Supporting information for:

Structural and Spectroscopic Properties of Assemblies of Self-Replicating Peptide Macrocycles

Pim W.J.M. Frederix, Julien Idé, Yigit Altay, Gaël Schaeffer, Mathieu Surin, David Beljonne, Anna S. Bondarenko, Thomas L.C. Jansen, Sijbren Otto and Siewert J. Marrink

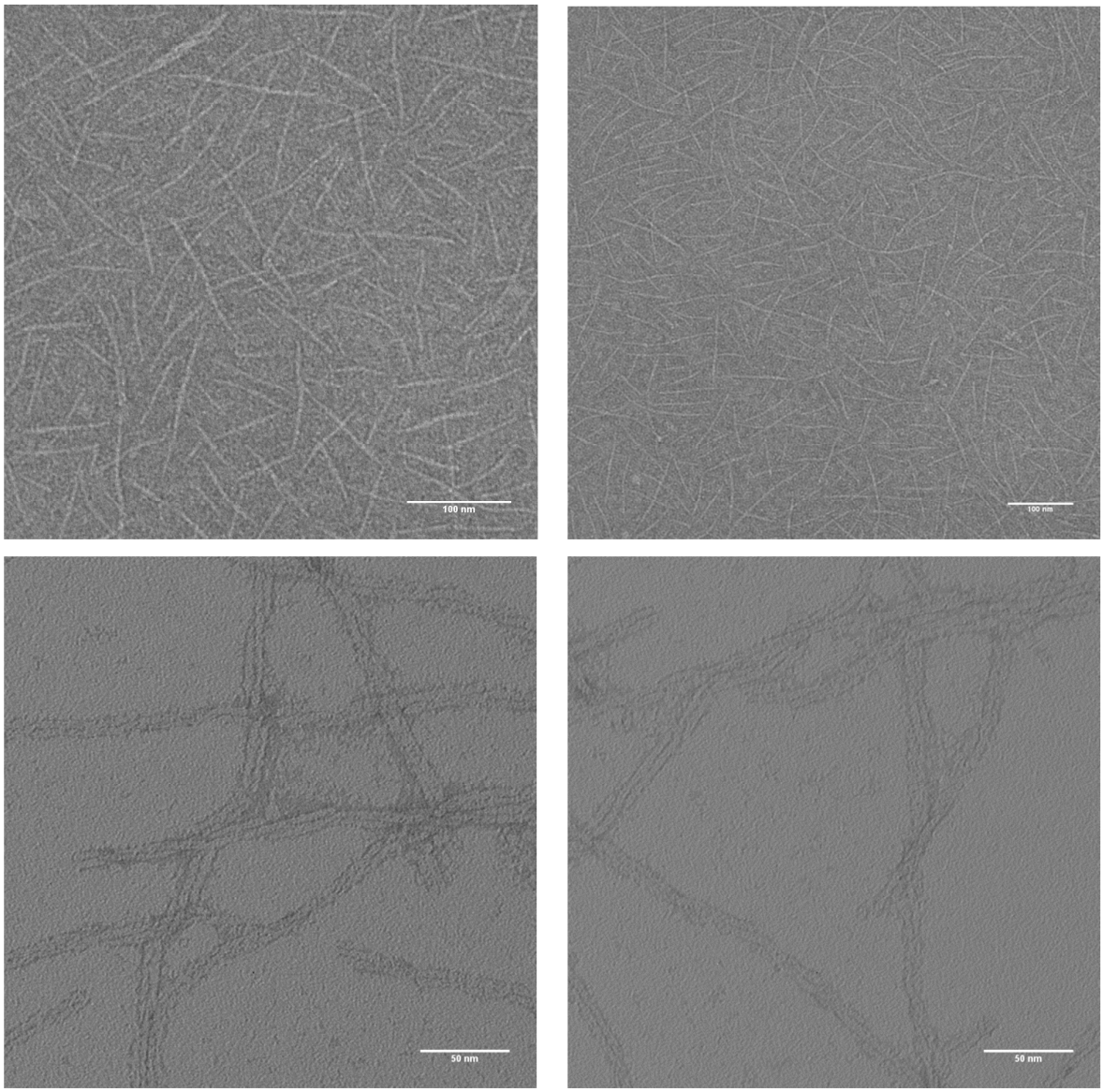


Fig. S1. Additional transmission electron micrographs of the FK hexamer. Top: single fibers. Scale bars 100 nm. Bottom: double fibers. Scale bar 50 nm.

