A conserved spider silk domain acts as a molecular switch that controls fibre assembly

Hagn, F., Eisoldt, L., Hardy, J. G., Vendrely, C., Coles, M., Scheibel, T., & Kessler, H. (2010). A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature*, *465*(7295), 239-242. https://doi.org/10.1038/nature08936

Published in:
Nature

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

A conserved spider silk domain acts as a molecular switch that controls fibre assembly

Franz Hagn, Lukas Eisoldt, John G. Hardy, Charlotte Vendrely, Murray Coles, Thomas Scheibel & Horst Kessler

Nature

Received

20 May 2009

Accepted

11 February 2010

A huge variety of proteins are able to form fibrillar structures[1], especially at high protein concentrations. Hence, it is surprising that spider silk proteins can be stored in a soluble form at high concentrations and transformed into extremely stable fibres on demand[2, 3]. Silk proteins are reminiscent of amphiphilic block copolymers containing stretches of polyalanine and glycine-rich polar elements forming a repetitive core flanked by highly conserved non-repetitive amino-terminal[4, 5] and carboxy-terminal[6] domains. The N-terminal domain comprises a secretion signal, but further functions remain unassigned. The C-terminal domain was implicated in the control of solubility and fibre formation[7] initiated by changes in ionic composition[8, 9] and mechanical stimuli known to align the repetitive sequence elements and promote β-sheet formation[10, 11, 12, 13, 14]. However, despite recent structural data[15], little is known about this remarkable behaviour in molecular detail. Here we present the solution structure of the C-terminal domain of a spider dragline silk protein and provide evidence that the structural state of this domain is essential for controlled switching between the storage and assembly forms of silk proteins. In addition, the C-terminal domain also has a role in the alignment of secondary structural features formed by the repetitive elements in the backbone of spider silk proteins, which is known to be important for the mechanical properties of the fibre.

Figure 1

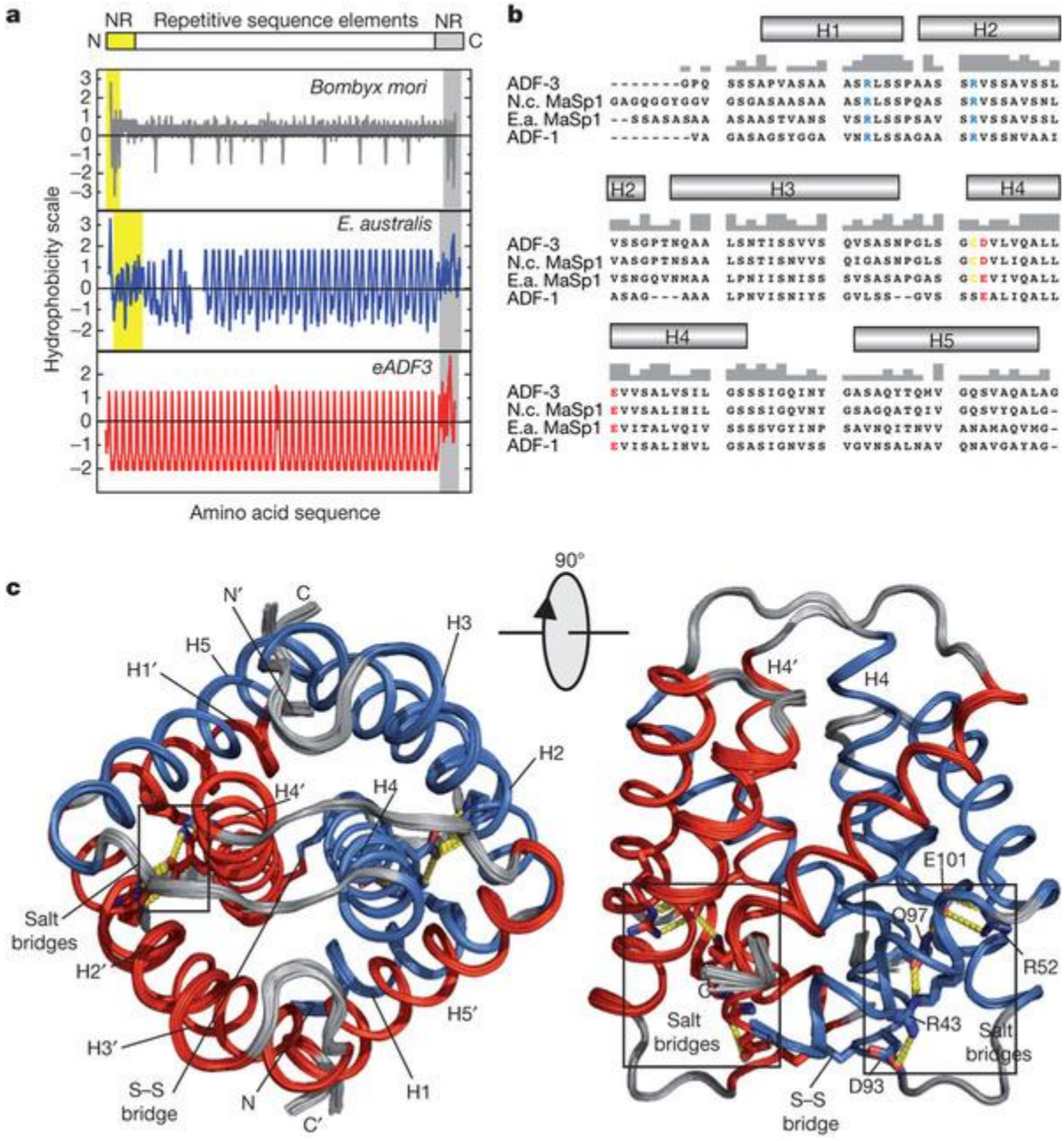

a
NR
Repetitive sequence elements
NR
N
C
Bombyx mori
E. australis
eADF3
Hydrophobicity scale
Amino acid sequence
b
H1
H2
H3
H4
H5
ADF-3
N.c. MaSp1
E.a. MaSp1
ADF-1
c
90°
N′
C
H5
H1′
H3
H2
H4′
H4
Salt bridges
H2′
H5′
H3′
S–S bridge
N
C′
H1
E101
O97
R52
R43
D93
Salt bridges
S–S bridge

