Biophysical Journal, Volume 98

Supporting Material

The assembly pathway of a designed α-helical protein fiber

Bromley et al.

The assembly pathway of a designed α-helical protein fiber

Elizabeth H. C. Bromley,[†] Kevin J. Channon,[†] Patrick J. S. King,[†] Zahra N. Mahmoud,[†] Eleanor F. Banwell,[† ††] Michael F. Butler,[‡] Matthew P. Crump,[†] Timothy R. Dafforn,[§] Matthew R. Hicks,[¶] Jonathan D. Hirst,[‖] Alison Rodger,[¶] and Derek N. Woolfson,[†**]*

† School of Chemistry, University of Bristol, Bristol BS8 1TS, UK
‡ Unilever Corporate Research, Colworth Science Park, Sharnbrook, Bedford, MK44 1LQ, UK
§ School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK,
¶ Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK
‖ School of Chemistry, University of Nottingham, Nottingham, NG7 2RD, UK
** Department of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK
†† Tokyo Institute of Technology, Global Edge Institute, 4259-S2-17, Nagatsuta, Midori–ku, Yokohama, Kanagawa, 2268501, Japan.

*To whom correspondence should be address: D.N.Woolfson@bristol.ac.uk

CONTENTS

1. KINETICS VIA LINEAR DICHROISM SPECTROSCOPY

SAFs exhibit an LD spectrum when shear aligned. The spectrum indicates that both the peptide backbone (210 nm) and the tyrosine residues (280 nm and 240 nm) become oriented, Fig. S1 A. By recording LD spectra during fibrillogenesis, we applied the same analysis used for CD spectra and generated fractional completion curves, Fig. S1 B. The LD data show an increase in the alignment of both the Tyr residues and the backbone sometime after the initial increase in the CD signal. This is expected as the fibers must become long and numerous before aligning. The data also showed that the backbone and aromatic signals arise at the same time, implying that for fibers large enough to align, the Tyr residues are in a specific, fixed conformation. The LD spectra continued to show an increase during the slower phase of the CD completion confirming that the system continued to develop during this time period.

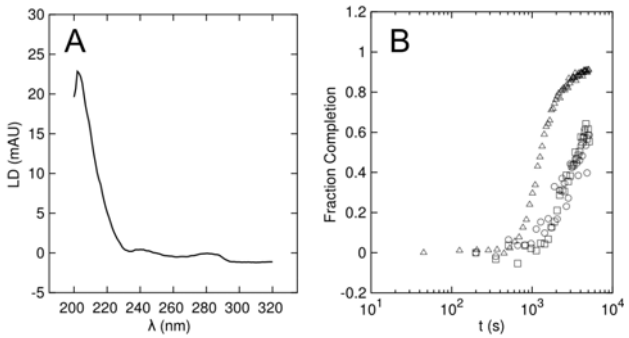

Figure S1. Linear dichroism spectroscopy of SAF formation. (A) LD spectrum of a mature 100 μM (of each peptide) sample. (B) Comparison of the fractional completion as a function of time for CD (open triangles), backbone LD (open circles) and aromatic LD (open squares).

2. CHIRAL SCATTERING IN CD SPECTRA

We investigated the nature of the distortion of the CD signal by collecting spectra with a range of sample detector distances. As this distance decreases, more of the light that has been scattered is collected, Fig. S2A. Figure S2 B shows CD spectra taken at various acceptance angles and shows that as we collect more of the scattered light the signal begins to appear more α-helical. This means that light is being scattered by the sample in a manner dependent on chirality. The existence of chiral scattering indicates that large α-helical assemblies that span the length scale of the wavelength of incident light are present, producing interfaces between the achiral solvent and mesoscopic regions of chiral material (1, 2). This is fully consistent with the previously published structural data for the SAF system (3).

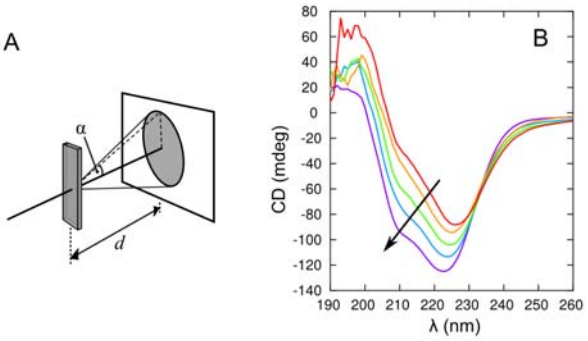

Figure S2: Influence of scattering on CD spectra: (A) Illustration of set-up for determining variation of CD spectra with acceptance angle. (B) CD spectra of SAFs (100 μM concentration path length 1 mm) taken with acceptance angles of 2.2°, 3.5°, 4.4°, 5.9° and 8.7° as calculated from the distance to the detector and the detectors side. As the angle of light accepted increases the signal moves towards the canonical alpha helical signal.

3. ANALYTICAL ULTRACENTRIFUGATION DATA

The individual SAF peptides (p1 and p2a) were sedimented under standard conditions at 100 μM in 10 mM MOPS buffer, pH 7.4 at speeds of 43,000–60,000 rpm, Fig S3. SAF-p1 gave a single species molecular weight of 6500 Da, and SAF-p2a gave a weight of 4500 Da.

