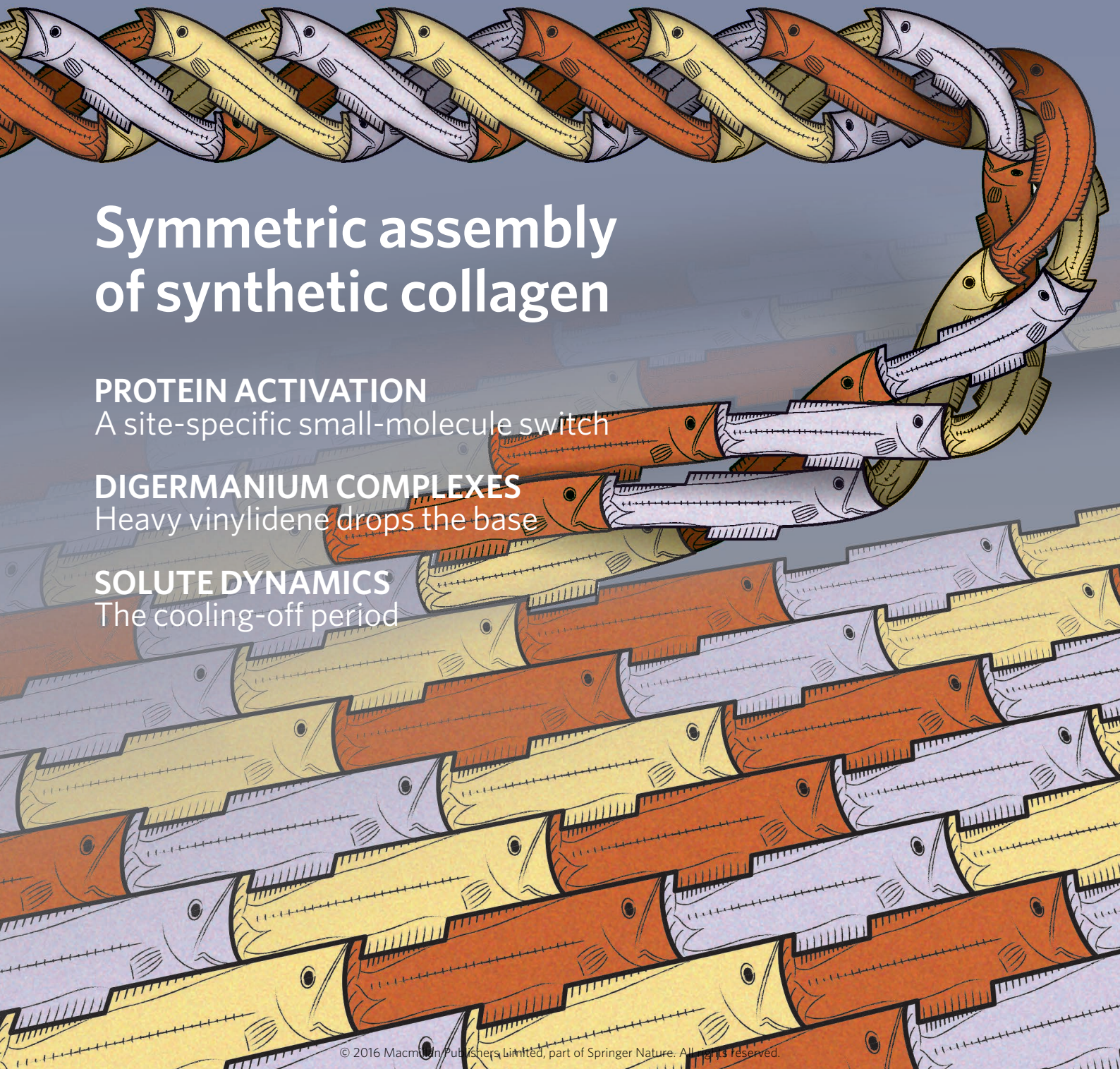
NOVEMBER 2016 VOL 8 NO 11
www.nature.com/naturechemistry

nature chemistry



Symmetric assembly of synthetic collagen

PROTEIN ACTIVATION
A site-specific small-molecule switch

DIGERMANIUM COMPLEXES
Heavy vinylidene drops the base

SOLUTE DYNAMICS
The cooling-off period

ARTICLES
PUBLISHED ONLINE: 11 JULY 2016 | DOI: 10.1038/NCHEM.2556
nature chemistry

Peptide tessellation yields micrometre-scale collagen triple helices

I. Caglar Tanrikulu[1], Audrey Forticaux[2], Song Jin[2] and Ronald T. Raines[1,2]*

Sticky-ended DNA duplexes can associate spontaneously into long double helices; however, such self-assembly is much less developed with proteins. Collagen is the most prevalent component of the extracellular matrix and a common clinical biomaterial. As for natural DNA, the ~10^3-residue triple helices (~300 nm) of natural collagen are recalcitrant to chemical synthesis. Here we show how the self-assembly of short collagen-mimetic peptides (CMPs) can enable the fabrication of synthetic collagen triple helices that are nearly a micrometre in length. Inspired by the mathematics of tessellations, we derive rules for the design of single CMPs that self-assemble into long triple helices with perfect symmetry. Sticky ends thus created are uniform across the assembly and drive its growth. Enacting this design yields individual triple helices that, in length, match or exceed those in natural collagen and are remarkably thermostable, despite the absence of higher-order association. The symmetric assembly of CMPs provides an enabling platform for the development of advanced materials for medicine and nanotechnology.

Strict base-pairing rules enable the DNA double helix to store, recall and replicate biochemical information. The ability of DNA strands to associate in a sequence-specific manner also allows DNA to form duplexes with single-stranded overhangs. These 'sticky ends' enable duplexes with complementary overhangs to anneal and thereby extend the double helix. The ensuing cohesive assemblies can far exceed the size of constituent strands and underlie the fields of molecular biology[1,2], synthetic biology[3] and DNA nanotechnology[4]. Similarly, the sticky-ended self-assembly of certain α-helical peptides can yield long, although thick, bundles[5,6].

Similar to that of DNA and α-helical bundles, the structure of collagen features intercoiled strands. As collagen is the predominant structural protein in animals[7,8], non-human collagen is the most-common biomaterial in the clinic, but its use can be complicated by allergic reaction and pathogen transmission[9]. Natural collagen contains a high level of post-translational modification and its biosynthesis is regulated tightly[10], complicating the heterologous production of human collagen, which has strands of ~10^3 residues (~300 nm in length). Conversely, chemical synthesis can offer only collagen-mimetic peptides (CMPs) of ~30 residues (~10 nm), which have little practical utility[11]. To endow CMPs with utility requires their assembly into 'human-scale' collagen.