Figure 2

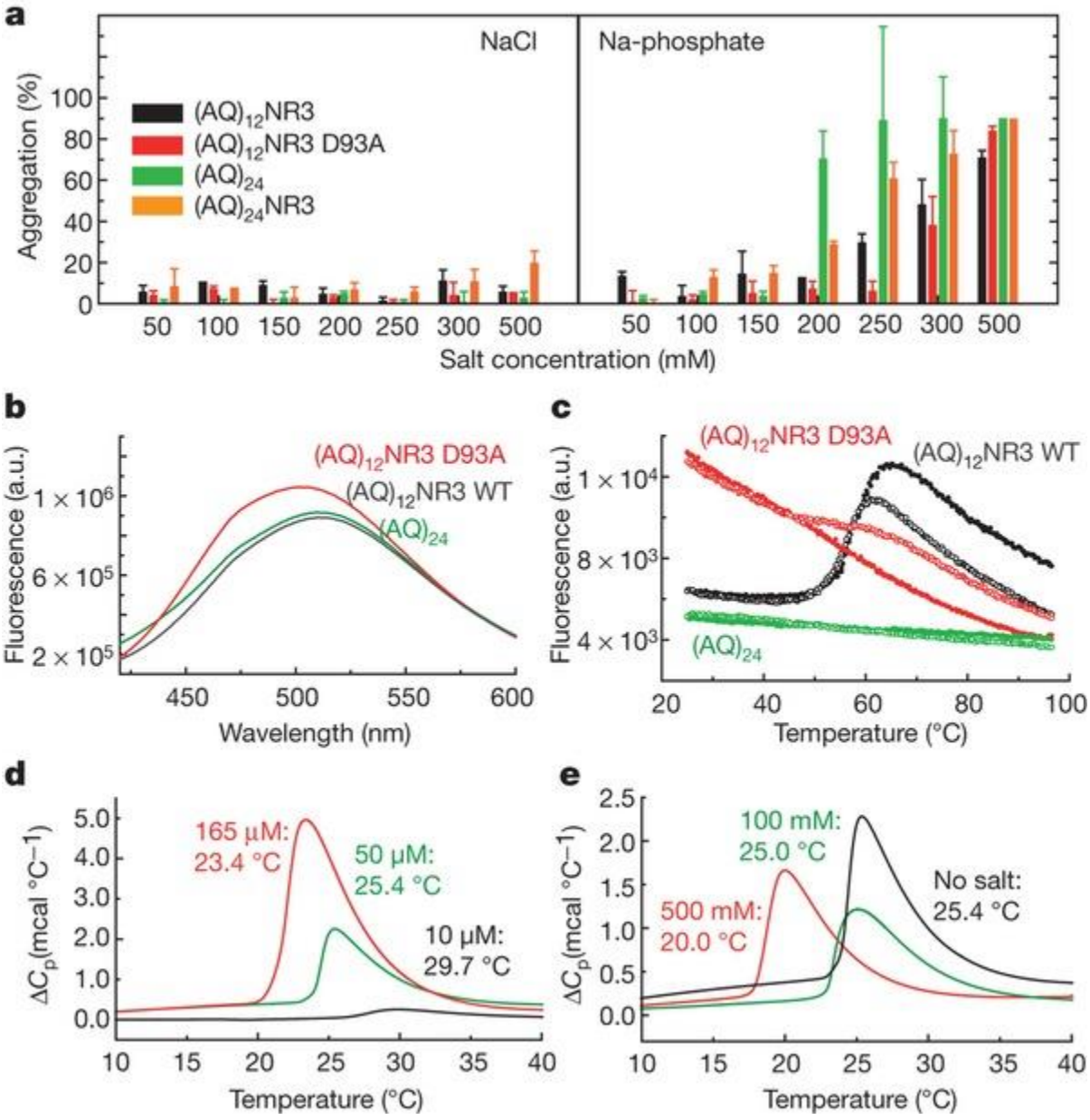

a
NaCl
Na-phosphate
Aggregation (%)
(AQ)12NR3
(AQ)12NR3 D93A
(AQ)24
(AQ)24NR3
Salt concentration (mM)
b
Fluorescence (a.u.)
(AQ)12NR3 D93A
(AQ)12NR3 WT
(AQ)24
Wavelength (nm)
c
Fluorescence (a.u.)
(AQ)12NR3 D93A
(AQ)12NR3 WT
(AQ)24
Temperature (°C)
d
ΔCp(mcal °C−1)
165 µM: 23.4 °C
50 µM: 25.4 °C
10 µM: 29.7 °C
Temperature (°C)
e
ΔCp(mcal °C−1)
100 mM: 25.0 °C
500 mM: 20.0 °C
No salt: 25.4 °C
Temperature (°C)

Figure 3

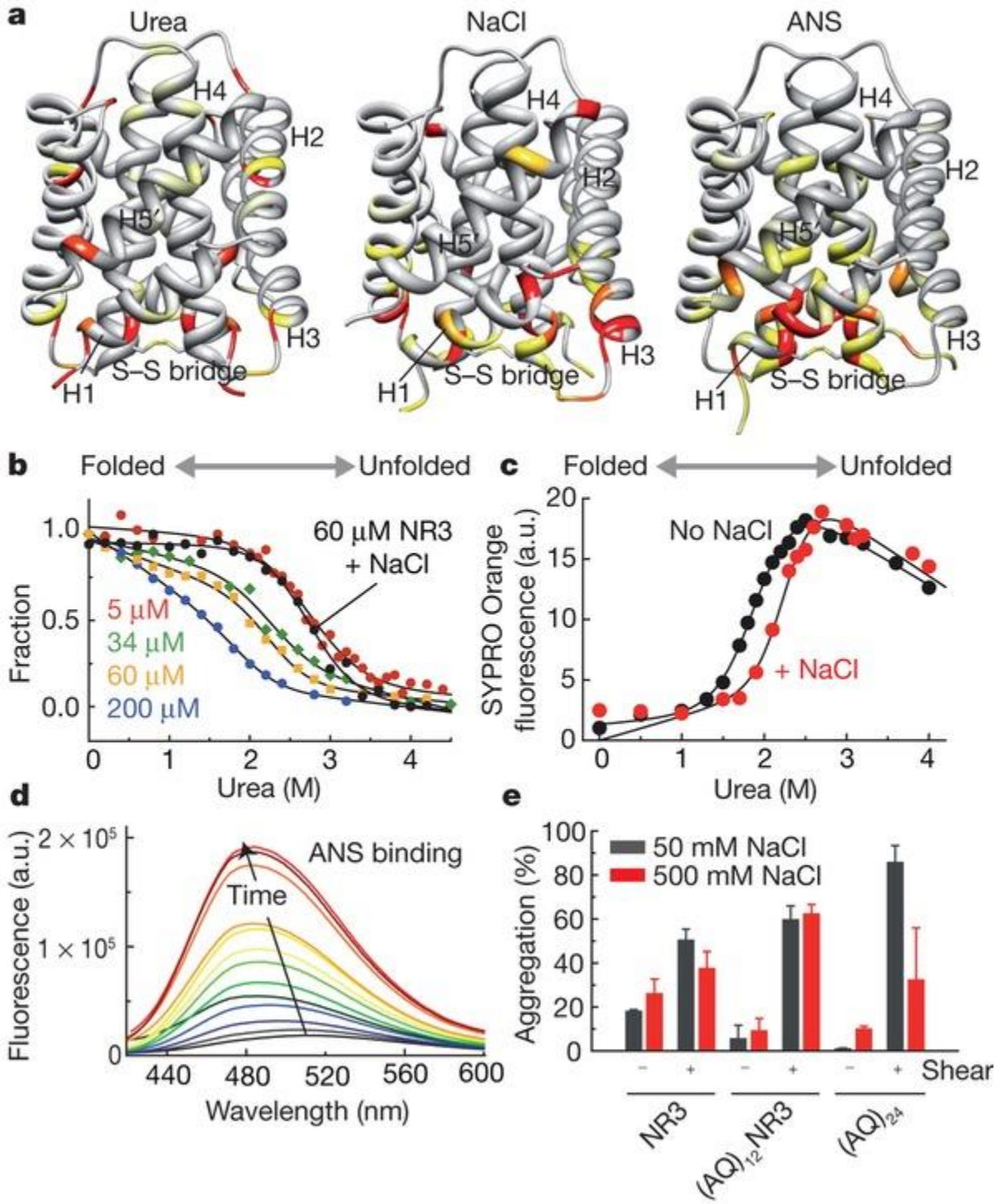
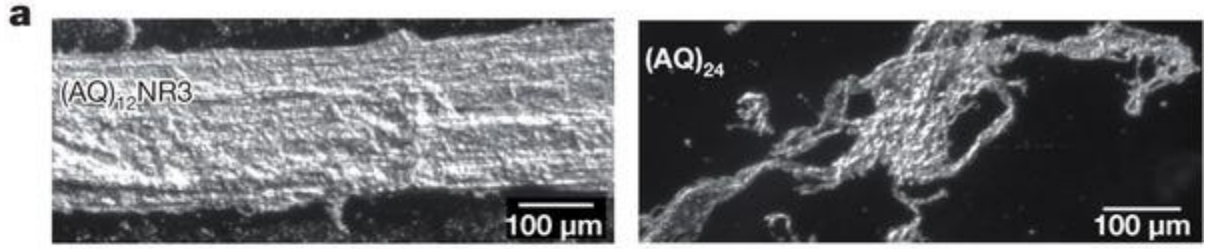
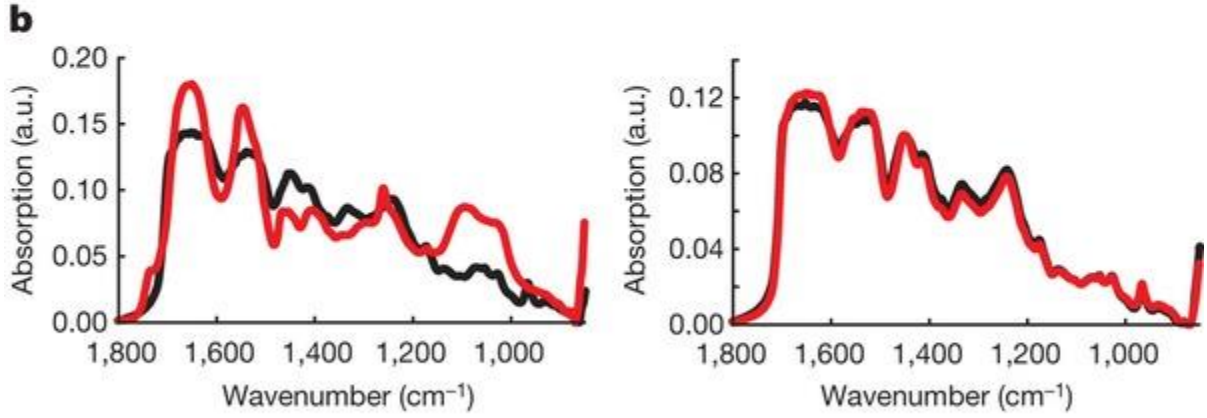

