

Supporting Information

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^{†*}, 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, UK

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

Methods

Protein expression and purification

A plasmid with cDNA of human SOD1 C6A/F50E/G51E/C57S/C111A/C146S (pwtSOD1^{ΔC}) in pET17b (synthesized by GenScript) was used for expression in *E. coli* BL21(DE3) pLysS cells.¹ Uniformly ¹⁵N-labeled pwtSOD1^{ΔC} was produced and purified essentially as described previously.² Apo pwtSOD1^{ΔC} was obtained by extensive dialysis against 0.12 M sodium acetate, 10 mM EDTA, pH 3.8, MWCO 6,000-8,000 (Spectrum Laboratories, Inc.), followed by extensive dialysis against 10 mM NaPO₄, pH 7.0 to refold the protein.

NMR sample preparation

To isolate monomeric apo pwtSOD1^{ΔC} a size-exclusion chromatography step was performed using a Superdex 75 analytical column (GE Healthcare) with 10 mM NaPO₄. The sample was concentrated through spin filters MWCO 10,000 (Millipore) directly before the NMR experiments. The final sample concentration was determined by absorbance at 280 nm using a NanoDrop 2000/2000c spectrophotometer (Thermo Scientific) using an extinction coefficient of 5500 M⁻¹ cm⁻¹. The NMR sample (600 μl) contained 1.42 mM pwtSOD1^{ΔC}, 10 mM NaPO₄ pH 7.0, 1 mM EDTA and 10 % D₂O.

SAXS sample preparation

Monomeric apo pwtSOD1^{ΔC} samples (freshly isolated by size-exclusion chromatography) were prepared by mixing ¹⁵N-enriched protein with an equal amount of unlabelled protein to yield 5 mg/ml of pwtSOD1^{ΔC} dissolved in 10 mM Na-PO₄, 1 mM EDTA, pH 7.0. Individual samples of 35 μl were stored in PCR tubes and incubated at 37°C. Two individual samples incubated for 0 days (i.e., freshly prepared) or 19 days were transported on ice to the beamline.

SAXS data collection

SAXS data were collected at the P12 beamline at the European Molecular Biology Laboratory on the Petra III storage ring of DESY, Hamburg. Bovine serum albumin (BSA) (2.5 mg/ml) was used as standard. Shortly before measurements each sample was centrifuged to concentrate it at the bottom of the reaction tube, and then gently mixed by pipetting up and down a few times. Samples were loaded by the automated loading system and scattering data

were collected at 37°C on a PILATUS detector with a q -range of $0.067 < q < 5.01 \text{ nm}^{-1}$ by acquiring 14 individual exposures of 0.045 s each. Data reduction was done automatically at the beamline and manually inspected afterwards. Buffer subtraction and normalization by concentration was done manually using the software packages Raw 1.6.0³ and AtSAS⁴.

Lysine cross-linking and mass spectrometry

Samples were prepared as described above for SAXS experiments. 4.8 μl aliquots were withdrawn from samples incubated for 0 and 19 days. To detect oligomers each sample was lysine cross-linked using bis(sulfosuccinimidyl) suberate (BS3), which has a linker length of 11 Å. Cross-linking was performed by reacting 50 μM pwtSOD1^{ΔC}, 3 mM BS3, 50 mM Na-PO₄, pH 8.0 for 15 min at room temperature. The reaction was quenched with 30 mM Tris pH 8.0 for 1 h at room temperature. Samples of cross-linked pwtSOD1^{ΔC} were stored at -20°C for later mass spectrometry analysis.

MS spectra were acquired using an Autoflex Speed MALDI TOF/TOF mass spectrometer (Bruker Daltonics) in linear detection mode. Samples were prepared on a MALDI stainless steel plate by adding 1 μl sample (150 μM pwtSOD1^{ΔC}, 5% formic acid) followed by 0.5 μl intact BSA standard (Bruker Daltonics) and 0.5 μl 5 mg/ml α -cyano-4-hydroxy cinnamic acid, 80% acetonitrile, 0.1% TFA. Spectra were internally calibrated using quadratic mode with the +1/+2/+3 charged m/z signals of intact BSA (m/z : 66431.0, 33216.0 and 22144.3). Note that the calibration is not linear and the mass-uncertainty is not constant over the full m/z range measured. With internal standards of intact proteins it is difficult to include additional calibration signals, as overlap with the pwtSOD1^{ΔC} signals will occur.