The assembly of both DNA strands into double helices and α-helical peptides into coiled coils relies on side-chain–side-chain interactions. In contrast, the triple-helical association of collagen proceeds through main-chain–main-chain interactions, influenced only weakly by the primary sequence. The structure of the collagen triple helix is defined by the repeating Xaa–Yaa–Gly tripeptide units (XYG repeats) of its strands, in which proline (Pro (P)) and 4-hydroxyproline (Hyp (O)) are abundant in the Xaa and Yaa positions, respectively, and a glycine residue (Gly (G)) is essential in every third position[12]. The resulting sequence favours the formation of left-handed polyproline-type II helices, which associate into right-handed triple helices through main-chain–main-chain hydrogen bonds. A tight, compact association is attained by a single-residue shift that places one Xaa, Yaa and Gly residue from each strand at every helical cross-section, which creates three distinct strand 'registers' (Fig. 1a). The resulting scaffold directs the Xaa and Yaa side chains away from the central triple-helical axis. The radiating side chains allow for extensive functionalization[13], but limit the options for CMP assembly. Moreover, a 'blunt-ended' association state is accessible to all CMPs, regardless of sequence.

Owing to the secondary role of the Xaa and Yaa residues in triple-helical association, the design of a CMP that avoids blunt-ended trimers has been a challenge. During the past decade, our group and others have reported on strategies that enable CMPs to form extended structures[14–21]. Each of these systems, however, relies on the tedious crosslinking of three strands to enforce sticky ends or yields thick fibres of limited utility. Here we show that well-defined, productive sticky ends for triple-helical assemblies necessitate the implementation of symmetry. We derive rules for the design of CMPs that tessellate along the triple-helical axis with uniform sticky ends and self-assemble into individual long triple helices, and we use these rules to generate human-scale collagen by the self-assembly of a single peptide.

Symmetric-assembly design

The recent discovery of a stabilizing 'axial' salt bridge in the collagen triple helix has enabled CMP association to be controlled through peptide sequence[22]. These salt bridges link a lysine residue (Lys (K)) in the Yaa position and an aspartic acid residue (Asp (D)) in the Xaa position three residues down on the strand that occupies the next register (Fig. 1a)[23]. Hence, permutations of $(PKG)_4$, $(DOG)_4$ and $(POG)_4$ blocks within a 36-mer have been used as cationic, anionic and neutral domains that weaken the blunt-ended association and strengthen the sticky-ended assembly of a CMP (Fig. 1b). The resulting constructs vary greatly in their morphology—some form higher-order structures, such as hydrogels[19] and birefringent rods[21], and others remain as amorphous aggregates[21]. This heterogeneity is partially attributable to the association of unpaired Lys and Asp residues with other strands or triple helices (for example, as in Fig. 1c).

We reasoned that engaging all the Lys and Asp residues in interstrand salt bridges could ensure maximal stability for CMP hybridization and assembly, widen the energetic gap between the sticky- and

[1]Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. [2]Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. *e-mail: rtraines@wisc.edu