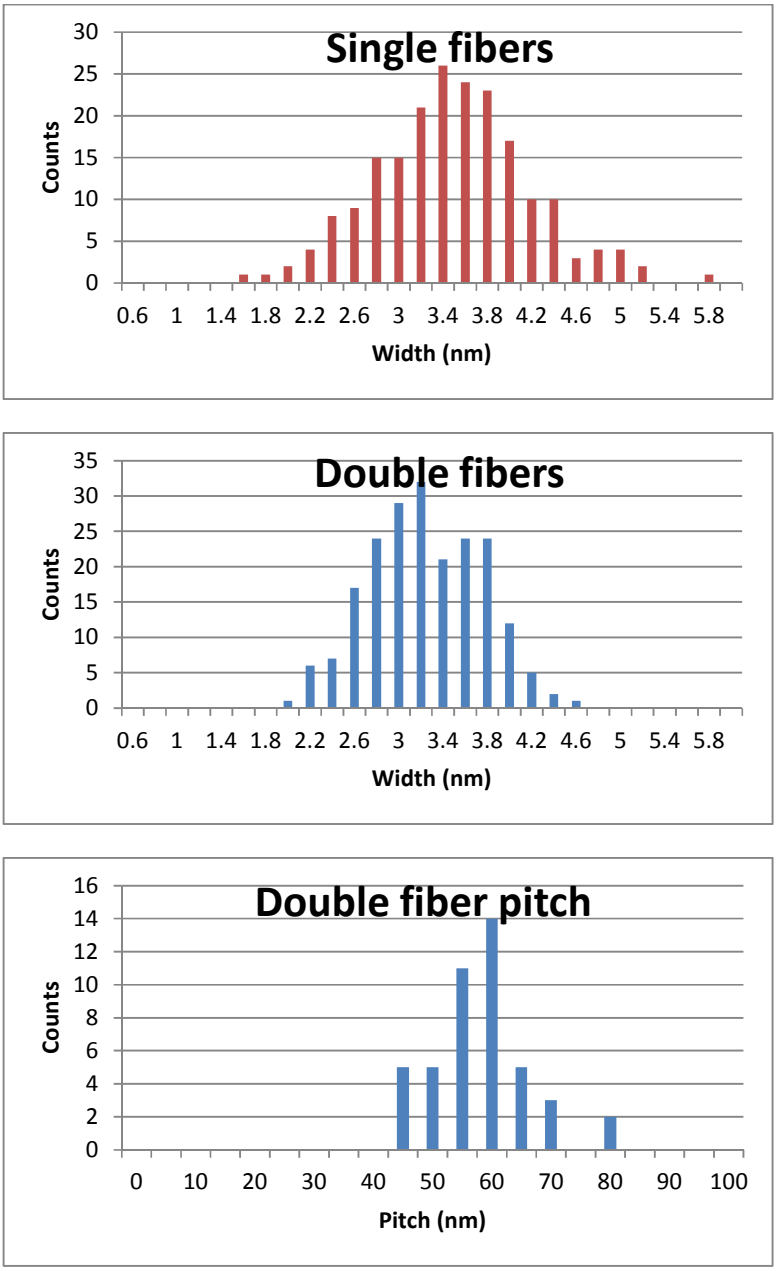


Fig. S2 Statistics on fibers of **1**$_6$. Width and pitch distribution measured from the micrographs in Fig. S1 and main text Fig. 2. The average width is 3.4 ± 0.7 nm for single fibers (top panel, 200 measurements from 5 images) and 3.1 ± 0.5 nm for double fibers (middle panel, 205 measurements from 7 images). Note that we estimate that in both cases the measuring error is larger than the standard deviation of the fiber width.The pitch observed is 56 ± 8 nm and was calculated from any full period twists that were visible across the 7 images of double fibers (lower panel, 45 measurements).

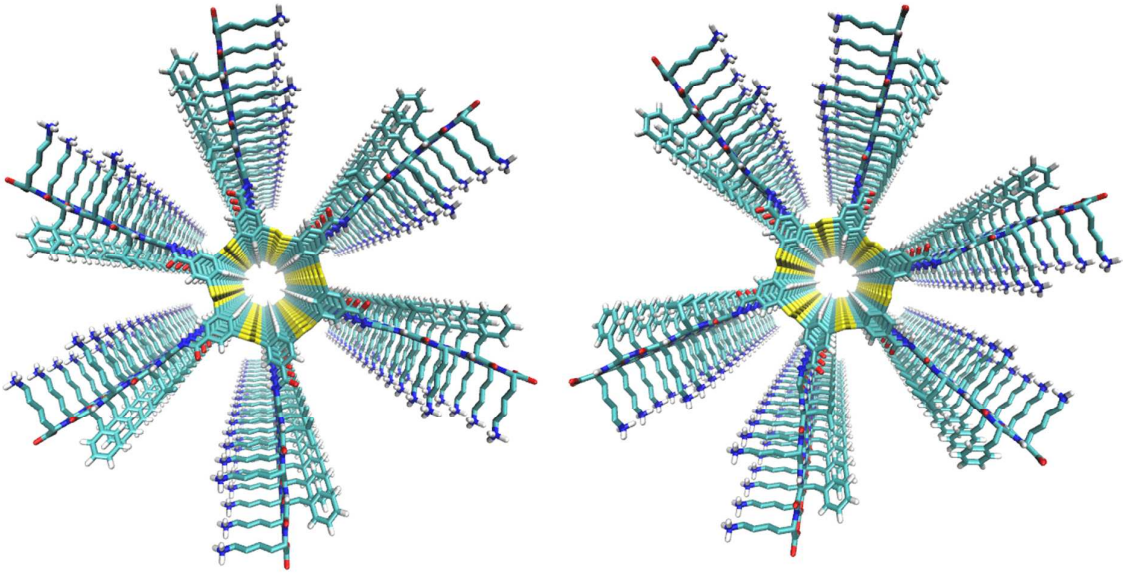


Fig. S3. Perspective view of starting configuration of ‘cartwheel’ (left) and ‘pairwise’ (right) conformations in the GROMOS54a8 force field. Water is not displayed for clarity, but is also present in the center of the macrocycles.

