

# Supporting online Material for Computational Design of Virus-like Protein Assemblies on Carbon Nanotube Surfaces

## 1. Methods

### 1.1 Design of SWNT-binding peptide template.

Selection rule 2: matching of SWNT periodicity to  $\alpha$ -helical function-group periodicity. In selection rule 1, we chose the  $C_\alpha H_2$  of Gly or the  $C_\alpha H-C_\beta H_3$  group of Ala as the functional groups compatible with SWNTs and the  $\alpha$ -helix as our structural unit. According to selection rule 2, functional-group symmetry present within the structural unit and its assembly should ideally match the symmetry of the target surface. Since the target surfaces are chiral SWNTs, their natural symmetry is that of a screw. Thus, to match this with  $\alpha$ -helices, we considered screw-like helical structures, e.g. coiled coils. The Crick parameterization describes the coiled-coil backbone as:

$$\begin{aligned}x(t) &= R_o \cos(\omega_o t + \varphi_o) + R_l \cos(\omega_o t + \varphi_o) \cos(\omega_1 t + \varphi_1) - R_l \cos(\alpha) \sin(\omega_o t + \varphi_o) \sin(\omega_1 t + \varphi_1) \\y(t) &= R_o \sin(\omega_o t + \varphi_o) + R_l \sin(\omega_o t + \varphi_o) \cos(\omega_1 t + \varphi_1) + R_l \cos(\alpha) \cos(\omega_o t + \varphi_o) \sin(\omega_1 t + \varphi_1) \\z(t) &= \frac{R_o \omega_o t}{\tan(\alpha)} - R_l \sin(\alpha) \sin(\omega_1 t + \varphi_1)\end{aligned}$$

where  $R_o$  and  $R_l$  and the superhelical and helical radii,  $\omega_o$  and  $\omega_1$  are the superhelical and helical frequencies,  $\varphi_o$  and  $\varphi_1$  are the superhelical and helical phase offsets, and  $\alpha$  is the superhelical pitch angle (the angle between the superhelical curve and the primary bundle axis, here the Z-axis). The helical curve formed by contact points between a SWNT and either the  $C_\alpha$  atom of Gly or  $C_\beta$  atom of Ala, positioned at every seventh residue of an ideal coiled coil, can be written in parametric terms as:

$$\begin{aligned}x'(t) &= R_o \cos(\omega_o t + \varphi_o) + R_l' \cos(\omega_o t + \varphi_o) \cos(\varphi_1) - R_l' \cos(\alpha) \sin(\omega_o t + \varphi_o) \sin(\varphi_1) \\y'(t) &= R_o \sin(\omega_o t + \varphi_o) + R_l' \sin(\omega_o t + \varphi_o) \cos(\varphi_1) + R_l' \cos(\alpha) \cos(\omega_o t + \varphi_o) \sin(\varphi_1) \\z'(t) &= \frac{R_o \omega_o t}{\tan(\alpha)} - R_l' \sin(\alpha) \sin(\varphi_1)\end{aligned}$$

where  $R'_1$  is the distance from the center of the helix to the point of contact between the functional group and SWNT (approximately  $R_1$  plus the distance to the center of the function group, plus the radius of the functional group). Note that the contribution from  $\omega_1$  disappears in the above equations, because by construction we are looking at residues on the coiled coil that have identical helical phase. The pitch angle formed by the above curve (the SWNT contact angle) can be calculated as the angle between its gradient vector and the Z-axis:

$$\begin{aligned} \cos(\alpha_{\text{contact}}) &= \frac{\partial x'/\partial t}{\sqrt{(\partial x'/\partial t)^2 + (\partial y'/\partial t)^2 + (\partial z'/\partial t)^2}} = \\ &= \frac{R_0}{\tan(\alpha) \sqrt{R_1^2 + 2R_o R'_1 + \frac{R_o^2}{\sin^2(\alpha)} - R_1^2 \sin^2(\alpha) \sin^2(\varphi_1) - 4R_o R'_1 \sin^2(\varphi_1/2)}} \end{aligned} \quad (1)$$

$R_o$  and  $R'_1$  are given by the radius of the SWNT of interest and the size of functional group chosen,  $\varphi_1$  is relatively fixed as this should be the phase of the **a**-position residue pointing into the peptide/SWNT interface, and the coiled-coil pitch angle  $\alpha$  is a variable parameter. To look for a perfect match to the SWNT periodicity, we considered the common SWNTs (5,6), (5,7) and (3,8). For each, we considered all pairs of symmetry-related benzenoid rings, each defining a specific screw helix on the surface of the SWNT, located within 20 groups of each other. Each such helix gave a specific pitch angle  $\alpha_{CNT}$  and the equation above was numerically solved to find the coiled-coil parameter  $\alpha$  that would give a SWNT contact angle equal to  $\alpha_{CNT}$ . The distance between the two symmetry-related benzenoid groups was also compared to the distance between symmetry-related functional groups of the coiled coil. Best matches were produced with SWNT (3,8), although ideal agreement of periodicity tended to require fairly “tight” coiled-coil pitch angles (e.g.  $-30^\circ$ )(S1). Therefore, erring on the side of selecting natural coiled-coil geometries, we chose to set the coiled-coil pitch angle  $\alpha$  directly to the pitch angle of the SWNT, which has the effect of matching the overall periodicity of the coiled coil, rather than that of a specific position, to that of the SWNT (equivalent to considering the effective average phase  $\varphi_1$  in equation 1; see Fig. 1E in main manuscript).

Selection rule 3: building a unit cell with designable interfaces. Given the radial size of targeted SWNTs and the sizes of interacting functional units chosen, reasonable coiled-coil

bundles would contain between five and seven chains (based on the empirical observation we had previously made that superhelical radius is approximately equal to  $1.24 \cdot n + 2.4 \text{ \AA}$ , where  $n$  is the number of chains in the bundle (S1)). We thus focused on an anti-parallel hexameric topology.

We utilized our extended version of the Crick parameterization(S1) to systematically search the space of coiled-coil parameters for a suitable unit cell. Helical parameters were not varied and were set at their ideal values,  $R_1 = 2.26 \text{ \AA}$ ,  $\omega_1 = 102.8571^\circ$ , and helical rise per residue,  $d$ , was set to  $1.51 \text{ \AA}$ . Superhelical pitch angle  $\alpha$  was set to  $\pi/6 - \alpha_{chiral}$ , where  $\alpha_{chiral}$  is the chiral angle of the SWNT in question. For (3,8), (5,7) and (5,6) SWNT species considered, this corresponds to a shallow left-handed helix of benzenoid rings (resulting pitch angles  $-14.7^\circ$ ,  $-5.5^\circ$ , and  $-5.6^\circ$ , respectively), appropriate for natural coiled coils. The remaining parameters were chosen in two steps. First,  $R_o$  and  $\varphi_1$  were selected by optimizing the van der Waals interaction energy between the targeted SWNT and a single helix of a coiled-coil bundle built along the SWNT axis. This search was performed via unconstrained minimization in Matlab, using function *fminunc*. Note that selecting  $R_o$  also specifies  $\omega_o$ , as the fixed helical rise per residue connects the two parameters via  $R_o \omega_o = d \cdot \sin(\alpha)$ . The remaining parameters to choose, given the  $D_3$  symmetry of the bundle, were superhelical phase offset between neighboring helices,  $\Delta\varphi_o$ , and helical axial offset  $\Delta z$  (S1). These parameters were selected via a grid search by optimizing the designability of the resulting two helix-helix interfaces as outlined below. The grid search was performed with four focusing runs and a 5-by-5 grid considered each time. Initially, the range of values considered for  $\Delta z$  was between  $-2.64$  and  $2.64 \text{ \AA}$ , and for  $\Delta\varphi_o$  it was between  $-80^\circ$  and  $-40^\circ$  (the value of  $-60^\circ$  gives equal superhelical phase offsets between all monomers). At the end of each focusing run, the optimal combination of parameters was chosen and the search space was expanded by focusing into the 25% of the range considered in the current run for each parameter, symmetrically around the currently optimal parameter values. Thus, with four focusing runs, the accuracy of the  $\Delta z$  parameter was  $0.02 \text{ \AA}$ , and that for  $\Delta\varphi_o$  was  $0.16^\circ$ .

Assessing designability (fishing). We define as designable those protein-protein interfaces with three-dimensional geometries that occur frequently in natural and/or synthetic proteins and additionally tolerate a reasonable degree of variation in their sequences. Therefore, in

order to assess the designability of a given helix-helix interface from a coiled coil, we sought to search for instances of helix-helix interactions in the PDB that are close in geometry to the query interface. To do this, we employed a distance map-based search technique, which we refer to as *fishing*. A complete description of the method is to be published elsewhere. Briefly, we define a distance map to be an  $N$ -by- $N$  matrix of  $C\alpha$ -to- $C\alpha$  distances within a protein, where  $N$  is the number of residues in the protein(S2). Distance maps were pre-computed for all entries in the Protein Data Bank (PDB) and stored in a sparse-matrix format, where distances above 25 Å were not considered. Next, for every query complex, we similarly built a distance map to use as a query. Because of the periodic nature of the coiled coil, there is no need to search for a long segment of a helix-helix interaction, as the local environment repeats exactly laterally. Therefore, for each candidate interface we considered only a two-heptad section as a query. The query map was compared to the pre-computed maps, looking for optimal agreement in the sense of the norm of the distance difference matrix between the query and each match. The top 1000 matches were preserved, and resulting  $C\alpha$  RMSD over the matching region was calculated for each match. As a metric of designability of an interface, we used the number of matches with  $C\alpha$  RMSD below 0.8 Å. To limit the effects of bias and redundancy present in the PDB, we searched for matches over a subset of the PDB pruned to have no more than 30% sequence identity between pairs of structures.

## 1.2 Computational Protein Design.

Choice of backbone. Upon carrying out the procedure above for the (3,8) SWNT, the optimal backbone structure identified was found to be within a remarkable 0.9 Å  $C\alpha$  RMSD (over 156 residues of the hexamer) of the central coiled-coil region of DSD (PDB code 1GU6; see Fig. S5). This was discovered as a result of reviewing structural matches obtained as part of the designability analysis of the final (3,8) template. Because of such a very close match to a crystallographic backbone, the backbone of the central coiled coil from DSD rather than the parametric structure was used as a template for protein design.

Amino-acid alphabet. A powerful advantage of assessing designability via fishing is that besides knowing that a backbone template is designable, one can also gather statistics on the sequence distribution among the structures forming the given geometry. The results of this analysis for the two interfaces in the optimal design for SWNT (3,8) are shown in Fig. S6. The **d** and **e** positions of the hexameric coiled coil were fixed to Ala and Leu, respectively, based

on this analysis. Interestingly, these are precisely the amino acids occurring at the corresponding coiled-coil positions in DSD. Although sequence analysis strongly suggested that the **a** position should be Leu, this choice was not made as this was the SWNT-contacting residue. The identity of this position was fixed at the very beginning of the POSI procedure to either Ala or Gly. The amino-acid alphabet allowed in the remaining positions is summarized in Table S1.