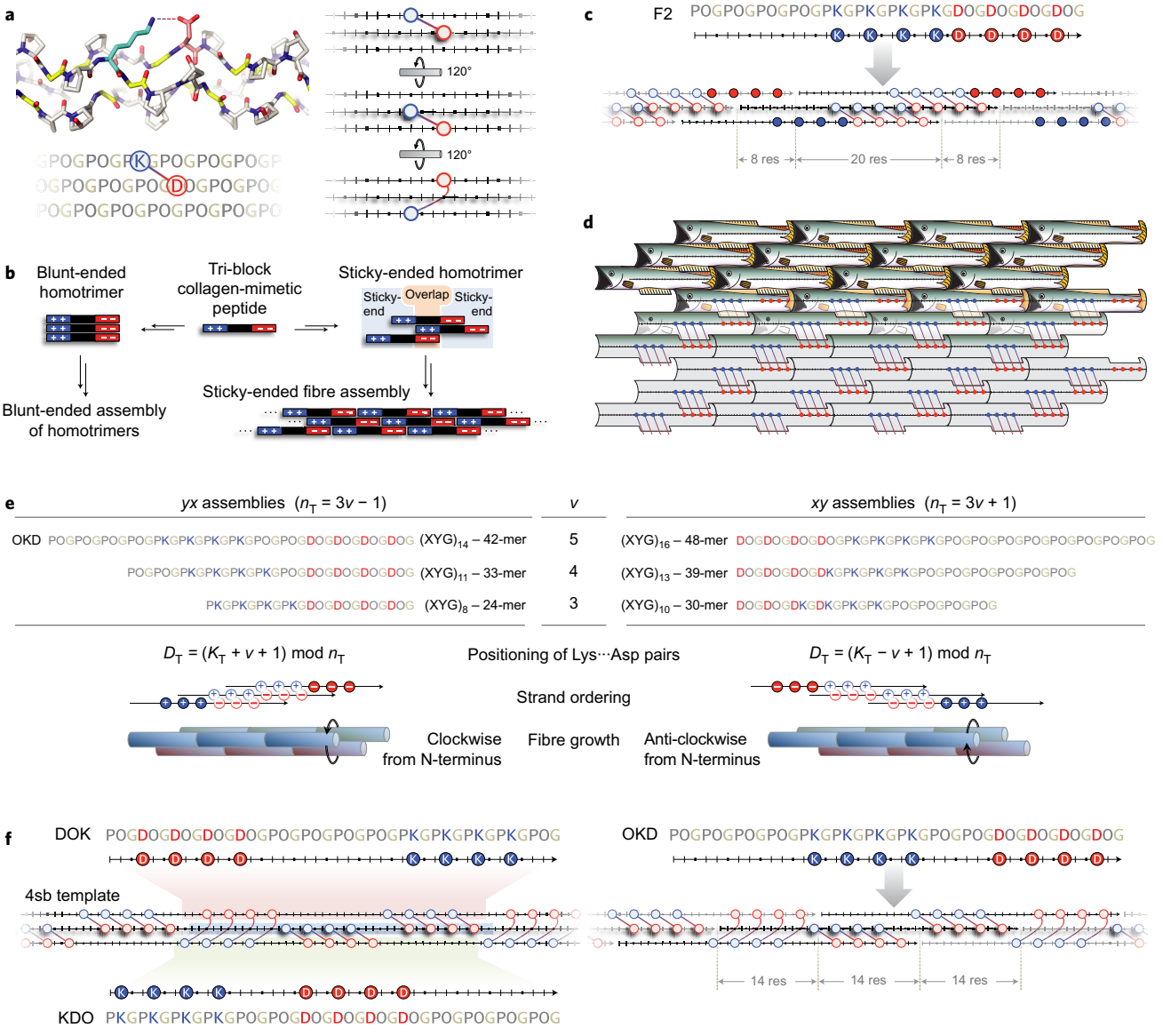


Figure 1 | Design of symmetric CMP self-assemblies. a, Molecular structure of the collagen triple helix and depictions of an axial salt bridge. The model, built on PyMOL v1.3 and based on Protein Data Bank 3u29, presents an interstrand Lys (cyan)···Asp (red) salt bridge over a $[(POG)_n]_3$ background with Gly residues shown in yellow. In the triple helix, salt-bridging residues are indicated by interconnected blue and red circles. Pro/Hyp and Gly residues on a CMP are depicted simply as bars and dots on a line to emphasize the Lys···Asp salt bridges. **b**, Fibrillar self-assembly of tri-block CMPs through blunt-ended[15] and sticky-ended[19] pathways. **c**, Self-assembly of the F2 peptide[21]. Unpaired Lys and Asp residues implicated in unintended associations are indicated with filled circles. **d**, Principles of plane geometry that inspire artistic tessellations can also inspire the design of symmetric CMP assemblies. The edges that each tessellating fish shares with its surroundings are indistinguishable. Similarly, assemblies with fully satisfied charge pairs can be achieved when the chemical environment of every CMP in the assembly (grey tiles) is designed to be identical. **e**, Design rules for the symmetric assembly of CMPs. The two classes of symmetric assembly require different relationships between K_T and D_T, which are the positions of XYG units that host Lys and Asp residues of an interacting Lys···Asp pair to ensure a symmetric strand association. The examples have four Lys···Asp pairs per CMP. **f**, Symmetric assembly of 4sb-derived peptides. DOK, KDO and OKD peptides are shown on the 4sb template, highlighted in red, green and blue, respectively. The perfect charge pairing and uniform 14-residue interstrand stagger ensured by the symmetry rules are demonstrated with the OKD assembly.

blunt-ended associations and limit unintended interactions. We realized that such perfect pairing of charges is possible only if the design incorporates elements of symmetry. A symmetric assembly would require every peptide to engage in identical interactions with its neighbours, analogous to the tessellation of uniform tiles described by the geometer H. S. M. Coxeter[24] and depicted by the artist M. C. Escher[25] (Fig. 1d). Symmetric assembly has been invoked as a design principle in other biomolecular contexts[26,27]. Here, symmetry would ensure that each CMP engages in an identical number of salt bridges, and would require that the same number of residues offset any two immediate neighbours in the assembly. Thus, the required sticky ends are of uniform size across the assembly at $n/3$ residues for an n-residue CMP. This rule, in combination with the requirement for a single-residue stagger between XYG repeats on neighbouring strands, restricts the peptides that are amenable to symmetric assembly to those that are $n_T = 3v \pm 1$ XYG repeats in length, where n_T is the number of XYG repeats on the CMP and v is a positive integer. Symmetric assemblies are therefore not possible for a 36-residue CMP ($n_T = 12$) (Supplementary Fig. 1).

Sequences that satisfy symmetric assembly requirements fall into two distinct classes (Fig. 1e). The *yx* class assembles through peptides of size $n_T = 3v - 1$ and accepts new building blocks in a