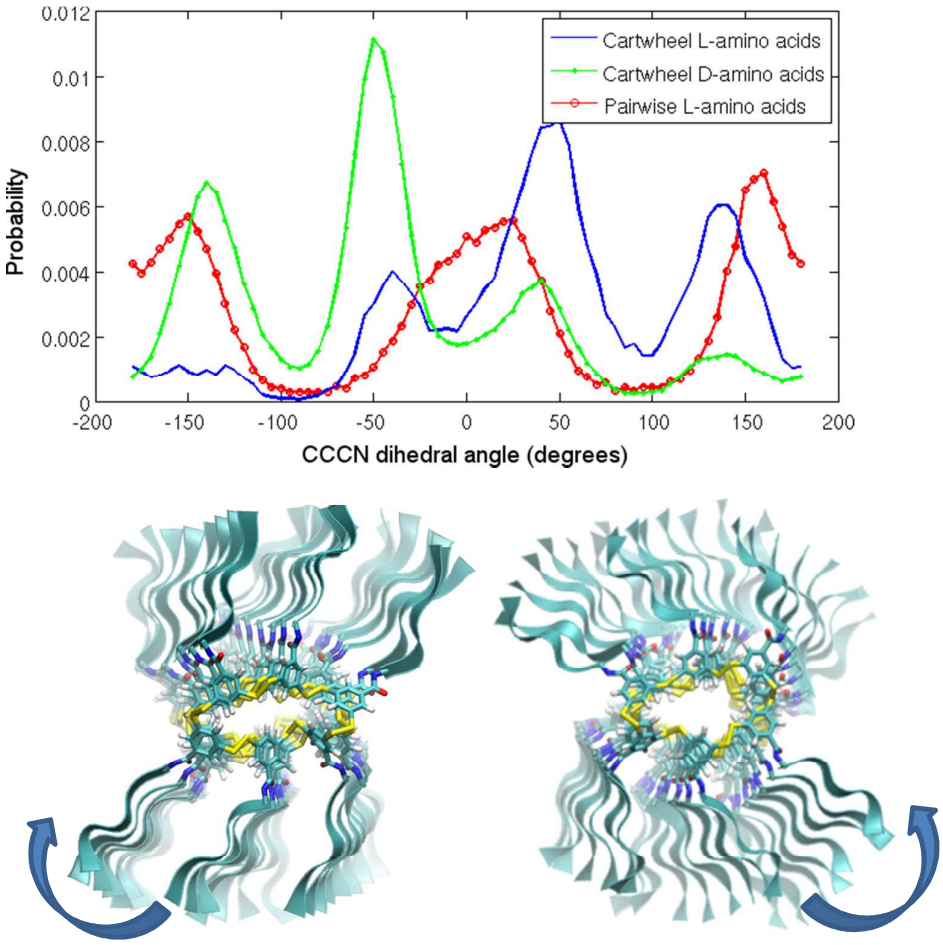


Fig. S4. Top: Distribution of the C-C-C-N dihedral of the first amide bond (head group to glycine) from the average of the last 1.5 ns of 10 independent 3 ns MD simulations for the cartwheel conformations with L- or D-amino acids. Also shown is the distribution of the same dihedral for the pairwise conformation from the last 5 ns of the simulation of the structure in main text Fig. 3D. Bottom: representative snapshots of the L (left) and D (right) structures after 3 ns of which the distribution was calculated, with their chirality indicated.

