Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/9/e1600307/DC1

Supplementary Materials for

Computationally designed peptides for self-assembly of nanostructured lattices

Huixi Violet Zhang, Frank Polzer, Michael J. Haider, Yu Tian, Jose A. Villegas, Kristi L. Kiick, Darrin J. Pochan, Jefferey G. Saven

Published 9 September 2016, *Sci. Adv.* **2**, e1600307 (2016)
DOI: 10.1126/sciadv.1600307

This PDF file includes:

Supplemental Materials and Methods: Computational Design of Peptides with supplemental table of molecules and supplemental figures

Design of homotetrameric helix bundle motif

An antiparallel tetrameric coiled coil with D_2 point group symmetry was selected as the building block for the designed nanomaterials. Structures and sequences of the tetramer were *de novo* designed computationally to identify a single tetrahelical motif that would be robust with respect to variation of exterior residues, residues which could be subsequently designed for specific assemblies.

The construction of the homotetrameric helical (coiled coil) structures used a mathematical model that describes such structures with a small number of geometric parameters.[39] Modifications were made to include rotation of peptide planes (comprising backbone N, C_α and C atoms).[40] Each sequence contained 29 residues to allow approximately four helical heptad repeats. In the model, the superhelix refers to the helical bundle structure formed by the coiling of four alpha helices. An ensemble of the four alpha helical peptides was created by varying a set of associated geometrical parameters: the super-helical radius *R*, the super-helical phase α (α=45° when the neighboring helices are equal-distant from each other), the relative displacement of the ends of the helices parallel to the super-helical axis Z, the minor-helical phase θ (rotation of the alpha helix about its axis), and the super-helical pitch *P* (fig. S1).
Given a set of these geometric parameters that specify the coordinates of a single helix, the remaining three helices were constructed by performing symmetry operations consistent with the D_2 point group.

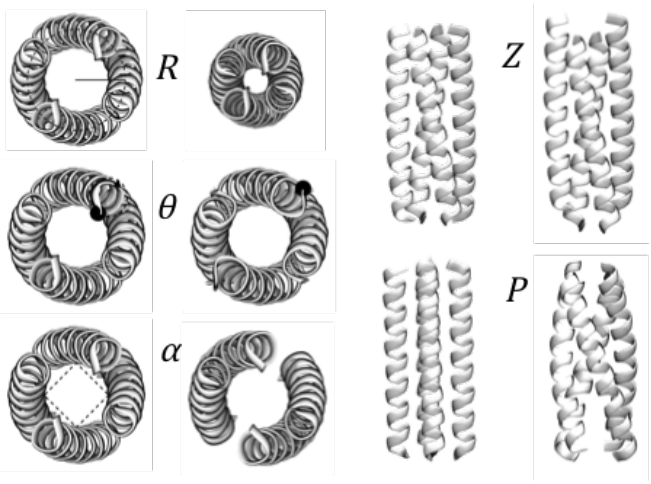

fig. S1. Representative backbone configurations of the helix bundle motif building-block illustrating variation of the geometric parameters associated with the bundle.

Monte-Carlo simulated annealing (MCSA) was used to sample parameters and identify helical bundle sequences and structures as local minima on an energy surface. Let $\{R, \alpha, Z, P, \theta\}$ denote the set of the parameters that characterize a particular backbone scaffold configuration. These parameters were confined within values associated with natural coiled-coil tetramers (*R*: 6.8Å -- 7.8Å, α: 35° -- 55°, *Z*: -1.5Å -- 3Å, *P: -118* --∞).[39] The values of θ were grouped into three categories of the bundle, based upon the location of heptad positions in the structure[41]: θ = 10° -- 30° (interior heptad positions: a, d); θ = -10° -- 10° (interior heptad positions: a, d, e); θ = 30° -- 50° (interior heptad positions: a, d, f).

For each generated helical structure, a statistical sequence design methodology was used to calculate the probabilities of hydrophobic amino acids (A, V, I, L, M, F, Y and W) at interior residue positions. Sites 1 and 28 were located at the end of the helical motif and solvent exposed, thus defined as exterior positions. To estimate the probabilities, a

statistical thermodynamic theory was applied, wherein an entropy or its Legendre transform are optimized subject to constraints on the sequences.[33,42] Herein, the CHARMM22 force field[43] was used. As done in previous work, an average internal energy over the ensemble of sequences was calculated, and its conjugate temperature β^{-1} was specified such that $\beta=0.5$ mol/kcal. A helix propensity scoring function, E_h,[44] was employed and constrained to values expected for helical peptides of the chosen length.[45] For 29-residue helical segments in a database of natural protein structures, E_h takes on values in the range -12.1 *kcal/mol* to -7.5 *kcal/mol*; in the design calculations, values of E_h = -8.88 *kcal/mol*, -9.75 *kcal/mol*, and -10.18 *kcal/mol* were applied in separate Monte Carlo samplings. A symmetry assumption was applied to leverage the symmetry of the D_2 point group.[33,46] For each sampled tetrahelical configuration, an average (internal) energy per peptide was calculated using the site-specific probabilities of the amino acids. This average energy was used as the objective function in 5000-step Monte Carlo searches with exponential annealing schedules.