NMR experiments

NMR experiments were performed at 37°C without spinning on an Agilent VNMR5 DirectDrive spectrometer operating at field strength of 11.7 T and equipped with a triple resonance probe. ¹⁵N DOSY-HSQC 3D experiments were performed as described,⁵ using a diffusion time (Δ) of 500 ms, encoding/decoding time δ of 1 ms, and gradient strengths G_{max} of 5.66, 11.31, 16.97, 22.62, 28.28, 33.93, 39.56, 45.24, 50.90, 56.55 G/cm, resulting in a total experiment time of 18.6 hours. The applied gradient strengths were calibrated against water diffusion at 25°C. The ¹⁵N DOSY-HSQC experiment was performed on the same sample at time points of 0.06 (1.4 h), 2.52, 5.69, 8.21, 12.68, 5.20, 18.01, 28.05, 30.63, and 41.74 days. The experiment was conducted with array precedence order of gradient strength over phase over number of increments (total $n_i = 256$). In this way the time evolution of the sample during the total experimental time (18.6 h) is averaged across all ¹⁵N-DOSY-HSQC experiments.

Data analysis

Spectra were processed in NMRPipe⁶ and analysed in CcpNMR Analysis.⁷ Diffusion coefficients were calculated using equation (S1):⁸

$$\frac{S}{S_0} = e^{(-D \cdot \Delta (2 \cdot \gamma \cdot s \cdot G_{\text{max}} \delta)^2)} \quad (\text{equation S1})$$

where S is the diffusion attenuated signal, S_0 is the reference signal obtained at the weakest gradient amplitude, D is the diffusion coefficient, s is the gradient pulse shape (hard), Δ is the diffusion time and γ is the proton gyromagnetic ratio.

PARAFAC was performed on partial volumes determined by summing the intensities in a 3 x 3 points window centred at the peak maximum using the N-way toolbox⁹ in Matlab (MathWorks). To minimize the effect of intensity variations between peaks with different dynamic properties, data was scaled to unit variance, i.e. the signal intensity of each signal was divided by the standard deviation of that signal over all time points. In order to produce physically meaningful data the analysis was constrained to produce only positive

contributions. PARAFAC models including up to 5 factors were computed. Anomalous behavior was identified for 3 weak signals from histidines, which were removed from further analysis. The analysis identifies four main species characterized by different sizes and time evolution using various quality assessment tools, such as residuals and core consistency diagnostics¹⁰ (Table S1). Factor 4 comprises only 3 signals. In order to obtain improved descriptors for factors 1–3, a new set of models were made excluding the 3 signals with significant contribution to factor 4 (Fig. 1). The reproducibility of the time and gradient strength profiles were assessed by creating 10 models using 50% randomly chosen signals, as described by the individual curves in Figure 1 of the main text. For factors 1 to 3 the curves were calculated in the model using 3 factors excluding the three signals contributing to factor 4. For factor 4 the curves were calculated in a 4-factor model, and in this case only models including 2 or more of the 3 signals were used. The analysis was repeated on unscaled data resulting in essentially unaltered time evolution and diffusion coefficients for all factors (not shown), however due to the intensity variations between peaks with different dynamic properties the contributions as a function of signal number are less interpretable.

Determination of Stokes radii

Stokes' radii (r_s) are calculated from the translational diffusion coefficient using the Stokes-Einstein relation:

$$D = \frac{k_B T}{6\pi\eta r_s} \quad (\text{equation S2})$$

where η is the solvent viscosity, D is the translational diffusion coefficient, k_B is Boltzmann's constant, and T is the temperature.

The relative molecular weight of state II and III were calculated from r_s and state I as follows (assuming spherical geometry):

$$\frac{M_{w,II}}{M_{w,I}} = \left(\frac{r_{s,II}}{r_{s,I}}\right)^3 \quad (\text{equation S3})$$

$M_{w,i}$ is the molecular weight of state i .

Interpretation of diffusion coefficients in terms of oligomeric order (molecular size) can be compromised if the viscosity of the sample changes over time as fibrils are formed.¹¹ However, in the present reaction the diffusion coefficients of the four states are constant (within error) over time, indicating that the sample viscosity is not significantly affected (Figure 2B of the main text).