Sequence optimization. Standard computational-design techniques were used to engineer HXT-Ala and HXT-Gly. A molecular-mechanics energy function based on the EEF1 implicit solvent model(S3), along with the rotamer library by the Richardsons and co-workers(S4), were used to construct a table of rotamer self and pair energies. The space of design sequences was searched using a simulated annealing algorithm. Two states were considered – a monomeric helix state, which served as a proxy for the dissociated state, and the desired hexameric state. For each sequence considered, a combined Dead-end Elimination/A\* branch-and-bound search algorithm was used to find the lowest-energy rotamer-based structure in each state(S5, S6, S7, S8, S9, S10). These structures were then subjected 10 steps of Steepest Descent and 10 steps of Adopted Basis Newton-Raphson continuous energy minimizing in CHARMM(S11) (using the EEF1 force-field(S3)), while keeping the backbone fixed. The final energy resulting from this minimization was used as the energy score for each sequence in each state, and the difference between one sixth of the hexamer-state energy and the monomer-state energy was used as the objective function for minimization. After ten parallel cycles of simulated-annealing Monte Carlo search (1,500 sequences considered in each), solutions were ranked by objective-function value. Top 10 solutions, different in the identity of only well solvated residues, were manually analyzed for potential solubility concerns (none were thought to have them) and to look for apparent favorable charge patterning to specify the anti-parallel orientation. Solution number 6 was selected with the sequence: AEGESALESGQQALEKGGQNALQSGRQALKA (as well as its Ala@**a** version). A second phase of design was performed based on this initial solution, focusing on two specific regions of the narrow interface of the hexamer (see Table S2 for the region of focus and amino-acid alphabet). An identical design approach was taken in this case as in the first phase. Manual analysis of the top 10 solutions from this phase, looking for optimized packing in the narrow interface, led to the selection of the second solution with sequence: AEGESALEYGGQQALEKGGQLALQAGRQALKA. Finally, in accordance with our selection rules,

the unit cell has to be able to cover the surface of interest in a periodic manner. Thus, we analyzed end-to-end interactions formed by adjacent hexamers on the SWNT and repacked the designed sequence in this new context. The resulting structure revealed favorable hydrogen bonding and potential salt bridge interactions at the complex-complex interface, so no additional changes to the sequence were made. The above sequence, and a version with Alanine at the SWNT-contacting core **a** position, were called HXT-Gly and HXT-Ala, respectively:

f	g	a	b	c	d	e	f	g	a	b	c	d	e	f	g	a	b	c	d	e	f	g									
ALA	GLU	GLY	GLU	SER	ALA	LEU	GLU	TYR	GLY	GLN	GLN	ALA	LEU	GLU	LYS	GLY	GLN	LEU	ALA	LEU	GLN	ALA	GLY	ARG	GLN	ALA	LEU	LYS	ALA		HXT-Gly
ALA	GLU	ALA	GLU	SER	ALA	LEU	GLU	TYR	ALA	GLN	GLN	ALA	LEU	GLU	LYS	ALA	GLN	LEU	ALA	LEU	GLN	ALA	ALA	ARG	GLN	ALA	LEU	LYS	ALA		HXT-Ala

**1.3 DSD variants.** The original structure of DSD is shown in Fig. S4A. The central region of the structure forms a hexameric coiled coil. We identified a single heptad position in the core of this coiled coil, the **a** position, occurring four times in each chain, which when substituted with an amino acid with a small sidechain, e.g. Ala or Gly, produces pores in the core of a radius  $\sim 4.7$  Å, a size appropriate to fit a single SWNT like those of (3,8), (5,7) or (5,6) chiralities (whose atom-center radii are 3.86, 3.74, and 4.09 Å, respectively). For example, complexes between DSD-Ala, DSD-Gly and the (5,6) nanotube are shown in Fig. S4B.

**1.4 General Peptide Synthesis and Purification.** Peptides were synthesized using Fmoc chemistry on an Applied Biosystems 433A and a Protein Technology Symphony synthesizer at 0.10 mmole scales. Products were cleaved from the rink amide MBHA resin (Novabiochem, substitution level of 0.56 mmole/g) in a mixture of trifluoroacetic acid (TFA)/triisopropyl silane/H<sub>2</sub>O (95:2.5:2.5 v/v) at room temperature under N<sub>2</sub> flow for 2 hours. After filtration the solvent was evaporated under a stream of N<sub>2</sub>. The crude peptides, precipitated with cold diethyl ether (Aldrich), were dried *in vacuo*. Purification proceeded by preparative reverse phase high performance liquid chromatography (HPLC, Varian ProStar 210) using preparative C4 column (Vydac) and a linear gradient of buffer A (99.9% H<sub>2</sub>O and 0.1 % trifluoroacetic acid) and buffer B (90% acetonitrile, 9.9% H<sub>2</sub>O and 0.1 % trifluoroacetic acid). Purity was assessed using HP1100 analytical HPLC system (Hewlett Packard) with a C4 column (Vydac). Molecular mass of all peptides was confirmed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry using Bruker Microflex 3.1 (flex control 1.3) and a Voyager Biospectrometry Workstation (PerSeptive Biosystems). All peptides were purified to levels exceeding 95% purity.

**1.5 Peptide Crystallization.** Crystals were grown by the hanging drop vapor diffusion technique at 291 K. The 2 $\mu$ L drop consisted of 1:1 v/v mixture of 1mg/ml protein solution in 20mM sodium phosphate/100mM NaCl pH 7.5 buffer and a reservoir solution of 75mM HEPES sodium pH 7.5, 0.6 M sodium phosphate monobasic monohydrate, 0.6M potassium phosphate monobasic, and 25%v/v Glycerol. X-ray diffraction quality crystals were obtained in 24 hours.

**1.6 Data collection and processing.** Diffraction data were collected for a flash-frozen crystal, on a diffractometer equipped with Cu K(alpha) radiation from a Rigaku rotating-anode generator and Mar image plate detector (300mm). Data sets were indexed and integrated with MOSFLM(S12, S13) and scaled using SCALA(S14) (Collaborative Computational Project, Number 4, 1994). Diffraction data were found recorded to a maximum resolution of 2.44 Å (Table S3).

**1.7 Structure solution and refinement.** Structure determination was carried out by the molecular replacement method using program *PHASER*(S15, S16). The Matthews coefficient(S17)  $V_m=3.58 \text{ \AA}^3/\text{Da}$  suggested a monomeric helix in the asymmetric unit if space group was I422 or I4<sub>1</sub>22 and a dimer if the space group was I4 or I4<sub>1</sub>. The systematic absence of axial reflections along axis I, revealed a 4<sub>1</sub> screw axis. The pseudo-precession images produced from the program *HKLVIEW*(S18) indicated 422 symmetry. Thus, initially molecular replacement was considered in the higher symmetry space group I4<sub>1</sub>22 using polyaniline  $\alpha$ -helix as the search model. The solution obtained had scores RFZ=3.7, TFZ=5.9 and LLG=69. The solution model was subjected for rigid body refinement and followed by iterative model building and restrained refinement protocols implemented in *Auto Build* module of *PHENIX*(S19). All side chains were traced in the electron density map. However, further refinement was stalled at unacceptable  $R_{\text{work}}/R_{\text{free}}=0.3128/0.3658$  and residuals could not be improved further. At this stage, the possibility of I4<sub>1</sub> was considered and dimers were located in asymmetric unit using the polyaniline  $\alpha$ -helix as a molecular replacement probe. The resulting model with scores RFZ=4.1 TFZ=3.5 LLG=38, RFZ=3.9 TFZ=9.2 LLG=102 was considered as potential solution. The iterative model building and restrained refinement resulted a complete tracing of the side chains in the electron density map. Further refinement could not improve the statistics  $R_{\text{work}}/R_{\text{free}}=0.2983/0.3572$ . The refinement were carried out using *REMAFC5*(S20) in *CCP4*(S18) program suite and model building using *COOT*(S21).

The H-test results,  $|H|=0.021$  (0.50 for untwinned, 0.0 for 50% twinned),  $H^2=0.001$  (0.33 for untwinned, 0.0 for 50% twinned), indicated merohedral twinning with the twin law (h, -k, -l), where  $H=|I_1-I_2|/|I_1+I_2|$ ,  $I_1$  and  $I_2$  are twin related acentric reflections. The cumulative distribution of H(S22, S23) and Britton plots(S24, S25) estimated twin domain fraction ( $\alpha$ ) to be 0.480 and 0.453 respectively.

As the estimated twin fraction was close to 0.5, the data were not detwinned for further refinement. The refinement was carried out by refining both parameters of the model and twin fraction. The *PHENIX*(S18) refinement protocol, which implements this option, was used. Model statistics resulting upon the convergence of this refinement strategy were:  $R_{\text{work}}/R_{\text{free}}=0.2438/0.2628$  and the twin fraction of 0.497. During the course of the refinement, several water molecules were located in map.

**1.8 Analytical Ultracentrifugation (AUC).** Oligomerization states of native DSD and designed peptides were determined by equilibrium sedimentation performed at 25 °C using a Beckman XL-I analytical ultracentrifuge. HexCoil-Gly was excluded from this analysis as it showed no evidence of being folded by CD (Fig. S9). Peptide solutions were prepared in a buffer of 20 mM sodium phosphate, 100 or 200 mM NaCl, at pH 7.5. Equilibrium radial concentration gradients at four different rotor speeds (35, 40, 45, and 50 KRPM for DSD-based designs and 30, 35, 40, and 45 KRPM for HexCoil-Ala designs) were acquired as absorbance scans (at 230 nm for DSD-based peptides and 280 nm for HexCoil-Ala). These data were analyzed by global curve-fitting to a monomer-dimer-hexamer, monomer-dimer-tetramer, or monomer-tetramer equilibrium models (in addition to a single-species fits) using a modified global fitting routine in Igor Pro (Wavemetrics), and best-fitting model was accepted. The partial specific volume(S26) and aqueous solution molar extinction coefficient of peptides were calculated using the program Sednterp. Sedimentation equilibrium data for each equilibrium model were fit to the following equation(S27):

$$Abs(r) = E + \sum_i i \cdot \epsilon \cdot l \cdot \frac{c(r_o)^i}{K_{d,i}} \cdot \exp\left[\frac{i \cdot M(1 - \bar{v}\rho)\omega^2}{2RT}(r^2 - r_o^2)\right]$$

where the sum is over all oligomerization states being fit (e.g. for monomer-dimer-hexamer,  $i$  goes over 1, 2, and 6),  $E$  is the baseline (zero concentration) absorbance,  $c(r_o)$  is the concentration of monomer at  $r_o$ ,  $\epsilon$  is the molar extinction coefficient of the monomer at 230 nm or 280 nm,  $l$  is the optical path length,  $M$  is the molecular weight of monomeric peptide,



$\bar{v}$  is the partial specific volume,  $\rho$  is the solvent density, and  $K_{d,i}$  is the monomer- $i$  mer dissociation constant. Figs. S7 shows sedimentation equilibrium profiles of DSD-Gly and DSD-Ala, along with corresponding species distribution plots consistent with monomer-hexamer equilibria. Fig. S8 shows sedimentation equilibrium profiles of HexCoil-Ala peptide and its species distribution plot that is consistent with a monomer-tetramer equilibrium. This is consistent with the observed tetramer in the crystal structure of HexCoil-Ala, formed by two approximately crystallographically symmetric dimers. All data are also consistent with results from size exclusion chromatography, shown in Fig. S10 (see below for details).

