

LETTERS

Self-assembly of spider silk proteins is controlled by a pH-sensitive relay

Glareh Askarieh^{1,2*}, My Hedhammar^{3*}, Kerstin Nordling³, Alejandra Saenz⁴, Cristina Casals⁴, Anna Rising³, Jan Johansson³ & Stefan D. Knight²

Nature's high-performance polymer, spider silk, consists of specific proteins, spidroins, with repetitive segments flanked by conserved non-repetitive domains^{1,2}. Spidroins are stored as a highly concentrated fluid dope. On silk formation, intermolecular interactions between repeat regions are established that provide strength and elasticity^{3,4}. How spiders manage to avoid premature spidroin aggregation before self-assembly is not yet established. A pH drop to 6.3 along the spider's spinning apparatus, altered salt composition and shear forces are believed to trigger the conversion to solid silk, but no molecular details are known. Miniature spidroins consisting of a few repetitive spidroin segments capped by the carboxy-terminal domain form metre-long silk-like fibres irrespective of pH⁵. We discovered that incorporation of the amino-terminal domain of major ampullate spidroin 1 from the dragline of the nursery web spider *Euprosthenois australis* (NT) into mini-spidroins enables immediate, charge-dependent self-assembly at pH values around 6.3, but delays aggregation above pH 7. The X-ray structure of NT, determined to 1.7 Å resolution, shows a homodimer of dipolar, antiparallel five-helix bundle subunits that lack homologues. The overall dimeric structure and observed charge distribution of NT is expected to be conserved through spider evolution and in all types of spidroins. Our results indicate a relay-like mechanism through which the N-terminal domain regulates spidroin assembly by inhibiting precocious aggregation during storage, and accelerating and directing self-assembly as the pH is lowered along the spider's silk extrusion duct.

Recombinant mini-spidroins of major ampullate spidroin (MaSp) 1 from *Euprosthenois australis* consisting of four repetitive co-segments with (4RepCT) or without (4Rep) the C-terminal domain, self-assemble within 30 min independent of pH changes in the 6.9–6.3 interval that is relevant to the spinning duct⁶. We therefore speculated that the N-terminal domain is responsible for the pH-sensitivity of native spidroins. Here, 4RepCT and 4Rep, but also mini-spidroins encompassing the N-terminal domain (NT), denoted NT4RepCT and NT4Rep (for details, see Supplementary Table 1), were used to probe the effects of pH and salt on self-assembly and fibre formation (Fig. 1, Supplementary Figs 1 and 2). At pH 7, the NT-mini-spidroins self-assemble into macroscopic structures more slowly (~2 h) than constructs lacking NT (<30 min) (Fig. 1a). This inhibitory effect of the NT domain is even more pronounced at high (pH 8) and low (pH 3) pH levels, where the NT-mini-spidroins require days to assemble (data not shown). In contrast, at pH 6, assembly of NT-mini-spidroins is very fast (<5 min) compared to constructs lacking NT, for which there is no difference in the rate of assembly at pH 6 or 7 (Fig. 1). In line with the ability of the C-terminal domain to promote silk formation⁶, only mini-spidroins that contain the C-terminal

domain form silk-like fibres, although all mini-spidroins that contain a repetitive part assemble into macroscopic structures. At the low concentrations (~1 mg ml⁻¹) used herein, complete maturation into silk-like fibres occurs later (hours) than the first appearance of macroscopic structures. Turbidimetry measurements, which detect much smaller assemblies than the macroscopic structures analysed in Fig. 1a, show that NT-mini-spidroins start to self-assemble within seconds around pH 6.3 (Fig. 1b). At pH 3 as well as at pH 8, NT-mini-spidroins do not form assemblies detectable by turbidimetry within hours (data not shown).

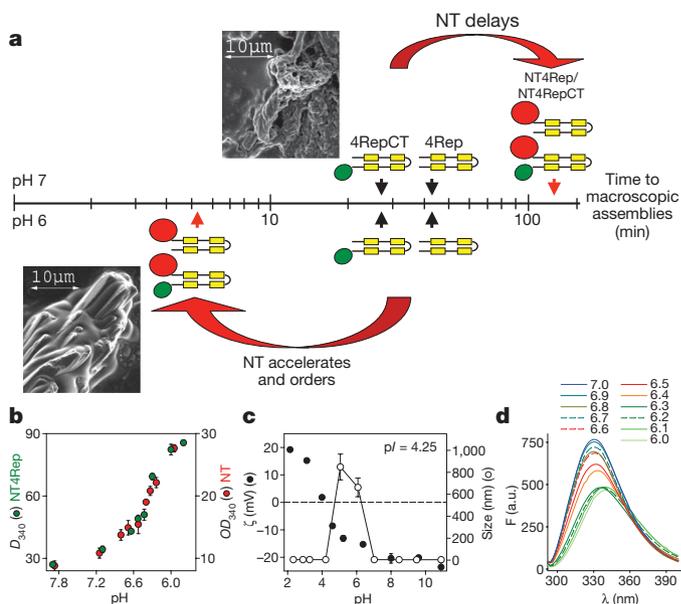


Figure 1 | pH-dependent assembly of NT and mini-spidroins. **a**, Self-assembly of mini-spidroins with NT (NT4RepCT or NT4Rep) or without NT (4RepCT or 4Rep) at pH 7 (above time scale) or at pH 6 (below time scale). The arrows indicate when macroscopic structures were first detected. The scanning electron micrographs are representative for early structures of 4RepCT at both pH 6 and 7 (figure above time axis) and of NT4RepCT at pH 6 (figure below time axis), respectively. The self-assembly process results in continuous fibres only for mini-spidroins including the C-terminal domain (NT4RepCT and 4RepCT). **b**, Turbidimetry of NT and NT4Rep at different pH values measured immediately after sample preparation. Mean values (\pm s.d., $n = 3$) of NT4Rep (green) and NT (red). **c**, Hydrodynamic radii determined by dynamic light scattering and zeta-potentials determined by electrophoretic mobility of NT at different pH values (\pm s.d., $n = 3$). **d**, Tryptophan fluorescence of NT between pH 7.0 and 6.0. The changes are reversible, and in the presence of 0.3 M NaCl no change in fluorescence is observed. a.u., arbitrary units.