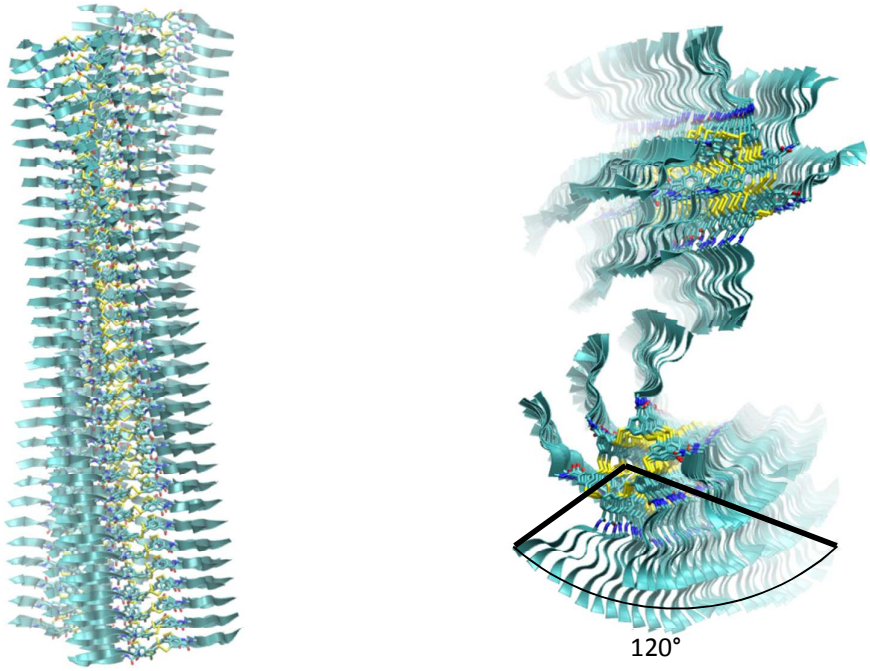


Fig. S5. Side view (left) and top view (right) of a bundle of two left-handed helical fibers. Each fiber consists of 32 hexameric macrocycles. The simulation was set up using two copies of the endpoint of a single fiber simulation in close proximity and was performed for 10 ns using the protocol described in the Methods section. Interfiber contacts are of both ionic (saltbridge) and van der Waals nature (Phe-Phe side chain contacts). The pitch for 32 molecules over a length of 15.4 nm is 120°, giving a full helical pitch of 46 nm.

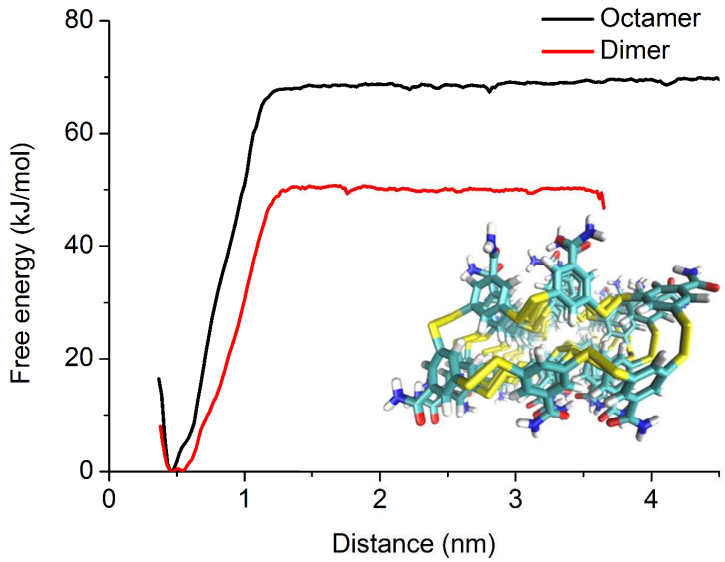


Fig. S6. Umbrella sampling free energy profile of truncated disulfide macrocycles as function of distance between the center of mass (COM) of the top macrocycle to the COM of the second macrocycle in a stack of either 8 (octamer) or 2 (dimer) macrocycles. The inset shows a top view of the stack of 8 macrocycles. Unfortunately, umbrella sampling of the dissociation/association of full macrocycle $\mathbf{1}_6$ was found to be subject to strong overestimation of the binding free energy due to the undersampling of the many degrees of freedom of a full peptide macrocycle using 10 ns sampling windows.

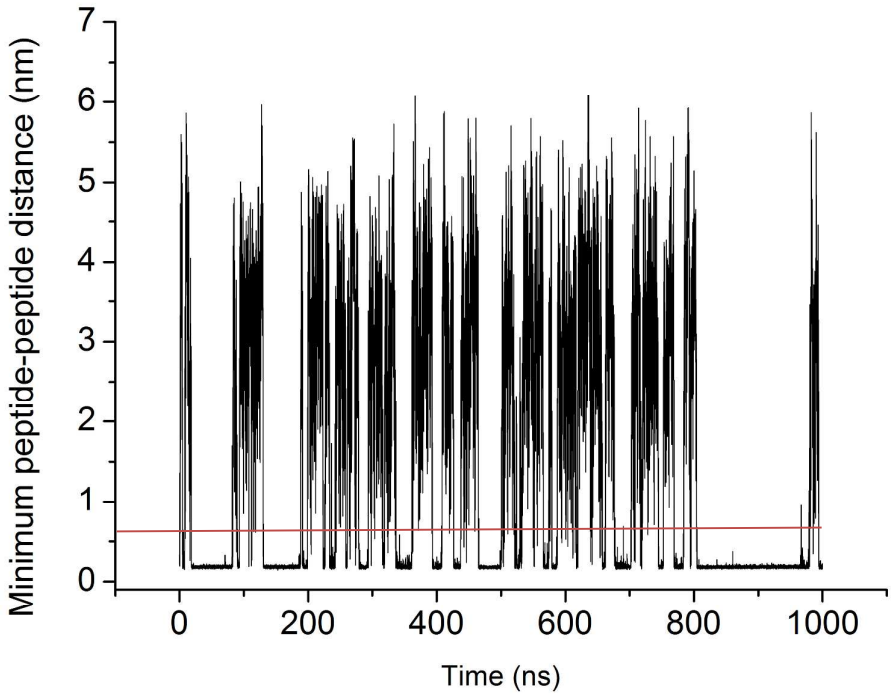


Fig. S7.Minimum distance between two Ac-Gly-Leu-Lys-Phe-Lys peptides as function of time. The red line indicates the distance cut-off used to calculated the binding free energy.