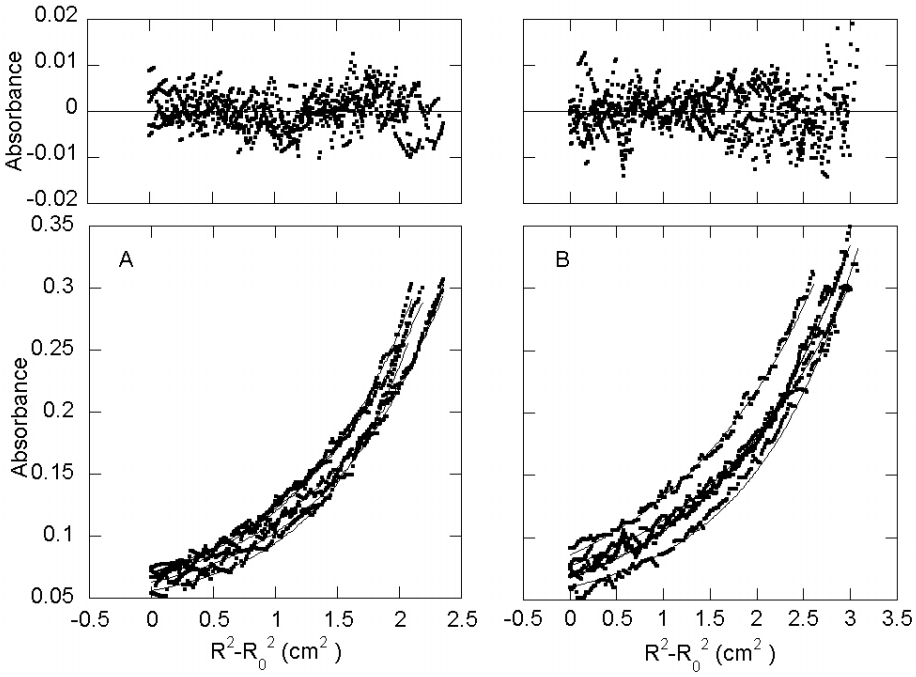

Figure S3: (A) Single ideal species fit for SAF-p1 giving a molecular weight of 6500 Da (predicted dimer mass 6348 Da). (B) Single ideal species fit for SAF-p2a giving a molecular weight of 4500 Da (predicted monomer mass 3325 Da, predicted dimer mass 6650 Da)

Fiber-containing solutions were prepared at peptide concentrations of 75, 100 and 150 μM in 10 mM MOPS buffer, pH 7.4 and spun at 3000 rpm. and at speeds in the range 50000–60000 rpm. The low-speed data indicated that 30–40 μM of each peptide remained in solution once the fibers had sedimented. This remaining material was then centrifuged at higher speeds to determine its oligomeric state.

The single ideal species fit produced a calculated molecular weight of 5230 (compared to an average monomer weight of 3249 and a dimer weight of 6498). The hetero-dimeric association fit gave a dissociation constant of 7–10 μM, in good agreement with the result from the CD spectroscopy experiments (main text).

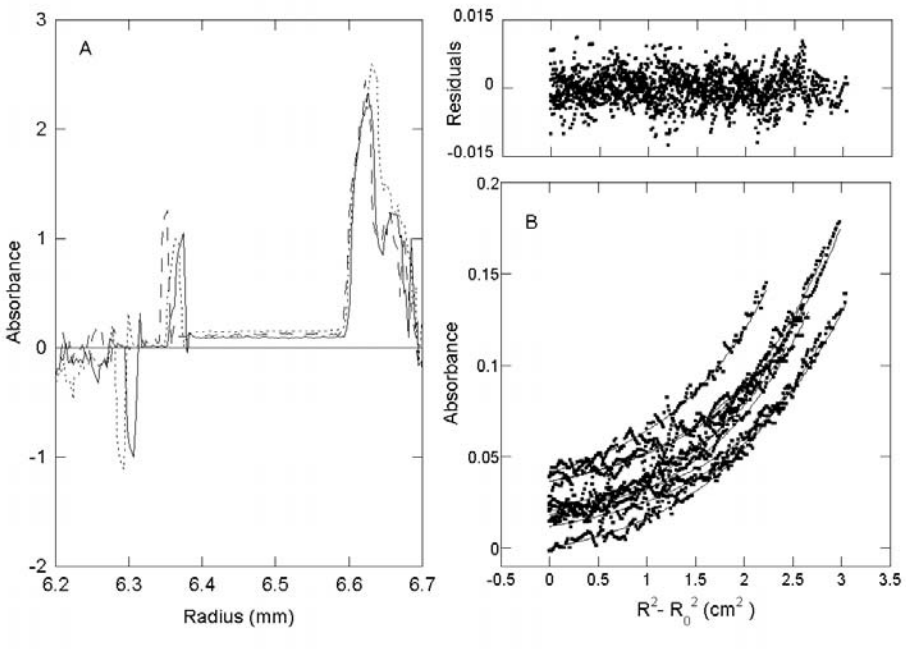

Figure S4: (A) Absorbance as a function of radius at 3000 rpm for SAF mixtures: 75 μM of each peptide (solid line), 100 μM (dashed line) and 150 μM (dotted line). (B) High speed (50,000–60,000 rpm.) data for the remaining soluble peptide. The fit is to a monomer–dimer equilibrium with K_D = 7–10 μM.