Ten lowest energy structures were selected. The sequence considered in each case comprised the most probable amino acid at each residue position. For the resulting structures, void volumes were assessed by CastP.[47] The final candidate was chosen because its total and maximum void volumes are comparable to those of 15 crystalized antiparallel homo-tetramers. This helix bundle structural motif, which comprised the tetrahelical structure and identities of 11 interior hydrophobic amino acids (fig. S2), was used in all the subsequent designs targeting distinct assemblies with predetermined nanostructures (Fig. 1, A to D).

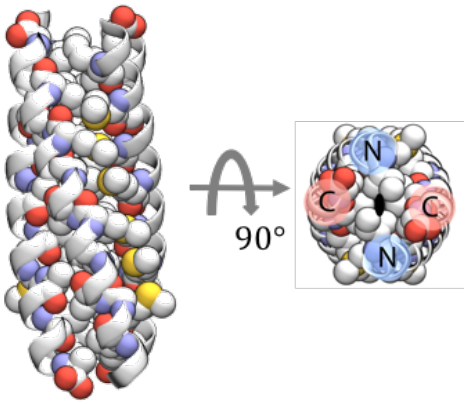

fig. S2. Side and top view of the selected low-energy helix bundle motif with the most probable amino acids at the interior sites shown in space-filling representations: N terminus is depicted in blue and C terminus in red circles

Design of helix bundles targeting materials with predetermined nanostructures

The exterior residue positions of the helix bundles were determined so that self-assembly of the bundle motifs would produce assemblies with chosen, distinct nanostructures. The targeted assemblies were an isolated helix bundle (expected to remain soluble and not form higher order lattice assemblies in solution) and three distinct, two-dimensional lattices (Fig. 2A). Layered lattice symmetries were selected to satisfy the following criteria: 1) the lattice contains internal positions with D_2 point group symmetry on which the bundle motifs were positioned and 2) the lattice contains no axes of skew symmetry perpendicular to each layer. The space groups P222, P422, and P622 were selected. Each is a layer space group, and only a single, isolated layer was used in the design calculations.

Peptide sequences were identified that are compatible with the targeted nanostructure. The statistical mechanical approach mentioned previously was applied and symmetry

assumptions were imposed to incorporate point and space group symmetries. For each targeted material, peptide sequence design was performed with the designed helix bundle building block motif, allowing 18 natural amino acids (all natural amino acids but Pro and Cys) at each of 18 variable exterior residue positions. The designed soluble, non-lattice forming helix bundle is labeled BNDL_1 (Fig. 1A), and the quality of the structure was validated by Molprobity.[48] For design in the context of one of the chosen lattice symmetries, unit cell parameters were modulated to vary the spacing between neighboring bundle motifs in the search for low-energy lattice structures. For P222, the unit cell dimensions *a* and *b* were varied. For P422 and P622, there is only one unique unit cell dimension *a*. For P622, an additional set of distinct lattice structures containing a 90° rotation about an interior C_2 symmetry axis was also considered. A grid search was used (at 0.1 Å increment) to construct the sequence-structure energy landscapes for each lattice nanostructure. For each lattice, the average energy over the sequence probabilities was calculated using the statistical design approach. Low energy candidates were selected from local minima on the landscape of each type of lattice. If Trp or Tyr were not among the most probable amino acid at any of the variable residues, sites where Trp or Tyr were probable were constrained as one of these two amino acids in a subsequent calculations; these residues were introduced to facilitate determination of peptide concentrations. The numbers of lowest-energy sequences selected for further characterization were 4, 5, and 14 for the P222, P422, and P622 lattices respectively. The candidates' Molprobity scores fall within the range observed for 53 solved structures of coiled-coil proteins from the CC+ database. In the last step, PDBePISA[49] was used to assess the candidates' potential to form desired assemblies

and the resultant protein-protein interactions at the bundle-bundle interfaces. Based upon these assessments, a total of 9 designed sequence candidates were chosen for experimental synthesis and characterization (Table 1).

Small angle neutron scattering (SANS):

SANS curves for BNDL_1 were fit using the SasView software[50]. A cylinder fit was performed using the equations below for the form factor of a right circular cylinder with a uniform scattering length density.[51,52] The form factor, P(q), is represented by the following:

$$P(q) = \frac{scale}{V} \int_0^{\pi/2} f^2(q,\alpha) \sin\alpha \, d\alpha + bkg$$

$$f(q) = 2(\Delta\rho)V * \frac{\sin(q\ L\ cos\alpha/2)}{q\ L\ cos\alpha/2} * \frac{J_1(q\ r \sin\alpha)}{(q\ r \sin\alpha)}$$

$$V = \pi r^2 L$$

where $J_1(x)$ is the first order Bessel function. Alpha, α, is defined as the angle between the cylinder axis and the scattering vector, q. The integral over alpha averages the form factor over all possible orientations of the cylinder with respect to q. L is the cylinder length while r is the cylinder radius. The model fit produced a length = 34.4 +/- 1.1 Å and radius = 10.4 +/- 0.2 Å for BNDL_1 as shown in Fig. 2D.