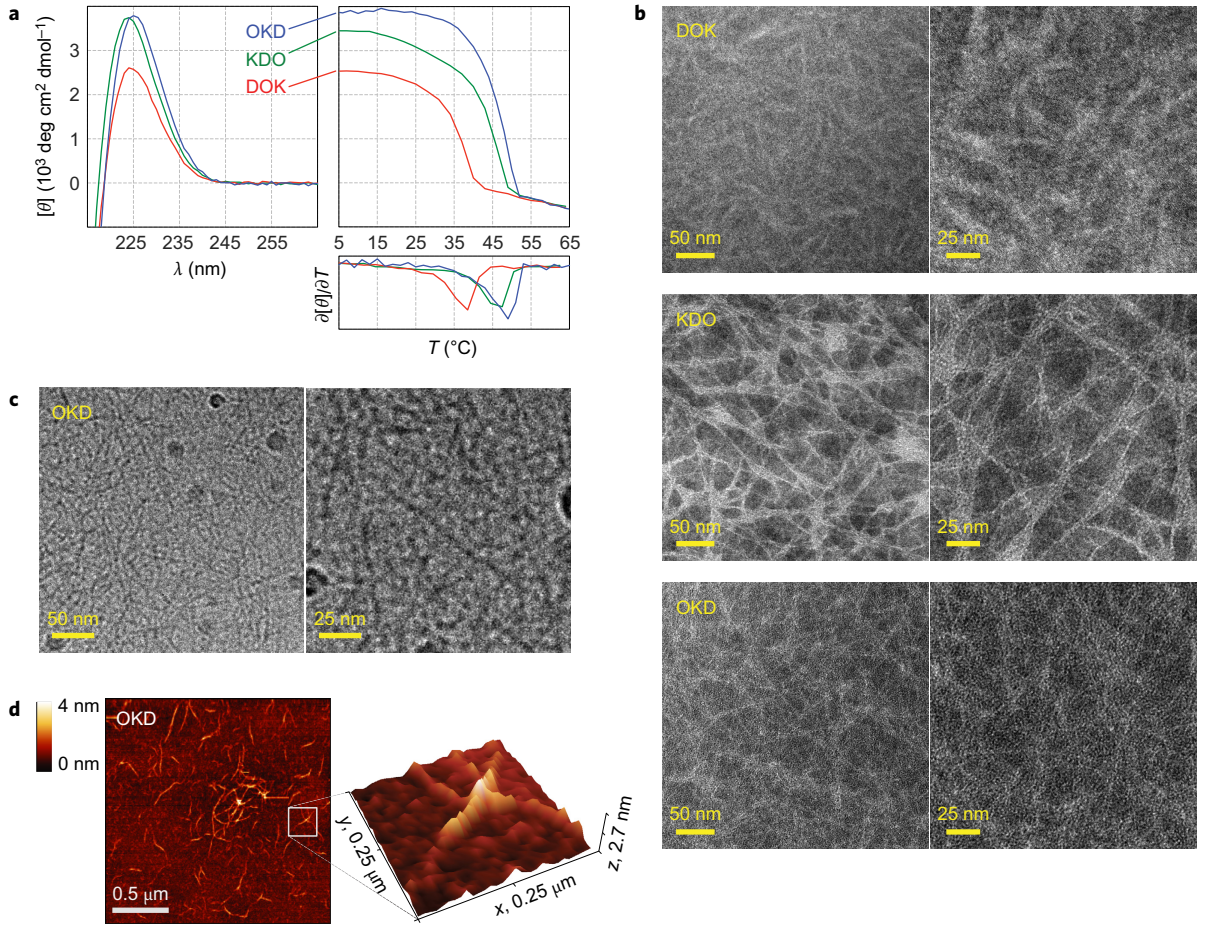


Figure 2 | Characterization of 4sb-derived CMP self-assemblies. a, CD spectra and thermal denaturation curves at 0.6 mg ml^{-1} confirm triple-helical assemblies. All the samples were prepared in 10 mM sodium phosphate buffer, pH 7.0, unless noted otherwise. **b**, TEM images of peptide assemblies reveal nanofibres for all the permutations. **c**, Vitreous-ice cryo-TEM images of OKD nanofibres (0.3 mg ml^{-1}). Triple helices do not form bundles in solution. Circular spots are ethane artefacts introduced during flash freezing. **d**, OKD nanofibres (0.03 mg ml^{-1}) imaged by AFM on freshly cleaved mica. Nanofibres retain their 1.2 nm diameter, even when overlaid. Colour scale refers only to the 2D AFM image.

clockwise order around the triple-helical axis during N- to C-terminal growth, whereas peptides of size $n_T = 3v + 1$ (the *xy* class) assemble anticlockwise. Although v represents the number of XYG repeats that offset neighbouring CMPs for both classes, the sign of the offset and ordering of strands differ. Thus, $n_T = 3v - 1$ assemblies favour the placement of PKG repeats ahead of DOG repeats on the peptide sequence, and vice versa. Among the candidates that allow four fully paired Asp and Lys residues per CMP, we opted to study the 42-residue system (*yx*, $v = 5$) as it enables the segregation of charged blocks in the assembled state.

The selected assembly can be interpreted as a homotrimer of three infinite strands, each offset by a 14-residue 'super stagger' with respect to its neighbours. We name this infinite homotrimer, which bears four salt bridges every 14 residues, the '4sb template' (Fig. 1f). Any 42-residue section taken out of this master template satisfies symmetric-assembly design requirements equally, and should be able to re-establish all contacts on the template on assembly. We chose to synthesize and assess three 42-mers (DOK, KDO and OKD (Fig. 1f)), which represent three extremes for charge placement on a 4sb-type peptide. As intended, symmetric design provides identical chemical environments within the assembly for all three permutations and allows uninterrupted CMP tessellation with fully paired Asp and Lys residues (Supplementary Fig. 2).

Results

Characterization of symmetric CMP assemblies. We employed methods established previously for CMP synthesis[14], purification[28] and sample preparation[28,29]. Annealed peptides were not turbid and remained soluble at all concentrations tested (up to 3.5, 5.0 and 3.0 mg ml^{-1} for DOK, KDO and OKD, respectively). Thermal denaturation experiments revealed 'melting' temperatures (T_m) above 37 °C for all the assemblies. All the peptides displayed a strong circular dichroism (CD) signature for collagen at ~225 nm, indicative of triple-helix formation (Fig. 2a); self-assembly, confirmed by analytical ultracentrifugation (AUC), yields an increase in size with an increase in thermostability (Supplementary Fig. 3). The DOK peptide, which keeps the $(POG)_4$ section mid-sequence, is the least-stable permutation ($T_m = 38$ °C), whereas the KDO and OKD peptides with negatively and positively charged central sections form assemblies with higher T_m values (46 and 49 °C, respectively). The denaturation curves are increasingly cooperative for the more-stable peptides, and display transitions reminiscent of cooperative assemblies that follow a nucleation–elongation model[30]. Although the surprisingly high thermostability of 4sb-like assemblies ($T_m = 38$–49 °C) over those of 36-mers reported previously ($T_m = 15$–25 °C (Supplementary Fig. 4)) could be attributable to the $(POG)_2$ extension, such an addition is expected to increase the T_m only by ~8 °C (ref. 22). Thus, we

Table 1 | Assembly characteristics of 4sb-derived peptides.

Peptide	T_m (°C)*		Average oligomerization by AUC†	TEM imaging of assemblies	
	I = 20 mM	I = 200 mM		Nanostructure and bundling	Feature dimensions (nm)
OKD	49	39	>188	Single TH nanofibres	2.3 ± 0.4
KDO	46	38	>61	Single TH nanofibres, intermittent bundling	3.8 ± 0.7
DOK	38	31	>47	Multiple TH nanofibres, no higher-level organization	7.9 ± 2.2
DOKctrl	27	31	3‡	None	-

*Values of T_m were determined at both 20 and 200 mM ionic strength (I). †Estimates for a lower-bound assembly size were obtained from a monomer + multimer model fit to AUC data at 15,000 r.p.m., 4 °C. ‡DOKctrl data do not agree with a monomer + multimer model, fitting best to a trimer + multimer model dominated by the trimeric species (see the Supplementary Information and Supplementary Fig. 3 for discussions). TH, triple helix.

believe that the additional interstrand salt bridges and uniform sticky ends enabled by our design play a major role in improving the assembly stability.

Another expected outcome of our strategy is the absence of a higher-order association of nanofibres, which we confirmed by negative-stained transmission electron microscopy (TEM) imaging. The 4sb-like fibres are <10 nm in diameter, are oriented and distributed irregularly and exhibit features that are as fine as those of any synthetic collagen characterized previously (Fig. 2b). Interestingly, the trends among 4sb-like assemblies for thermostability and nanostructure overlap remarkably well (Table 1). DOK assemblies, which have the lowest thermostability (by CD) and smallest assembly size (by AUC), also form the widest nanofibres (mean ± s.d. = 8 ± 2 nm). Comparison with previous TEM studies of single triple helices of natural collagen (~4 nm)[31,32] suggests that DOK fibres might be formed through the association of multiple triple helices. KDO and OKD nanofibres exhibit diminishing diameters at 3.8 ± 0.7 nm and 2.3 ± 0.4 nm, respectively, consistent with a single triple helix. Whereas KDO and OKD both produce intertwined networks, KDO nanofibres have a greater tendency for bundling, and commonly remain aligned for 30–50 nm before separating.