4. NMR SPECTROSCOPY KINETICS DATA

The large size of the fibers means that nuclear magnetic resonance (NMR) signals from incorporated peptide were no longer detected. Fibrilization kinetics could therefore be followed by observing the decrease in signal from soluble peptide. During fibrilization, the amount of soluble peptide fell to 32% of the starting concentration. NMR spectra of the supernatant of a mature SAF sample after centrifugation were similar.

Kinetics were followed by integration of methyl and aromatic NMR peaks at each time point shown and compared to the CD kinetics, Fig. S5 C.

The incorporation of peptide into fibers measured by NMR spectroscopy closely coincides (within the experimental error) with the formation of α-helix measured by CD spectroscopy, confirming that both techniques monitor the formation of fibers.

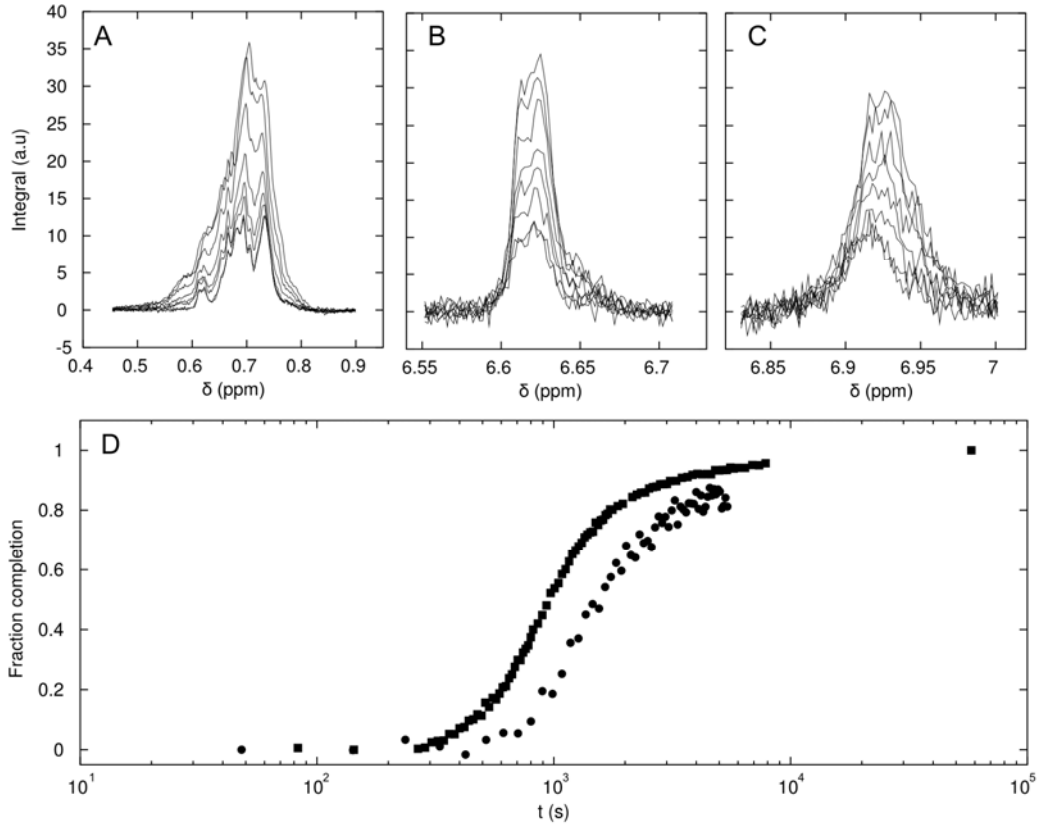

Figure S5: Kinetics of SAF assembly monitored using NMR spectroscopy. (A) Reduction in peptide methyl signal during fibrillogenesis. (B,C) Reduction in peptide tyrosine signal during fibrillogenesis. (D) Kinetics of SAF assembly monitored using NMR spectroscopy and circular dichroism spectroscopy.

5. PROLINE-SCANNING OF SAF PEPTIDES

In total, four mutants of each of the SAF peptides were made. Each mutant had a different *d*-position leucine replaced by a proline residue. The sequences of all the mutants are shown below, with the mutant proline residues highlighted.

```
SAF-p1-pro1:   KIAAPKQKIASLKQ EIDALEYENDALEQ
SAF-p1-pro2:   KIAALKQKIASPKQ EIDALEYENDALEQ
SAF-p1-pro3:   KIAALKQKIASLKQ EIDAPEYENDALEQ
SAF-p1-pro4:   KIAALKQKIASLKQ EIDALEYENDAPEQ

SAF-p2a-pro1:  KIRRPKQKNARLKQ EIAALEYEIAALEQ
SAF-p2a-pro2:  KIRRLKQKNARPKQ EIAALEYEIAALEQ
SAF-p2a-pro3:  KIRRLKQKNARLKQ EIAAPEYEIAALEQ
SAF-p2a-pro4:  KIRRLKQKNARLKQ EIAALEYEIAAPEQ
```

Figure S6: Sequences of SAF peptides, proline-scanned at *d* positions.