a
Urea
NaCl
ANS
H4
H2
H5′
H3
H1
S–S bridge
b
Folded
Unfolded
Fraction
1.0
0.5
0.0
60 μM NR3
+ NaCl
5 μM
34 μM
60 μM
200 μM
0
1
2
3
4
Urea (M)
c
Folded
Unfolded
SYPRO Orange
fluorescence (a.u.)
20
15
10
5
0
No NaCl
+ NaCl
Urea (M)
d
Fluorescence (a.u.)
2 × 10⁵
1 × 10⁵
0
ANS binding
Time
440
480
520
560
600
Wavelength (nm)
e
Aggregation (%)
100
80
60
40
20
0
50 mM NaCl
500 mM NaCl
–
+
Shear
NR3
(AQ)12 NR3
(AQ)24

Figure 4

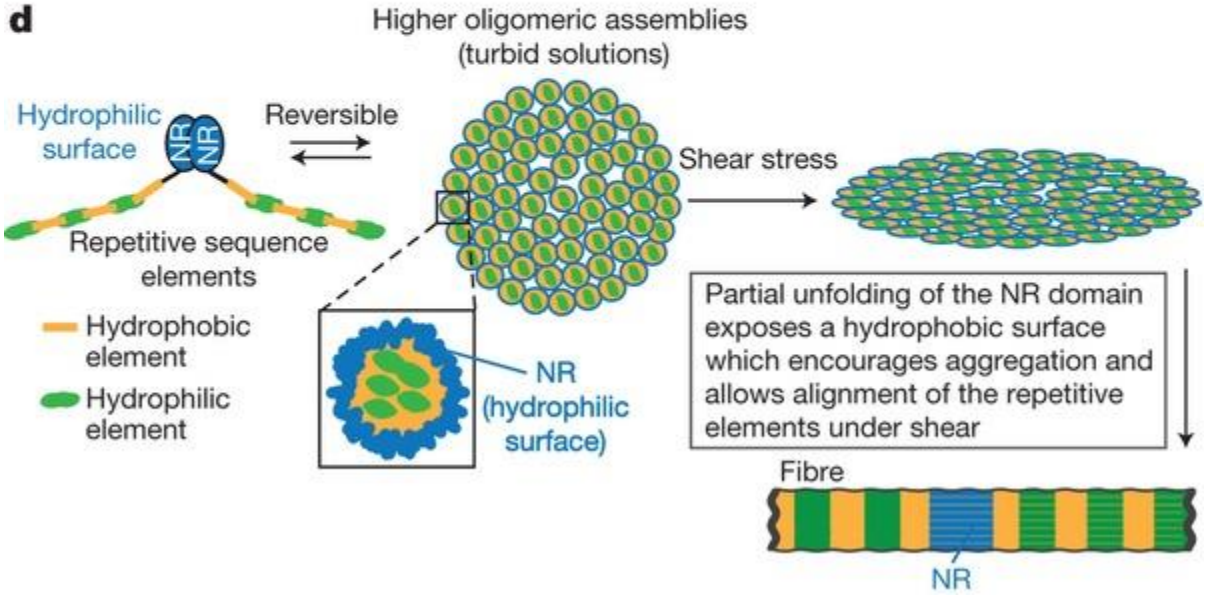




c

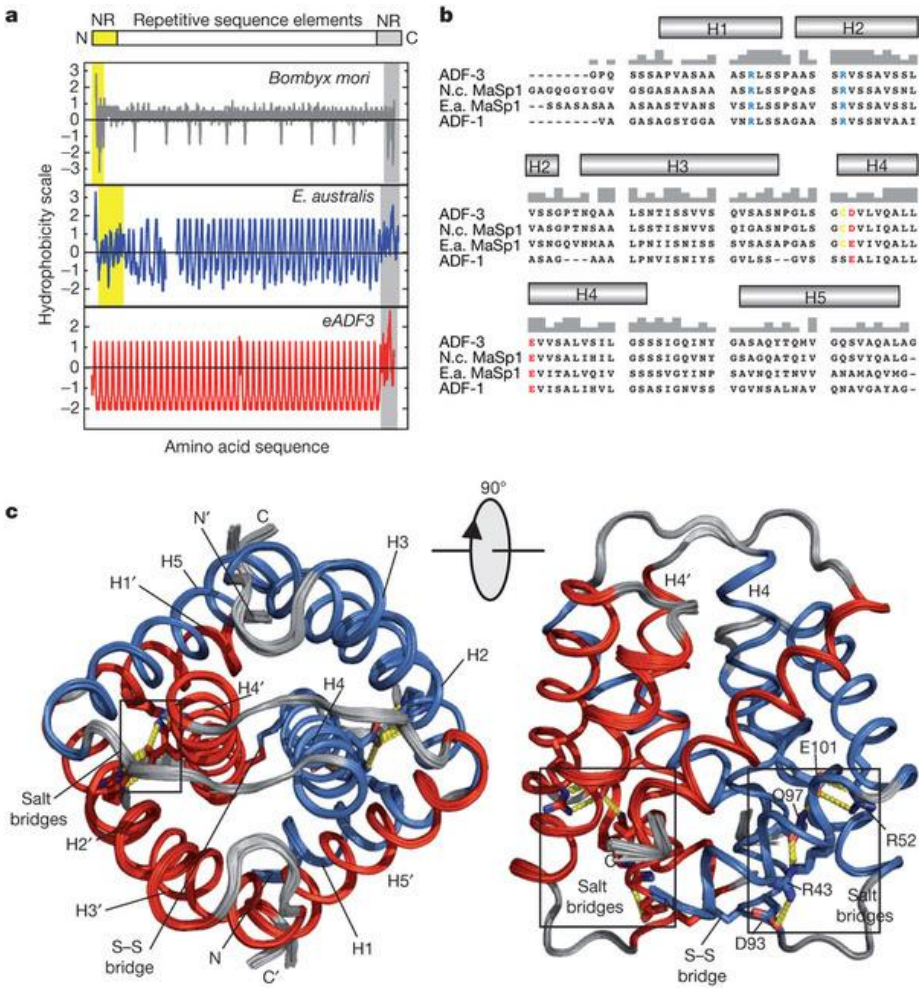
β-sheet content (%)

	Salt-induced aggregation			Shear-induced aggregation		
pH	$(AQ)_{24}$	Wild type	D93A	$(AQ)_{24}$	Wild type	D93A
7	32 ± 6	26 ± 3	35 ± 2	32/32/32 ± 1	28/28/25 ± 1	35/34/35 ± 1
6	41 ± 2	32 ± 4	32 ± 1	30/32/31 ± 2	36/33/30 ± 2	28/29/29 ± 1
2	38 ± 1	34 ± 3	35 ± 3	35/33/36 ± 2	31/33/32 ± 2	30/33/30 ± 2



Silk proteins are amphiphilic (Fig. 1a) and the C-terminal non-repetitive (NR) domain of such proteins produced by orb-weaving spiders is one of the most conserved regions (Fig. 1b). To determine the structure of the C-terminal NR domain (NR3) of *Araneus diadematus* fibroin 3 (ADF-3), we used high-resolution NMR with an optimized protein construct (Fig. 1c and Supplementary Tables 1 and 2). The structure is a new protein fold composed of a parallel-oriented dimeric five-helix bundle in which the longest helix (helix 4) is the main dimerization site (Fig. 1c, Supplementary Fig. 1). The single cysteine residue of each monomer involved in intermolecular disulphide formation is located at the N-terminal end of helix 4. The dimerization interface consists predominantly of hydrophobic residues. Packing of these residues is facilitated by a slight right-handed twist between helices 4 in both monomers. Helix 1 of one monomer and helix 5 of the second monomer interact in a clamp-like manner. Two salt bridges (R43–D93 and R52–E101) are located within each monomer, which fix helices 1 and 2 at one side of helix 4 (Fig. 1c). These salt bridges use the only charged residues of the entire known sequence of ADF-3 and are located in the most conserved parts of the NR domain (Fig. 1b).