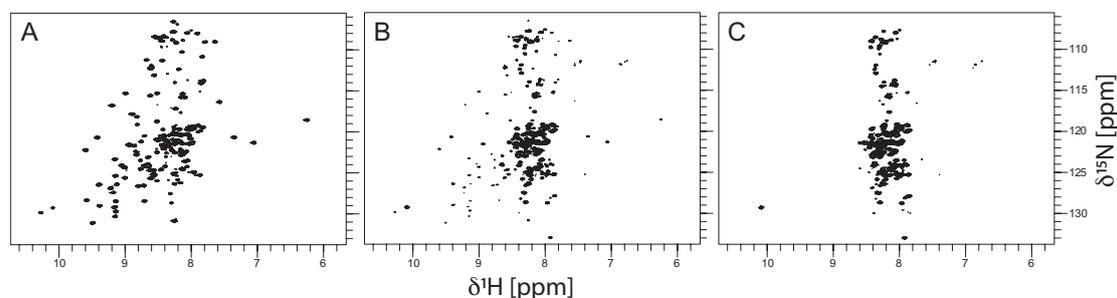


Figure S1. ^1H - ^{15}N DOSY-HSQC spectra at lowest gradient strength (5,655 G/cm) at time point 1.4 hours (A) 15.2 days (B) and 41.7 days (C).