The structure and stability of OKD nanofibres are most impressive. Although individual OKD nanofibres are not isolable through dilutions, their bundling is rare and triple-helical stretches can reach 0.3 μm in length. The properties of OKD assemblies remain largely the same when hydrated, as evidenced by cryo-TEM, a technique in which the solution structure is preserved by flash freezing in liquid ethane at −196 °C prior to imaging. Solvated nanofibres appear to be distributed randomly, and are intertwined extensively without forming bundles (Fig. 2c). These solvated nanofibres have a uniform diameter (3.8 ± 0.5 nm) that is slightly thicker than those recorded for dried samples, probably because of tightly associated water and buffer molecules. Individual OKD nanofibres observed

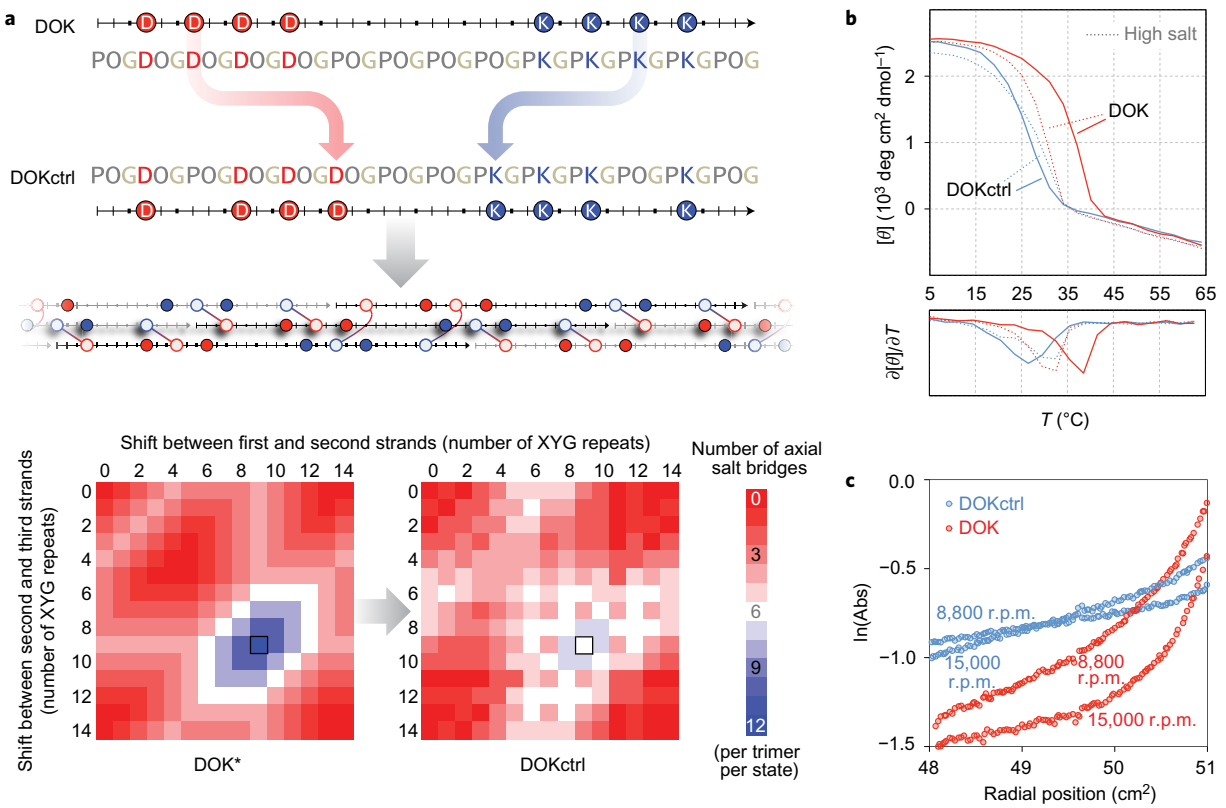


Figure 3 | Perturbing the strand-association landscape abolishes assembly. a, CMP sequences and strand-association landscapes for DOK and DOKctrl. Mutations to DOK destabilize its unique optimum for self-assembly and result in a landscape that lacks any defining features. Symmetric association states for DOK and DOKctrl are marked with black borders. *This landscape describes strand association not only for DOK, but for all 4sb-derived peptides. **b**, CD thermal denaturation curves for DOK and DOKctrl (0.6 mg ml^{-1}). Data acquired at a low ionic strength (20 mM, solid line) and at a high ionic strength (200 mM, dashed line) point to diverging assembly pathways. **c**, Equilibrium gradients during ultracentrifugation indicate different assembly regimes for DOK (steep gradients) and DOKctrl (shallow gradients) at 0.3 mg ml^{-1}.

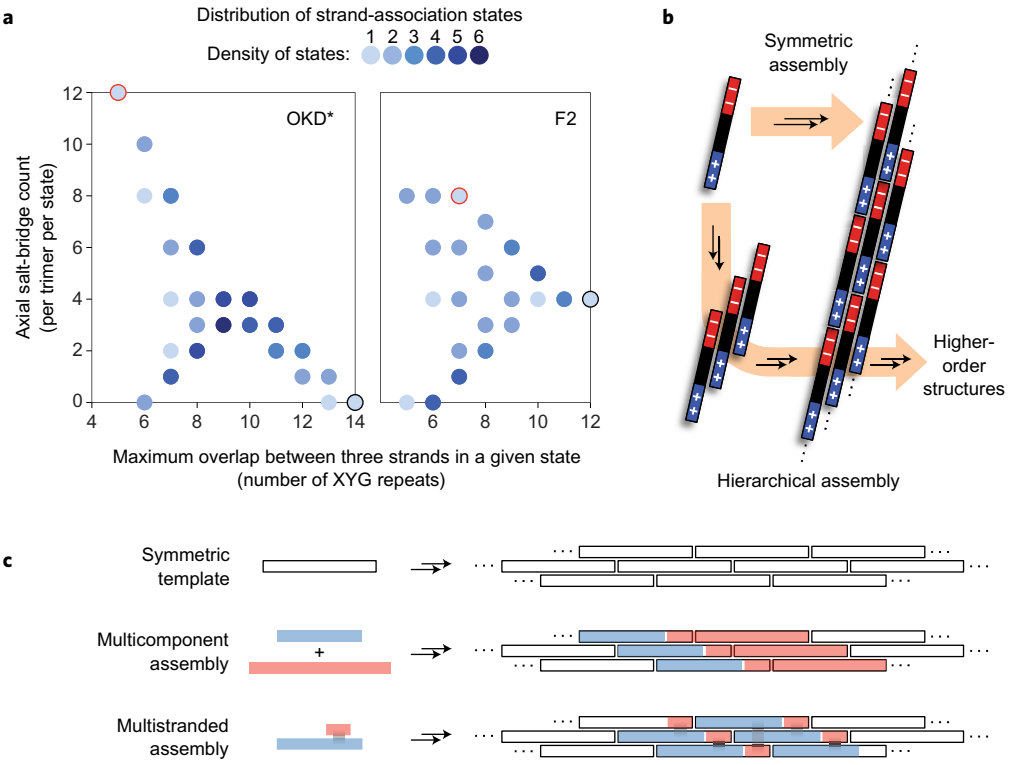


Figure 4 | Benefits of symmetric assembly. a, The distribution of strand-association states available to OKD and F2 peptides. States that lead to blunt ends (black outline) and to the dominant sticky-ended assemblies (red outline) are highlighted. *This distribution of states is shared with all 4sb-derived peptides. **b**, Comparison of symmetric versus hierarchical assembly. Symmetrically designed CMPs can interact directly with the growing triple helix, without a need to form sticky-ended homotrimeric intermediates *a priori*. **c**, Design of multicomponent fibres on a symmetric-assembly template. The strand-association pattern set by the symmetric assembly of a CMP (empty bars) can be used to generate complex building blocks (red and blue bars), each of which can achieve symmetric assembly.