**1.9 Size Exclusion Chromatography.** Size-exclusive gel filtration elution profiles were obtained on a Superdex 75 10/300 GL column on a GE Healthcare FPLC system (Amersham Pharmacia Biosystems). Peptides (at 50  $\mu$ M) were prepared in a buffer of 20 mM sodium phosphate, 100 mM NaCl, pH 7.5 at room temperature. 200  $\mu$ L of each sample were loaded and eluted with the same buffer. The column was equilibrated in 20 mM sodium phosphate, 100 mM NaCl, pH 7.5 with a mobile phase flow rate of 0.5 mL/min, and the absorbance at 214 nm and 280 nm was recorded. Calibration curves were obtained using the molecular-weight standard kit, MWGF70 6,500-66,000. The elution volume parameter  $K_{av}$  was calculated for each standard as:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

where  $V_e$  is elution volume for the peptide,  $V_o$  is column void volume measured with Blue Dextran, and  $V_t$  is total bed volume. The calibration curve was prepared by plotting the obtained  $K_{av}$  values versus logarithms of the corresponding standard molecular weights. Molecular weights of designed peptides were then determined by calculating their  $K_{av}$  values and interpolating from the standard curve.

**1.10 Circular Dichroism (CD) Spectroscopy.** Spectra were acquired on a Jasco 810 Spectropolarimeter. Samples of SWNT/peptide suspensions were prepared as described below. Measurements of CD spectra were performed in 1 mm path-length quartz cells, in the far UV region from 190 nm to 260 nm (see Fig. S9). For the purpose of calculating Mean-Residue Ellipticity (MRE) of peptide-SWNT suspensions, before or after filtration to remove unbound peptides, accurate determination of the peptide concentration was necessary. As

SWNTs absorb significantly in the UV-visible range, UV-visible spectra were separately measured for SWNTs in the common surfactant sodium dodecyl benzene sulfonate (SDBS), SDBS alone, peptide/SWNT suspensions in aqueous buffer, and peptide alone in aqueous buffer (first two spectra blanked against air, and second two against aqueous buffer). Linear regression of the peptide/SWNT suspension spectrum (in the range between 230 and 400 nm) based on the spectra of its components then provided the contribution of peptide alone and hence peptide concentration. CD spectra of peptide-SWNT suspensions in Fig. S9C-D are built using concentrations determined in this way. These concentrations were also internally consistent with the shapes and magnitudes of the measured CD spectra, considering that the CD spectrum of the unfiltered peptide-SWNT suspension should be decomposable into contributions from the filtered-suspension spectrum the peptide-alone spectrum, in proportion to the relative concentration of the two species.

**1.11 Carbon Nanotube Binding Assay with Designed Peptides.** A stock solution of SWNTs in glycerol was prepared by adding 20 mg of nanotubes from SouthWest NanoTechnologies, Inc. (SWeNT SG65 #000-0013, Texas) to a 20 ml solution of 1% by weight glycerol (Sigma Aldrich, Missouri). The mixture was bath-sonicated (Cole Palmer Sonicator, model 08849) for approximately 20 minutes in order to produce homogeneously-dispersed SWNTs. For each peptide to test, 100  $\mu$ L of this SWNTs stock solution was centrifuged (Thermo Electron Corporation, Legend 14) at 14,000 g for 10 minutes and the supernatant discarded. Centrifugation was repeated twice more, each time adding 1 ml of DI water beforehand and discarding the supernatant afterwards. The resulting SWNT pellet (eventually 0.1 mg/mL in mixture solution) was mixed with 1 ml of 1 mg/ml peptide solution in 20 mM sodium phosphate, 100 mM NaCl, pH 7.5. The sample was then tip-sonicated (Misonix XL- 2000, 1/8th inch tip) in an ice bath 30 minutes, slowly increasing sonication power up to 10 W. Ice-bath cooling was to prevent excessive sample heating and destabilization of protein structure. The sonicated samples were warmed up to room temperature and centrifuged at 14,500 g for 90 minutes (Thermo Electron Corporation, Legend 14). As a reference sample for measurement of 2D-PL and absorption spectra, an SWNT suspension in sodium deoxy cholate (SDOC) in water was prepared with 10 mL of 1% by weight solution of SDOC and 10 mg of SWNTs by bath-sonicating for 20 minutes.

**1.12 UV/Vis and Far-IR Absorption Spectroscopy.** The absorbance spectra of the SWNT suspensions were measured using a custom-built system that uses a PIActon/Roper Scientific 0.5 meter TriVista Spectrometer, a stable tungsten-halogen white light source (Ocean Optics LS-1), computer-controlled filters to block second-order diffraction signals, and a Princeton Instruments PIXIS CCD or OMA-V InGaAs linear array detectors. Liquid samples were placed in a Suprasil cuvette with a 1.0 cm path length, using a SDOC solution or DI water as a reference, for SDOC or peptide SWNTs, respectively. Spectra from 380 to 1000 nm were collected with the PIXIS CCD, and the partly overlapping infrared part of the spectrum from 850 nm – 1400 nm was collected using the OMA-V detector. The first two stages of the spectrometer were used in a subtractive configuration, employing mirror-imaged positive and negative first-order grating diffractions, to achieve a tunable bandpass filter that augmented the performance of the tunable filter wheel. Spectra were then dispersed by the third spectrometer stage onto the detectors.

**1.13 2-Dimensional Photoluminescence (2D PL) Spectroscopy.** Photoluminescence (PL) excitation-emission maps were obtained using the aforementioned Trivista spectrometer and cameras, but configured rather differently. Tunable excitation of 10 nm bandwidth was obtained by passing the output of a Newport-Oriel Xenon arc lamp through a second-order blocking, computer-controlled filter wheel and either (a) a 0.125 meter single-stage Oriel Cornerstone monochromator or (b) a pair of 0.5 meter Trivista stages used as a subtractive bandpass filter. The excitation intensity was recorded using a beam splitter, a diffuser, an Ocean Optics USB 2000 detector, all of which was calibrated using a NIST-traceable light source (Ocean Optics LS-1-CAL) and calibrated multi-mode. The signal from the excited sample was collected at 90° from the direction of excitation and passed into the third stage of the Trivista spectrometer and onto the PIXIS or OMA-V detector, as appropriate. The excitation wavelength was filtered away as necessary before reaching this spectrometer stage. PL maps were generated by exciting samples over the range of 500-690 nm in 10 nm steps and recording the emission spectrum at each step.

**1.14 Transmission Electron Microscopy (TEM).** Droplets of liquid sample, containing peptide/SWNT/gold cluster mixture, were dropcast onto a C-flat holey carbon grid (Electron Microscopy Sciences, 400-mesh). TEM was conducted on FEI-Technai T12 Transmission electron microscope operating at an accelerating voltage of 80 kV, and images were collected

using a Gatan CCD image system. For higher-magnification images, a JEOL 2010 TEM operating at an accelerating voltage of 120 kV was used (Fig. 4D-E in main manuscript).

**1.15 High-Resolution Transmission Electron Microscopy (HRTEM).** The HRTEM measurement was conducted on a JEOL 2010F field emission gun (FEG) HRTEM. The images were taken in alpha 1 mode, with an accelerating voltage of 197 kV. The imaging device was a Gatan image filter (GIF), and the software used was Gatan Digital Micrograph. Images in Fig. 4F and 5H (in the main manuscript) were produced by HRTEM. The lattice apparent on the spots corresponding to gold particles in these images indicates that atomic resolution was reached (lattice cells correspond to individual Au atoms).

**1.16 TEM-image Spacing Analysis.** A stringent test of our structural model is whether the inter-gold cluster distances observed in TEM images are consistent with those predicted based on the structural model of SWNT/protein/gold assembly (Fig. 4G in main manuscript). Further, it is important to establish that cluster displacements occur in phase, that is to say that the gold clusters form a single repeating structure with a given pattern, rather than being randomly scattered. To ascertain both of these aspects without operator bias, we performed an auto-correlation based analysis of our TEM images (each image taken was 1024 x 1024 pixels). The method was implemented in Matlab R2010a. Upon an initial filtration step (a median filter with a mask of 10-by-10 pixels, Matlab function *medfilt2*), the 2D auto-correlation of each image was calculated. Besides the trivial central peak, the auto-correlation image contained additional peaks corresponding to characteristic repeats within each image. To simplify the problem of peak identification, the user was allowed to specify the locations of centers of all apparent peaks using the mouse. Distances between the central trivial peak and these additional peaks were then calculated and automatically converted to physical distances using the scale bar for reference. The presence of non-trivial peaks in the auto-correlation function indicates a consistent in-phase spacing between gold clusters, confirming that these clusters are forming a single structured pattern rather than randomly dispersed. Further, the average inter-cluster distance derived from this analysis is in excellent agreement with the one predicted from the structural model (see Fig. S1). Based on our structural model of DSD-Gly/SWNT/gold cluster complex (Fig. 4G in main manuscript), inter-particle distance between neighboring clusters on the same hexamer is 63 Å, or 52 Å after averaging over possible projection angles, and the distance between cluster on adjacent hexamers is 48 Å. The

average distance from the auto-correlation analysis is  $\sim 47$  Å, which is in excellent agreement with these numbers.

