G. T. A Spectroscopic Characterization of a Monomeric Analog of Copper, Zinc Superoxide-Dismutase. *Eur. Biophys. J.* **1994**, *23*, 167–176.
- (13) Khan, M. A. I.; Respondek, M.; Kjellström, S.; Deep, S.; Linse, S.; Akke, M. Cu/Zn Superoxide Dismutase Forms Amyloid Fibrils under Near-Physiological Quiescent Conditions: The Roles of Disulfide Bonds and Effects of Denaturant. *ACS Chem. Neurosci.* **2017**, *8*, 2019–2026.
- (14) Lindberg, M. J.; Byström, R.; Boknäs, N.; Andersen, P. M.; Oliveberg, M. Systematically Perturbed Folding Patterns of Amyotrophic Lateral Sclerosis (ALS)-Associated SOD1 Mutants. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 9754–9759.
- (15) Furukawa, Y.; Kaneko, K.; Yamanaka, K.; O'Halloran, T. V.; Nukina, N. Complete Loss of Post-Translational Modifications Triggers Fibrillar Aggregation of SOD1 in the Familial Form of Amyotrophic Lateral Sclerosis. *J. Biol. Chem.* **2008**, *283*, 24167–24176.
- (16) Rakhit, R.; Crow, J. P.; Lepock, J. R.; Kondejewski, L. H.; Cashman, N. R.; Chakrabarty, A. Monomeric Cu,Zn-Superoxide Dismutase Is a Common Misfolding Intermediate in the Oxidation Models of Sporadic and Familial Amyotrophic Lateral Sclerosis. *J. Biol. Chem.* **2004**, *279*, 15499–15504.
- (17) Khare, S. D.; Caplow, M.; Dokholyan, N. V. The Rate and Equilibrium Constants for a Multistep Reaction Sequence for the Aggregation of Superoxide Dismutase in Amyotrophic Lateral Sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15094–15099.
- (18) Ferrage, F.; Zoonens, M.; Warschawski, D. E.; Popot, J. L.; Bodenhausen, G. Slow Diffusion of Macromolecular Assemblies by a New Pulsed Field Gradient NMR Method. *J. Am. Chem. Soc.* **2003**, *125*, 2541–2545.
- (19) Engelsen, S. B.; Bro, R. PowerSlicing. *J. Magn. Reson.* **2003**, *163*, 192–197.
- (20) Arnesano, F.; Banci, L.; Bertini, I.; Martinelli, M.; Furukawa, Y.; O'Halloran, T. V. The Unusually Stable Quaternary Structure of Human Cu,Zn-Superoxide Dismutase 1 Is Controlled by Both Metal Occupancy and Disulfide Status. *J. Biol. Chem.* **2004**, *279*, 47998–48003.
- (21) Michaels, T. C. T.; Saric, S.; Lazell, H. W.; Curk, S.; Arosio, P.; Meisl, G.; Dear, A. J.; Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Linse, S.; Knowles, T. J. P. Kinetic Characterization of Oligomer Populations during the Aggregation of Alzheimer's A β 42 Peptide. **2019**, Submitted.
- (22) Sun, X.; Dyson, H. J.; Wright, P. E. Kinetic Analysis of the Multistep Aggregation Pathway of Human Transthyretin. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E6201–E6208.
- (23) Proctor, E. A.; Fee, L.; Tao, Y.; Redler, R. L.; Fay, J. M.; Zhang, Y.; Lv, Z.; Mercer, I. P.; Deshmukh, M.; Lyubchenko, Y. L.; Dokholyan, N. V. Nonnative SOD1 Trimer Is Toxic to Motor Neurons in a Model of Amyotrophic Lateral Sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 614–619.
- (24) Baldwin, A. J.; Anthony-Cahill, S. J.; Knowles, T. P. J.; Lippens, G.; Christodoulou, J.; Barker, P. D.; Dobson, C. M. Measurement of Amyloid Fibril Length Distributions by Inclusion of Rotational Motion in Solution NMR Diffusion Measurements. *Angew. Chem., Int. Ed.* **2008**, *47*, 3385–3387.
- (25) Lindberg, M. J.; Normark, J.; Holmgren, A.; Oliveberg, M. Folding of Human Superoxide Dismutase: Disulfide Reduction Prevents Dimerization and Produces Marginally Stable Monomers. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15893–15898.