with tapping-mode atomic force microscopy (AFM) appear to be 1.2 ± 0.2 nm thick and up to 0.5 μm long (Fig. 2d). The fibre diameter is maintained even when two fibres cross on the mica surface, as overlaid fibres are detected at twice the height of single fibres. Overall, data on OKD assemblies are indicative of nanofibres composed of a single triple helix. Hence, the surprising stability of OKD assemblies must result from interstrand rather than extra-triple helical interactions, which provides the first demonstration of the strength attainable from sticky-ended CMP assembly.

OKD, KDO and DOK are designed to self-assemble by engaging in identical interactions specified by the 4sb template and to generate nearly equivalent uninterrupted triple-helical segments (Supplementary Fig. 5). Yet, the positioning of $(PKG)_4$ and $(DOG)_4$ blocks on the peptides influences both assembly stability and nanostructure. The low stability of DOK assemblies could be related to a lack of charge contacts between neighbours that interact most closely in the DOK assemblies (Supplementary Fig. 6), which could destabilize sticky ends in particular. A more general explanation becomes apparent on consideration of the triple-helical stability near CMP termini in an assembly. POG repeats exhibit the highest triple-helical propensity, followed by DOG and then PKG repeats[33]. The thermostability of assemblies improves when sequences that better accommodate the collagen fold are placed near strand termini, which keeps those termini associated closely with the growing nanofibre. In contrast, termini that associate loosely with the growing assembly could compromise the structure (translated to larger diameters) and charge pairing (translated to increased bundling). As CMP termini are displayed at the edges of the assembly as sticky ends, improved triple-helical pre-organization at the termini could also enhance the assembly kinetics. Interestingly, these benefits require sequences of low triple-helical propensity to be central on the peptide. This behaviour is dichotomous to that of blunt-ended triple helices, in which destabilization to central residues is most damaging to the structure and stability[34].

Sequence control over self-assembly. Numerous sticky-ended association states can be generated for any CMP by varying the stagger between neighbouring strands. The design of a CMP that leads to a sticky-ended assembly requires careful positioning of the Lys and Asp residues on the sequence so that a minimal number of association states are stabilized maximally. Symmetric assembly enables perfect charge pairing for a unique state on the strand-association landscape, and thus readily yields an exceptional design (Fig. 3a).

Altering the placement of Lys and Asp residues can disrupt the ideal landscape for CMP assembly (Fig. 3a). By swapping two pairs of residues on the DOK sequence, we created DOKctrl, a CMP that maintains the sequence composition and charge distribution of DOK but has a featureless strand-association landscape. Interestingly, this minor sequence modification is expected to diminish only the prospect for sticky-end formation, but not blunt-ended association. The stark differences observed between DOK and DOKctrl structures provide experimental confirmation for the sticky-ended assembly of DOK. Whereas DOKctrl ($T_m = 27$ °C) forms a less-stable assembly than does DOK (38 °C) at low ionic strength ($I = 20$ mM), increasing the ionic strength to 200 mM with NaCl removes this difference ($T_m = 31$ °C for both DOK and DOKctrl) (Fig. 3b). Divergent responses to weakening Coulombic interactions suggest disparate modes of association. AUC provides further support for this idea by revealing two distinct association regimes for DOK (>47-mer) and DOKctrl (trimers) (Fig. 3c and Supplementary Fig. 3). In addition to providing evidence for

sticky-ended assembly, the design of nearly identical CMPs that are productive (DOK) and unproductive (DOKctrl) for assembly also demonstrates the tight control that sequence exerts over nanostructure at a level not seen before for collagen-based nanomaterials.

Discussion

There are 66 unique strand-association states available to 4sb-like peptides (Fig. 4a). Examination of these states reveals that diminishing the triple-helical overlap between three strands is always balanced with additional axial salt bridges as the trimer advances from a blunt-ended to a sticky-ended configuration. This balance is not accessible for F2, a similarly constructed CMP that does not support symmetric assembly[21]. The accepted association state for F2 (Fig. 1c) supports the longest uninterrupted triple helix among five top-performing states, yet its short overhangs bear no opportunity for additional axial contacts through further self-assembly. Therefore, assembly through this state requires the formation of sticky-ended homotrimers *a priori*, consistent with a hierarchical-assembly model[19,21] (Fig. 4b). In contrast, symmetric assemblies extend uniformly and free of 'weak' regions, which allows CMP building blocks to add directly to the fibre. Moreover, because the symmetric-association state is unique, its stabilization does not lead to unintended competing states with matching performances (Supplementary Fig. 7). Hence, the symmetric-assembly algorithm both simplifies the design process and increases its likelihood of success.

Symmetric assembly is possible for all CMPs of size $n_T = 3v \pm 1$ through a unique strand-association state, which can be stabilized through axial salt bridges or by other means. In this work, we applied these principles to a self-assembling single-peptide system. Symmetry would, however, benefit any system in which peptide building blocks appear on every strand of the assembled triple helix. Indeed, the patterns set by peptide tessellations provide a blueprint for the design and construction of currently inaccessible self-assembling systems that involve multiple CMPs. Co-assembling CMPs could be designed on a symmetric-assembly template so that, although peptides violate the template individually, the template would be reinstated on co-assembly (Fig. 4c). Such units could also be stapled together to generate complex building blocks[28]. The benefits of symmetry can transfer to other systems that are based on a symmetric template, just as shapes created by merging adjacent pairs of tessellating tiles with p1 symmetry (or dividing such tiles) also tessellate (Fig. 1d and Supplementary Fig. 2). Further, the design of multiple self-assembling peptides built on incompatible templates (for example, an *xy* and a *yx* template) could produce sets of CMPs that assemble independently, even when mixed. In theory, elements that allow the programmed association of triple helices[35,36] can also be encoded into CMPs. Our current work enables us to produce 'human-scale' triple helices. An expanded CMP toolkit would enable the control of size, strength and orientation of synthetic collagen fibrils that mimic connective tissues (such as skin and cartilage) in which multiple types of fibrillar collagens interact[37].