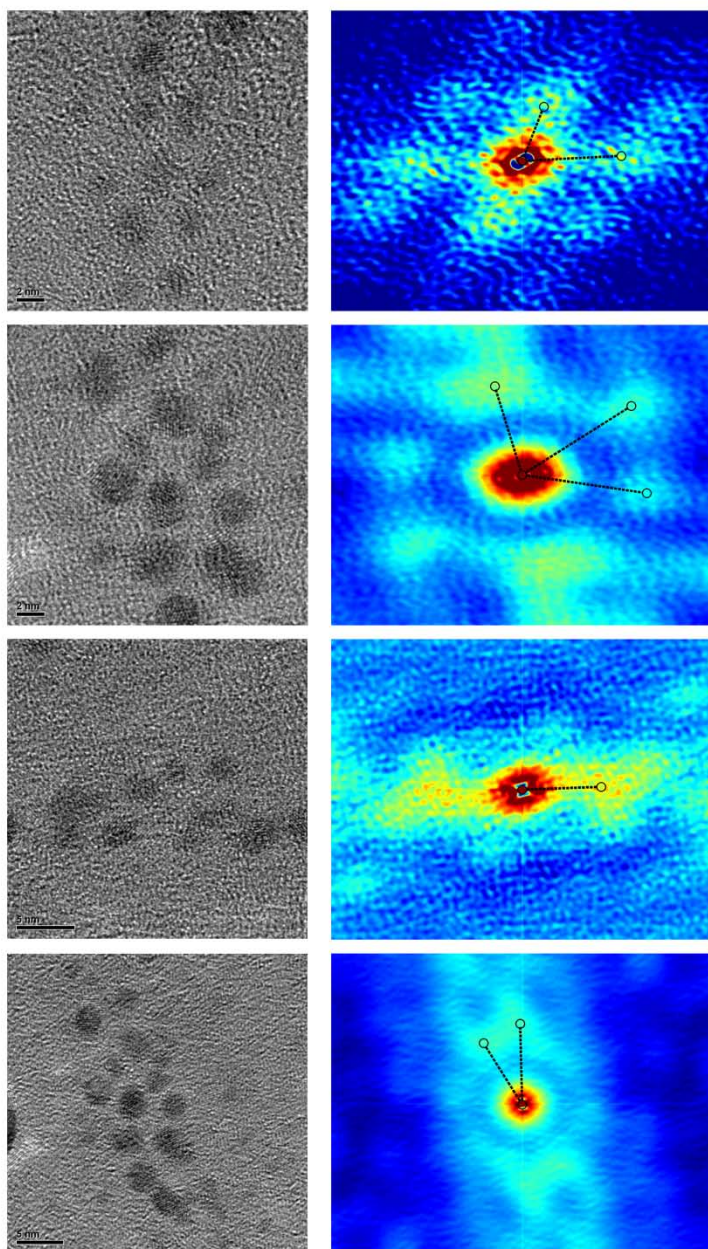
**1.17 Gold cluster growth on DSD/SWNT complex.** The goal was to nucleate the growth of small-sized gold nanoparticles using a single cysteine residue in DSD-Gly. Cys-mutated DSD-Gly/SWNT suspensions were prepared as detailed above, followed by treatment with 1 mM of Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) to reduce the sulfhydryl group of Cys. Next, the mixture was purified by centrifugal filtration (Microcon Ultracel YM-30, 30,000 MWCO, MILLIPORE) and reconstituted into 100  $\mu$ L of buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.4). Consequently, precursor ( $\text{HAuCl}_4$ , chloroauric acid, 1.5 mM) and reducing agent ( $\text{NaBH}_4$ , sodium borohydride, 1.5 mM) were added to the peptide/SWNT suspension and incubated at 4 °C for 12~18 hours.

**1.18 Size Distribution of Gold clusters.** The distribution of gold cluster sizes was measured by analyzing TEM images in Matlab R2010a. Initially, a median filter with a window of size 10-by-10 pixels was applied to each 1024-by-1024 image. Identification of gold clusters and calculation of their sizes proceeded through an iteration of two steps. First, the user was presented with the filtered image and asked to outline a box that fully contained (with significant margins) a perceived cluster. Here we rely on the superior power of human vision and image processing compared to computer-based ones. However, to remove human bias from the estimation of cluster size, an automated procedure was adopted. The sub-image presumed to contain a gold cluster was systematically cross-correlated with binary images (of the same size) of a circle of increasing radius, and the offset of highest correlation as well as the maximum normalized correlation coefficient (from Matlab function *normxcorr2*) noted for each. The circle radius producing the largest correlation coefficient was taken as the radius of the gold cluster and the corresponding offset dictated the center of each cluster for illustration purposes. The procedure was repeated until all apparent clusters were covered in all TEM images. Fig. S2 shows the locations and sizes of gold clusters identified by this algorithm. Fig. S3 plots the histogram of cluster radii (from 113 data points) and the best-fit Gaussian probability density.

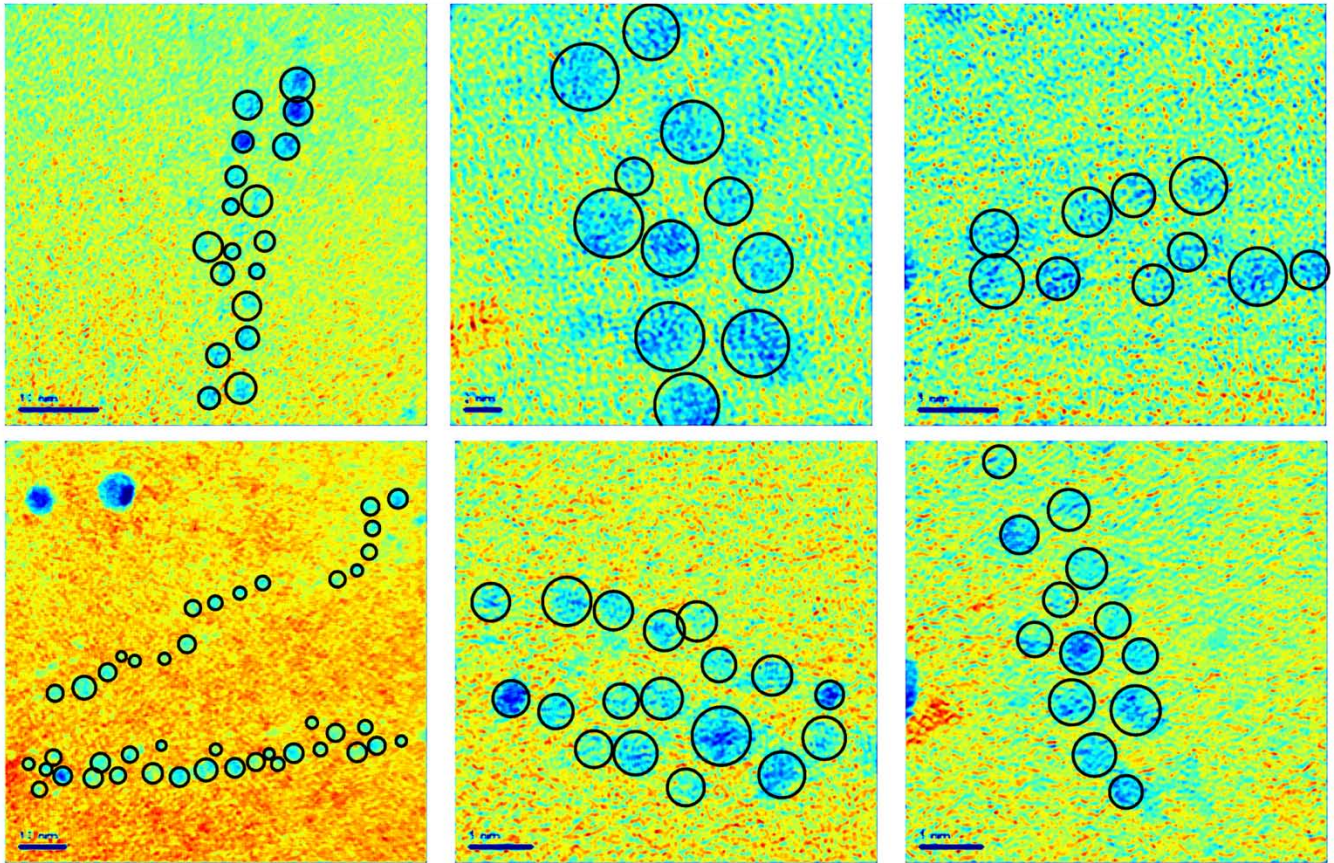
**1.19 Control Experiment with Permuted Versions of Designed Peptides.** Peptides cHexCoil-Gly and cHexCoil-Ala (see Fig. S12) were synthesized and purified as detailed in

section 1.4. SWNT/peptide suspensions were created following exactly the same procedure as in section 1.11, except that sonication power was increased up to 8 W. To control for any variability and to provide internal standards, the same procedure was applied to cHexCoil-Gly, cHexCoil-Ala, as well as HexCoil-Gly and HexCoil-Ala at the same time. UV/visible spectra of all four resulting SWNT/peptide suspensions were recorded from 200 nm to 900 nm in a 1-cm cell, following a 10-fold dilution. These spectra are shown in Fig. S13.

## 2. Supporting Figures

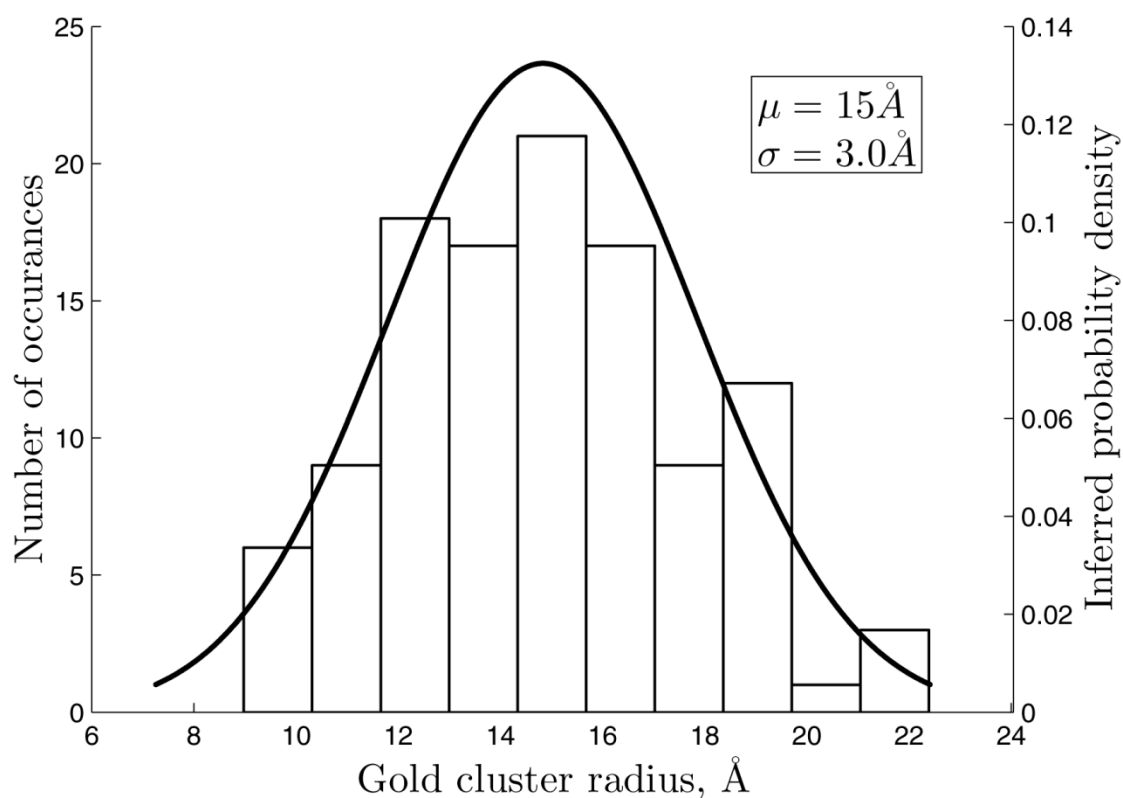


**Figure S1** Examples of TEM images of gold/SWNT/protein assembly and resulting auto-correlation images (right; only the middle 40% x 40% of the auto-correlation image is shown). The peaks manually identified are shown with black circles and the distances measured are shown with black dashed lines. The mean, median and standard deviation of all distances thus measured are 47.3 Å, 47.6 Å, and 8.6 Å, respectively. Importantly, though these averages came from 13 distance measurements, each measurement implicitly contains information from all inter-particle distances in each image. The average distance expected from the structural model is 52 Å.