¹Department of Chemistry, Oslo University, 1033 Blindern, 0315 Oslo, Norway. ²Department of Molecular Biology, Uppsala BioCenter, SLU, Biomedical Centre, P.O. Box 590, SE-751 24 Uppsala, Sweden. ³Department of Anatomy, Physiology and Biochemistry, SLU, Biomedical Centre, P.O. Box 575, SE-751 23 Uppsala, Sweden. ⁴Department of Biochemistry and Molecular Biology I & CIBER Enfermedades Respiratorias, Complutense University of Madrid, 28040-Madrid, Spain.

*These authors contributed equally to this work

To investigate if the pH-sensitivity of NT-mini-spidroins is intrinsic to NT, this domain alone was analysed using dynamic light scattering, electrophoretic mobility, turbidimetry, size-exclusion chromatography, circular dichroism and fluorescence spectroscopy. NT exists as a highly soluble ($>210 \text{ mg ml}^{-1}$) dimer at pH 7 (ref. 6), but even at low concentrations (0.8 mg ml^{-1}) instantly forms large complexes with a hydrodynamic size of $\sim 700 \text{ nm}$ as the pH approaches 6.3 (Fig. 1b, c). Assembly of NT is readily reversed at a pH above 6.4 or below 4 (Fig. 1c), and is also counteracted by high levels of salt. The kinetic effects of NT on fibre formation of NT-mini-spidroins (Fig. 1a) are diminished by high levels of salt (Supplementary Fig. 2). The pH and salt dependence of NT, NT4Rep and NT4RepCT self-assembly, and lack thereof for 4Rep and 4RepCT, indicates that spidroin assembly depends on titratable groups of NT. Assembly of NT occurs without major changes in zeta potential, and disassembly takes place around the isoelectric point (Fig. 1c). At the same time, limiting the effect of any long-range repulsive interactions by high levels of salt does not allow assembly of NT above pH 6.4 or below pH 4. These results indicate that assembly of NT around pH 6.3 requires specific complementary surfaces that do not exist at higher or lower pH values.

The NT dimer, with a molecular weight of 28.4 kDa⁶ and a radius of gyration of $2.2 \pm 0.2 \text{ nm}$, determined by dynamic light scattering (Fig. 1c) and small-angle X-ray scattering (unpublished data), does not dissociate under any tested condition. This, and the single melting transition observed by differential scanning calorimetry⁶ indicates cooperative subunit unfolding and dissociation, as expected from a tight dimer interaction. Moreover, circular dichroism⁶, 8-anilino-naphthalene-1-sulphonic acid fluorescence and melting temperature (data not shown) show no significant changes in the pH interval 7–3, with or without salt, indicating that the NT structure and dimer integrity is maintained under all tested conditions. However, fluorescence of the only tryptophan of NT, Trp 10, shifts with a similar, reversible, pH dependence as the increase in turbidity (Fig. 1d), indicating that the environment around this residue is changed upon assembly.

We determined the X-ray structure of NT to 1.7 Å resolution (Supplementary Table 2). NT folds as an up-and-down five-helix bundle, with no structural homologues found using the Dali algorithm⁷. In the asymmetric unit, two NT molecules pack together in a tight homodimer (Fig. 2) mainly through helix-stacking interactions involving helices H2, H3 and H5 (Supplementary Fig. 3). Approximately 20% of the subunit surface area ($\sim 1,300 \text{ Å}^2$ per subunit) is buried in the dimer interface. The interface is largely hydrophobic ($\sim 70\%$ of the accessible surface area) and conserved throughout all known spidroins (Fig. 2, Supplementary Figs 3 and 4). Two conserved sequence motifs, AxxxAxASS (residues 68–76) and TTGxxNxxF (residues 108–116), are situated at the dimer interface (H3) and at one end of a shallow crevice across the poles of the dimer, respectively (Fig. 2a, b and Supplementary Figs 3 and 4). The AxxxAxASS motif is important for maintaining the integrity of the dimer with symmetry-related inter-subunit contacts, enabling H3 from the two subunits to pack closely together (Supplementary Fig. 3). A solution structure of a fragment of a spidroin N-terminal domain that after refolding forms a monomer in the presence of detergent micelles has been reported⁸. This structure is also α -helical but is otherwise quite different from ours.

Each subunit in the NT dimer has a markedly non-uniform charge distribution with a clear dipole moment (Supplementary Fig. 5a) that is expected to be maintained in other spidroins (Supplementary Fig. 5b). Acidic and basic residues are clustered at opposite ends of the five-helix bundle, with the subunit dipole axis approximately parallel to the helices. The two dimer subunits are arranged with their dipole axes in opposite directions (Supplementary Fig. 5a). Most of the charged residues are exposed in the dimer; three (His 6, Arg 60 and Lys 65) basic residues at the positive pole of each subunit and five (Asp 39, Asp 40, Glu 84, Glu 85 and Asp 134) of the eight acidic residues at the negative subunit poles (Fig. 2a, d). Glu 79 and Glu 119 are not exposed, but positioned in the near vicinity of the other negative residues. Hence