Representative analytical HPLC for designed sequences:

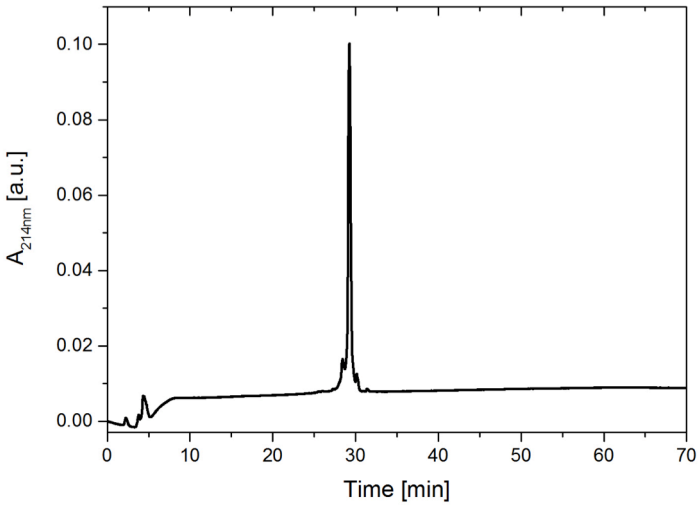

fig. S3. P422_1 analytical HPLC.

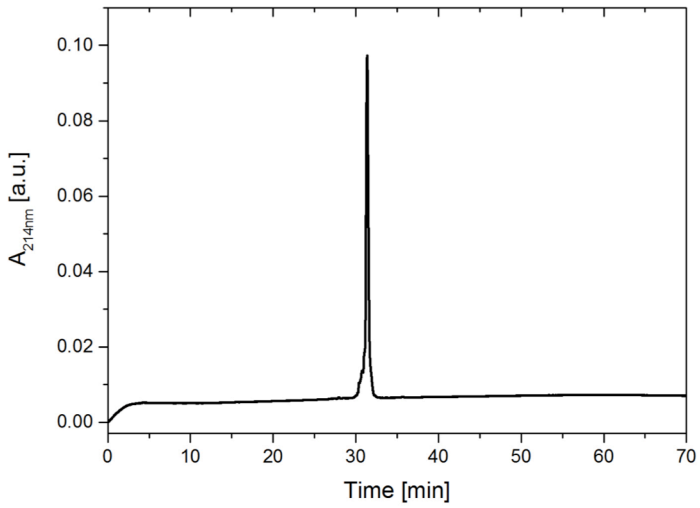

fig. S4. P422_1_ac analytical HPLC

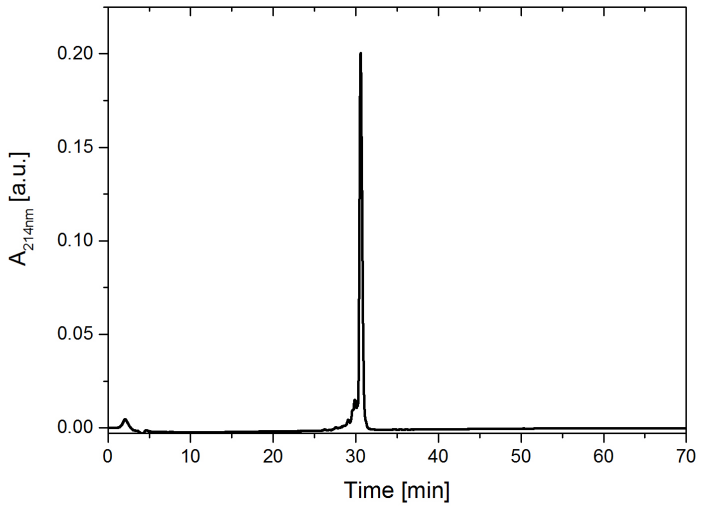

fig. S5. P222_9 analytical HPLC

Representative ESI-mass spectrometry for designed sequences:

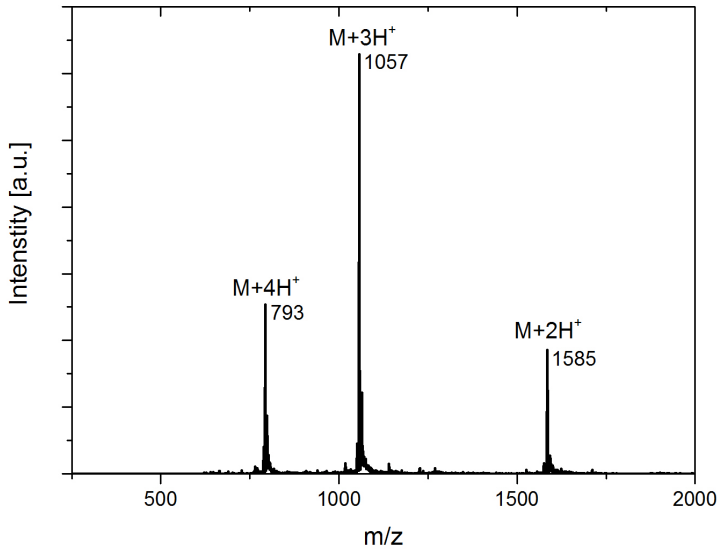

fig. S6. P222_9 sequence, 3168 Da

Calculated mass: [M+2H$^+$] 3170 Da, [M+3H$^+$] 3171 Da, [M+4H$^+$] 3172 Da.