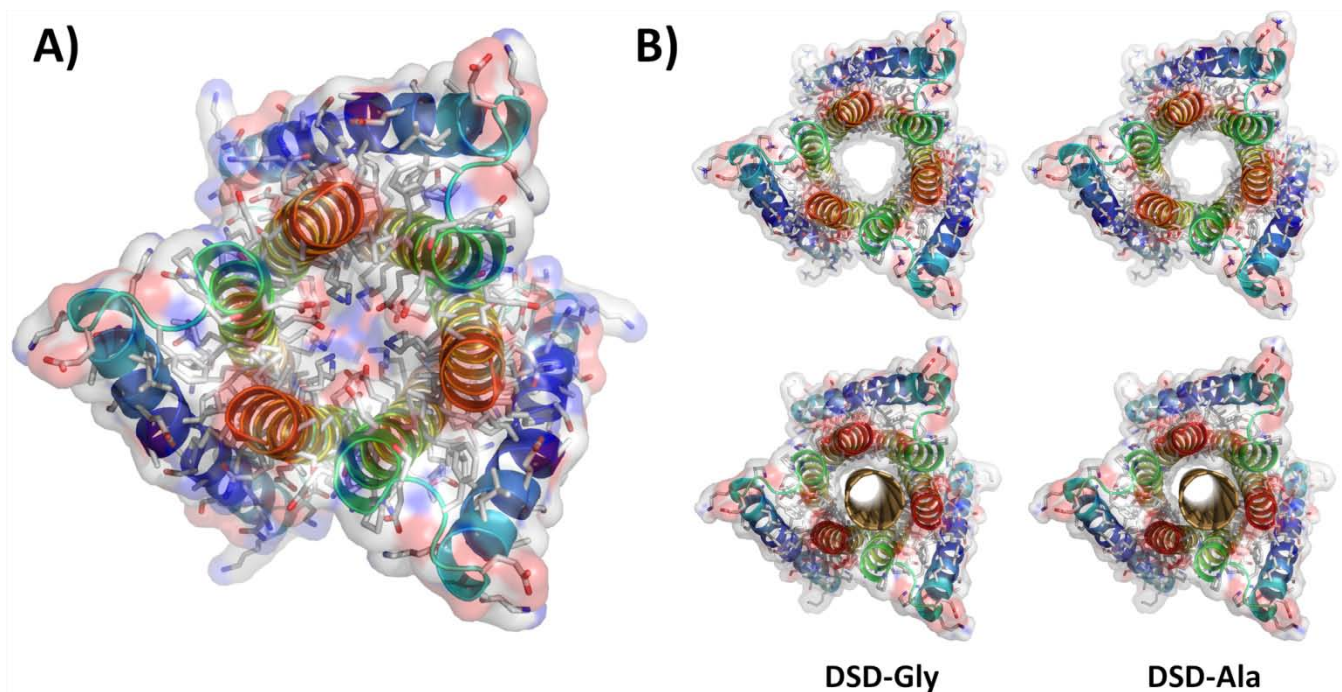


**Figure S2** Shown in false color are some sample TEM images. Black circles show the location and sizes of clusters identified by the semi-automated procedure outlined above. Circle radius and center is chosen automatically, but the general area where a cluster is observed is chosen by a human operator.

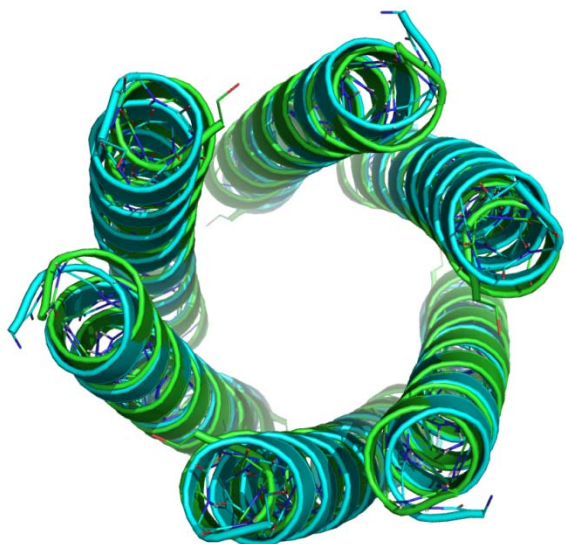




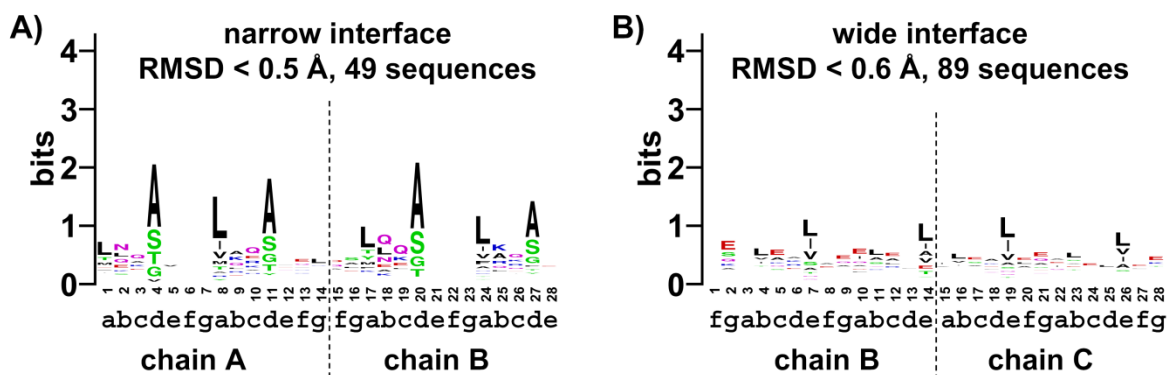
**Figure S3** Distribution of gold-cluster radii derived by applying the procedure outlined above to all TEM images. The bar plot shows the histogram of 113 measured radii (y-axis on the left) and the solid line represents the best-fit Gaussian probability density. The mean and standard deviation of the latter are 15 Å and 3.0 Å, respectively.



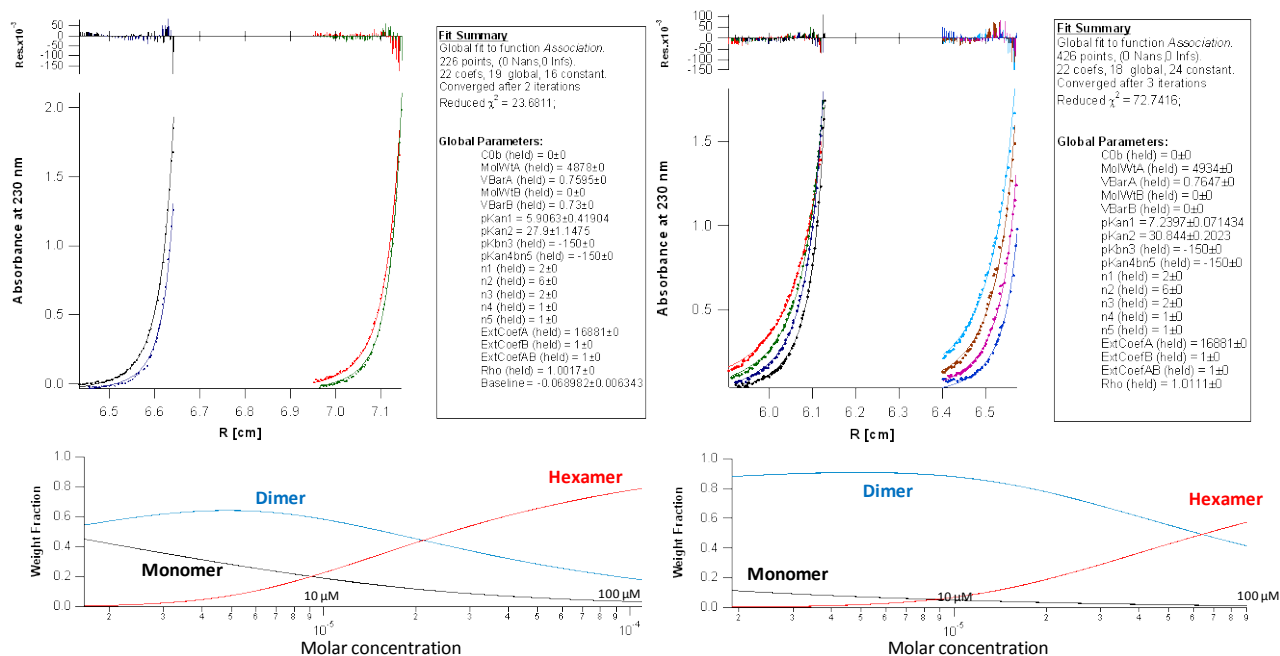
**Figure S4** Structures of DSD and DSD-based designs. **A)** Crystal structure of DSD (PDB code 1G6U). Backbone is colored blue-to-red from the N-terminus to the C-terminus. The van der Waals surface is shown in a semi-transparent manner, illustrating that the core is filled. **B)** Structures of DSD-Ala and DSD-G (top left and right, respectively) and the same two structures with (6,5) SWNT inserted into the core (bottom left and right, respectively). The van der Waals surface is shown in a semi-transparent manner, illustrating the pore that forms by introducing either a Gly or Ala at the **a** position.