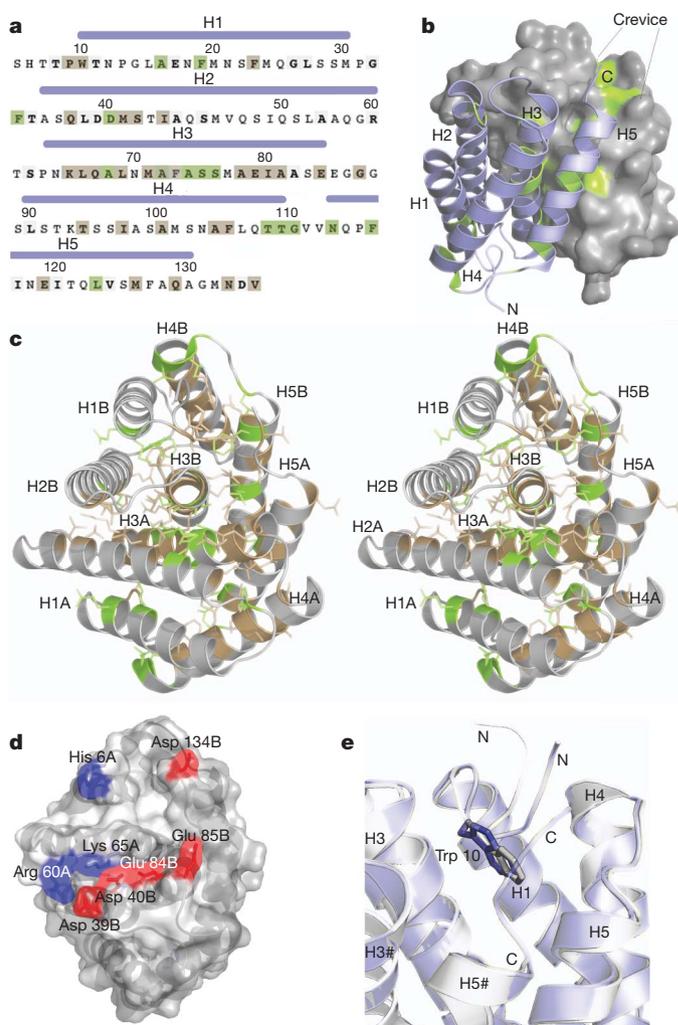


Figure 2 | Structural and conserved features of spidroin N-terminal domain. **a**, Amino acid sequence of NT. Numbering as in the deposited structure. Secondary structure elements are labelled H1–H5. Strictly conserved residues and residues appearing in $>60\%$ of sequences (Supplementary Fig. 4) are shown with green and brown background, respectively. **b**, Overall structure of NT homodimer. Subunits A (grey) and B (blue) are shown in surface and ribbon representations, respectively. Strictly conserved residues are highlighted in green. **c**, Stereo ribbon diagram of the NT dimer showing the distribution of conserved residues (sticks colour coded as in **a**). **d**, Surface representation of the exposed charged residues on NT dimers, with acidic (red) and basic (blue) residues shown as sticks. (Viewed at 90° rotation compared to **b**). **e**, Superposition^{13,14} of the subunits showing conformational differences in the N- and C-terminal regions (chain A in grey, chain B in blue). The hash-sign indicates structural elements of the neighbouring subunit.

each end of the dimer displays one (highly) negatively and one juxtaposed (less) positively charged region.

Given that assembly of NT at pH ~ 6.3 involves protonatable groups, and the distribution of such groups on the dimer surface, we speculated that the charged surfaces at each end of the NT dimer, and in particular groups with pK_a values near the transition point, might play a role in the assembly process. The only residue in NT that would normally have a pK_a value close to the pH 6.3 transition point is the non-conserved His 6. A deletion mutant (NT $_{\Delta\text{His}6}$) shows similar characteristics as NT, and self-assembles at the same pH transition point (Supplementary Fig. 6a). Self-assembly of NT $_{\Delta\text{His}6}$ is again readily reversible, can be inhibited with salt (data not shown), and is not accompanied by any significant structural changes (Supplementary Fig. 6b). Hence, the observed pH effects are not caused by His 6, nor is His 6 required for assembly.

Clustering of acidic residues, as in NT, is expected to increase the side-chain pK_a values. The local environment may shift the pK_a of acidic or basic amino acid side-chains by as much as several pH

units^{9,10}. In the NT structure, two of the three conserved acidic residues (Glu 79 and Glu 84) are predicted^{11,12} to have significantly increased pK_a values. Glu 84 is involved in a clear example of a charge clustering effect together with the third conserved acidic residue Asp 40 (Fig. 2a, d). The carboxylate side chains of Asp 40 and Glu 84 form a tight 'handshake' interaction¹⁰. To probe the role of Asp 40, Glu 79 and Glu 84 in the assembly process, they were individually replaced by the respective isosteric, non-titratable residues Asn and Gln, and Asp 40 and Glu 84 were also mutated together. The X-ray structures for the NT_{D40N}, NT_{E79Q} and NT_{D40NE84Q} mutants were very similar to the wild-type NT structure (NT_{D40N} r.m.s.d. 0.200 Å/252 C α positions; NT_{E79Q} r.m.s.d. 0.516 Å/255 C α positions; NT_{D40NE84Q} r.m.s.d. 1.067 Å/253 C α positions; r.m.s.d., root mean squared deviation). As for wild-type NT, assembly of all mutants was triggered around pH 6.3 (Supplementary Fig. 7). Hence none of these residues alone, nor Asp 40 and Glu 84 together, are essential for determining the pH at which transition from soluble dimers to larger assemblies occurs. However, NT_{E84Q}, NT_{D40N} and NT_{D40NE84Q} are apparently less soluble than wild-type NT and have altered assembly properties compared to wild-type NT. At $\sim 1 \text{ mg ml}^{-1}$ these mutants are turbid already at pH 8, and the turbidity increases further at pH 6.4–6.3 (data not shown). At $\sim 10 \text{ mg ml}^{-1}$ both NT_{D40N} and NT_{D40NE84Q} form gels already at pH 8. Assembly of NT_{E84Q} and NT_{D40NE84Q} is irreversible, and NT_{D40N} and NT_{D40NE84Q} do not show the shift in Trp fluorescence between pH 7 and 6 that accompanies wild-type NT assembly (Fig. 1e and Supplementary Fig. 8). Asp 40 and Glu 84 are thus required for wild-type-like solubility and reversible assembly. These two conserved acidic residues are likely to be directly involved in assembly interactions, or in generating conformational changes that are required for functional assembly.

The core of the NT five-helix bundle is essentially identical in the two subunits of the dimer, but significant differences are observed for the non-helical N- and C-terminal regions (Fig. 2e). The observed pH-induced shift in tryptophan fluorescence (Fig. 1e) confirms that self-assembly is accompanied by changes in the N-terminal region of NT that position the Trp 10 side chain in a less restricted and more polar environment. In our structures, Trp 10 occupies a wide pocket at one end of a conserved crevice cutting through the dimer interface across the basic pole of the subunits (Fig. 2). This crevice is formed by the beginning of H3, the loop connecting H4 and H5 (TTGxxNxxF motif), and the N-terminal region (residues 6–11) of one subunit, and the end of H3 and H5 in the second subunit. In one of the subunits, the C-terminal linker that would connect the N-terminal domain to the following repetitive spidroin segments, packs into the crevice of the neighbouring subunit (Fig. 2e). In the second subunit, conformational changes in the subunit's N-terminal region, in particular around Pro 9 and Trp 10, close off the crevice and prevent binding of the C-terminal linker (Fig. 2e). The C-terminal residues of this subunit are disordered and have not been modelled in the structure. Docking of the linker is expected to orient and restrict the movement of the initial segments of the repetitive region with respect to the N-terminal domain. Our data does not allow us to determine whether self-assembly restricts the organization of the linker region to maintain it exclusively in the free or in the docked conformation. However, it is clear that the observed conformational differences may affect the ability of the repetitive segments to interact and convert into silk, since docking into the crevice will lock and direct the repetitive segments closest to the N-terminal domain, whereas they are expected to be much more mobile when the linker is free. Possibly, assembly of NT dimers is accompanied by orientation and ordering of the repetitive regions into silk-like fibres, which could be the reason why early assemblies of NT4RepCT formed at pH 6 look more ordered than 4RepCT (Fig. 1a).