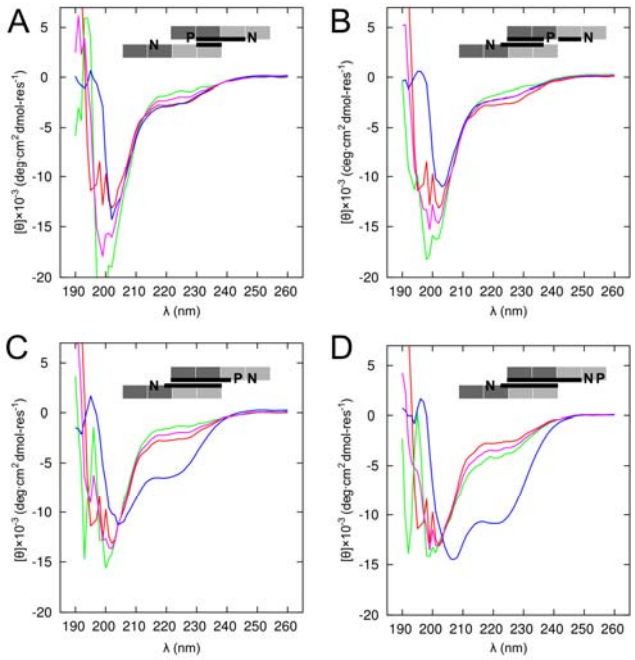

Figure S7a: CD spectra of SAF-p1 proline mutants with and without SAF-p2a. All peptides at and in 10 mM MOPS. In each plot shows SAF-p2a (red), SAF-p1-proX (green), SAF-p1-proX mixed with SAF-p2a (blue) and the mean of the spectra from SAF-p1-proX and SAF-p2a (purple). A) SAF-p1-pro4. B) SAF-p1-pro3. C) SAF-p1-pro2. D) SAF-p1-pro1.

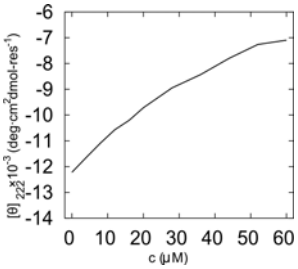

Figure S7b: The CD signal at 222 nm was measured as a function of NaCl concentration for 100 μM SAF-p1-pro1 mixed with 100 μM SAF-p2a in 10 mM MOPS. These data indicate that as the ionic strength increases, helicity is lost from the system. This implies that the specific interactions leading to helix formation, as well as the non-specific interactions that lead to onward aggregation are reduced by the addition of salt.

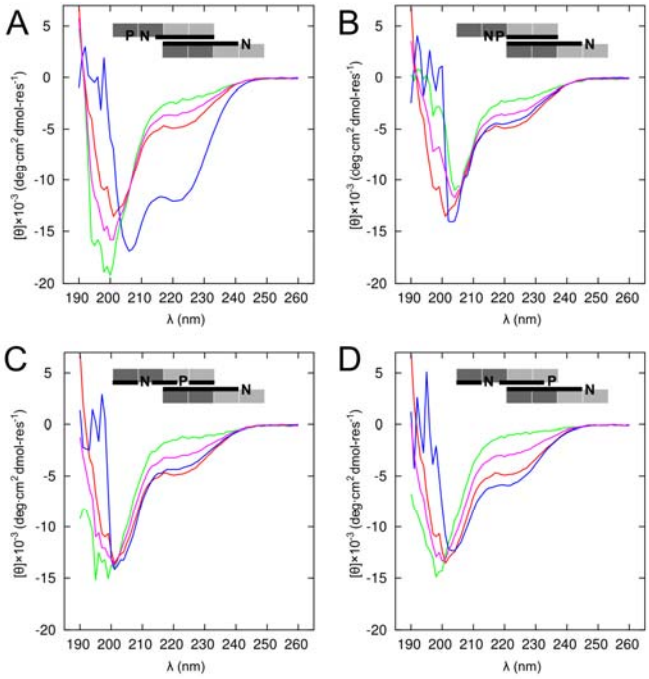

Figure S8: CD spectra of SAF-p2a proline mutants with and without SAF-p1. All peptides at 100 μM and in 10 mM MOPS. In each plot shows SAF-p1 (red), SAF-p2-proX (green), SAF-p1mixed with SAF-p2a-proX (blue) and the mean of the spectra from SAF-p1 and SAF-p2a-proX (purple). (A) SAF-p2a-pro4. (B) SAF-p2a-pro3. (C) SAF-p2a-pro2. (D) SAF-p2a-pro1.

Job plots of the SAF system and the proline-scanned systems that mimic the parent system were measured, and the binding stoichiometry in each case was found to be 1:1. Stoichiometry was determined by subtracting a linear background from the CD data[1] and then fitting the resulting points to a curve of the form

$$y(x) = \begin{cases} \dfrac{y_o}{x_o} x, & x < x_o \\ \dfrac{y_o}{1 - x_o}(1 - x), & x \geq x_o \end{cases}$$

where x_o and y_o are the x- and y-coordinates of the apex of the job plot. Fitting for x_o and y_o gives the position of the maximum and hence the stoichiometry. The Job plots imply that coiled-coil dimers form initially on mixing of SAF peptides, and that the stronger interaction (not containing asparagines at *a*) forms.