Figure 1: Sequence analysis and structure of the non-repetitive (NR) domain of ADF-3.

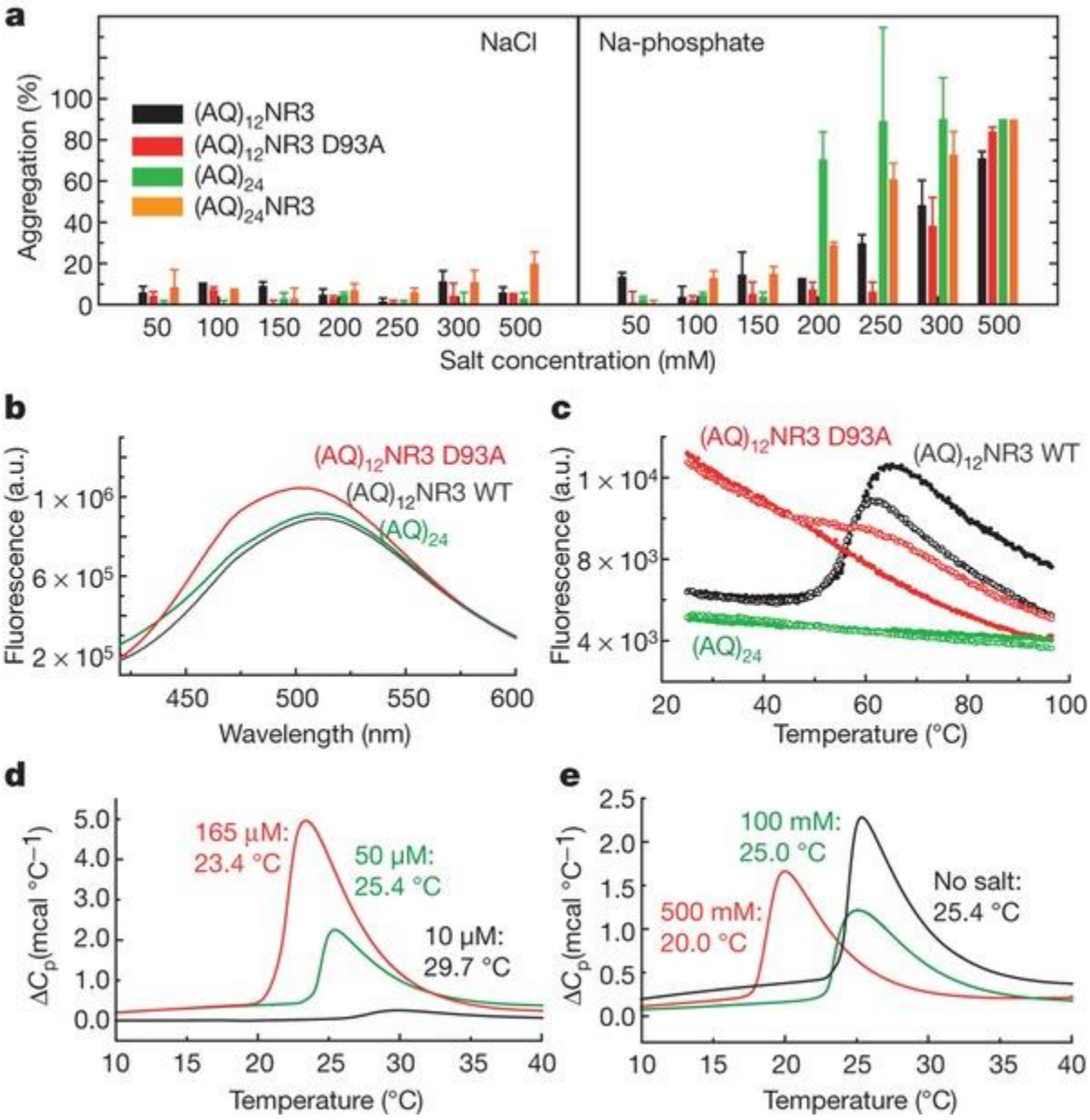


a, Hydrophobicity index of various silk proteins. b, Sequence alignment of the C-terminal NR domains of the major ampullate silk proteins (MaSp) of *Araneus diadematus*, *Nephila clavipes* and *Euprosthenops australis*, and minor ampullate silk fibroin 1 (ADF-1) of *Araneus diadematus* (top to bottom). Grey bars indicate the degree of conservation between these sequences. The conserved

charged residues are coloured. c, Overlay of the 20 best-energy structures of the C-terminal NR domain of ADF-3 having a root mean square deviation (r.m.s.d.) of 0.18 Å.

Using mutagenesis experiments, our data show that the charged residues are essential for the structural integrity of the NR domain, as deletion of the salt bridges results in significantly reduced thermal stability of the entire domain (Supplementary Fig. 2). Of particular note is that although the NR domain is relatively hydrophobic in comparison to the repetitive backbone of the protein (Fig. 1a), in the folded state the hydrophobic residues are buried and the solvent-exposed surface displays hydrophilic amino acids such as glutamine and serine, assuring its hydration (Supplementary Fig. 1b). At low pH (pH 2, where the salt bridges are disrupted) NR3 shows markedly increased hydrophobicity and decreased secondary structure content as probed with 8-anilinonaphthalene-1-sulphonic acid (ANS) fluorescence, circular dichroism (CD) and NMR spectroscopy (Supplementary Fig. 3a–d). In addition, a destabilized NR3 salt bridge mutant (D93A) of an engineered ADF-3 analogue, $(AQ)_{12}$NR3 containing 12 repeats of the 'A' and 'Q' repetitive sequence elements[16] (see Methods), also shows pronounced ANS binding (Fig. 2b) and molten globule-like thermal unfolding behaviour (a partially structured state that shows higher binding affinity to ANS than the native state), which can be stabilized by the addition of sodium chloride (Fig. 2c). This emphasizes the role of a correctly folded NR domain for silk protein storage and the inhibition of undesired aggregation.

Figure 2: Assembly and aggregation properties of our spider silk-like proteins.



a, Salt-induced protein aggregation. Error bars indicate standard deviation. b, Binding of hydrophobic ANS by proteins. a.u., arbitrary units; WT, wild type. c, Thermal transition experiments indicate molten globule-like behaviour of the D93A variant (red symbols), whereas the wild-type protein binds to the dye only during unfolding (black symbols). A structural transition of the D93A variant can be induced by sodium chloride (open symbols: in presence of 300 mM sodium chloride). Without the NR domain, no transition can be observed (green symbols). d, The concentration dependence of the LCST of $(AQ)_{24}$NR3. e, The sodium chloride dependence of the LCST of $(AQ)_{24}$NR3.