Our results present the first insights into pH regulation of spider silk formation at the atomic level. The charge distribution of the N-terminal domain provides a module, the assembly of which can be precisely

controlled by pH, allowing the intrinsic pH gradient of spider silk glands to regulate silk formation. The C-terminal domain, on the other hand, is indifferent to pH and affects silk formation by ordering assembly of repetitive segments into fibres. Our study provides an important stepping-stone for further understanding of the remarkably fine-tuned process of spider silk formation. Such understanding has implications for many important issues relating to protein self-assembly as well as to generation of novel bio-degradable materials.

METHODS SUMMARY

NT and mini-spidroins were produced essentially as described^{5,6}. Further details and full methods for crystal growth, structure determination and refinement, as well as analysis of fibre formation and self-assembly, and any associated references, are available in the online version of the paper at www.nature.com/nature.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 June 2009; accepted 24 February 2010.

- Candelas, G. C. & Cintron, J. A spider fibroin and its synthesis. *J. Exp. Zool.* **216**, 1–6 (1981).
- Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. & Hayashi, C. Y. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS ONE* **2**, e514.
- 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).
- Iltah, 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).
- Stark, M. *et al.* Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. *Biomacromolecules* **8**, 1695–1701 (2007).
- Hedhammar, M. *et al.* Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from *Euprosthenops australis*: implications for fiber formation. *Biochemistry* **47**, 3407–3417 (2008).
- Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
- 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).
- Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **284**, 13285–13289 (2009).
- Flocco, M. M. & Mowbray, S. L. Strange bedfellows: interactions between acidic side-chains in proteins. *J. Mol. Biol.* **254**, 96–105 (1995).
- Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins* **61**, 704–721 (2005).
- Gordon, J. C. *et al.* H++: a server for estimating pK_a s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, (Web Server issue) W368–W371 (2005).
- Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

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

Acknowledgements This work was supported by grants (to J.J., M.H. and S.D.K.) from the Swedish Research Council and FORMAS. We thank E. Andersson, K. Tars and D. Ericsson for collecting the high resolution wild-type X-ray diffraction data set, H. Bramfeldt and C. Aulin for help with collecting scanning electron microscopy images, L. Serpell, I. Andersson, A. Zavialov and T. Härd for comments on the manuscript, and a number of other colleagues for discussions. We thank the ESRF (Grenoble, France) beamline staff for help during data collection.

Author Contributions G.A. and M.H. contributed equally to this work. G.A., M.H. performed experiments and wrote the paper, K.N., A.S. performed experiments, C.C., A.R., J.J., S.D.K. discussed experiments and wrote the paper.

Author Information X-ray crystallographic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB) with PDB ID codes 3LR2, 3LR6, 3LR8, 3LRD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.D.K. (stefan.knight@molbio.slu.se) or J.J. (jan.johansson@afb.slu.se).

METHODS

Protein expression and purification. Expression vectors were constructed to produce NT and mutants thereof, NT4Rep, 4RepCT, NT4RepCT, 4Rep and the C-terminal domain respectively, as C-terminal fusions to His₆TrxHis₆ (His₆-Thioredoxin-His₆). The different vectors were used to transform *Escherichia coli* BL21(DE3) cells (Merck Biosciences) that were grown at 30 °C in Luria-Bertani medium containing kanamycin to an *OD*₆₀₀ of ~1, induced with isopropyl-β-D-thiogalactopyranoside, and further incubated for up to 4 h at room temperature. Lysis, immobilized metal affinity purification and proteolytic removal of the His₆TrxHis₆-tag was performed as described⁶.

Crystallization. A 137-residue construct encompassing NT plus four additional N-terminal residues originating from the vector was used for crystallization. Crystals were obtained by the hanging-drop method at 293 K. Two different crystal forms were obtained, both with two NT protomers in the asymmetric unit (Supplementary Table 2). Crystals used to collect high-resolution native data were grown in 22% (w/v) PEG 8000 and 0.2 M ammonium sulphate, using ~16 mg ml⁻¹ protein in 20 mM Tris-HCl, pH 7.0. Crystals used for derivative soaks were grown from 1.2–1.6 M ammonium sulphate and 3–10% (w/v) PEG 400. A useful Pt derivative was produced by soaking in 5 mM platinum ammonium nitrite for 1 h with subsequent back-soaking. All crystals were cryo-protected in mother liquor containing 25% glycerol and flash-frozen in liquid nitrogen until data collection. All mutant crystals crystallized in similar conditions as the derivative crystal. The same cryoprotectant and procedure was used before data collection.

Data collection, processing, scaling and structure determination. All data sets were collected at the European Synchrotron Radiation Facility, France, using an ADSC Q315 detector on beam lines ID14-2 (native, wild type), ID14-4 (three wavelength anomalous dispersion, wild type) (Supplementary Table 2), ID23-1 and ID23-2 (native, mutants). All data were processed with MOSFLM¹⁵ and scaled in SCALA^{14,16}. The high-energy remote anomalous data set provided the strongest anomalous signal and was sufficient for the initial phasing. The structure was solved using the PHENIX software suite^{17–20}, where eight heavy atom sites were initially found by HYSS and the experimental phases were obtained and subsequently improved using Solve and RESOLVE. AutoSol successfully built 235 out of the 274 residues of the dimer. For phasing of the wild-type and mutant native data sets by molecular replacement, one subunit was used as a search model using Phaser²¹.

Model building and refinement. Simulated annealing and refinement was performed in CNS as an initial step, followed by subsequent rebuilding in Coot²² and O²³, and refinement in REFMAC5²⁴. In early refinement steps, strict twofold non-crystallographic symmetry restraints were applied to the dimer in the asymmetric unit, but were gradually removed towards the end of the refinement. Translation, libration and screw rotation displacement (TLS) refinement^{25,26} was applied in final refinement cycles with the complete dimer divided into twenty TLS groups. The final wild-type native model consists of 252 amino acids (residues 6–130 in chain A and 6–134 in chain B), and 137 water molecules. Data and model statistics are summarized in Supplementary Table 2.