Equation S1 was used to calculate the free energy of dimerization, following the strategy of de Jong et al.[1]

$$G = -k_B T ln(\frac{n_1}{n_0}\left(\frac{v}{v_D} - 1\right)) \qquad \text{(eq. S1)}$$

Where n_1 and n_0 are the number of bound and unbound snapshots, respectively, v is the volume of the simulation box (658.5 nm^3) and v_D is the dimerization volume of a monomer; i.e. the volume of the space in which two molecule would be considered bound. In this case this is taken as the volume of a sphere with a radius of the average radius of gyration of the molecule (0.55 nm) + the cutoff of 0.6 nm.

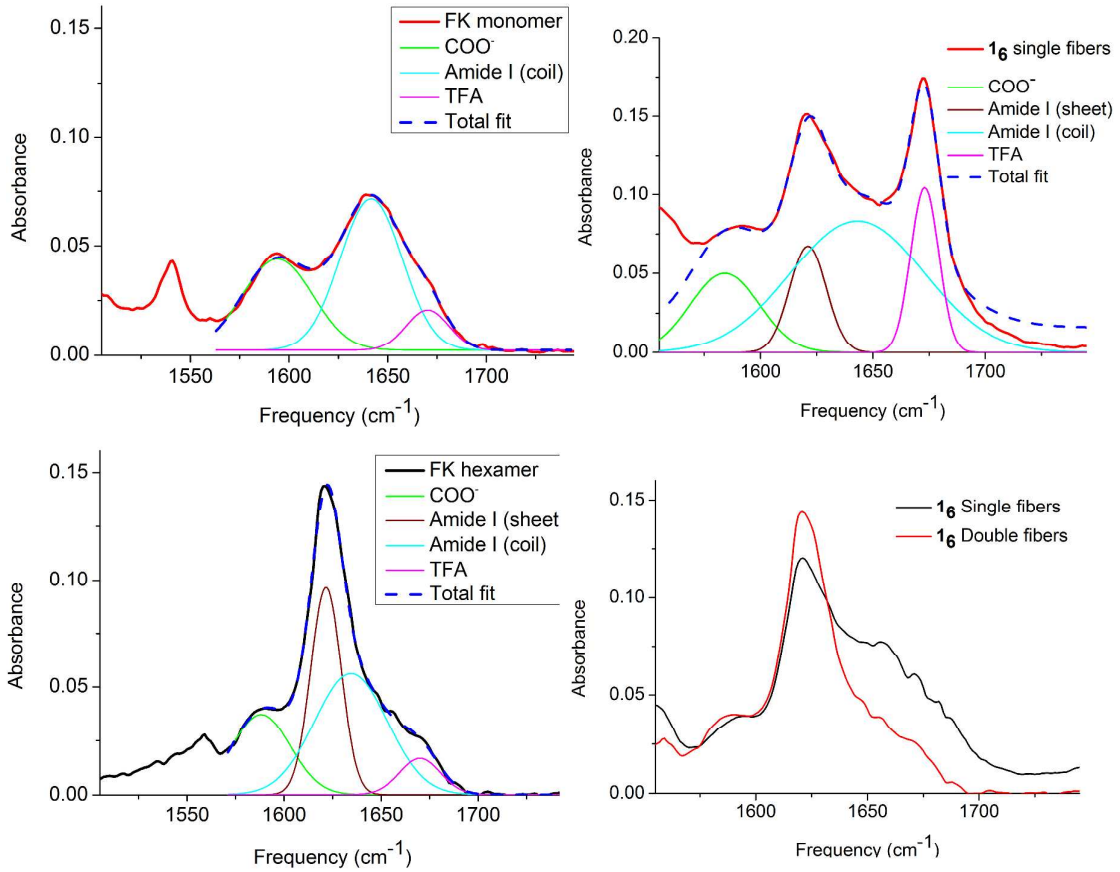


Fig. S8. Decomposition of experimental spectra fitted to absorptions with a Gaussian lineshape. Top left: Specrum of **1** fitted to three Gaussian functions. The fit is not perfect around 1600-1615 cm^{-1}, indicating the possibility of a small absorption originating from the first amide group in the peptide as discussed in the main text. However, the spectrum could not be reliably fitted to four Gaussian functions. Top right: Spectrum of single fibers of $\mathbf{1}_6$ with four Gaussian functions. The narrow peak at 1622 cm^{-1} and the wide peak around 1647 cm^{-1} indicate the presence of both β-sheet-like and less structured peptide chains, respectively. Bottom left: Spectrum of double fibers of $\mathbf{1}_6$ with four Gaussian functions. Bottom right: Overlay of the single and double fiber spectra, indicating the larger sheet-to-coil ratio in slowly oxidized double fibers.