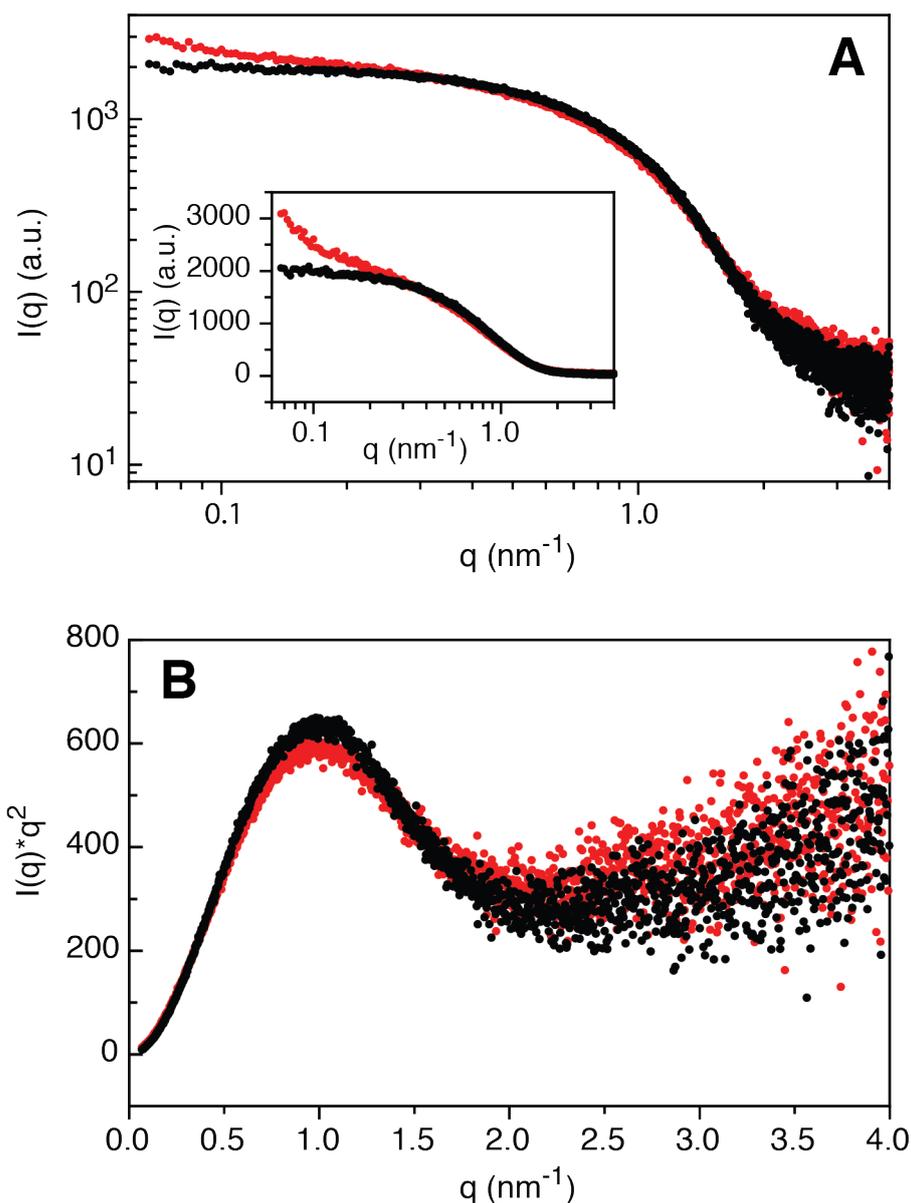


Figure S2. (A) Doubled logarithmic plot of SAXS profiles of pwtSOD1 $^{\Delta C}$ at day 0 (black) and day 19 (red) as function of the momentum transfer, $q = 4 \cdot \pi \cdot \sin(\theta)/\lambda$, where 2θ is the scattering angle between the incident beam and the direction of observation, and λ is the wavelength, $\lambda = 1.241 \text{ nm}$. Inset: linear-logarithmic plot of the same data. The scattering at low q -values increases at day 19, compared to day 0, characteristic for the formation of larger species. The sample at day 19 is no longer monodisperse. Guinier's approximation holds for data at day 0 and gives estimates of the extrapolated forward scattering $I_0 = 1964 \pm 3$ and radius of gyration $R_g = 19.1 \pm 0.1 \text{ \AA}$; q -range: $0.064 - 0.68$; qR_g limits: 0.12 (lower) and 1.30 (upper).⁴ Based on the extrapolated forward scattering (I_0) of BSA the molecular weight of the sample at day 0 is calculated to $M_w = 14001 \text{ Da}$, thus confirming that the sample is monomeric at the beginning of the experiment. The sample at day 19 is polydisperse and does not permit Guinier analysis. (B) Kratky plots of scattering data. The peak at $q = 0.98 \text{ nm}^{-1}$ indicates that both samples are dominated by a globular protein and the linear increase at higher q -values is diagnostic for disordered segments¹². At day 19, the contribution from species I (monomeric pwtSOD1 $^{\Delta C}$) is expected to be low based on the kinetics of the NMR analysis and the sample is dominated by states II, III and IV. Note: fully fibrillated samples are expected not to show a peak in the Kratky plot but to increase linearly at higher q -values.¹³⁻¹⁵