Conclusions

Despite its ancient origin, paramount importance in modern biology and repetitious architecture[12], the collagen triple helix has been a challenging target for controlled self-assembly. The symmetric assembly rules that we present arise from fundamental aspects of collagen structure and allow easy access to sticky-ended fibrillation. Nanofibres enabled by symmetric design owe their extraordinary stability solely to the sticky-ended association of peptides, and yield extended human-scale triple helices. Although the self-assembly 'alphabet' available for CMP-based nanostructures is crude in comparison with that of DNA/RNA, the ease of their interfacing with both biology[9,38] and nanotechnology[16,39] encourages their development.

Methods

Peptide synthesis. All CMPs were synthesized on polyethylene glycol-based resins on a Prelude peptide synthesizer from Protein Technologies using standard Fmoc-based methods at the Peptide Synthesis Facility of the University of Wisconsin–Madison (UW) Biotechnology Center (www.biotech.wisc.edu/services/peptidesynth). Condensation of Fmoc-ProHypGly-OH[14] or Fmoc-GlyProHyp-OH tripeptide segments from Bachem was employed wherever applicable. Fmoc removal was achieved in piperidine (20% v/v in DMF), and peptide building blocks (4 equiv.), activated through treatment with HATU (1-(bis(dimethylamino) methylene)-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate) and *N*-methylmorpholine, were coupled to the free amine of the growing chain for 60 minutes. Additions onto Pro or Hyp residues were performed through double couplings of 30 minutes. Peptides were cleaved from the resin and deprotected in 95:2.5:2.5 trifluoroacetic acid/triisopropylsilane/water (1.5–2.0 ml), precipitated from methyl *t*-butyl ether below 0 °C and isolated by centrifugation. The purification and characterization of the resulting peptides is described in the Supplementary Information.

Sample preparation. Peptide solutions were prepared in 10 mM sodium phosphate buffer, pH 7.0 ($I = 20$ mM), unless noted otherwise. Further introduction of NaCl to 180 mM yielded high ionic-strength samples ($I = 200$ mM) for CD analysis. To facilitate the formation of the thermodynamic assembly product, peptide solutions were denatured for 20 minutes at 65 °C, then annealed to 4 °C at a rate of −12 °C per hour. Samples were left at 4 °C for at least 48 hours before data acquisition.

CD spectroscopy. All CD data were acquired on 0.6 mg ml^{-1} (~160 μM) samples with a 202SF or 420 CD spectrophotometer from Aviv Biomedical at the UW Biophysics Instrumentation Facility (BIF; www.biochem.wisc.edu/bif). CD spectra were recorded at 4 °C with a 1 nm band-pass filter and an averaging time of 4 seconds in 0.1 cm path-length quartz cuvettes. For thermal denaturation experiments, the CD signal was monitored at 226 nm as the sample was heated at a rate of 12 °C per hour in 2 or 3 °C steps. The melting transition is indicated by a minimum on the temperature derivative of the melting curve ($\partial[\theta]/\partial T$), and values (±1 °C) of T_m were determined assuming a constant curvature in the immediate vicinity of the transition minimum.

AUC. Sedimentation equilibrium experiments were performed at the UW BIF with an XL-A analytical ultracentrifuge from Beckman Coulter equipped with an An-60 Ti rotor. Samples prepared at 0.6 mg ml^{-1} were diluted to 0.3 mg ml^{-1} prior to the experiments. The sample (100 μl) and matching buffer (110 μl) were placed in a cell with an Epon 12 mm double-sector charcoal-filled centrepiece from Beckman Coulter. Experiments were run at 4 °C for more than seven days at speeds of 8,800, 15,000, 25,000, 36,000 and 60,000 revolutions per minute (r.p.m.), and gradients recorded at 231 nm were monitored until superimposable four hours apart. Equilibrium gradients at 4 °C were modelled as single and multiple non-interacting species through nonlinear least-squares fits to the gradient data, using a buffer density of 1.0011 g ml^{-1} and a partial specific volume of 0.6955 ml g^{-1}, as calculated from the amino acid content[40]. Non-sedimenting baselines between 0.04 and 0.07 attenuance were applied for all the samples during the analysis, which was performed with programs written by D. R. McCaslin (UW BIF) for IGOR PRO software from WaveMetrics.

TEM. All the imaging was performed at the UW Materials Science Center (http://msc.engr.wisc.edu) with a Tecnai TF-30 TEM instrument (300 kV) from FEI. Filtered samples were allowed to adhere to a Quantifoil R1.2/1.3 holey-carbon mesh on copper grids at 4 °C for one minute. The excess was blotted with filter paper and the grid was allowed to dry for five minutes. A negative stain was applied for ten minutes by inverting the grid on a drop of a freshly prepared solution of phosphotungstic acid (2.0% w/w), adjusted to pH 6.0 with aqueous NaOH and filtered. The excess stain was removed with filter paper, and the grid was allowed to dry overnight prior to imaging. For vitreous-ice cryo-TEM imaging, a solution of OKD (0.3 mg ml^{-1}) was applied to a TEM grid, glow discharged for 30 seconds at 25 mA beforehand with a Pelco easiGlow unit from Ted Pella. The grid was allowed to equilibrate at 5 °C at 95% humidity, blotted and immersed in liquid ethane on a Vitrobot from FEI. The grid was transferred from liquid ethane to liquid nitrogen manually, and stored in liquid nitrogen until imaging. Acquisition was performed at −180 °C with a low dose and long exposures.

AFM. Samples (0.03 mg ml^{-1}) were applied onto freshly cleaved mica at 4 °C. After ten minutes, the peptide solution was washed with cold water for 30 seconds and excess water was removed with filter paper. The samples were dried overnight at 4 °C prior to imaging. AFM images were collected in tapping mode with an Agilent 5500 SPM instrument using NCHV-A probes from Bruker and processed using Gwyddion version 2.39 (http://gwyddion.net).

Strand-association landscapes. Association states for sticky-ended assemblies were enumerated through a set of purpose-written Python (v.2.7) scripts. The construction and properties of these landscapes are further discussed in the Supplementary Information.