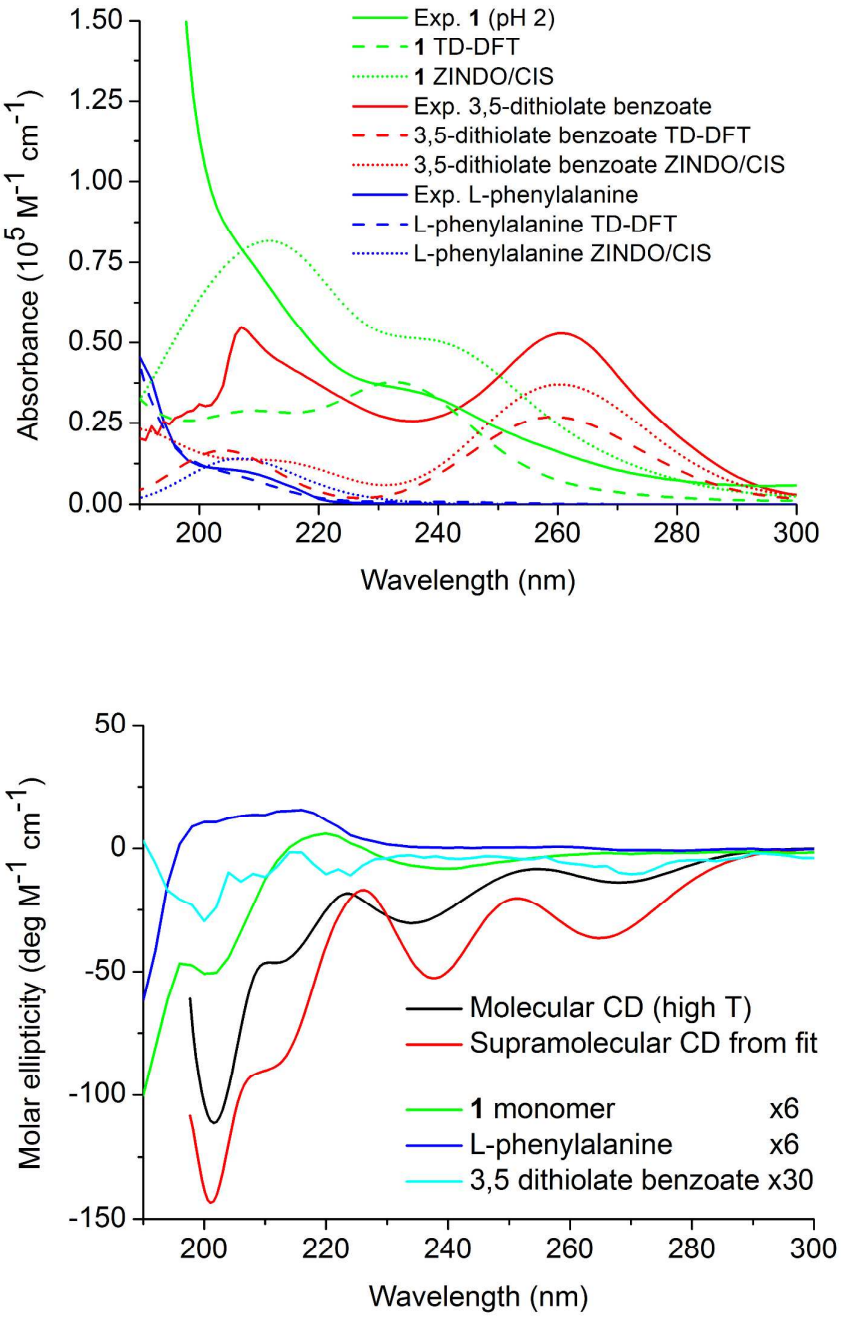


Fig. S9. Top: Experimental UV/Vis spectrum of **1** and its chromophore components 3,5-dithiolbenzamide and L-phenylalanine, overlaid with their TD-DFT (CAM-B3LYP/cc-pVDZ) and ZINDO/CIS calculation results. Results have been redshifted by 0.5 and 0.1 eV for TD-DFT and ZINDO, respectively. ZINDO results for **1** and Phe have been vertically scaled by a factor of 2 and 4, respectively, to facilitate comparison. Bottom: CD spectra for **1**, 3,5-dithiolate benzoate and L-phenylalanine overlaid with the supramolecular CD signal extrapolated from the T-controlled CD experiment. Phe and **1** spectra have been multiplied by 6 to reflect the number of monomers per hexamer. The head group spectrum has been scaled for clarity.