Sequence alignment. Published sequences of spider N-terminal domains were aligned using CLUSTALW (MACVECTOR). Sequences for alignment were taken from ref. 27 plus four new entries: *Latrodectus hesperus* (Lh_MaSp1; ABY67414), *Nephila clavipes* (Nc_MaSp1; ACF19411), *Latrodectus geometricus* (Lg_MaSp2; ABY67417) and *Latrodectus hesperus* (Lh_MaSp2; ABR68855).

Dynamic light scattering. The effect of pH and ionic strength on the hydrodynamic diameter of NT and mutants thereof was measured at 25 ± 0.1 °C in a Zetasizer Nano S from Malvern Instruments, equipped with a 633 nm HeNe laser. The buffers were filtered through nylon filters before use. The sample volume was 50 μl and ZEN2112 low glass cuvettes were used. The attenuation and measurement positions from the cuvette wall (4.65 mm) were kept constant for all analyses. Six scans were performed for each sample. All samples were analysed in triplicate. The hydrodynamic diameter (*d*_H) was calculated using the General Purpose algorithm in the Malvern software for dynamic light scattering analysis, which correlates the diffusion coefficient to the hydrodynamic diameter through the Stokes–Einstein equation. The viscosity and refractive index values of the solvent were obtained from the Malvern software. The Multiple Narrow Modes algorithm was used to verify the results obtained by the General Purpose method. All proteins were analysed at a concentration of 0.8 mg ml⁻¹ after centrifugation at 3,000 r.p.m. (400g) for 2.5 min. Centrifugation at 400g for 2.5 min was

performed to analyse the S-value and molecular mass of non-polymerized protein remaining in solution by sedimentation velocity and sedimentation equilibrium experiments, respectively.

Turbidimetry. Turbidity was estimated from the apparent absorbance at 340 nm of proteins (0.8 mg ml⁻¹) at different pH values at 25 °C, in an SLM 4800S spectrofluorimeter equipped with OLIS electronics and software. Measurements were done directly after sample preparation.

Protein electrostatic potential. The z-potential of wild-type NT and mutants thereof was determined by measuring the electrophoretic mobility using a Zetasizer Nano ZS (Malvern Instruments). The zeta potential in volts (ξ) was calculated by the Smoluchowski equation:

$$\xi = 4\pi \frac{\eta\mu}{D}$$

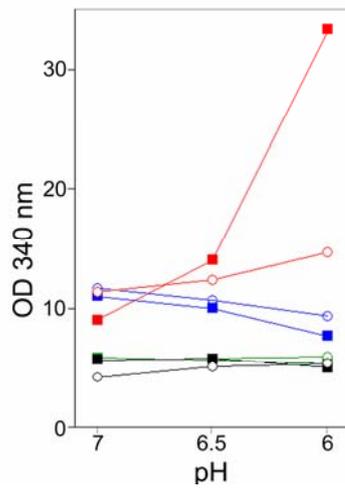
where η is the viscosity of the suspension at 25 °C, D is the dielectric constant of the solution at 25 °C and μ is the electrophoretic mobility of particles (μm s⁻¹ V⁻¹ cm⁻¹). Measurements were performed at a protein concentration of 0.8 mg ml⁻¹ in disposable Zeta cells. All measurements were done in at least triplicate.

Fluorescence. Fluorescence emission spectra were measured in a Hitachi F-400 spectrofluorometer using cells with 1-cm path lengths. Intrinsic tryptophan fluorescence was recorded between 290 and 400 nm (bandwidth 10 nm) by exciting the sample (5 μM) at 280 nm (bandwidth 5 nm). 1,1-bis-(4 anilino) naphthalene-5, 5'-disulfonic acid (bis-ANS) fluorescence experiments were performed on 2 μM of NT incubated with 20 μM bis-ANS for 20 min at 25 °C. The fluorescence spectra of bis-ANS from 400 to 600 nm (bandwidth 10 nm) were obtained with excitation at 395 nm (bandwidth 5 nm).

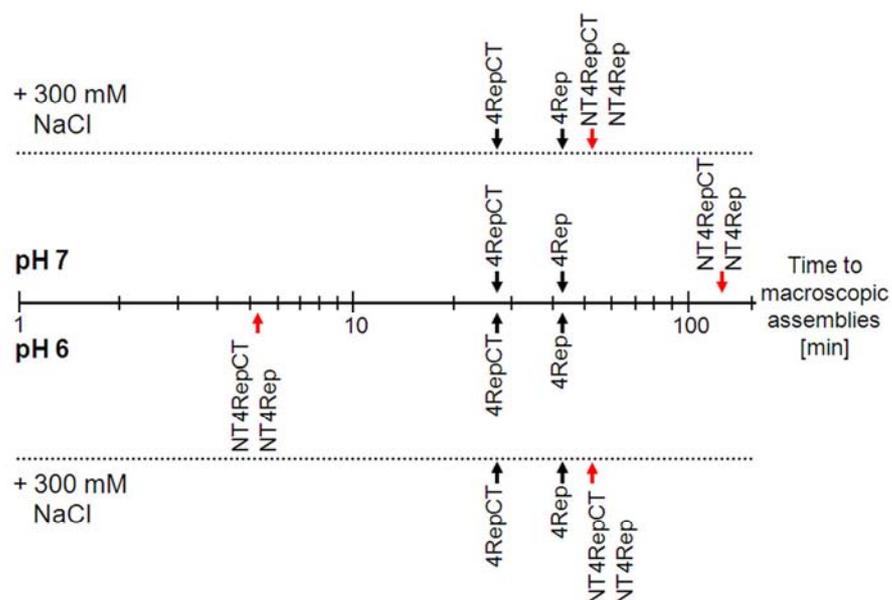
Circular dichroism (CD) spectroscopy. CD spectra from 250 to 190 nm and from 350 to 250 nm were recorded in 0.1 and 1.0 cm, respectively, path length quartz cuvettes using a J-810 spectropolarimeter (Jasco). The scan speed was 50 nm min⁻¹, response time 2 s and with an acquisition interval and bandwidth 1 nm. Temperature scans from 20 to 90 °C were performed using a cuvette equipped with a Teflon lid. Temperature elevation was 0.5 °C min⁻¹.