Received 13 January 2016; accepted 23 May 2016;
published online 11 July 2016

References

1. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in-vitro*. *Proc. Natl Acad. Sci. USA* **70,** 3240–3244 (1973).
2. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
3. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nature Rev. Genet.* **11,** 367–379 (2010).
4. Seeman, N. C. Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79,** 65–87 (2010).
5. Zimenkov, Y. *et al.* Rational design of a reversible pH-responsive switch for peptide self-assembly. *J. Am. Chem. Soc.* **128,** 6770–6771 (2006).
6. Banwell, E. F. *et al.* Rational design and application of responsive α-helical peptide hydrogels. *Nature Mater.* **8,** 596–600 (2009).
7. Brinckmann, J. Collagens at a glance. *Top. Curr. Chem.* **247,** 1–6 (2005).
8. Meyers, M. A., Chen, P.-Y., Lin, A. Y.-M. & Seki, Y. Biological materials: structure and mechanical properties. *Prog. Mater. Sci.* **53,** 1–206 (2008).
9. Chattopadhyay, S. & Raines, R. T. Collagen-based biomaterials for wound healing. *Biopolymers* **101,** 821–833 (2014).
10. Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **3,** a004978 (2011).
11. Fields, G. B. Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **8,** 1237–1258 (2010).
12. Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **78,** 929–958 (2009).
13. Siebler, C., Erdmann, R. S. & Wennemers, H. From azidoproline to functionalizable collagen. *Chimia* **67,** 891–895 (2013).
14. Kotch, F. W. & Raines, R. T. Self-assembly of synthetic collagen triple helices. *Proc. Natl Acad. Sci. USA* **103,** 3028–3033 (2006).
15. Rele, S. *et al.* D-Periodic collagen-mimetic microfibers. *J. Am. Chem. Soc.* **129,** 14780–14787 (2007).
16. Gottlieb, D., Morin, S. A., Jin, S. & Raines, R. T. Self-assembled collagen-like peptide fibers as templates for metallic nanowires. *J. Mater. Chem.* **18,** 3865–3870 (2008).
17. Cejas, M. A. *et al.* Thrombogenic collagen-mimetic peptides: self-assembly of triple helix-based fibrils driven by hydrophobic interactions. *Proc. Natl Acad. Sci. USA* **105,** 8513–8518 (2008).
18. Yamazaki, C. M. *et al.* A collagen-mimetic triple helical supramolecule that evokes integrin-dependent cell responses. *Biomaterials* **31,** 1925–1934 (2010).
19. O'Leary, L. E. R., Fallas, J. A., Bakota, E. L., Kang, M. K. & Hartgerink, J. D. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nature Chem.* **3,** 821–828 (2011).
20. Xu, F. *et al.* Compositional control of higher order assembly using synthetic collagen peptides. *J. Am. Chem. Soc.* **134,** 47–50 (2012).
21. Sarkar, B., O'Leary, L. E. R. & Hartgerink, J. D. Self-assembly of fiber-forming collagen mimetic peptides controlled by triple-helical nucleation. *J. Am. Chem. Soc.* **136,** 14417–14424 (2014).
22. Persikov, A. V., Ramshaw, J. A. M. & Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280,** 19343–19349 (2005).
23. Fallas, J. A., Gauba, V. & Hartgerink, J. D. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J. Biol. Chem.* **284,** 26851–26859 (2009).
24. Coxeter, H. S. M. Crystal symmetry and its generalizations. *Trans. R. Soc. Can.* **51,** 1–13 (1957).
25. Emmer, M. & Schattschneider, D. *M. C. Escher's Legacy: A Centennial Celebration* (Springer, 2003).
26. Crick, F. H. C. & Watson, J. D. Structure of small viruses. *Nature* **177,** 473–475 (1956).
27. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl Acad. Sci. USA* **98,** 2217–2221 (2001).
28. Tanrikulu, I. C. & Raines, R. T. Optimal interstrand bridges for collagen-like biomaterials. *J. Am. Chem. Soc.* **136,** 13490–13493 (2014).
29. Gauba, V. & Hartgerink, J. D. Surprisingly high stability of collagen ABC heterotrimer: evaluation of side chain charge pairs. *J. Am. Chem. Soc.* **129,** 15034–15041 (2007).
30. Smulders, M. M. J. *et al.* How to distinguish isodesmic from cooperative supramolecular polymerisation. *Chem. Eur. J.* **16,** 362–367 (2010).
31. Gelman, R. A., Williams, B. R. & Piez, K. A. Collagen fibril formation: evidence for a multistep process. *J. Biol. Chem.* **254,** 180–186 (1979).
32. Bai, H. Y., Xu, K., Xu, Y. J. & Matsui, H. Fabrication of Au nanowires of uniform length and diameter using a monodisperse and rigid biomolecular template: collagen-like triple helix. *Angew. Chem. Int. Ed.* **46,** 3319–3322 (2007).
33. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39,** 14960–14967 (2000).
34. Chen, Y.-S., Chen, C.-C. & Horng, J.-C. Thermodynamic and kinetic consequences of substituting glycine at different positions in a Pro-Hyp-Gly repeat collagen model peptide. *Biopolymers* **96,** 60–68 (2011).
35. Jiang, T., Xu, C., Zuo, X. & Conticello, V. P. Structurally homogeneous nanosheets from self-assembly of a collagen-mimetic peptide. *Angew. Chem. Int. Ed.* **53,** 8367–8371 (2014).
36. McGuinness, K., Khan, I. J. & Nanda, V. Morphological diversity and polymorphism of self-assembling collagen peptides controlled by length of hydrophobic domains. *ACS Nano* **8,** 12514–12523 (2014).
37. Wess, T. J. Collagen fibril form and function. *Adv. Protein Chem.* **70,** 341–374 (2005).
38. Rad-Malekshahi, M., Lempsink, L., Amidi, M., Hennink, W. E. & Mastrobattista, E. Biomedical applications of self-assembling peptides. *Bioconjugate Chem.* **27,** 3–18 (2016).
39. Kaur, P. *et al.* Three-dimensional directed self-assembly of peptide nanowires into micrometer-sized crystalline cubes with nanoparticle joints. *Angew. Chem. Int. Ed.* **49,** 8375–8378 (2010).
40. Durchschlag, H. & Zipper, P. Calculation of the partial volume of organic compounds and polymers. *Ultracentrifugation* **94,** 20–39 (1994).

Acknowledgements

We thank A. J. Ellison, B. M. Hoover, R. Biswas and S. Chattopadhyay for help with solution- and solid-phase peptide synthesis, S. A. Morin for help with nanocharacterization, M. D. Shoulders, F. W. Kotch and E. Emrah for discussions on the CMP assembly and R. W. Newberry for reviewing the manuscript. This study would not have been possible without expert advice and assistance from M. D. Boersma and N. Porcaro (UW Biotechnology Center) and from D. R. McCaslin (UW BIF). A.F. and S.J. were supported by Grant DMR-0832760 (National Science Foundation). This work was supported by Grant R01 AR044276 (National Institutes of Health).

Author contributions

I.C.T and R.T.R. conceived the project and planned the experiments. I.C.T. designed, synthesized and characterized the peptides in solution, and computed their association landscapes. A.F. imaged the peptide assemblies. All the authors analysed the data. I.C.T. and R.T.R. wrote the paper. All the authors proofread, commented on and approved the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.T.R.

Competing financial interests

The authors declare no competing financial interests.