Measured mass: [M+2H$^+$] 3170 Da, [M+3H$^+$] 3171 Da, [M+4H$^+$] 3172 Da.

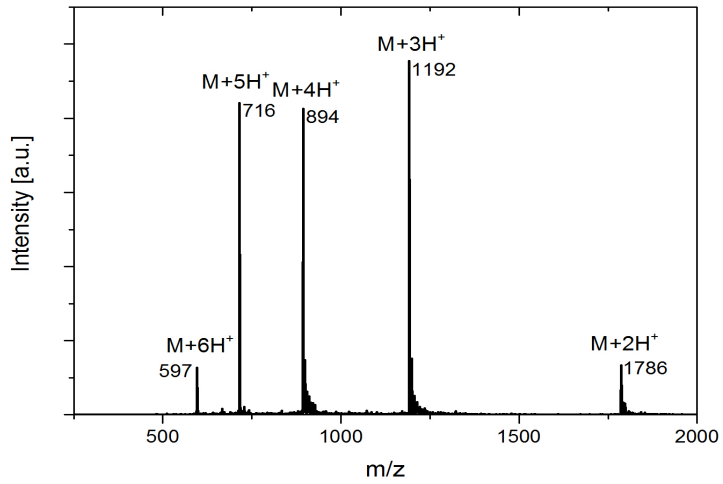

fig. S7. P422_1 sequence, 3572 Da.

Calculated mass: [M+2H$^+$] 3574 Da, [M+3H$^+$] 3575 Da, [M+4H$^+$] 3576 Da, [M+5H$^+$] 3577 Da, [M+6H$^+$] 3578 Da.

Measured mass: [M+2H$^+$] 3572 Da, [M+3H$^+$] 3576 Da, [M+4H$^+$] 3576 Da, [M+5H$^+$] 3580 Da, [M+6H$^+$] 3582 Da.

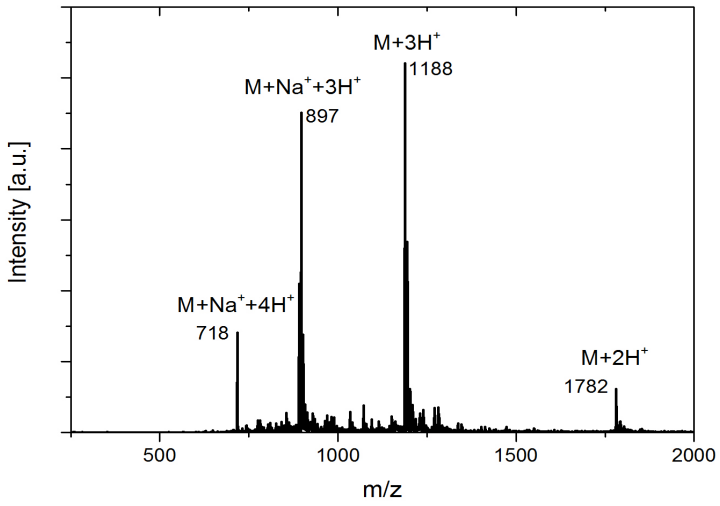

fig. S8. BNDL_1, 3560 Da.

Calculated mass: [M+2H$^+$] 3562 Da, [M+3H$^+$] 3563 Da, [M+Na$^+$+3H$^+$] 3586 Da, [M+Na$^+$+4H$^+$] 3587 Da.

Measured mass: [M+2H$^+$] 3564 Da, [M+3H$^+$] 3564 Da, [M+Na$^+$+3H$^+$] 3588 Da, [M+Na$^+$+4H$^+$] 3590 Da.

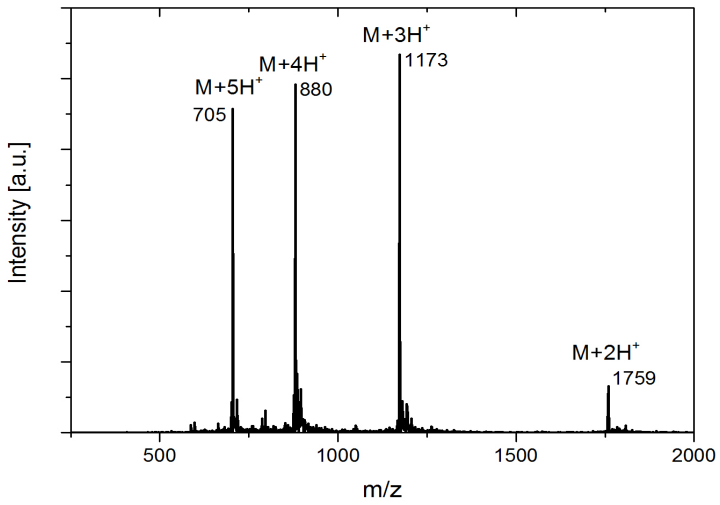

fig. S9. P622_6, 3517 Da.

Calculated mass: [M+2H$^+$] 3519 Da, [M+3H$^+$] 3520 Da, [M+4H$^+$] 3521 Da, [M+5H$^+$] 3522 Da.

Measured mass: [M+2H$^+$] 3518 Da, [M+3H$^+$] 3519 Da, [M+4H$^+$] 3520 Da, [M+5H$^+$] 3525 Da.