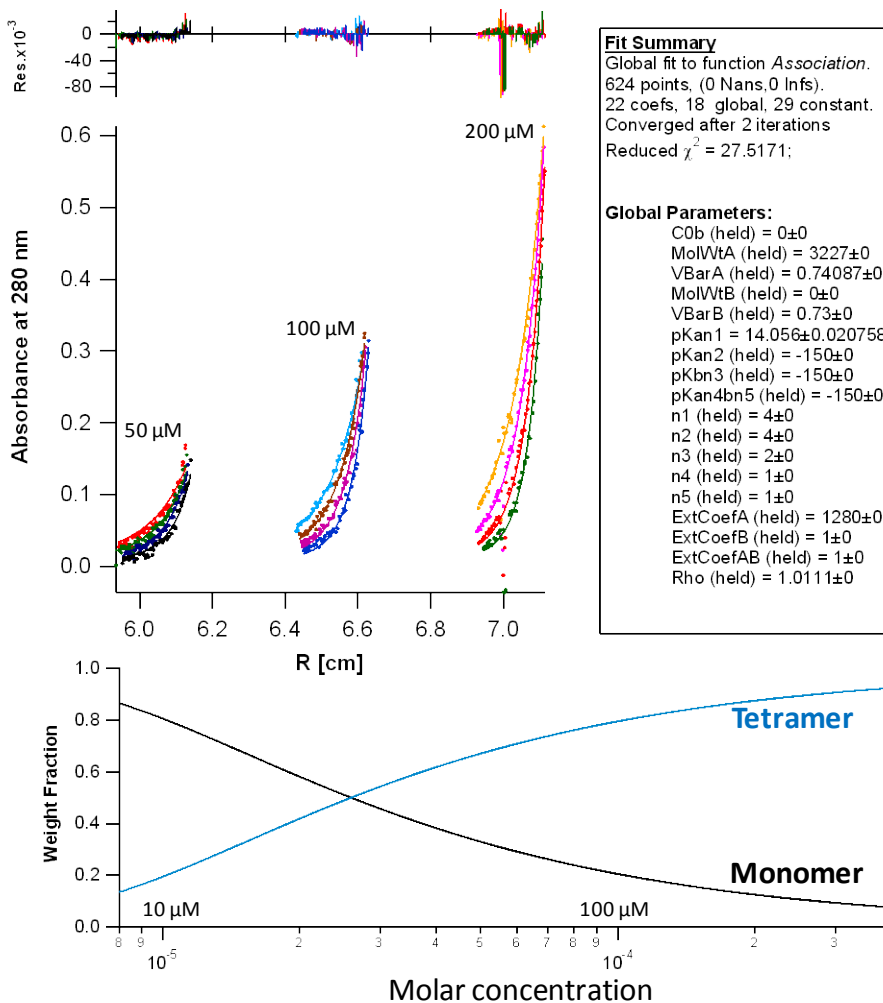
**Figure S5** Superposition of the optimal backbone against SWNT (3,8), shown in green, onto the central coiled-coil portion of DSD (PDB code 1GU6, residues 19-48), in cyan. The  $C_{\alpha}$  RMSD between the two structures over the middle 26 residues per monomer (156 residues in total for the hexamer) is 0.9 Å. 1GU6 arrived as the clear top score when fishing for a three-helix sub-bundle (two heptads per helix) of the (3,8)-optimal backbone.



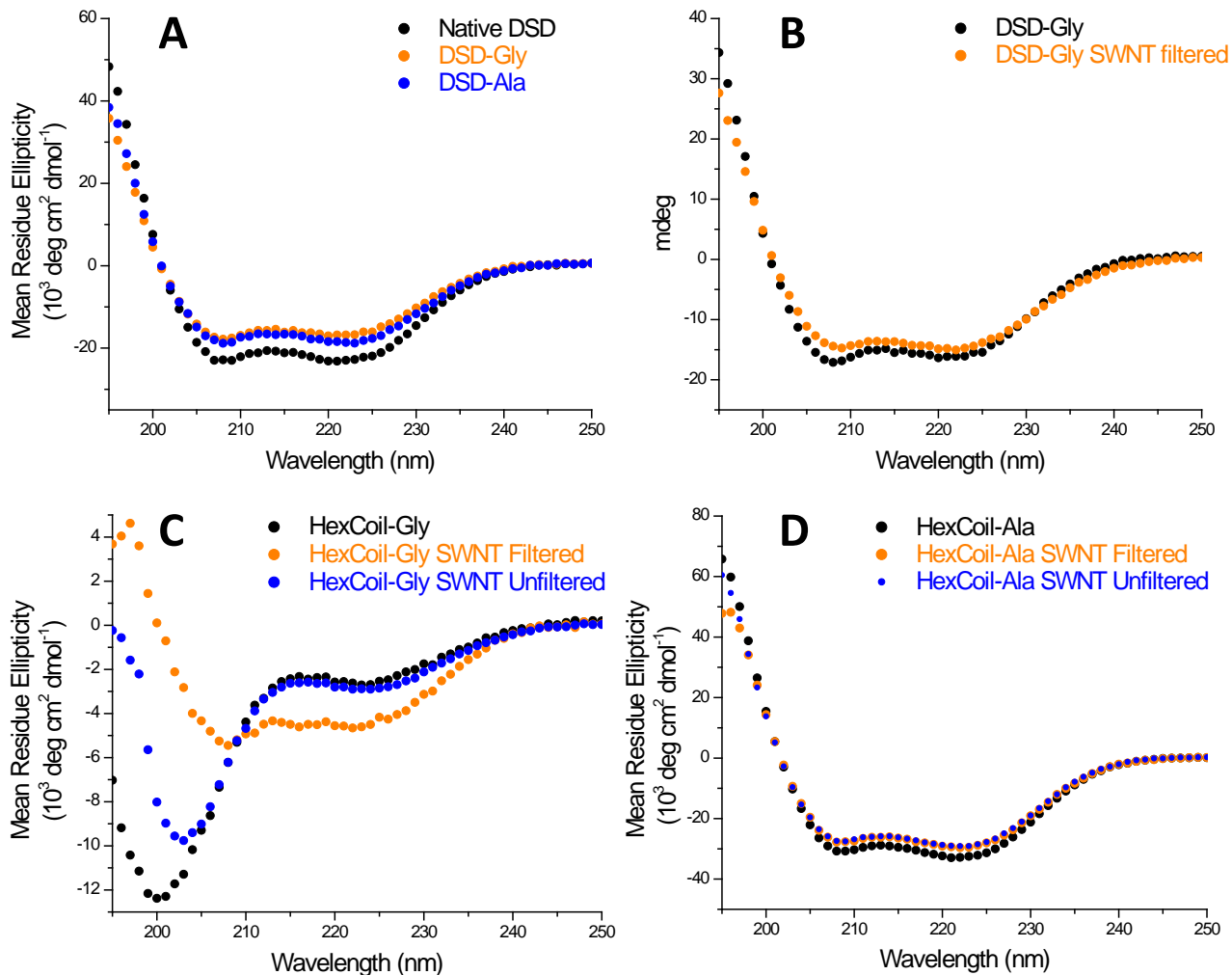
**Figure S6** Sequence distributions resulting from fishing for interfaces matching those in the (3,8)-optimal hexamer. **A)** and **B)** show results for searching for matches against the narrow and wide interfaces of the coiled coil, respectively (see Fig. S5). A two-heptad stretch of the corresponding helix-helix interface was considered in each case. Letters under the x-axis indicate heptad-position designations, in the context of the full hexamer. The  $C_{\alpha}$  RMSD cutoff used to select matching interfaces and the number of matches this produced is shown at the top of each panel. Sequence logo diagrams created with the standalone version of the weblogo software(S28).



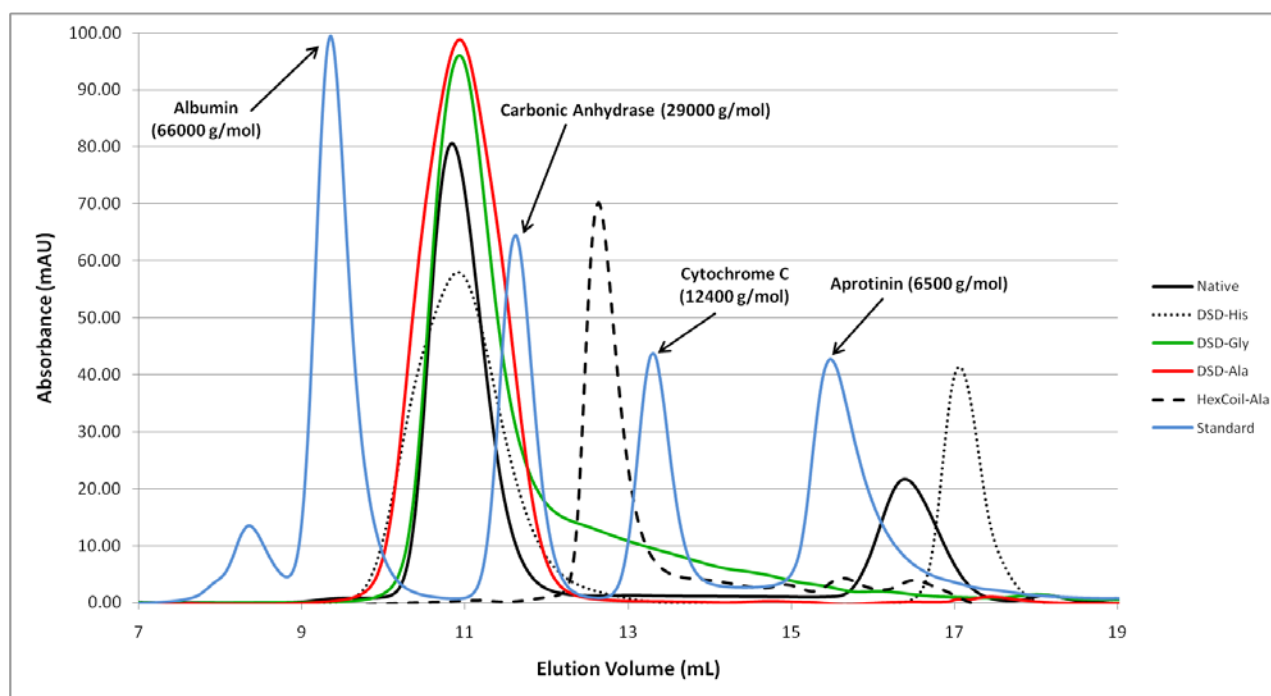
**Figure S7** Sedimentation equilibrium analytical ultracentrifugation (AUC) profiles for DSD-Gly (on the left) and DSD-Ala (on the right). The data fit well to a monomer-dimer-hexamer equilibrium model. A global fit to all data sets was performed, holding the partial specific volume at 0.7595 mL/g and 0.7647 for DSD-Gly and DSD-Ala, respectively, and buffer density at 1.0073 g/mL. A total of eight data sets were included for DSD-Ala: all combinations of 50  $\mu\text{M}$  or 100  $\mu\text{M}$  concentrations and rotor speeds of 35, 40, 45, and 50 KRPM; and a total of four data sets were included for DSD-Gly: concentration of 50  $\mu\text{M}$  with rotor speeds of 40 and 45 KPRM and concentration of 100  $\mu\text{M}$  with rotor speeds of 45 and 50 KPRM). The species distributions (bottom) show significant presence of dimer and hexamer (and a smaller amount of monomer) at both 50  $\mu\text{M}$  and 100  $\mu\text{M}$ , indicating an appropriate concentration regime for association determination.