Fibre formation and scanning electron microscopy (SEM). Conditions for fibre formation were essentially as described⁵. Approximately 25 μM of each protein was incubated in 20 mM Na phosphate buffer at pH 7 or 6, with or without 300 mM NaCl. At different time points samples were applied on SEM stubs, where they were air-dried and vacuum-coated with gold and palladium. The samples were photographed with a LEO 1550 FEG microscope (Carl Zeiss) using an acceleration voltage of 5 kV.

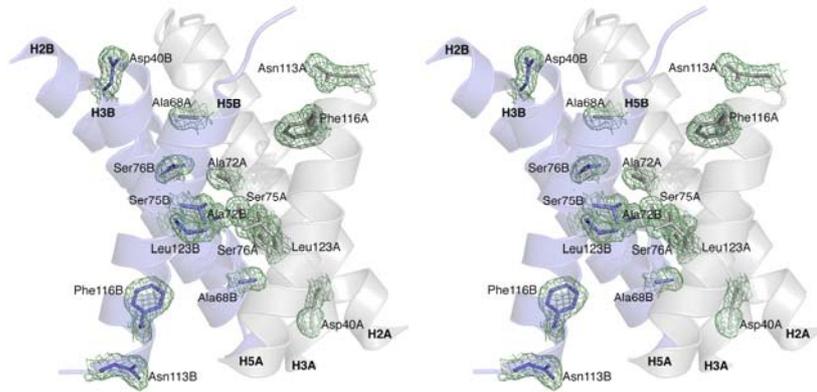
- Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* no. 26, (1992).
- Evans, P. R. SCALA. *Joint CCP4 ESF-EACBM Newsletter Protein Crystallography* 33, 22–24 (1997).
- Weeks, C. M. et al. Automatic solution of heavy-atom substructures. *Methods Enzymol.* 374, 37–83 (2003).
- Sheldrick, G. M. & Schneider, T. R. SHELXL: high-resolution refinement. *Methods Enzymol.* 277, 319–343 (1997).
- Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* 58, 1948–1954 (2002).
- Zwart, P. H. et al. Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* 426, 419–435 (2008).
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* 61, 458–464 (2005).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* 60, 2126–2132 (2004).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119 (1991).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* 53, 240–255 (1997).
- Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* 62, 439–450 (2006).
- Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* 39, 109–111 (2006).
- Rising, A., Hjalms, G., Engstrom, W. & Johansson, J. N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindrical spider silk proteins. *Biomacromolecules* 7, 3120–3124 (2006).



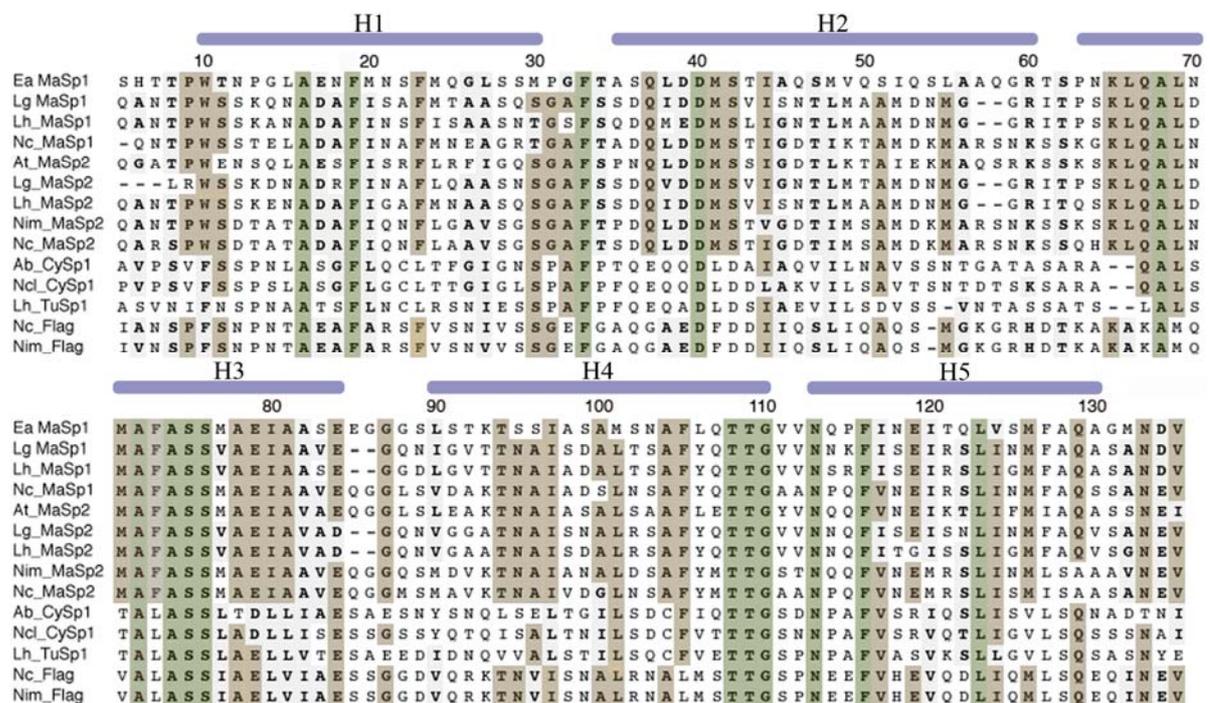
Supplementary figure 1. Turbidimetry of 4Rep (green) and 4RepCT (black) at pH 6-7, with (circles) and without (squares) 300 mM NaCl. As references NT (red) and CT (blue) are also included in the graph.