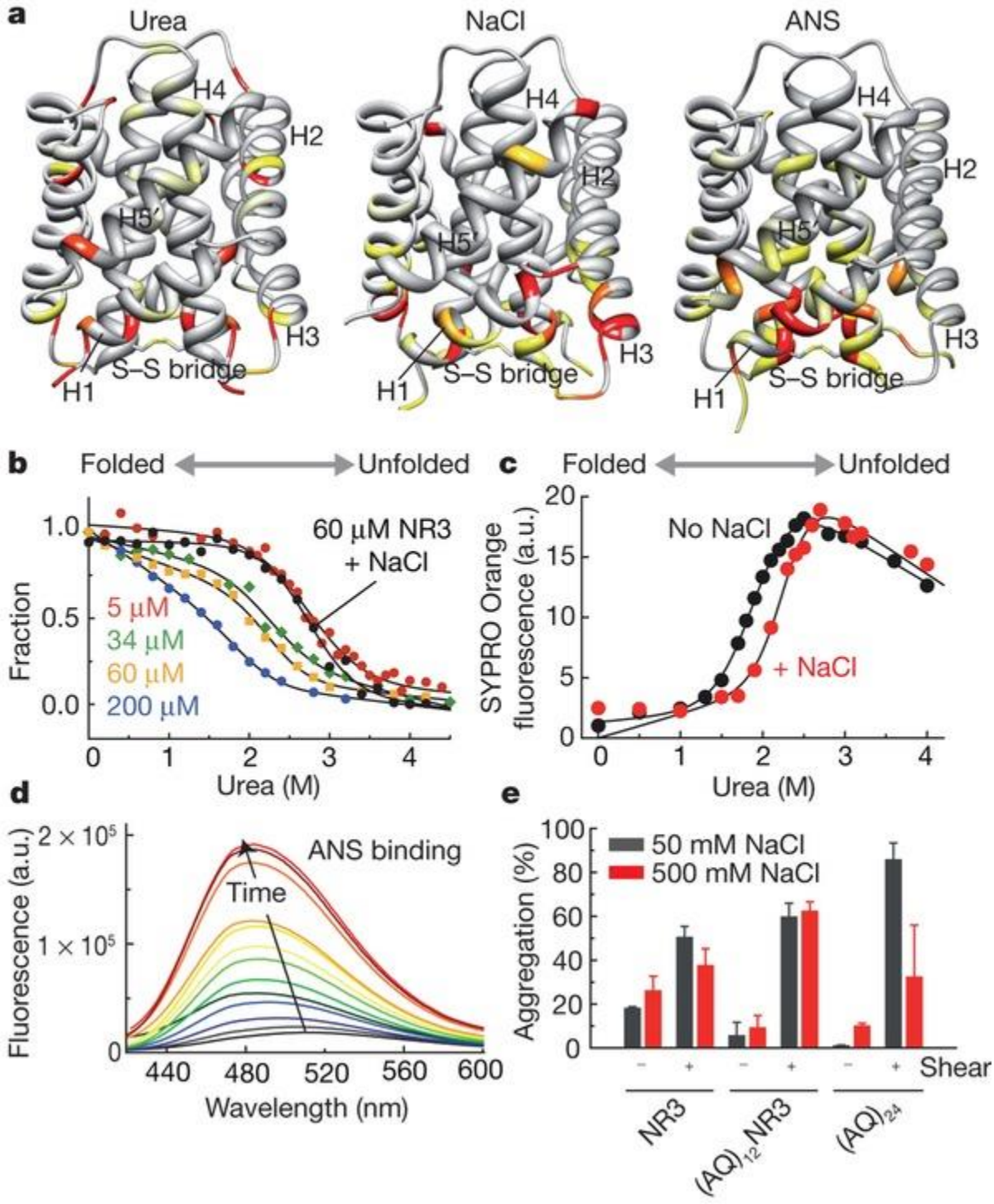
In vivo the transition from soluble protein to solid fibres involves a combination of chemical and mechanical stimuli (such as ion exchange, extraction of water, acidification and elongational flow[3, 13, 17]). Hence, we were interested in determining the effect of such stimuli on spider silk-like proteins, and used dimeric $(AQ)_{12}$NR3 and monomeric $(AQ)_{24}$ (an analogue of approximately the same molecular mass as the dimer without the NR carboxy-terminal domain). In the absence of shear forces (as silk proteins are stored in the lumen *in vivo*) the extent of protein aggregation *in vitro* is determined by the concentration of salt, and even at extremely high concentrations of sodium chloride (up to

500 mM) the level of aggregation is low (less than 15%), whereas in the presence of equivalent concentrations of sodium phosphate the level of aggregation is markedly higher (up to 100%) owing to the more kosmotropic nature of the phosphate anion (Fig. 2a). Such a finding explains why the proteins are stored inside the lumen in the presence of the relatively chaotropic sodium chloride (approximately 100–150 mM)[9], and emphasizes the requirement for ion exchange within the spinning duct.

Higher supramolecular assemblies are required to achieve efficient protein storage[8] and we have previously reported our proteins to be thermoresponsive, displaying fully reversible lower critical solution temperature (LCST, phase separation on increasing the temperature) behaviour owing to a self-assembly process[16]. This behaviour was displayed only for silk proteins incorporating the NR domain. Using differential scanning calorimetry (DSC) we show here that the LCST is decreased in the presence of higher concentrations of salt and/or protein (Fig. 2d, e), thereby stabilizing the proteins during storage.

We have also found that the NR domain has an impact on the fibre formation process. On titration of chemical denaturants (urea or guanidine hydrochloride) into solutions of ^{15}N-labelled NR3, NMR chemical shift changes (indicating structural variations or binding) monitored with ^{15}N HSQC (heteronuclear single quantum coherence) experiments could be observed mainly within helix 1 and around the salt bridges (Fig. 3a, Supplementary Fig. 4), therefore this region seems to unfold first. Titrations with sodium chloride demonstrated chemical shift changes in the same region (where the two salt bridges are located, Fig. 3a), which also shows the highest hydrophobicity in the protein as monitored with chemical shift changes upon the addition of ANS (Fig. 3a and Supplementary Fig. 4). Interestingly, at higher protein concentrations the protein shows a greater tendency to unfold (Fig. 3b). Additionally, SYPRO Orange (an ANS-like fluorescent dye) binding experiments[18] demonstrate that the partially folded state of NR3 is more hydrophobic than the natively folded or the unfolded state (Fig. 3c). It is also noteworthy that the presence of sodium chloride (present during protein storage) stabilizes the NR domain against thermal or chemical denaturation (Fig. 3b, c and Supplementary Fig. 5a) and can even compensate for the loss in stability during reduction of the inter-molecular disulphide bridge (Supplementary Fig. 5b).

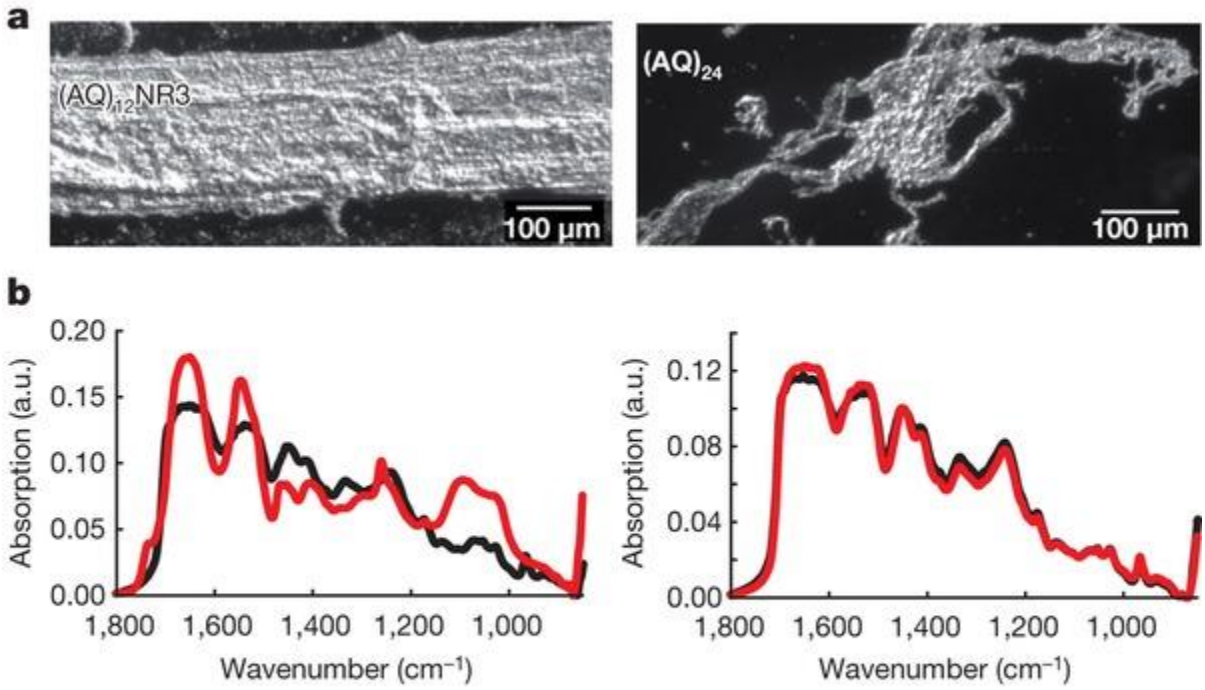
Figure 3: Stability and folding of spider dragline silk constructs.



a, NMR chemical shift perturbation experiments with urea, sodium chloride and ANS indicate that the structural elements around the salt bridges represent the most labile and hydrophobic region within the protein. b, Urea-induced unfolding of NR3 as determined by CD spectroscopy. c, Binding of hydrophobic SYPRO orange to NR3 during urea-induced unfolding. d, Shear-induced ANS binding of NR3. e, Aggregation of our spider silk-like proteins with and without shear stress. Shear stress leads to increased aggregation and is the most likely trigger for the assembly process in nature. Error bars represent standard deviation.