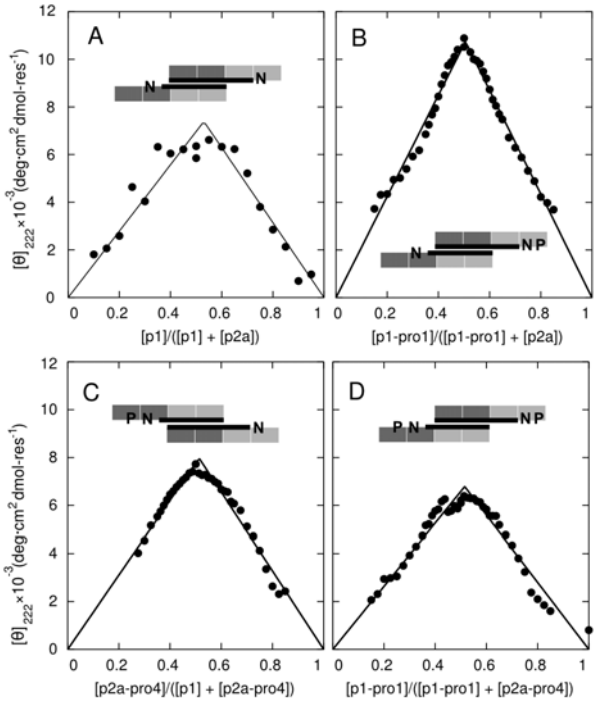

Figure S9: Job plots. A) SAF-p1 mixed with SAF-p2 B) SAF-p1-pro1 mixed with SAF-p2a. C) SAF-p1 mixed with SAF-p2a-pro4. D) SAF-p1-pro1 mixed with SAF-p2-pro4.

All of the proline mutants were subjected to AUC experiments when mixed with the parent partner, as well as some salient combinations of mutants. All gave small soluble oligomers. At lower concentrations the interaction of SAF-p1-pro1 with SAF-p2a tended towards that seen for SAF-p1 and SAF-p2a, indicating that the increase in molecular weight above the expected dimer was due to a weaker interaction—most likely a dimer-tetramer equilibrium.

[1] This is an acceptable approximation, since the first and last data points represent contributions from only SAF-p2a and SAF-p1, respectively.