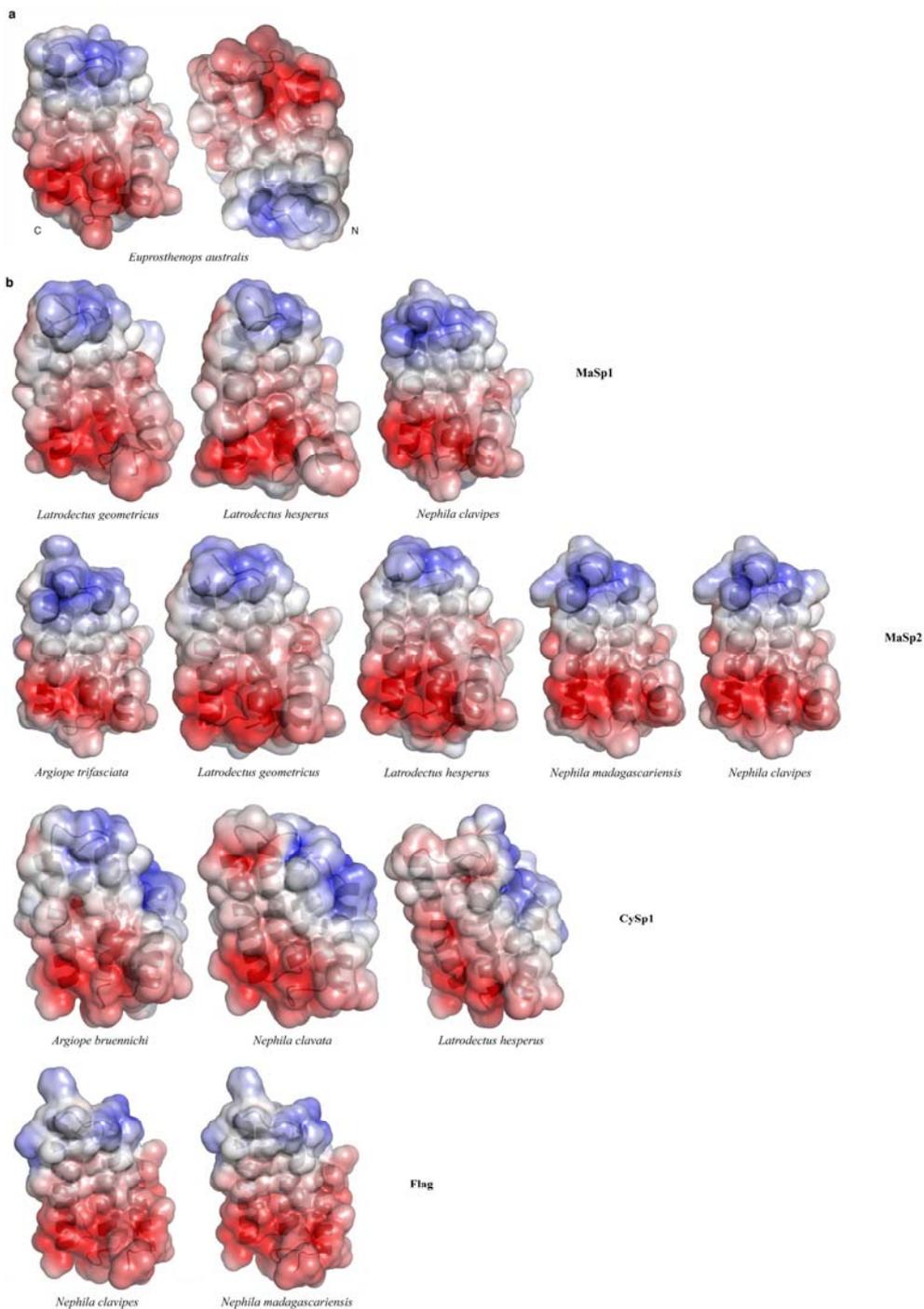
Supplementary figure 2. pH- and salt-dependent assembly of mini-spidroins. Self-assembly of mini-spidroins with NT (NT4RepCT or NT4Rep) or without NT (4RepCT or 4Rep) at pH 7 (above time scale) or at pH 6 (below time scale), with salt (uppermost and lowermost) and without salt (next to time scale). The arrows indicate when macroscopic structures were first detected, showing that at pH 7 the presence of NT delays self-assembly, while at pH 6 it accelerates self-assembly. However, at both pH 7 and 6, the presence of salt counteracts the respective effect conferred by NT.



Supplementary figure 3. Structure of the N-terminal dimer interface and its conserved features. Ribbon representations of the dimer interface, helices H2, H3 and H5, of subunits A in grey and B in blue. Strictly conserved residues are shown as sticks and superimposed on the $2F_o-F_c$ electron density map (contoured at 1σ ; for clarity the map has been masked to only show density surrounding the selected side chains). (Viewed at a 45° rotation compared to Figure 2a).

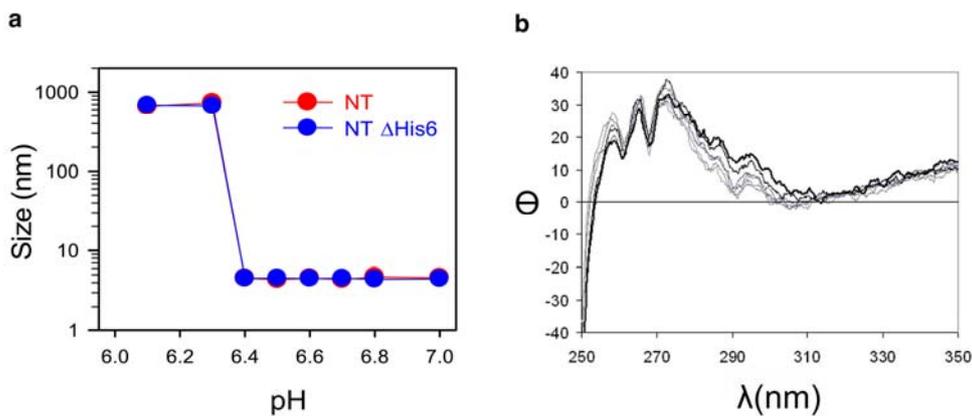


Supplementary figure 4. Sequence alignment of full-length N-terminal domains. The sequences included are, from top to bottom, *Euprosthonops australis* MaSp1 (NT), *Latrodectus geometricus* MaSp1, *Latrodectus hesperus* MaSp1, *Nephila clavipes* MaSp1, *Argiope trifasciata* MaSp2, *Latrodectus geometricus* MaSp2, *Latrodectus hesperus* MaSp2, *Nephila inaurata* *madagascariensis* MaSp2, *Nephila clavipes* MaSp2, *Argiope bruennichi* CySp1, *Nephila clavata* CySp1, *Latrodectus Hesperus* CySp1, *Nephila clavipes* Flag and *Nephila madagascariensis* Flag. Numbering as in the deposited structures. Secondary structure elements are labelled H1-H5. Strictly conserved residues and residues appearing in > 60% of sequences are shown with green and brown background, respectively.