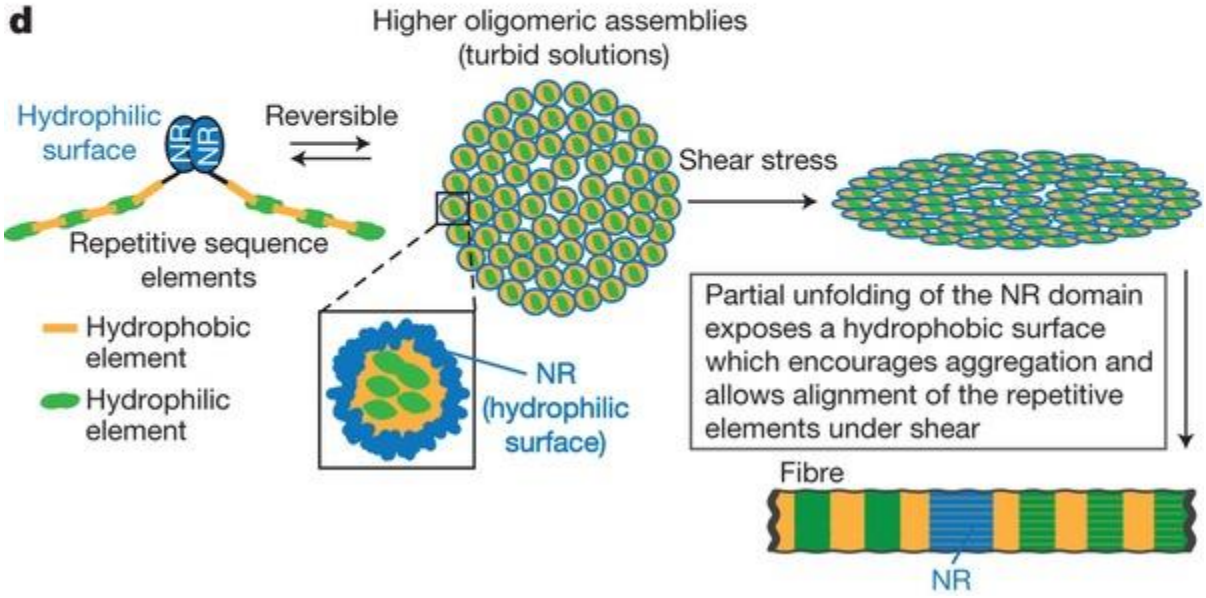
Previous rheological studies emphasize the role of the NR domain for shear-induced aggregation of ADF-3, where solutions of proteins with the NR domain exhibit significant viscosity changes under shear, whereas solutions of proteins without the NR domain exhibit no viscosity changes at various shear rates[14, 19]. ANS binding also increases under shear, indicating that the NR domain partially unfolds, thereby exposing hydrophobic surfaces (Fig. 3d and Supplementary Fig. 6), leading to oligomerization and ultimately β-sheet-rich fibre formation. In further *in vitro* protein aggregation assays under shear stress, we find that aggregation is vastly increased in comparison to corresponding assays in the absence of shear, and that the protein without the NR domain is much more prone to unspecific aggregation (Figs 3e and 4). Exposure of solutions of proteins without the NR domain to shear stress resulted in the formation of ill-defined β-sheet-rich aggregates, typically with dimensions of micro- to millimetre scale (Fig. 4a). By contrast, proteins incorporating the NR domain formed relatively well-defined fibrous aggregates, which optical microscopy showed to be composed of ordered (aligned) bundles of long fibrillar aggregates, implying that the NR domain has a role in stabilizing the protein against undesirable aggregation. Polarized Fourier transform infrared spectra of the fibrous aggregates formed from our protein without the NR domain ($(AQ)_{24}$) show that the β-sheets are randomly aligned, whereas the β-sheets within fibres formed from our protein with the correctly folded NR domain (that is, wild type and not the D93A mutant) were predominantly aligned with the long axis of the fibres (in analogy to naturally spun fibres) owing to a controlled assembly process (Fig. 4b, c).

Figure 4: Fibre assembly mechanism of dragline silk proteins.



c β-sheet content (%)

	Salt-induced aggregation			Shear-induced aggregation		
pH	$(AQ)_{24}$	Wild type	D93A	$(AQ)_{24}$	Wild type	D93A
7	32 ± 6	26 ± 3	35 ± 2	32/32/32 ± 1	28/28/25 ± 1	35/34/35 ± 1
6	41 ± 2	32 ± 4	32 ± 1	30/32/31 ± 2	36/33/30 ± 2	28/29/29 ± 1
2	38 ± 1	34 ± 3	35 ± 3	35/33/36 ± 2	31/33/32 ± 2	30/33/30 ± 2



a, A well-defined fibre (dry state) of $(AQ)_{12}$NR3 and ill-defined fibrous aggregate of $(AQ)_{24}$ formed under shear. b, Infrared absorption spectra of $(AQ)_{12}$NR3 and $(AQ)_{24}$ fibres recorded at 0° and 90° relative to the long axis of the fibrous aggregates (black vs red line). c, Estimated β-sheet content of the aggregates. The values of β-sheet content for the aggregates formed under shear were recorded

at 0°, 45° and 90° relative to the long axis of the fibrous aggregates. d, The silk proteins are stored as higher oligomeric assemblies. Exposure of these assemblies to shear and salting out leads to partial unfolding of NR3 and controlled fibre assembly.

To examine the influence of the length of the repetitive backbone on the behaviour of the proteins, we performed aggregation assays with $(AQ)_{24}$NR3, finding that it was significantly more prone to aggregation than the equivalent $(AQ)_{12}$NR3 analogue (Fig. 2a), which leads us to speculate that the α-helical N-terminal NR domain[4] (not included in our recombinant proteins) might also play an important function in protein storage and fibre assembly, especially in case of the natural system, where the proteins are larger.

Our data indicate that the C-terminal NR domain has a role in both spider silk protein storage and fibre assembly (Fig. 4d). During storage of spider silk proteins, the NR domain stabilizes a solution-competent state of the silk protein via the formation of higher supramolecular assemblies, and the C-terminal NR domain also acts as a trigger for the fast, controlled and efficient salting-out of the proteins in combination with the correct alignment of the β-sheet forming repetitive sequence elements during shear-induced assembly. The hydrophobic surface of the assembly state of the NR domain, in addition to the disulphide bridge, provides anchor points for the correct positioning of the repetitive sequence elements. This mechanism is reminiscent of other fibrillar proteins[20] like collagen[21], where the highly conserved terminal pro-peptides form intermolecular disulphide bridges[22] and regulate pro-collagen solubility and lateral assembly[23].