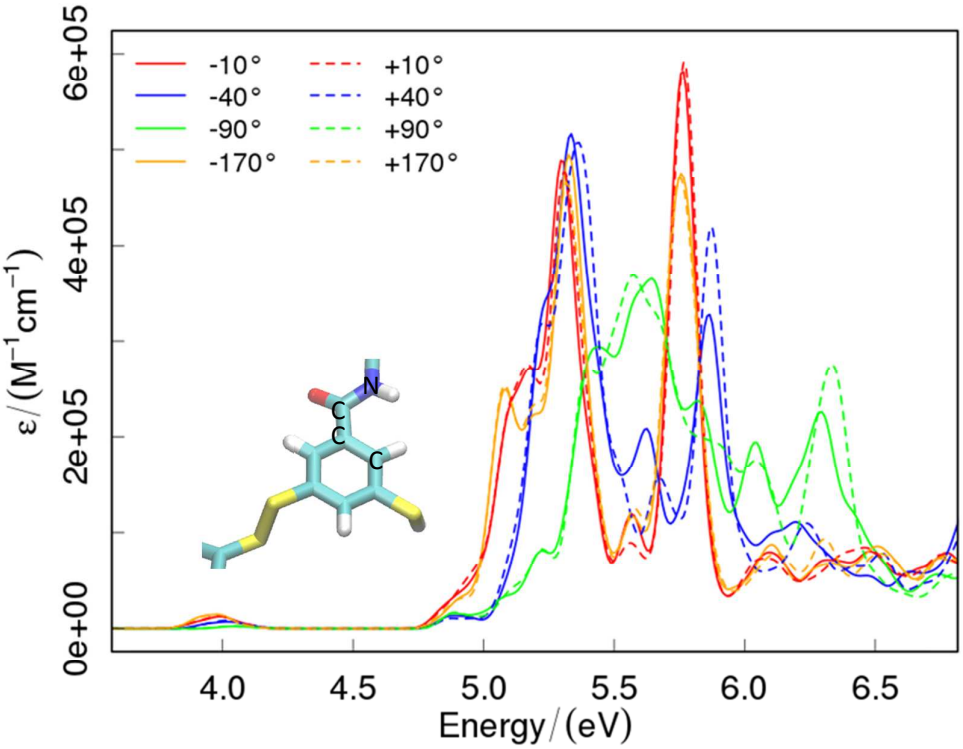


Fig S10. Calculated electronic absorption spectrum of a single macrocycle as function of the C-C-C-N dihedral angles between the aromatic head group and glycine. Note the increase in intensity at 4.0 eV (310 nm), 4.9-5.1 eV (253-243 nm) and 5.8 eV (213 nm) at angles close to 0 or 180°. Inset: definition of the C-C-C-N dihedral.

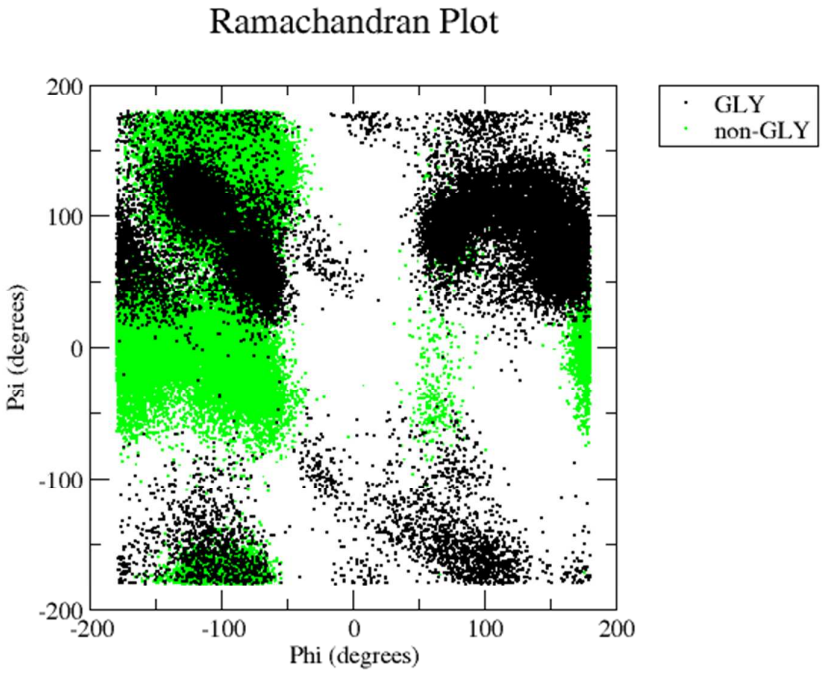


Fig. S11. Ramachandran plot for the first amide group in the peptide (head group to glycine connection, black) compared to the Ramachandran plot for the other amino acids (green). Data was extracted using the Gromacs rama tool from the last 25 ns of the trajectory of the system displayed in main text Fig. 3.

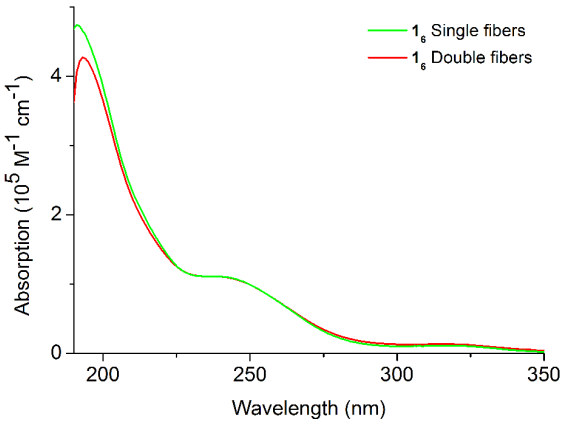


Fig. S12. UV/Vis spectra of double fibers of **1_6**. The single fiber spectrum has been normalized to the intensity at 235 nm and overlaid for comparison.