Representative circular dichroic spectroscopy data for designed sequences:

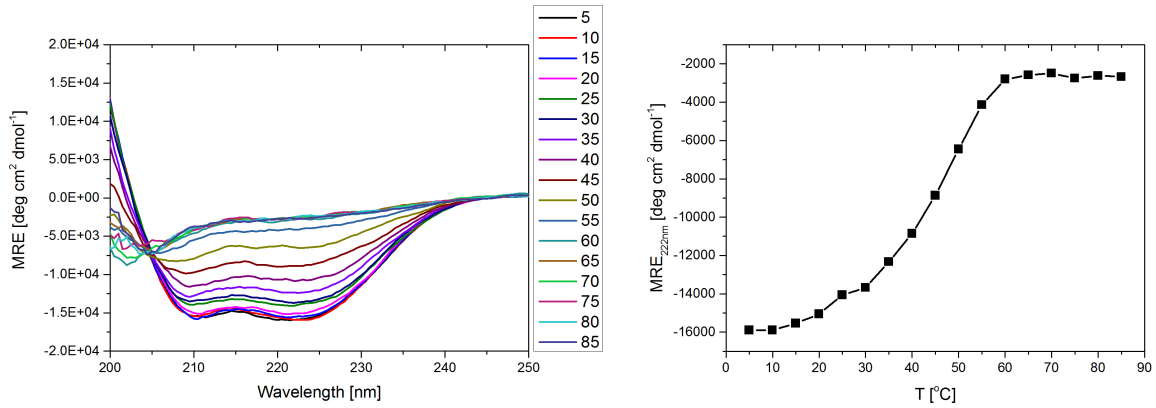

fig. S10. P222_9, 0.1 mM in Borate buffer pH10 on heating. Wavelength scans (left) at temperatures shown with different color. MRE at 222 nm wavelength plotted vs. temperature (right) showing melting of coiled coil helix bundle.

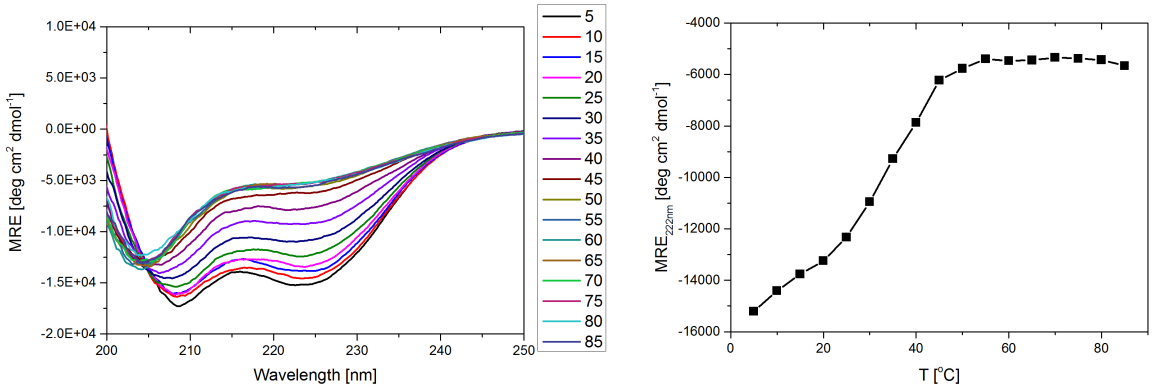

fig. S11. P622_6, 0.1 mM in phosphate buffer pH7 on heating. Wavelength scans (left) at temperatures shown with different color. MRE at 222 nm wavelength plotted vs. temperature (right) showing melting of coiled coil helix bundle.

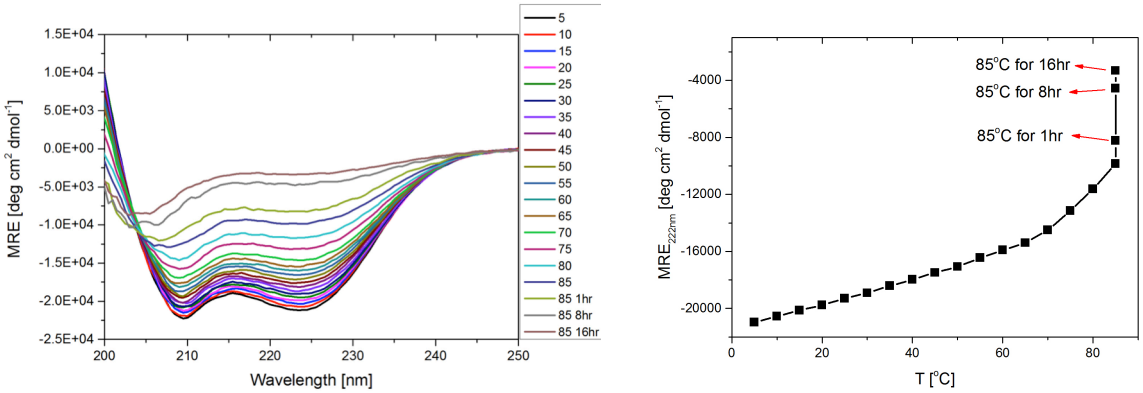

fig. S12. BNDL_1, 0.1 mM in borate buffer pH 10 on heating. Wavelength scans (left) at temperatures shown with different color. MRE at 222 nm wavelength plotted vs. temperature (right) showing melting of coiled coil helix bundle. The stability of the BNDL_1 nanostructure is revealed in the excess time needed at high temperature in order to melt the coiled coil structure.