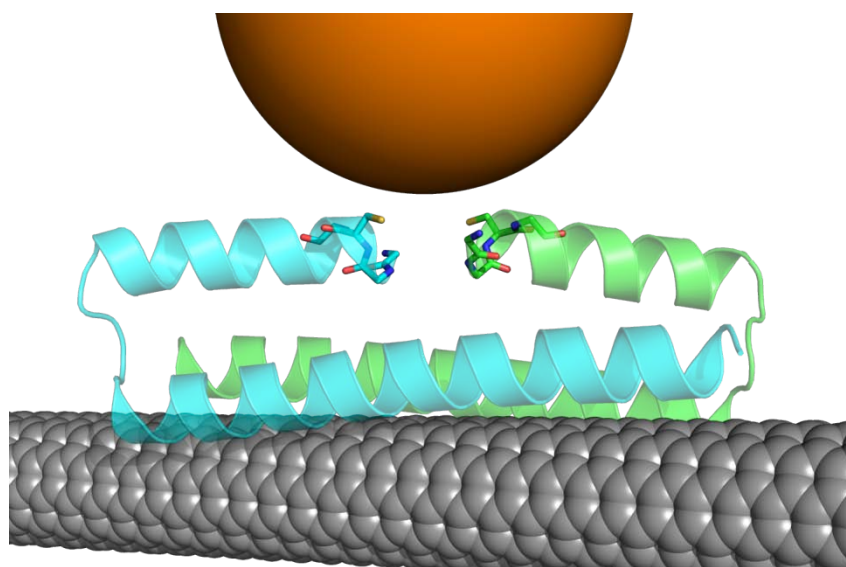
**Figure S8** Sedimentation equilibrium analytical ultracentrifugation (AUC) profile for HexCoil-Ala. The data fit well to a single-species model, producing a molecular weight of 12,671, in good agreement with the tetrameric molecular weight of 12,908, but a monomer-tetramer equilibrium model shown here produces a superior fit. A global fit to all data sets (collected at 25, 30, 35, 40 and 45 KRPM and 200, 100, and 50  $\mu\text{M}$  peptide concentration, 20 mM sodium phosphate, 100 mM NaCl, pH 7.5), was performed, holding the partial specific volume and buffer density at 0.749 mL/g and 1.0073 g/mL, respectively.



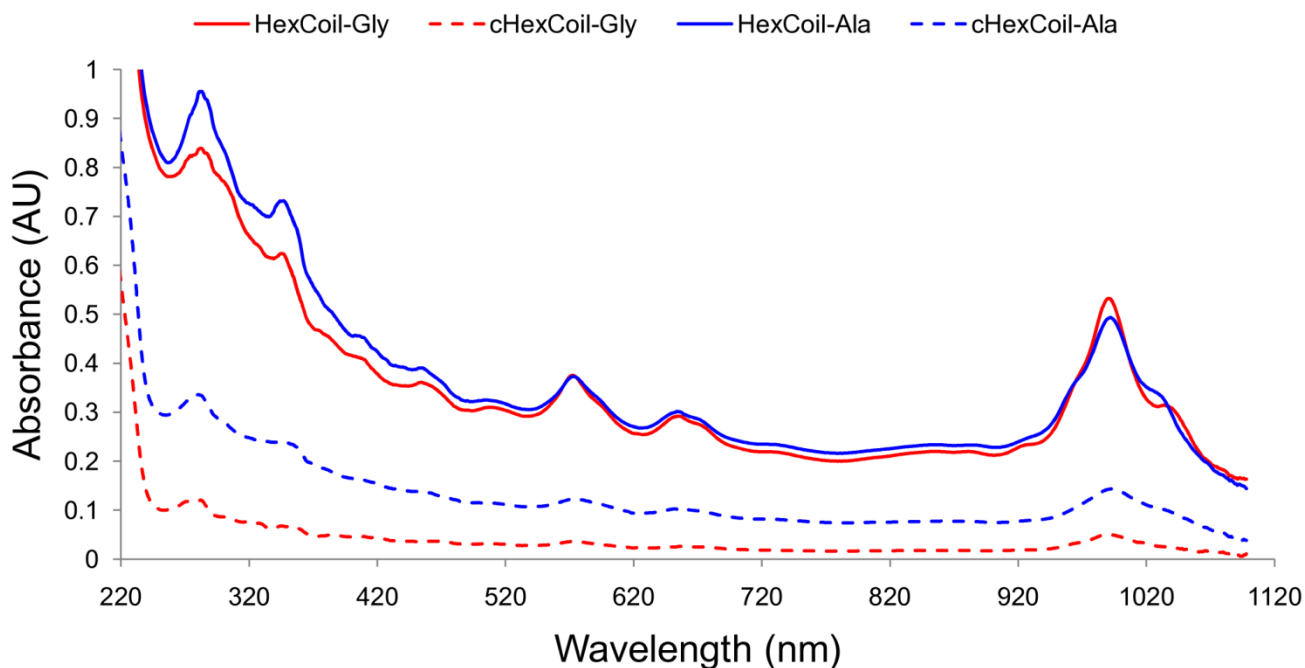
**Figure S9** Far UV circular dichroism (CD) spectra of peptides and SWNT-peptide suspensions in 20 mM sodium phosphate, 100 mM NaCl, pH 7.5 at 25 °C. **A)** CD spectra of native DSD, DSD-Ala and DSD-Gly at 20  $\mu$ M. **B)** Change in the CD spectrum upon introduction of SWNTs into solution of DSD-Gly. Because the shape of the spectrum did not change appreciably and DSD-Gly was expected to be well folded and helical without SWNTs by CD/AUC and FPLC, concentrations of peptide in SWNT/peptide suspension was not deconvoluted and the spectra are shown in millidegrees. **C)** CD spectra of 63  $\mu$ M HexCoil-Gly and a HexCoil-Gly/SWNT suspension containing 139  $\mu$ M HexCoil-Gly are shown in black and blue, respectively. Membrane filtration of the latter sample to remove unbound peptide produces a peptide/SWNT suspension with 40  $\mu$ M HexCoil-Gly (see text for method of concentration determination), and gives the CD spectrum shown in orange. **D)** CD spectra of 52  $\mu$ M HexCoil-Ala and a HexCoil-Ala/SWNT suspension containing 42.3  $\mu$ M HexCoil-Ala are shown in black and blue, respectively. Membrane filtration of the latter sample to remove unbound peptide produces a peptide/SWNT suspension with 47.8  $\mu$ M HexCoil-Ala, with a CD spectrum shown in orange.



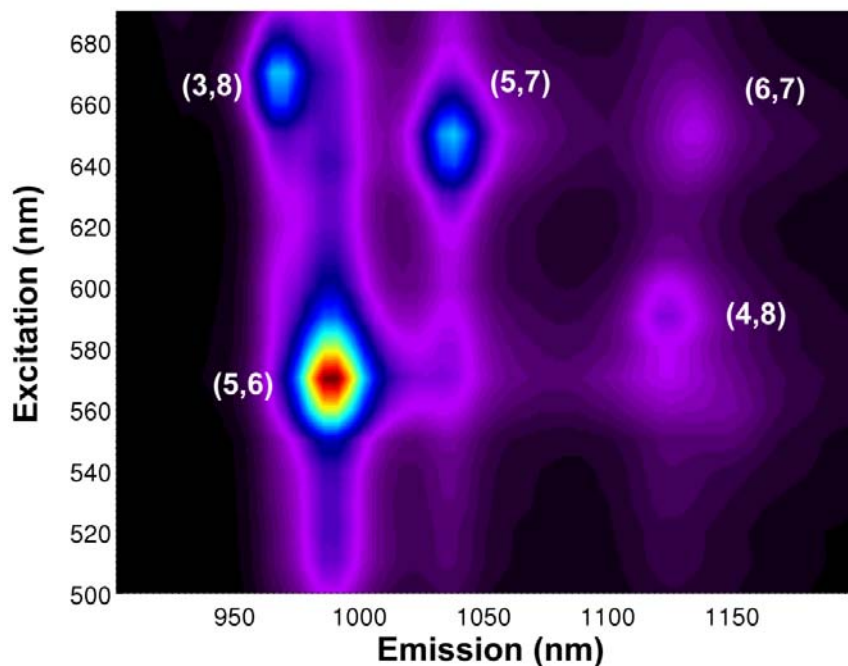
**Figure S10** Size exclusion chromatograms of 50  $\mu\text{M}$  solution of all DSD variant peptides, and 20  $\mu\text{M}$  of HexCoil-Ala at 220 nm. Data for HexCoil-Gly was not collected as it was not seen to be folded by CD (see Fig. S9C). A standard chromatogram, shown in blue, is normalized to have a maximum absorbance of 0.1. Peptides exist in equilibrium between the hexamer/tetramer associated state for DSD variants and HexCoil-Ala, respectively, and lower oligomeric states.