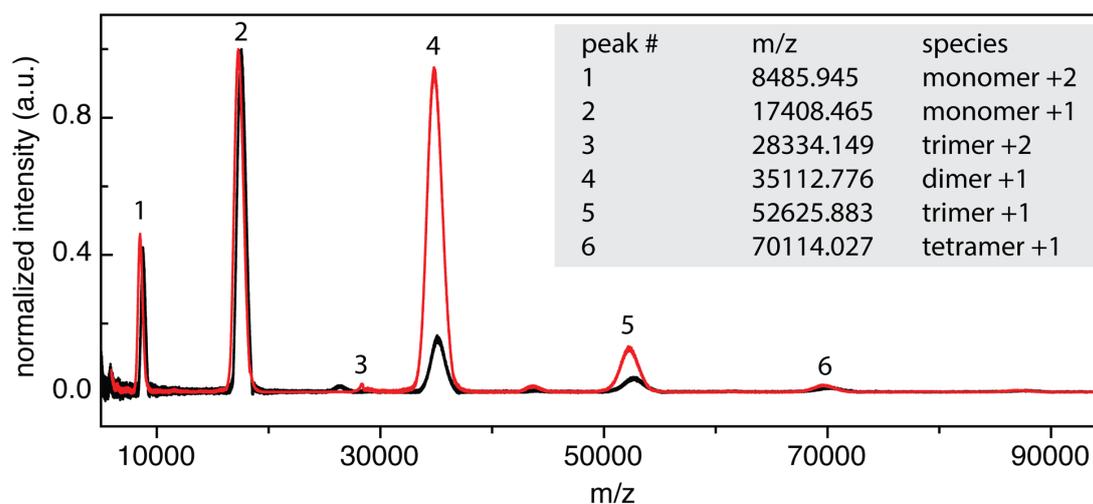


Figure S3. Identification of pwtSOD1^{AC} oligomers by mass spectrometry of intact lysine cross-linked pwtSOD1^{AC}; day 0 (black) and day 19 (red) normalized to the maximum intensity in each spectrum (monomer +1 peak) for comparison. The mass spectrometry analysis identifies different oligomeric states of pwtSOD1^{AC}: Peak #1 m/z = 8486 corresponds to monomer +2 charge (17408 Da/2 = 8704 Da), peak #2 m/z = 17408 corresponds to monomer +1 charge, peak #3 m/z = 28334 corresponds to trimer +2 charge (3 x 17408 Da + 2(3) x 170 Da)/2 = 26282(26367) Da), peak #4 m/z = 35113 corresponds to dimer +1 charge (2 x 17408 Da + 1(2) x 170Da = 34986(35156) Da), peak 5 m/z = 52626 corresponds to trimer + 1 charge (3 x 17408 Da + 2(3) x 170 Da = 52564(52734) Da) and peak #6 m/z = 70114 Da corresponds to tetramer + 1 charge (4 x 17408 Da + 3(4) x 170 Da = 70142(70312) Da). All peaks are wide, which will always be the case with MALDI-TOF/TOF spectra of intact proteins. In addition, the sample contains a mixture of ¹⁵N- and ¹⁴N-labeled protein and variation of lysine modifications, all of which will result in mass heterogeneity. The monomeric mass (m/z: 17408) is higher than expected based on the molecular weight of pwtSOD1^{AC} (¹⁴N: 15762.2666; ¹⁵N: 15963.7773). This may be explained by dead end modifications of lysine residues by BS3 cross-linker at 4 to 5 positions. In total, there are 11 lysine residues in pwtSOD1^{AC} and each BS3 cross-link will result in a mass addition of 170 Da and each dead-end in a mass addition of 348 Da.

Clearly, the sample obtained at day 19 contains a higher proportion of oligomeric states than does the sample at day 0. Note that the higher oligomers are expected to be underrepresented in these experiments as their capture relies on the formation of multiple cross-links within the same oligomer, i.e. if just one cross-link between two subunits of a tetramer is formed, the acidification, crystallization and laser energy will likely result in the observation of one dimer and two monomer equivalents; correspondingly if three subunits of a tetramer are cross-linked, then one trimer and one monomer will be detected. In addition, the signal intensities are not quantitative. The number of BS3 dead end modifications is likely to be higher on the monomeric species than in the oligomers as lysine residues may be buried in the subunit interface thereby making them less accessible for reaction with BS3.

Note also that the spectra were calibrated with internal BSA intact protein to obtain the best possible calibration, but the calibration will not be linear over the full m/z range measured. As a result it is expected that the monomer/oligomer masses cannot be exactly matched. In summary, the cross-linking MS spectrum unequivocally detects the formation of oligomers over time.

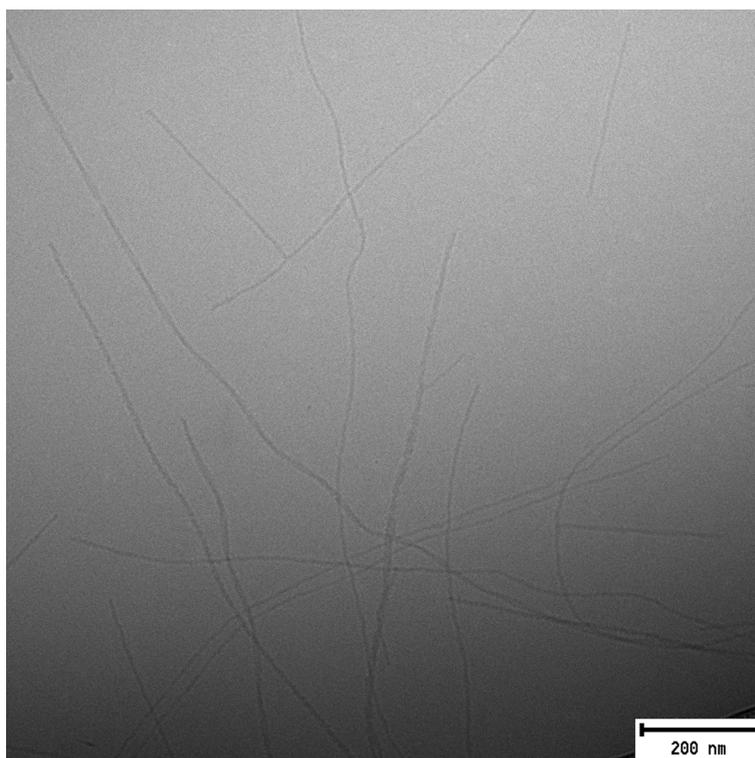


Figure S4. Fibril formation. Cryo transmission electron micrograph of the sample after the ^{15}N DOSY-HSQC time-series. The NMR sample was stored at 5°C prior to cryo-TEM experiments. A fraction of the sample was diluted 100 times in 10 mM Na-PO_4 pH 7.0 and prepared as a thin liquid layer on lacey carbon filmed copper grids and dropped into liquid ethane (-180°C) using a controlled environment vitrification system (CEVS). The sample was transferred using a Fischione Model 2550 cryo-transfer tomography sample holder into a JEOL JEM-2200FS transmission electron microscope and the micrograph was recorded with a 20eV energy selecting slit otherwise essentially as described.¹⁶