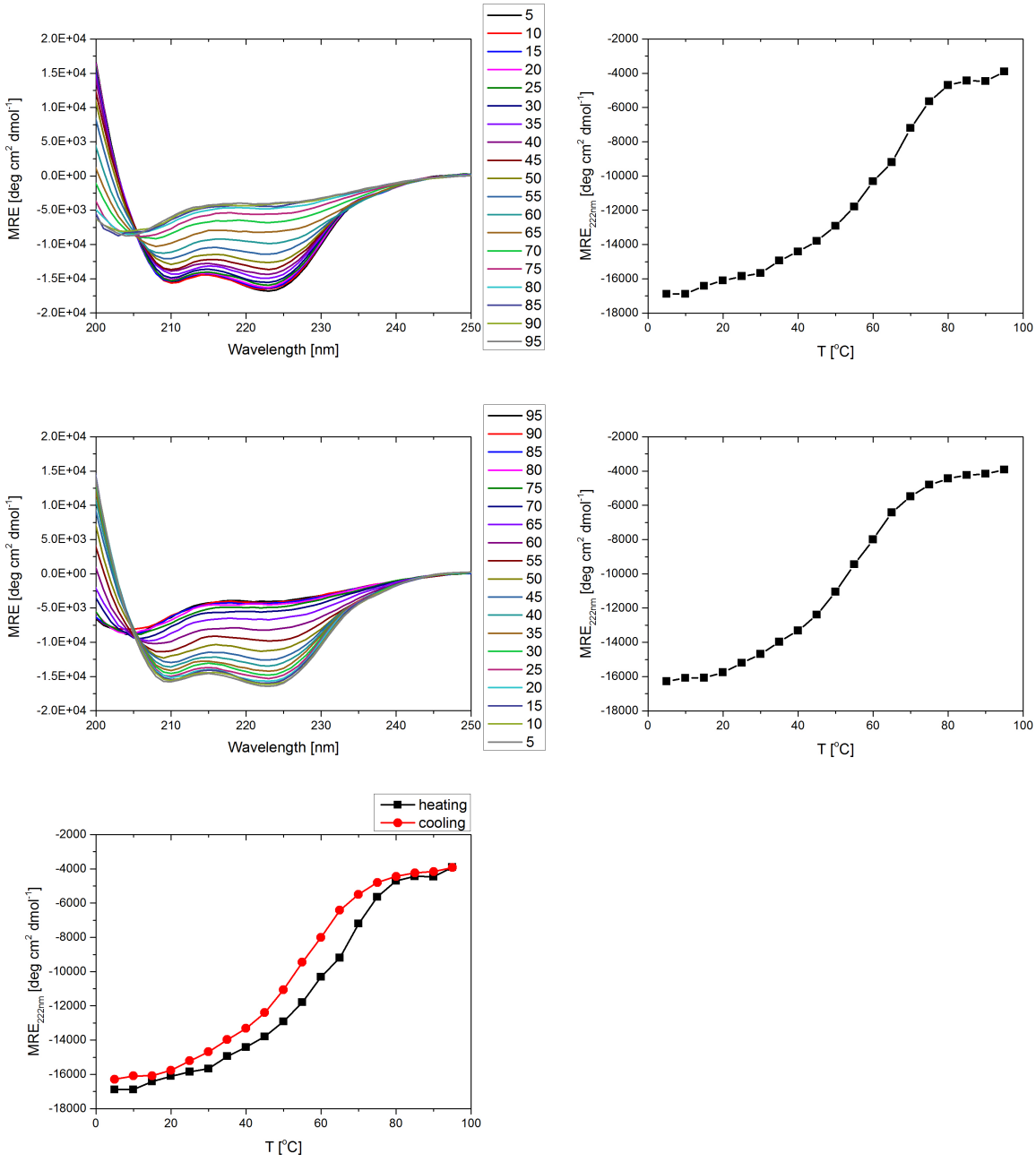

fig. S13. 0.1mM P422_1 in Borate buffer pH10. Top left: wavelength scan at different temperatures on heating. Top right: MRE at 222 nm on heating. Middle left: wavelength scan at different temperatures on cooling. Middle right: MRE at 222 nm on cooling. Bottom: overlay of MRE at 222 nm from both heating and cooling wavelength scans showing little difference in coiled coil melting/formation temperature range.