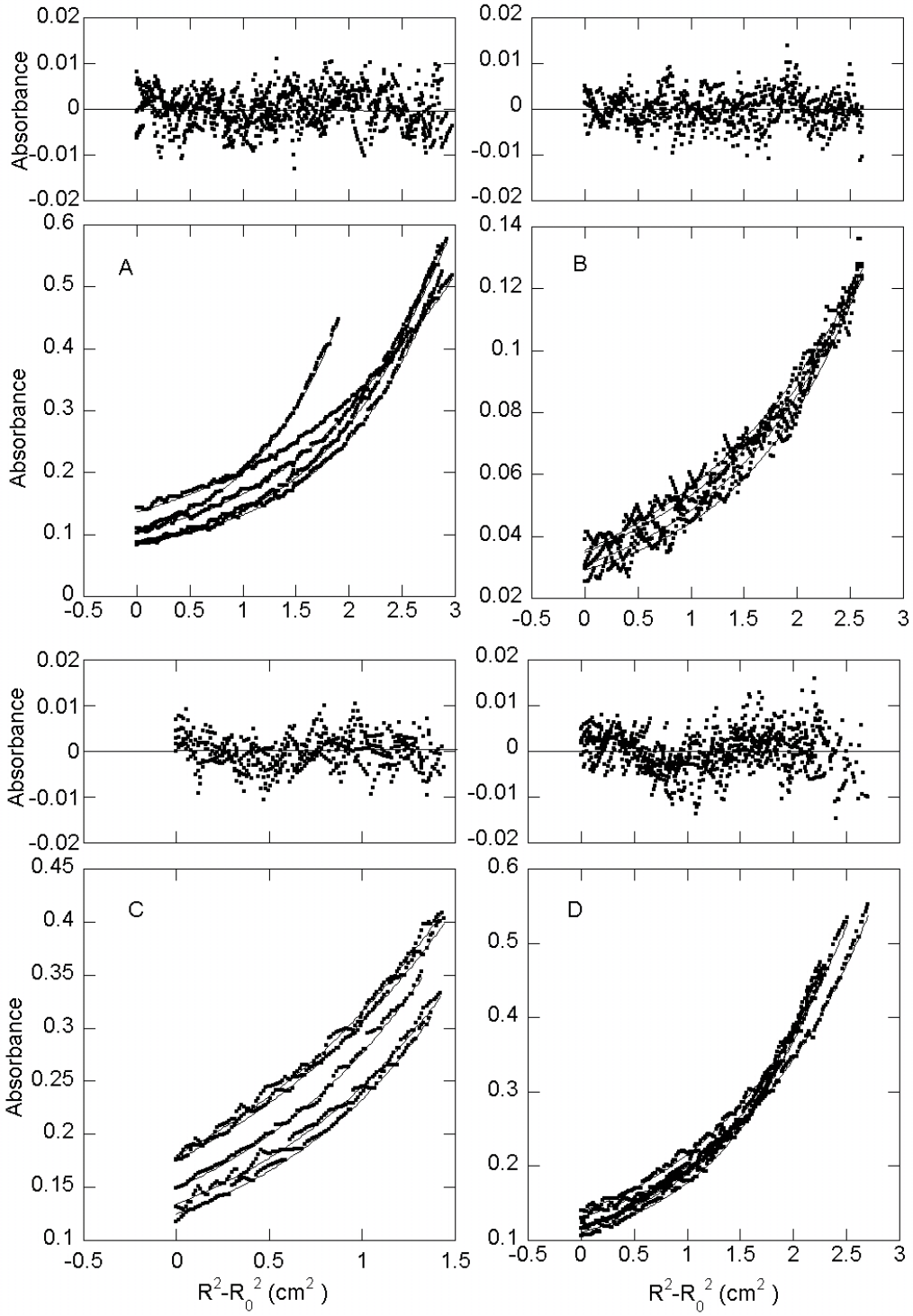

Figure S10: Single ideal species fits. (A) SAF-p1-pro1 mixed with SAFp2a at 100 µM giving 11,200 Da (average monomer mass 3242 Da). (B) SAF-p1 mixed with SAF-p2-pro4 at 100 µM giving 9650 Da. (C) SAF-p1-pro1 mixed with SAF-p2a at 30 µM giving 5300 Da. (D) SAF-p1-pro1 mixed with SAF-p2a-pro4 at 100 µM giving 10,500 Da (average monomer mass 3234 Da).

6. ANALYSIS OF KINETIC DATA TO GIVE PERCENTAGE COMPLETION VS. TIME

Usually, two-state transitions can be followed by measuring changes in spectral intensity at a single, particular wavelength. In the case of CD spectroscopy of an α-helical transition, this usually amounts to monitoring the change in molar ellipticity at 222 nm over time. However, the chiral scattering present in the SAF system causes a distortion of the CD spectrum at later times and results in a non-monotonic change in molar ellipticity at 222 nm that does not accurately reflect the underlying kinetics.

A more realistic view of the underlying kinetics may be recovered if it is assumed that the observed CD spectrum comprises some combination of the actual CD spectrum of the peptide and the contribution from the chiral scattering. If these two effects develop at similar rates, then the CD spectrum at any time can be well approximated by a linear combination of the initial spectrum (which we assume contains only un-fibrilized material) and the final spectrum (which contains some equilibrium between fibrilized and un-fibrilized material, with all the associated artefacts).

Polynomial functions that best fit the first and last spectra were determined by finding the order of polynomial that minimized the auto-correlation of the fit residuals. The intermediate spectra were then fitted using a linear combination of these two basis polynomials.

Thus, by finding the relative fractions of the initial and final spectra that contribute to an intermediate spectrum, we can make a reasonable estimate of how far to completion (α) the fibrilization process has progressed at that time.

As the system does not exhibit an exact isodichroic point, we tested the accuracy of the two state fit by plotting the residuals as a function of time, Fig. S11B. The data indicated that there is insufficient information to add a third state to the model and that the two-state model is therefore the best approximation.

In this analysis we have chosen the 24 hours spectrum as a reference for the final state of the system to reflect the fact that we observe fibrilization to be essentially complete at this point. If the same sample is left for a longer period, the spectrum is still found to change but this is due to some process of maturation that does not involve lengthening by monomer addition from solution. This stage of fiber growth remains to be fully characterized, but its contribution to the spectra up to 24 hours is taken to be minimal. The sensitivity of the analysis to the choice of time point for the final reference spectrum was checked by recalculating the dimensionality and nucleus size for a reference time point of 5000 seconds. The values obtained were in good agreement with those obtained using the 24 hour reference.

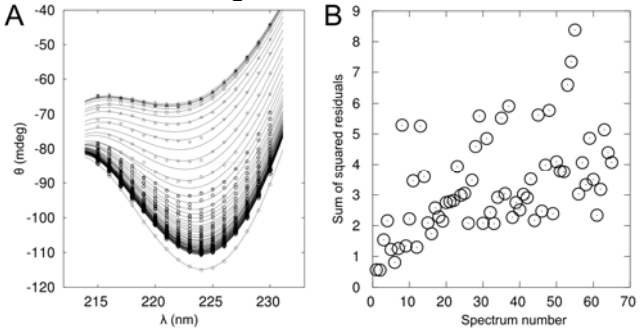

Figure S11: Fitting of CD spectra: (A) The original data and the linear two state fit for all of the spectra in a single kinetics run. (B) The residuals from the fit indicating that the two state approximation is a reasonable one.