Methods:

Cloning, protein production and purification

The initial construct used for NMR resonance assignment (NR3-T7) was produced and purified as described previously[24]. The gene of a shorter construct (NR3Δ19) used for NOESY experiments was inserted into a pET28a expression system (Merck Biosciences) encoding an N-terminal His_6-tag and expressed in *E. coli* BL21 (DE3) (Agilent Technologies) at 20 °C for 16–20 h. Purification was achieved by passage over a Ni-NTA column (GE healthcare) and a Superdex75 column. Between both columns the hexahistidine tag was removed by digestion with thrombin. Isotopically (^{13}C, ^{15}N) enriched protein samples were produced by growing bacteria in M9 medium supplemented with 1 g l^{-1} $^{15}NH_4Cl$ and 2 g l^{-1} ^{13}C glucose (Eurisotope). For NMR measurements, the proteins were concentrated up to 1 mM in 10 mM sodium phosphate pH 6.0. Addition of 1% (v/v) trifluoroethanol-d_3 (Eurisotope) was used for sample stabilization. The generation of a mixed ^{13}C, ^{15}N–^{12}C, ^{14}N dimeric sample was achieved by mixing equal amounts of labelled and unlabelled His_6-tagged protein samples and a subsequent addition of guanidine hydrochloride to a final concentration of 4 M and 50 mM DTT to achieve unfolding and reduction of the inter-molecular disulphide bridge. For refolding, the mixture was dialyzed against 50 mM Tris/HCl pH 8.0 resulting in an almost quantitative refolding yield. By this procedure, the monomers are allowed to exchange resulting in a 1:2:1 population of the homo-unlabelled, the hetero-labelled and the homo-^{13}C, ^{15}N-labelled species. Genes encoding proteins containing the repetitive elements $(AQ)_{12}$ and $(AQ)_{24}$ (A, hydrophobic polyalanine-rich motif: GPYGPGASAAAAAAGGYGPGSGQQ; Q, hydrophilic glutamine- and glycine-rich motif: GPGQQGPGQQGPGQQGPGQQ) were cloned into a pET29 expression vector (Merck Biosciences) encoding an N-terminal T7-tag and were expressed in *E. coli* strain BLR (DE3) (Agilent Technologies) at 25 °C for 3 h. Purification was carried out as described previously[24]. Site-directed mutagenesis experiments with the gene of NR3Δ19 were done using the QuikChange mutagenesis kit (Agilent Technologies).

Circular dichroism (CD) spectroscopy

CD spectra and stability measurements were carried out on a Jasco J-715 spectropolarimeter (Jasco). For spectra and thermal transitions, a protein concentration of 10 µM was used, and thermal unfolding was monitored by the CD signal at 222 nm in a 1 mm path length cuvette using a heating rate of 60 K h^{-1} in 10 mM sodium phosphate pH 6.0. For chemical unfolding experiments, samples of 5, 34, 110, 200 and 230 µM concentration were incubated with increasing amounts of urea overnight at 4 °C, and the CD signal at 222 nm was recorded with a path length of 0.1 or 1 cm, respectively, at 20 °C. A bandwidth of 5 nm and a response of 2 s were used. Fitting of the thermally and urea-induced unfolding profiles was done as reported previously[28] assuming a two-state folding mechanism.

NMR spectroscopy and structure calculation

Nuclear magnetic resonance (NMR) spectroscopic measurements were conducted at 298 K on Bruker DMX600, DMX750 and Avance900 spectrometers (Bruker Biospin). Backbone resonance assignment was achieved by a set of standard triple resonance experiments[25]. Side-chain assignment was carried out using a combination of CCH-COSY (correlation spectroscopy via the nuclei C,C,H[29]) and HCH-TOCSY

,(total correlation spectroscopy[30]) on double-labelled samples. Stereospecific assignment of prochiral $H^{C\beta}$ methylene and valine methyl groups and the resulting rotamer assignment were made using $^{3}J_{NH\beta}$ couplings observed in the HNHB experiment and patterns of NOESY connectivities. Distance data were derived from a set of three-dimensional NOESY experiments including ^{15}N-HSQC-NOESY, NNH-NOESY on a ^{15}N-labelled sample and ^{13}C-HSQC-NOESY and heteronuclear edited 3D-CCH NOESY and 3D-CNH NOESY experiments[26]. For obtaining inter-subunit NOE distance restraints, we used a ^{14}N,^{12}C-filtered/^{13}C edited two-dimensional-NOESY experiment on a mixed ^{13}C, ^{15}N–^{12}C, ^{14}N labelled sample. NOESY cross-peaks in the three-dimensional spectra were converted into distance ranges after rescaling according to the corresponding HSQC intensities. Cross-peaks were divided into the following four classes: strong, medium, weak and very weak, which resulted in restraints on upper distance of 2.7, 3.2, 4.0 and 5.0 Å, respectively. Lower distance restraints were also included for very weak and absent sequential H^{N}–H^{N} cross-peaks using a minimum distance of 3.2 Å and for medium intensity or weaker sequential and intraresidual H^{N}-H^{α} cross-peaks using a minimum distance of 2.7 Å. Allowance for the use of pseudo atoms (using r^{-6} averaging) were added for methyl groups (0.8 Å for one methyl and 1.5 Å for two methyls) and nonstereospecifically-assigned methylene groups (0.7 Å). Dihedral angle restraints were derived for backbone φ and ψ angles based on C^{α}, C^{β}, C′ and H^{α} chemical shifts using the program TALOS[31]. Restraints were applied for predictions consistent with the input chemical shifts, $^{3}J_{HNH\alpha}$ coupling constants measured from an HNHA experiment and the observed NOE pattern within sequential residues. Accepted predictions were applied using the tolerance calculated by the program ± 5°. The $^{3}J_{HNH\alpha}$ couplings were also applied as direct coupling constant restraints on the backbone φ angles. Hydrogen bond restraints were applied for 73 residues in secondary structural elements with low hydrogen exchange rates, as measured with MEXICO experiments[32, 33] and where donor and acceptor were consistently identified in preliminary calculations. The restraints were applied with the inclusion of pseudo covalent bonds as described previously[34]. Structures were calculated with XPlor-NIH[27] using standard protocols. Experimental restraints were applied to only one monomer, with non-crystallographic symmetry restraints over the backbone of ordered residues (S34–L138) used to assure the symmetry of the dimer. Sets of 50 structures were calculated and a final set of 20 structures chosen on the basis of lowest restraints violations. An average structure was calculated and regularized to give a structure representative of the ensemble. The quality of the regularized average structure was assessed with the program PROCHECK-NMR[35]. Titration experiments were performed with 200–500 µM NR3 in 10 mM sodium phosphate pH 6.0 at 293 K using standard ^{15}N-HSQC experiments. Chemical shift deviations were calculated as the mean of ^{1}H and ^{15}N chemical shift differences using following equation: $\Delta\delta_{av} = \left(\sqrt{\Delta\delta_{H}^{2} + (\Delta\delta_{N}/5)^{2}}\right)/2$.

Fluorescence spectroscopy

Fluorescence spectra of 10 µM ANS and SYPRO Orange in 10 mM sodium phosphate pH 6.0 were recorded with a Fluoromax-4 spectrofluorometer (Horiba Jobin Yvon) using excitation at 350 nm (5 nm bandwidth) and by collecting emission spectra from 420 to 600 nm with a bandwidth of 5 nm. With SYPRO Orange, excitation was set to 485 nm and emission was recorded from 520 to 700 nm. Thermofluor experiments[36] were performed with a real-time PCR machine (Agilent Technologies) by detection of ANS fluorescence (10 µM ANS was used). Heating rate was 30 K h^{-1}, protein concentration was 50 µM and sample volume was 50 µl in a 96-well plate. Five individual measurements were done for each protein sample and solvent condition.

Differential scanning calorimetry (DSC)

DSC experiments were performed using a VP-DSC Micro-Calorimeter (MicroCal). Standard measurements were performed at a protein concentration of 50 µM in 10 mM Tris/HCl, pH 8.0, with a heating rate of 20 K h^{-1}. Evaluation of the LCST behaviour was done using the ORIGIN software (MicroCal).

Rotation-induced aggregation

Before the experiment the proteins were dissolved in guanidine thiocyanate solution (6 M) and then dialyzed against 10 mM Tris/HCl, pH 8.0. Then the proteins were diluted into buffer containing 10 mM Tris/HCl, pH 8 with 50 mM NaCl. Final protein concentrations were adjusted to 40 µM of NR3, 14 µM of AQ_{12}NR3 and 8 µM of AQ_{24}. The samples were incubated for 12 h at 25 °C rotating with 25 r.p.m. After rotation, the clearly visible aggregates were transferred onto glass slides for optical microscopy (DMI3000B, Leica). The samples were then centrifuged at 125,000*g* to remove non-visible aggregates, and the amount of remaining soluble protein in the supernatant was determined using an ultraviolet spectrometer (NanoDrop ND1000, Thermo Fischer).