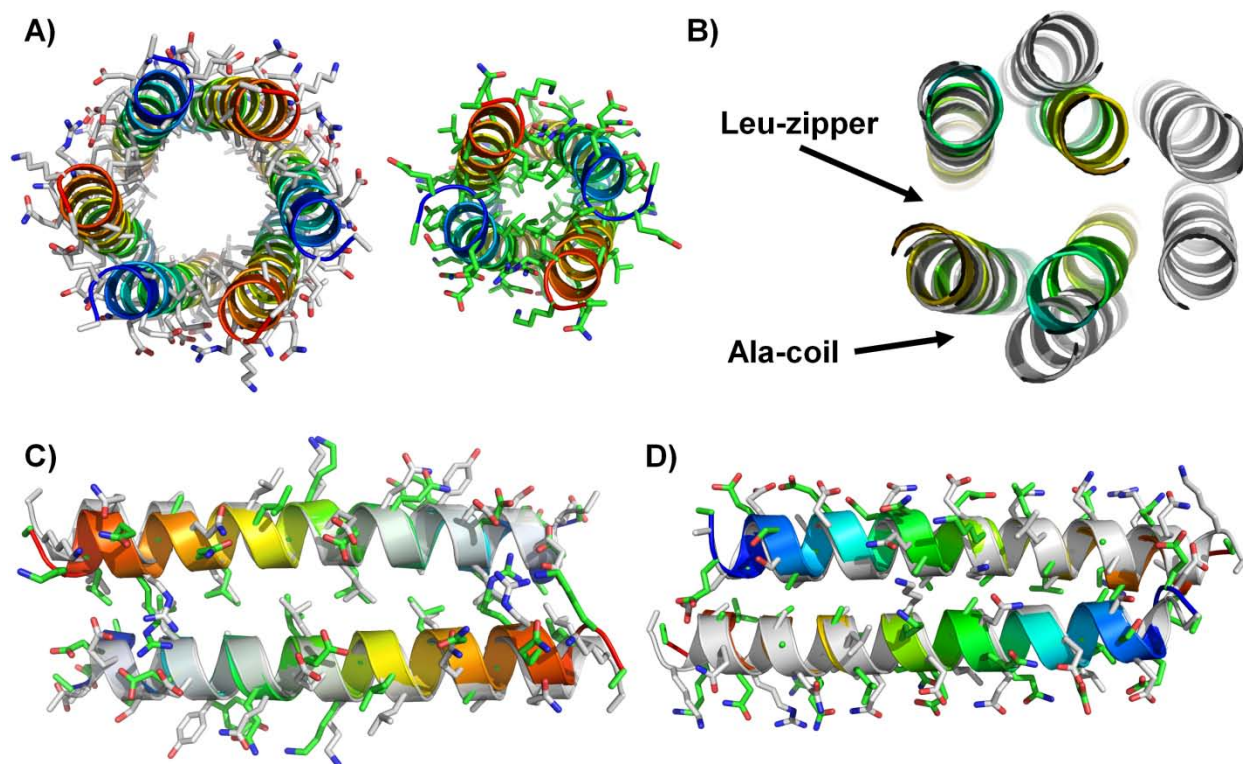
**Figure S11** A detailed view of the modeled assembly between SWNT, gold particles (here depicted as a sphere of 30  $\text{\AA}$  diameter), and Cys-modified DSD-Gly peptide. Cys residues were placed in position 3 of the peptide, creating a convergent gold-binding site. For simplicity, only two chains of the hexameric DSD-Gly are shown.



**Figure S13** UV/Vis spectra of peptide/SWNT suspensions produced by the two *de novo* designed peptides, HexCoil-Ala and HexCoil-Gly, and their corresponding control peptides, cHexCoil-Ala and cHex-Coil-Gly. Spectra were recorded upon a 10-fold dilution of the suspension obtained by sonication/centrifugation. The control peptides display a significantly reduced capacity to solubilize SWNTs compared to their corresponding design peptides, with the effect being particularly stark for cHexCoil-Gly.



**Figure S14** 2D-PL spectrum of SWNT suspension produced with sodium deoxy cholate (SDOC).



**Figure S15** Comparison of intersubunit interaction geometries between the designed SWNT-wrapped hexamer and the tetrameric crystal structure of HexCoil-Ala on its own. **A)** Designed hexamer (left) showing the hydrophobic pore that would be solve-exposed in the absence of SWNT, whereas the crystallographic tetramer (right) has a tightly-packed hydrophobic core (blue-to-red indicates N-to-C terminal direction). **B)** Superposition of the Leu-zipper like interface of the hexamer and the tetramer shows that lower oligomerization state in solution is achieved primarily by adjusting the relative phase of the Ala-coil (“narrow”) interface. **C)** and **D)** show the optimal superposition of the Leu-zipper and Ala-coil interfaces, respectively, between the designed hexamer (gray backbone) and the crystal structure (blue-to-red backbone). Superimposing the central 24 (out of 30) residues per chain results in backbone RMSDs of 1.5 and 1.9 Å for the two interfaces, respectively. To clarify the view, side-chains of residues Gln11 and Phe9 are hidden in **C)** and **D)**, respectively.



**Table S1** Amino-acid alphabet for the first phase of HXT design

HXT positions	Heptad	Amino-acid alphabet	Comments
4, 11, 18, 25	<b>B</b>	L, E, D, R, K, Q, N, S, F, Y	
8, 15, 22	<b>f</b>	E, D, R, K, Q, N, S, F, Y, L	
5, 19	<b>c</b>	A, S, T, L, F, Y, N, Q	
3, 10, 17, 24	<b>a</b>	A, G	Ala for HXT-Ala, and Gly for HXT-Gly. Position that contacts SWNT.
6, 13, 20, 27	<b>d</b>	A	from fishing
7, 14, 21, 28	<b>e</b>	L	from fishing
1	<b>f</b>	A	Pro in DSD, changed to Ala for helix stability
30	<b>g</b>	A	Was Gly in DSD, changed to Ala for stability
2, 9, 16, 23	<b>g</b>	fixed to native DSD	Taken from DSD: Glu, Ser, Lys, Ser
12, 26	<b>c</b>	fixed to native DSD	Taken from DSD: Gln, Gln
29	<b>fF</b>	fixed to native DSD	Taken from DSD: Lys

**Table S2** Amino-acid alphabet for the second phase of HXT design

HXT positions	Heptad	Amino-acid alphabet	Comments
5	<b>c</b>	S, A, L, V, N, Q, H, W, M, F, Y	
9	<b>g</b>	S, A, L, V, N, Q, H, W, M, F, Y	Previously fixed to DSD-native identities
12	<b>c</b>	Q	Flexible conformation, fixed amino-acid identity
19	<b>c</b>	N, A, L, V, I, Q, H, W, M, F, Y, E, K, R	
23	<b>g</b>	S, A, L, V, N, Q, H, W, M, F, Y	Previously fixed to DSD-native identities
remaining	-	taken from solution 6 of first phase	Amino-acid identities and conformations fixed

**Table S3** Data collection, processing and refinement statistics

<b>Data collection and processing</b>	
Space group	I4 <sub>1</sub>
Cell dimensions a, b, c (Å)	45.529, 45.529, 93.293
Resolution (Å)	40.92-2.44 (2.57-2.44)*
R <sub>merge</sub>	0.067(0.168)
I/σ(I)	20.5 (10.8)
Completeness (%)	97.3(81.8)
Multiplicity	7.4 (7.1)
Total number of (observation/unique)	26019/3496
<b>Refinement</b>	
Resolution (Å)	18.89-2.45
Number of reflections	6816
Twin fraction (α) estimated and refined	0.48 (S(H) plot), 0.45 (Britton plot), 0.497(refined)
R <sub>work</sub> /R <sub>free</sub>	0.2438/0.2628
Number of atoms	464
Proteins	446
Ligand/ion	
Waters	18
B-factors (Å <sup>2</sup> )	
Proteins	27.5
Ligand/ion	
Waters	30.6
R.M.S. deviation	
Bond lengths (Å)	0.01
Bond angle (°)	1.09
<b>Ramachandran plot</b>	
Residues in	
Most favorable region (%)	100.0
Additional allowed region (%)	0.0
Generously allowed region (%)	0.0

\*Highest resolution shell is shown in parenthesis.

$R_{work} = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$  where  $F_{obs}$  and  $F_{calc}$  are calculated observed and calculated structure factor amplitudes respectively,  $R_{free}$  was calculated as  $R_{work}$  using 10.0% of the randomly selected unique reflections that were not included in the structure refinement.

### 3. Reference for supporting materials

- S1. G. Grigoryan, W. F. Degrado, *J Mol Biol* **405**, 1079 (January 2011, 2010).
- S2. I. G. Choi, J. Kwon, S. H. Kim, *Proc Natl Acad Sci U S A* **101**, 3797 (Mar 16, 2004).
- S3. T. Lazaridis, M. Karplus, *Proteins* **35**, 133 (May 1, 1999).
- S4. S. C. Lovell, J. M. Word, J. S. Richardson, D. C. Richardson, *Proteins* **40**, 389 (Aug 15, 2000).
- S5. J. Desmet, M. De Maeyer, B. Hazes, I. Lasters, *Nature* **356**, 539 (1992).
- S6. A. R. Leach, A. P. Lemon, *Proteins* **33**, 227 (Nov 1, 1998).
- S7. R. F. Goldstein, *Biophys J* **66**, 1335 (May, 1994).
- S8. D. B. Gordon, S. L. Mayo, *Structure* **7**, 1089 (Sep 15, 1999).
- S9. I. Lasters, M. De Maeyer, J. Desmet, *Protein Eng* **8**, 815 (Aug, 1995).
- S10. N. A. Pierce, J. A. Spriet, J. Desmet, S. L. Mayo, *J Comp Chem* **21**, 999 (2000).
- S11. B. R. Brooks *et al.*, *J Comput Chem* **30**, 1545 (Jul 30, 2009).
- S12. H. R. Powell, *Acta Crystallogr D Biol Crystallogr* **55**, 1690 (Oct, 1999).
- S13. A. G. Leslie, *Acta Crystallogr D Biol Crystallogr* **55**, 1696 (Oct, 1999).
- S14. P. Evans, *Acta Crystallogr D Biol Crystallogr* **62**, 72 (Jan, 2006).
- S15. R. J. Read, *Acta Crystallogr D Biol Crystallogr* **57**, 1373 (Oct, 2001).
- S16. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr D Biol Crystallogr* **61**, 458 (Apr, 2005).
- S17. B. W. Matthews, *J Mol Biol* **33**, 491 (Apr 28, 1968).
- S18. *Acta Crystallogr D Biol Crystallogr* **50**, 760 (Sep 1, 1994).
- S19. P. D. Adams *et al.*, *Acta Crystallogr D Biol Crystallogr* **66**, 213 (Feb).
- S20. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr D Biol Crystallogr* **53**, 240 (May 1, 1997).
- S21. P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (Dec, 2004).
- S22. T. O. Yeates, *Acta Crystallogr A* **44 (Pt 2)**, 142 (Mar 1, 1988).
- S23. T. O. Yeates, *Methods Enzymol* **276**, 344 (1997).
- S24. R. G. Fisher, R. M. Sweet, *Acta Cryst* **A36**, 755 (1980).
- S25. D. Britton, *Acta Cryst* **A28**, 296 (1972).
- S26. H. Durchschlag, P. Zipper, *Progress in Colloid and Polymer Science* **94**, 20 (1994).
- S27. J. Lebowitz, M. S. Lewis, P. Schuck, *Protein Sci* **11**, 2067 (Sep, 2002).
- S28. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res* **14**, 1188 (Jun, 2004).