7. MOTIVATION OF KINETIC MODELS

Once $\alpha(t)$ has been calculated from the individual CD spectra, a plot of $\alpha(t)$ vs. t represents the kinetics of fiber-formation. The formation of a SAF can be viewed as equivalent to a crystallization process, and the model for crystallization in metals, developed by Avrami(4), may be applied to fit $\alpha(t)$ and extract parameters relating to the nucleus size and the dimensionality of growth. The Avrami model has three basic premises:

1. There are fixed number of "germ nuclei" initially present in the system. These are not able to nucleate the new phase, but may spontaneously transform into a state that can.
2. Any germ nucleus has a constant probability per unit time of transforming into a "growth nucleus". Once this transformation occurs, growth will proceed from this nucleation point.
3. Once nucleated, the radius of the growing crystal will expand linearly with time (perhaps engulfing germ nuclei as it goes).

In the context of a phase-transition in a metal, this terminology makes sense, but to draw an analogue with the SAF system we can ignore the idea of the germ nuclei and focus on the generation and growth of nuclei. In the Avrami model, the rate of generation of growth nuclei decreases exponentially with time; since it is proportional to the number of germ nuclei, which are constantly being used up. In the SAF system, the rate of formation of nuclei will be dominated by the rate at which enough monomeric units can spontaneously aggregate to form the nucleus. For a nucleus containing n monomers at a concentration c, this will occur with a rate proportional to c^n. This does not fit with the Avrami model *generally*, but if we restrict ourselves to early times, when most of the system is in the soluble state, then the concentration is not changing rapidly and $c(t)^n \rightarrow c_o^n$, where c_o is the initial monomer concentration. Similarly, in the Avrami model, if we look only at early times then the number of germ nuclei remains large in relation to the rate at which they are being used and the rate of generation of growth nuclei is thus constant—the Avrami model is therefore applicable to the SAF system in this regime. Under these assumptions we can go ahead and fit $\alpha(t)$ to the Avrami model at early times in the fibrilization process.

Using the three-component model described above Avrami shows that, at early times when the number of growing nuclei is small, the fraction of transformed material, α, is described by the relation (equation 27(4))

$$\alpha(t) = 1 - \exp(-Kpt^4), \qquad \text{(S1)}$$

where, K is a constant relating to the growth rate and geometric considerations, p is the probability that a growth nucleus will form from a germ nucleus and t is the time. In the case of the SAFs, as described above, p is proportional to c_o^n, so that

$$\alpha(t) = 1 - \exp(-Kc_o^n t^4) \qquad \text{(S2)}$$

at early times, where we have subsumed the ¼ into the constant K. This particular form of the Avrami model describes a three-dimensional growth, but it may be extended to include growth of any lower dimensionality by giving it the form (equations 29–32(4))

$$\alpha(t) = 1 - \exp(-\mathrm{K}c_o^{\,n}\, t^{d+1}) \qquad \text{(S3)}$$

where, d is the dimensionality of growth.

Avrami also shows how this model can be used to describe cases where the number of growing fibers is large relative to the rate of spontaneous nucleation. This is exactly the case when a SAF solution is "seeded" with a large number (relative to the number of nuclei spontaneously formed in the un-seeded case) of pre-formed fiber fragments. In this case, the relation becomes (Equation 32(4))

$$\alpha(t) = 1 - \exp(-Kt^d) \qquad \text{(S4)}$$

In this case, the spontaneous nucleation rate ($c_o^{\,n}$) is not relevant to the kinetics, since spontaneous nucleation is assumed to be vastly slower than growth from seeds (which will rapidly decrease the amount of soluble monomer and thus even more rapidly decrease the spontaneous nucleation rate). All nucleation is assumed to happen at $t = 0$ with the addition of the seed material. As a consequence, the assumptions that limited the applicability of S3 to the early part of the kinetics are no longer necessary and the whole curve can be fitted using S3 in the case of a seeded kinetics.

8. DETAILED MATERIALS AND METHODS

Peptide Synthesis and Purification

Peptides were synthesized by standard Fmoc-based solid-phase peptide synthesis. Amino acids were obtained from Novabiochem (Merck KgaA, Darmstadt, Germany) and the peptide was synthesized using a Liberty CEM Microwave synthesizer. Fluorescein labelled SAF-p1 was synthesized by covalent attachment of fluorescein to a p1 mutant (aloc-protected lysine for glutamine substitution at position 14). While still in the solid phase and fully N-terminally and side chain protected, the aloc group was selectively removed using $Pd(PPh_3)_4$ (0.25eq) and HoBt (10eq) in THF, stirring overnight at RT. The fluorophore was coupled using carboxyfluorescein (10eq), PyBop (10eq) and DIPEA (10eq) in DMF, stirring at RT for 2hrs. Purification was performed by reversed-phase HPLC and peptide identity and purity was determined by MALDI-TOF mass spectrometry and analytical HPLC. Peptide stocks were prepared from the purified fractions, and were lyophilized before being stored at -80 °C. Stocks for use in subsequent experiments were prepared in ultrapure water. Peptide concentrations were determined in solution by UV absorption at 274 nm, $\varepsilon = 1405\ mol^{-1}cm^{-1}$.