Analytical ultracentrifugation data for peptide BNDL_1:

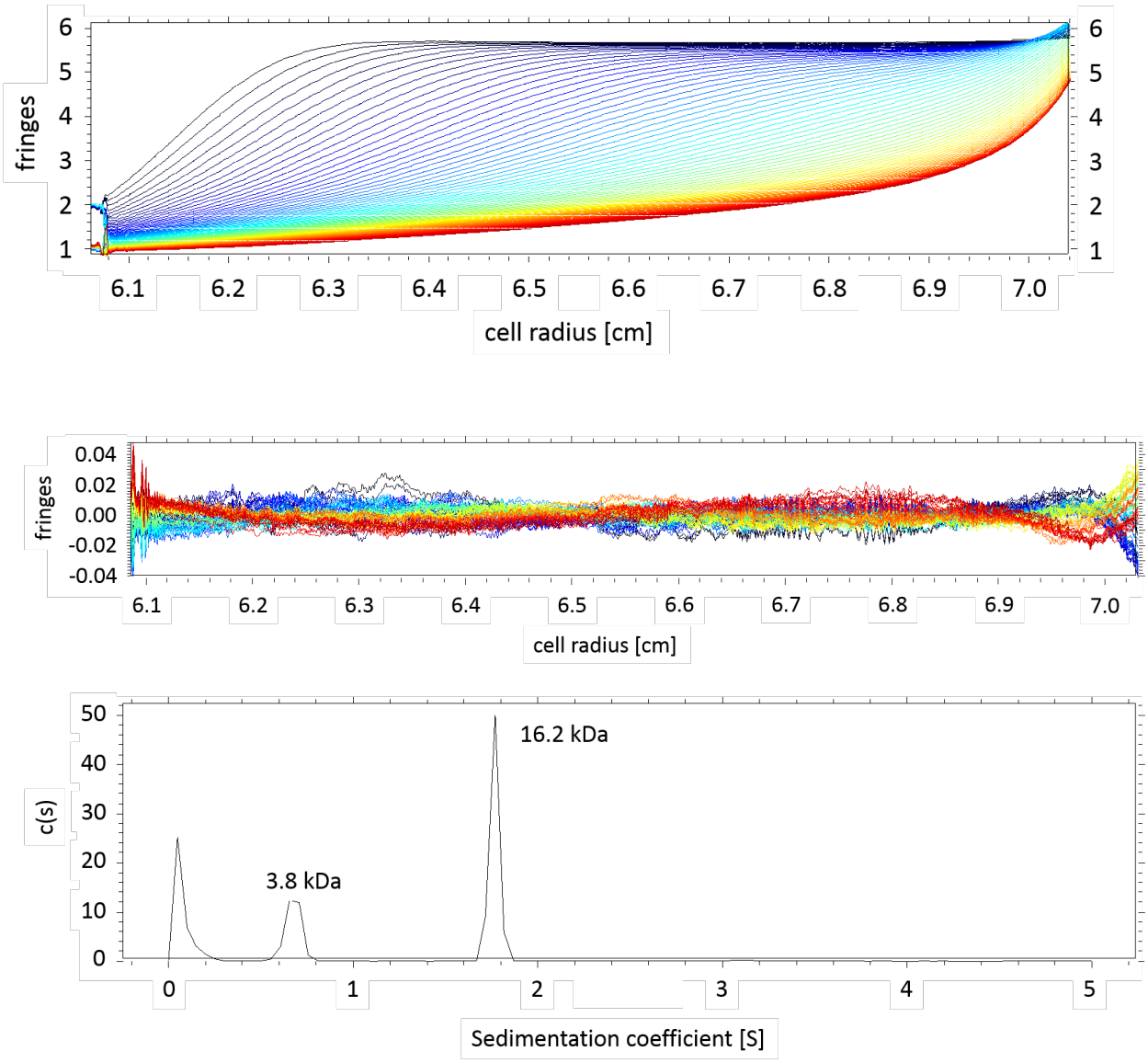

fig. S14. Analytical ultracentrifugation data and analysis of BNDL_1. Top: Experimental interference pattern including best fit of the sedimentation velocity data from 6.086 to 7.032 cm. Sample BNDL_1 at 0.5 mM in borate buffer (pH10). Middle: Residual plot of the fitted sedimentation velocity data along the radius of the cell. Bottom: Distribution plot showing one solution population at ~ 3.8 kDa consistent with a peptide monomer and one at ~16.2 kDa consistent with a peptide tetramer or pentamer. The sedimentation coefficient, S, is directly proportional to M_f (buoyancy mass). The peak at very low S values is most likely due to residual TFA from the synthesis and purification of the peptide. Statistics of the data fit are: n =98116, root-mean-square deviation, rmsd =0.0061, sum of squares of the regression, SSR = 3.62, signal/rmsd > 1000.

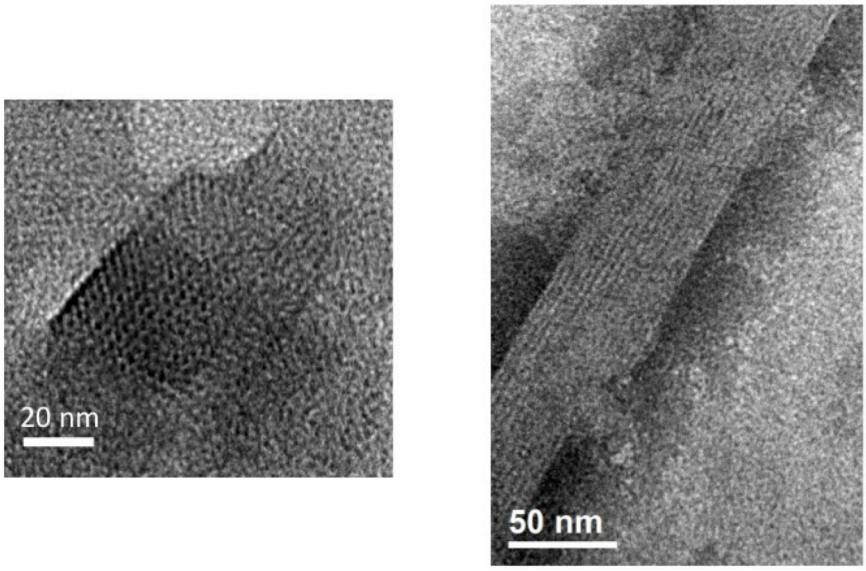

fig. S15. High mag TEM of 1.0 mM P222_9 ambiently cooled to room temperature from 80°C at pH 7 (left) and pH10 (right).