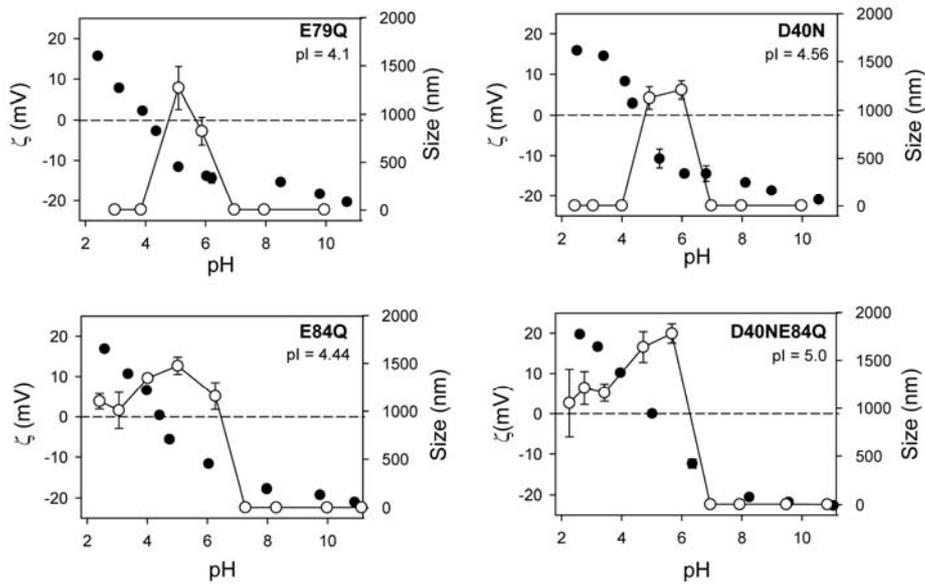


Supplementary figure 5. Preserved non-uniform charge distribution of all spider N-terminal domains. **a**, Electrostatic surface potential at +5 kT/e (blue) and -5 kT/e (red) of the high-resolution NT structure, viewed perpendicular to the dimer interface of subunits A (left) and B (right). **b**, Electrostatic surface potential of all known N-terminal domains shown at +5

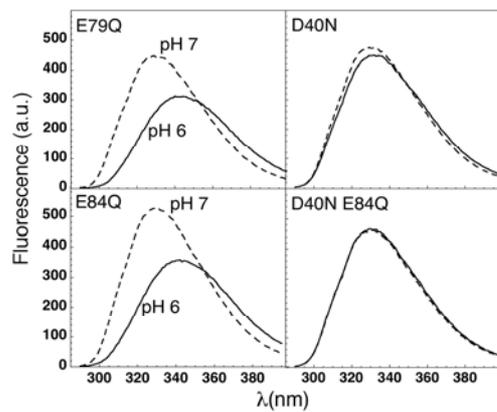
kT/e (blue) and -5 kT/e (red). The view is the same as for NT subunit A in **a**. The homology models are listed in the same order as they appear in the sequence alignment of Supplementary Figure 4. (Figures were made with PyMol; DeLano Scientific, San Carlos, CA).



Supplementary figure 6. Self assembly of NT Δ His6. **a**, Dynamic light scattering showing change of particle size of wt NT and NT Δ His6 as a function of pH. **b**, Near-UV CD spectra of NT Δ His6 at pH 6 (light grey), pH 7 (black) and intermediate pH values.



Supplementary figure 7. Self assembly of NT mutants. Dynamic light scattering and zeta-potentials of NT mutants as a function of pH.



Supplementary figure 8. Tryptophan fluorescence of NT mutants.

Tryptophan fluorescence of NT mutants at pH 7 and 6.

Supplementary table 1. Properties of the different protein constructs used.

Protein	Size (kDa)	Theoretical pI
NT	14.2	4.3
NT _{ΔHis6}	14.2	4.1
NT _{D40N}	14.2	4.5
NT _{E79Q}	14.2	4.4
NT _{E84Q}	14.2	4.4
NT _{D40NE84Q}	14.2	4.6
NT4Rep	26.9	4.8
NT4RepCT	39.6	6.4
4Rep	12.2	9.2
4RepCT	23.4	8.9
CT	12.4	6.1

Supplementary table 2. Unit cell parameters, data collection, phasing, and refinement statistics.

	Native	Platinum derivative
Unit-cell parameters		
Space group	P3 ₂ 21	P6 ₁ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	68.4, 68.4, 97.9	91.7, 91.7, 143.4
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 120	90, 90, 120
Data collection		
Wavelength (Å)	0.9330	1.0675
Resolution (Å)	59-1.7	79-2.7
Observed reflections	1037689	681062
Unique reflections	35243	18634
Completeness (%)	87.5 (84.1)	100 (100)
Multiplicity	5.9 (6.0)	29.6 (31.1)
<i><I / σI></i>	8.4 (1.9)	6.1 (2.1)
<i>R</i> _{merge} (%)	6.0 (41.1)	10.4 (36.5)
Phasing		
Resolution (Å) / Bijvoet pairs		3.0 / 5889
Anomalous signal		0.0743
No. of sites		8
FOM		0.47
Refinement		
Resolution (Å)	59-1.7	
No. reflections		
Total	29725	
Work set	27832	
Test set	1893	
<i>R</i> _{work} / <i>R</i> _{free} (%)	16.0 / 20.5	
Number of atoms		
Protein	1989	
Water	127	
Average B-factor (Å ²)		
Protein (all atoms)	23.3	

Protein (main chain)	22.3
Water	31.2
R.m.s.d. stereochemistry	
Bond lengths (Å)	0.015
Bond angles (°)	1.366
R.m.s.d. B-factors (Å ²)	
Main chain	0.95
Side chain	2.12

Values for the highest resolution shell are shown in parenthesis. $R_{merge} = \sum_h \sum_i |I(h)_i - \langle I \rangle| / \sum_h \sum_i I(h)_i$, where $I(h)$ is the intensity of a reflection h , \sum_h is the sum over all reflections and \sum_i is the sum over i measurements of reflection h . $R_{work} = \sum |F_o - F_c| / \sum F_o$ where F_o and F_c are the observed and calculated structure factors respectively. R_{free} is calculated for a test set of reflections randomly excluded from refinement. B-factors are given with contribution from TLS tensors included. R.m.s.d. stereochemistry is the deviation from ideal values. R.m.s.d. B-factors is deviation between bonded atoms.