Circular Dichroism

CD measurements were made using a JASCO J-815 spectropolarimeter fitted with a Peltier temperature controller. Peptide solutions were prepared in 10 mM MOPS/NaOH at pH 7.4, and were examined in 1 mm quartz cuvettes. Spectra were recorded at 20 °C using 1 nm interval, 1 nm bandwidth and 2 sec response times. After baseline correction, ellipticities in mdeg were converted to molar ellipticities (deg cm^2 $dmol^{-1}$) by normalizing for the concentration of peptide bonds and pathlength.

The deviation of the CD spectrum of the mature fibres from that expected for a fully folded alpha-helix was attributed to chiral scattering. Deviations due to absorption flattening, birefringence, alignment and achiral scattering were found to be insignificant. This conclusion was supported by the fact that the distortion could be increased by reducing the solid angle of light accepted by the photomultiplier and returned towards the expected spectrum as the solid angle was increased. Time series of CD spectra were interpreted by fitting to a linear combination of the initial and final spectra.

Linear Dichroism

Couette flow experiments were performed using a Jasco J815 spectropolarimeter modified for LD measurements (Jasco UK, Great Dunmow, UK). Spectra were measured from 320 nm to 200 nm using a bandwidth of 1 nm, a data pitch of 1 nm, scanning speed of 50 $nm \cdot min^{-1}$ and a response time of 2 s. A total of 35 spectra were measured over a period of 0–5064 s. Measurements were carried out at room temperature (21°C), and baseline spectra of the buffer were subtracted from the sample spectra. The rotation speed for alignment of the samples was 2000 rpm. The micro-volume Couette cell used had a pathlength of 0.5 mm and was built in-house; equivalent models are commercially available (Kromatek, Great Dunmow, UK). All sample volumes were 60 µl. SAF peptides were at a concentration of 100 µM and were mixed immediately prior to the measurements.

Transmission Electron Microscopy

Samples for TEM were prepared in duplicate, in identical conditions to those used for CD measurements. The SAF peptides, p1 and p2a, were mixed together at a concentration of 100 μM of each peptide in 10 mM MOPS buffer, pH 7.4 and allowed to assemble at 20°C. The final volume for each sample was typically 200 μl. At each time point, 5 μl of the sample was applied to carbon-coated copper TEM grids (Agar Scientific, UK). Samples were immediately dried down using filter paper, then stained with 5 μl of 1% uranyl acetate. The grids were air-dried and observed using a JEOL JSM 1200EX transmission electron microscope coupled with a Soft Imaging System Megaview II digital camera, at an accelerating voltage of 120 keV, and at a magnification of 60000× and 120000×.

Nuclear Magnetic Resonance

All ^{1}H NMR data were collected on a Varian VNMRS 600 MHz spectrometer equipped with Z-pulsed field gradients and a triple resonance salt tolerant cryoprobe. NMR buffer consisted of 3-(N-morpholino)propanesulfonic acid dissolved in 99.9% D_2O. 1D NMR spectra were collected at 20°C. A series of 404 1D spectra (4 scans per spectrum) were collected over a period of 24 h to monitor fiber assembly. Selected peaks (0.5–0.85 ppm and 6.5–7.5 ppm) were integrated, normalized and plotted against time to produce kinetic plots of un-incorporated peptide concentration during fiber assembly.

Light Microscopy

Confocal images were obtained using a Perkin Elmer Ultraview ERS 6FE confocal system (63_HCX PL APO objective) attached to a Leica DMI6000 inverted epifluorescence microscope and a Hamamatsu C9100-50 EM-CCD camera. Fluorophores were excited using a 14 mW Ar laser (488 nm). Samples were initially mixed in Eppendorf tubes and immediately transferred to sealed optical cells. Images were recorded in situ.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were conducted at 20°C in a Beckman-Optima XL-I analytical ultracentrifuge fitted with an An-60 Ti rotor. Data were fitted simultaneously initially assuming a single ideal species model and secondly assuming a hetero-dimeric association using Ultrascan(5). The average partial specific volume for the systems and the density of the solvent (1.002 $g \cdot ml^{-1}$) were calculated using Sednterp(6).

9. REFERENCES

1. Bustamante, C., I. Tinoco, and M. F. Maestre. 1983. Circular differential scattering can be an important part of the circular dichroism of macromolecules. Proc. Natl. Acad. Sci. U. S. A. 80:3568-3572.
2. Patterson, C. W., S. B. Singham, G. C. Salzam, and C. Bustamante. 1985. Circular intensity differential scattering of light by hierarchical molecular structure. J. Chem. Phys. 84:1916-1921.
3. Papapostolou, D., A. M. Smith, E. D. T. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell, and D. N. Woolfson. 2007. Engineering nanoscale order into a designed protein fiber. Proc. Natl. Acad. Sci. U. S. A. 104:10853-10858.
4. Avrami, M. 1940. Kinetics of phase change. II Transformation-Time Relations for Random Distribution of Nuclei. J. Chem. Phys. 8 212-224.
5. Demeler, B. 2005. Modern Analytical Ultracentrifugation: Techniques and Methods. . D. J. Scott, S. E. Harding, and A. J. Rowe, editors. Royal Society of Chemistry 210-229.
6. Hayes, D. B., T. Laue, and J. Philo. 1995-1998. Sednterp. University of New Hampshire, U.S.A.