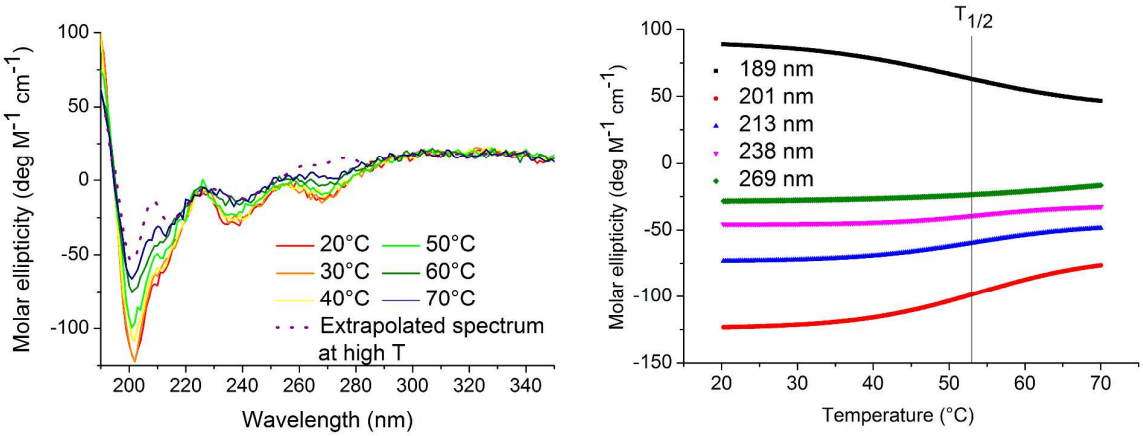


Fig. S13. Left: CD spectra of double fibers of **1_6** as temperature is increasing. Right: Molar ellipticity of double fibers of **1_6** as function of temperature. The average $T_{1/2}$ is indicated. Note that the 269 nm band did not fit well to a sigmoid curve, suggesting it may consist of two separate bands.

Parametrization of aromatic disulfide head group

For the peptide component of molecule **1**, the parameters available in the GROMOS54a8 force field were used. However, not all parameters for the aromatic disulfide head group of the peptide were available.

Bonded parameters:

As flexibility around the disulfide bond and the linker to the peptide strand was rendered important, the dihedrals considering these bonds were parametrised by comparing the MD single point energy to QM single point energies at the BP86/DZVP level of theory using GAMESS UK on the system of truncated $\mathbf{1}_2$ depicted in Fig. S14. Ryckaert-Bellemans (RB) or standard dihedral potentials were fitted to the difference in energy and added to the force field.

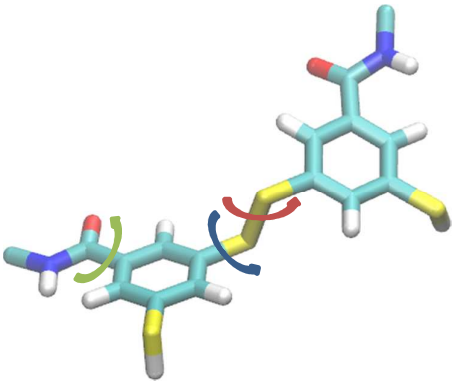


Fig. S14. Aromatic head group dimer used for parameterising dihedral angles in the GROMOS54a8 force field. The colors of the arrows correspond to the colors in Table S1.

Table S1. Dihedral angle parameters for the three dihedral angles indicated in Fig. S14.

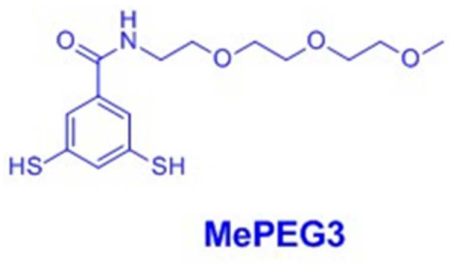
Dihedral (RB)	*C0*	*C1*	*C2*	*C3*	*C4*	*C5*
1 (CSSC) RB	-2.922	-5.231	13.782	8.33	19.196	-14.391
2 (CCSS) RB	-0.232	2.767	4.904	0.814	-9.281	-2.235
Dihedral	*phase*	*Force constant*	*multiplicity*			
3 (CCCO)	180	11.0	2			

Validation of non-bonded parameters:

CH_3Cl : H_2O partitioning was measured on a model system of the aromatic dithiol head group attached to a triethylene glycol tail (Fig. S15) by UPLC. The equivalent partitioning was calculated using thermodynamic integration using the default GROMOS54a8 atom types. Good agreement was found between experimental and calculated results.

$\Delta\Delta G_{CH3Cl \rightarrow water}$, experimental: 3.3 ± 1.3 kJ/mol

$\Delta\Delta G_{CH3Cl \rightarrow water}$, calculated: 3.6 ± 0.8 kJ/mol



MePEG3

Fig. S15. Model system used for parameterisation of non-bonded parameters of the dithiol aromatic head group. Synthesis of MePEG3 is described elsewhere.[3]

Supplementary references

(1) De Jong, D. H.; Schäfer, L. V.; De Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C.; Grubmüller, H. Determining Equilibrium Constants for Dimerization Reactions from Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 1919–1928.

(2) Gupta, V. P.; Keiderling, T. A. Vibrational CD of the Amide II Band in Some Model Polypeptides and Proteins. *Biopolymers* **1992**, *32*, 239–248.

(3) Komáromy, D.; Stuart, M. C. A.; Monreal Santiago, G.; Tezcan, M.; Krasnikov, V. V.; Otto, S. Self-Assembly Can Direct Dynamic Covalent Bond Formation toward Diversity or Specificity. *J. Am. Chem. Soc.* **2017**, *139*, 6234–6241.