Fourier transform infrared (FTIR) spectroscopic studies

FTIR spectra of fibres were recorded on a liquid-nitrogen-cooled Bruker Tensor 27 FTIR spectrometer, fitted with a liquid-nitrogen-cooled Bruker Hyperion microscope with IR polarizers supplied by Optometrics Corporation. Spectroscopy was performed on fibres formed during the previously described aggregation assay that were washed with water to remove salts, air dried and placed on highly polished optical grade calcium fluoride disks (Crystal GmbH). Spectra of all aggregates were recorded in absorbance mode at 21 °C, with a 4 cm^{-1} resolution and 600 scans (corrected for background and atmosphere using OPUS software), and polarized FTIR spectra of aggregates of micrometre scale (and larger) were recorded at three different orientations (0°, 45° and 90° relative to the long axis of the fibrous aggregates) to determine if any of the secondary structural elements were aligned with the fibres.

References

1. Dobson, C. M. Protein folding and misfolding. Nature 426, 884–890 (2003).

2. Gosline, J. M., Guerette, P. A., Ortlepp, C. S.&Savage, K. N. The mechanical design

of spider silks: from fibroin sequence to mechanical function. J. Exp. Biol. 202,

3295–3303 (1999).

3. Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. Nature 410,

541–548 (2001).

4. Rising, A., Hjalm, G., Engstrom, W. & Johansson, J. N-terminal nonrepetitive

domain common to dragline, flagelliform, and cylindriform spider silk proteins.

Biomacromolecules 7, 3120–3124 (2006).

5. Motriuk-Smith, D., Smith, A., Hayashi, C. Y. & Lewis, R. V. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. Biomacromolecules 6, 3152–3159 (2005).

6. Sponner, A. et al. The conserved C-termini contribute to the properties of spider silk fibroins. Biochem. Biophys. Res. Commun. 338, 897–902 (2005).

7. Ittah, S., Cohen, S., Garty, S., Cohn, D.&Gat, U. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules 7, 1790–1795 (2006).

8. Jin, H. J. & Kaplan, D. L. Mechanism of silk processing in insects and spiders. Nature 424, 1057–1061 (2003).

9. Knight, D. P. & Vollrath, F. Changes in element composition along the spinning duct in a Nephila spider. Naturwissenschaften 88, 179–182 (2001).

10. Knight, D. P. & Vollrath, F. Liquid crystals and flow elongation in a spider's silk production line. Proc. Roy. Soc. Lond. B266 519–523 (1999).

11. Knight, D. P., Knight, M. M. & Vollrath, F. Beta transition and stress-induced phase separation in the spinning of spider dragline silk. Int. J. Biol. Macromol. 27, 205–210 (2000).

12. Lefe`vre, T., Boudreault, S., Cloutier, C. & Pe´zolet, M. Conformational and orientational transformation of silk proteins in the major ampullate gland of Nephila clavipes spiders. Biomacromolecules 9, 2399–2407 (2008).

13. Heim, M., Keerl, D. & Scheibel, T. Spider silk: from soluble protein to extraordinary fiber. Angew. Chem. Int. Ed. 48, 3584–3596 (2009).

14. Rammensee, S., Slotta, U., Scheibel, T. & Bausch, A. R. Assembly mechanism of recombinant spider silk proteins. Proc. Natl Acad. Sci. USA105, 6590–6595 (2008).

15. Lin, Z., Huang, W., Zhang, J., Fan, J. S. & Yang, D. Solution structure of eggcase silk protein and its implications for silk fiber formation. Proc. Natl Acad. Sci. USA 106, 8906–8911 (2009).

16. Exler, J. H., Hu¨mmerich, D.&Scheibel, T. The amphiphilic properties of spider silks are important for spinning. Angew. Chem. Int. Ed. 46, 3559–3562 (2007).

17. Hardy, J. G. & Scheibel, T. R. Silk-inspired polymers and proteins. Biochem. Soc. Trans. 37, 677–681 (2009).

18. Lavinder, J. J., Hari, S. B., Sullivan, B. J. & Magliery, T. J. High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J. Am. Chem. Soc. 131, 3794–3795 (2009).

19. Ve´zy, C., Hermanson, K. D., Scheibel, T. & Bausch, A. R. Interfacial rheological properties of recombinant spider-silk proteins. Biointerphases 4, 43–46 (2009).

20. Lupas, A. Coiled coils: new structures and new functions. Trends Biochem. Sci. 21, 375–382 (1996).

21. Brown, J. C. & Timpl, R. The collagen superfamily. Int. Arch. Allergy Immunol. 107, 484–490 (1995).

22. Sacca`, B., Renner, C. & Moroder, L. The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics. J. Mol. Biol. 324, 309–318 (2002).

23. Lees, J. F. & Bulleid, N. J. The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. J. Biol. Chem. 269, 24354–24360 (1994).

24. Huemmerich, D. et al. Primary structure elements of spider dragline silks and their contribution to protein solubility. Biochemistry 43, 13604–13612 (2004).

25. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectr. 34, 93–158 (1999).

26. Diercks, T., Coles, M. & Kessler, H. An efficient strategy for assignment of crosspeaks in 3D heteronuclear NOESY experiments. J. Biomol. NMR 15, 177–180 (1999).

27. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. 160, 65–73 (2003).

28. Privalov, P. L. Stability of proteins: small globular proteins. Adv. Protein Chem. 33, 167–241 (1979).

29. Ikura, M., Kay, L. E. & Bax, A. Improved three-dimensional 1H-13C-1H correlation

spectroscopy of a 13C-labeled protein using constant-time evolution. J. Biomol. NMR 1, 299–304 (1991).
30. Kay, L. E., Ikura, M. & Bax, A. Proton–proton correlation via carbon–carbon couplings: a three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon 13. J. Am. Chem. Soc. 112, 888–889 (1990).
31. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13, 289–302 (1999).
32. Gemmecker, G., Jahnke, W. & Kessler, H. Measurement of fast proton-exchange rates in isotopically labeled compounds. J. Am. Chem. Soc. 115, 11620–11621 (1993).
33. Koide, S., Jahnke, W. & Wright, P. E. Measurement of intrinsic exchange-rates of amide protons in a 15N-labeled peptide. J. Biomol. NMR 6, 306–312 (1995).
34. Truffault, V. et al. The solution structure of the N-terminal domain of riboflavin synthase. J. Mol. Biol. 309, 949–960 (2001).
35. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291 (1993).
36. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. 357, 289–298 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is in memoriam of R. Rudolph who triggered spider silk research in the lab of T.S.Wewant to thank D. Hu¨mmerich and J. Exler for initial work on the project, S. Quedzuweit for sample preparation in early stages of the project, M. Heim and M. Suhre for cloning work, and H. Krause for mass spectrometry. This work was supported by the Center for Integrated Protein

Science Munich (CIPSM) (to H.K. and T.S.) and the Deutsche Forschungsgemeinschaft SCHE 603/4-3 (to T.S.). F.H. was supported by the Elitenetzwerk Bayern, CompInt. J.G.H. was supported by the Alexander von Humboldt Foundation.

Author Contributions F.H. designed research, performed cloning work, purified proteins, performed protein folding studies, recorded NMR experiments, performed structure calculations and wrote the manuscript; L.E. performed cloning work, purified proteins and performed aggregation assays. J.G.H. performed aggregation assays, characterized the fibres and wrote the manuscript; C.V. performed cloning work and purified proteins; M.C. was involved in structure calculation and provided software tools for structure calculation and analysis; T.S. designed research and wrote the manuscript; H.K. designed research and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The resonance assignment obtained was deposited at the BMRB data bank under accession code 16249 and the atomic coordinates of the best 20 structures plus a regularized average structure have been deposited at the Protein Data Bank under accession code 2khm. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.K. (horst.kessler@ch.tum.de) or T.S. (thomas.scheibel@uni-bayreuth.de).