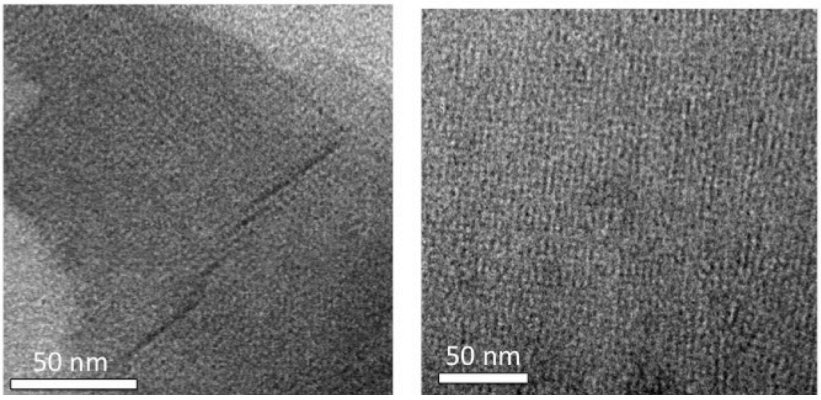

fig. S16. High mag TEM of (left) lattice of P222_9_Ac at pH 7 ambiently cooled to room temperature from 80°C and (right) P422_1_Ac at pH 8 quenched to 40°C from 80°C.

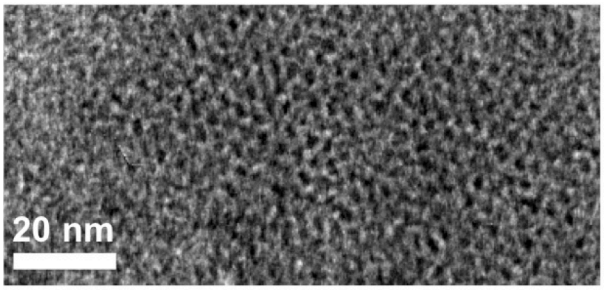

fig. S17. High mag TEM of lattice of P222_9_6Gly at pH 7 ambiently cooled to room temperature from 80°C

Computational Investigation Of The Lattices Formed By P422_1 and P222_1

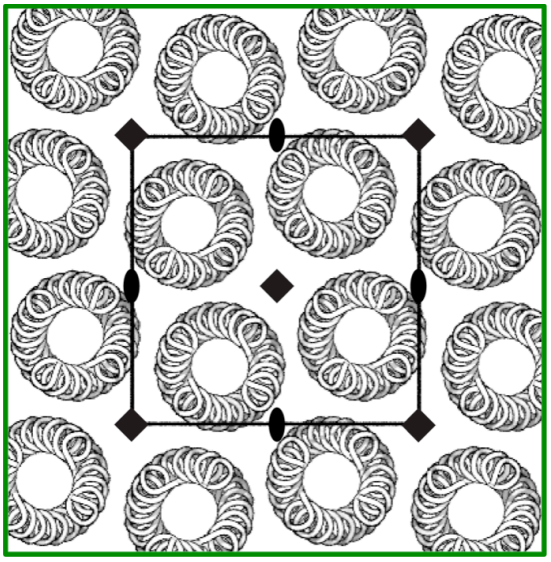

fig. S18. Putative structure of assembly of P422_1 helix bundles packed with P4 symmetry, which is a local minimum within the structure energy landscape with respect

to variation of the unit cell parameter. The dimension of the unit cell is a = b = 4.25 nm, consistent with that observed experimentally.

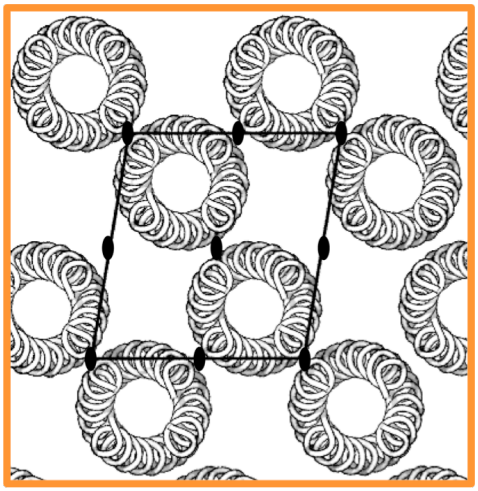

fig. S19. Putative assembly structures of P222_1 assemblies with P2 symmetry located at a local minimum of the structure energy landscape with respect to variation of the unit cell parameters. The dimensions of the unit cell is within the range of those observed experimentally: a = 3.24 nm, b = 3.03 nm and γ= 99.6°.