Table S1 PARAFAC model statistics

Signals used	Number of factors	Number of iterations	Error	Concordia
220 (all but 3 weak His signals)	1	6	9304	100.0
	2	12	1638	94.9
	3	29	862	58.9
	4	33	747	4.4
	5	60	653	0.3
217 (all but 3 weak His signals and 3 small peptide signals)	1	6	9089	100.0
	2	12	1498	96.3
	3	27	752	56.7
	4	45	660	22.3

References

- (1) 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. <https://doi.org/10.1021/acschemneuro.7b00162>.
- (2) 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.* **2009**, *106*, 18273–18278. <https://doi.org/10.1073/pnas.0907387106>.
- (3) Hopkins, J. B.; Gillilan, R. E.; Skou, S. BioXTAS RAW: Improvements to a Free Open-Source Program for Small-Angle X-Ray Scattering Data Reduction and Analysis. *J. Appl. Crystallogr.* **2017**, *50*, 1545–1553. <https://doi.org/10.1107/S1600576717011438>.
- (4) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; et al. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. <https://doi.org/10.1107/S1600576717007786>.
- (5) 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. <https://doi.org/10.1021/ja0211407>.
- (6) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A 277–293. *J. Biomol. NMR* **1995**, *6*, . <https://doi.org/10.1007/BF00197809>.
- (7) 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. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins Struct. Funct. Bioinforma.* **2005**, *59*, 687–696. <https://doi.org/10.1002/prot.20449>.
- (8) Ferrage, F.; Zoonens, M.; Warschawski, D. E.; Popot, J.-L.; Bodenhausen, G. Slow Diffusion of Macromolecular Assemblies Measured by a New Pulsed Field Gradient NMR Method [*J. Am. Chem. Soc.* **2003**, *125*, 2541–2545]. *J. Am. Chem. Soc.* **2004**, *126*, 5654–5654. <https://doi.org/10.1021/ja033464g>.
- (9) Andersson, C. A.; Bro, R. The N-Way Toolbox for MATLAB. *Chemom. Intell. Lab. Syst.* **2000**, *5*, 1–4. [https://doi.org/10.1016/S0169-7439\(00\)00071-X](https://doi.org/10.1016/S0169-7439(00)00071-X).
- (10) Engelsen, S. B.; Bro, R. PowerSlicing. *J. Magn. Reson.* **2003**, *163*, 192–197. [https://doi.org/10.1016/S1090-7807\(03\)00125-3](https://doi.org/10.1016/S1090-7807(03)00125-3).
- (11) Nicoud, L.; Lattuada, M.; Yates, A.; Morbidelli, M. Impact of Aggregate Formation on the Viscosity of Protein Solutions. *Soft Matter* **2015**, *11*, 5513–5522. <https://doi.org/10.1039/c5sm00513b>.
- (12) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Quarterly Reviews of Biophysics.* **2007**, *40*, 191–285. <https://doi.org/10.1017/S0033583507004635>.
- (13) Oliveira, C. L. P.; Pedersen, J. S. Structures of Aggregating Species by Small-Angle X-Ray Scattering. In *Amyloid Fibrils and Prefibrillar Aggregates: Molecular and Biological Properties*; **2013**. Wiley - VCH Verlag GmbH. <https://doi.org/10.1002/9783527654185.ch5>.
- (14) Otzen, D. E. *Amyloid Fibrils and Prefibrillar Aggregates: Molecular and Biological Properties*; **2013**. Wiley - VCH Verlag GmbH. <https://doi.org/10.1002/9783527654185>.
- (15) Sitar, S.; Aseyev, V.; Kogej, K. Microgel-like Aggregates of Isotactic and Atactic Poly(Methacrylic Acid) Chains in Aqueous Alkali Chloride Solutions as Evidenced by Light Scattering. *Soft Matter* **2014**, *10*, 7712–7722. <https://doi.org/10.1039/c4sm01448k>.
- (16) Munke, A.; Persson, J.; Weiffert, T.; De Genst, E.; Meisl, G.; Arosio, P.; Camerup, A.; Dobson, C. M.; Vendruscolo, M.; Knowles, T. P. J.; et al. Phage Display and Kinetic Selection of Antibodies That Specifically Inhibit Amyloid Self-Replication. *Proc. Natl. Acad. Sci.* **2017**, *114*, 6444–6449. <https://doi.org/10.1073/pnas.1700407114>.