Peptide tessellation yields micron-scale collagen triple helices

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

[1]Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

[2]Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

Abstract

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. Like natural DNA, the $\sim10^3$-residue triple-helices ($\sim$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 micron 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 match or exceed those in natural collagen in length and are remarkably thermostable, despite the absence of higher-order association. 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 associate 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]. Likewise, the sticky-ended self-assembly of certain α-helical peptides can yield long, though thick, bundles[5,6].

Similar to that of DNA and α-helical bundles, the structure of collagen features intercoiled strands. Collagen is the predominant structural protein in animals[7,8]. Accordingly, 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 $\sim 10^3$ residues ($\sim$300 nm in length). Conversely, chemical synthesis can offer only collagen-mimetic peptides (CMPs) of $\sim$30 residues ($\sim$10 nm), which have little practical utility[11]. Endowing 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 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 favors 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, creating 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 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 fibers 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 those 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 occupying 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 blunt-ended association while strengthening sticky-ended assembly of a CMP (Fig. 1b). The resulting constructs vary greatly in their morphology, some forming higher-order structures, such as hydrogels[19] and birefringent rods[21], and others remaining as amorphous aggregates[21]. This heterogeneity is partially attributable to the association of unpaired Lys and Asp residues with other strands or triple helices (*e.g.*, Fig. 1c).

We reasoned that engaging *all* 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 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 neighbors, 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 neighbors 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 neighboring strands, restricts peptides that are amenable to symmetric assembly to those that are $n_T = 3\nu \pm 1$ XYG-repeats in length, where n_T is the number of XYG repeats on the CMP and ν 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 $n_T = 3\nu - 1$ and accepts new building blocks in clockwise order around the triple-helical axis during N- to C-terminal growth, while peptides of size $n_T = 3\nu + 1$ (the *xy*-class) assemble counter-clockwise. Although ν represents the number of in XYG-repeats that offset neighboring CMPs for both classes, the sign of the offset and ordering of strands differ. Thus, $n_T = 3\nu - 1$ assemblies favor the placement of PKG-repeats ahead of DOG-repeats on the peptide sequence, and *vice versa*. Among candidates that allow four fully paired Asp and Lys residues per CMP, we opted to study the 42-residue system (*yx*, $\nu = 5$) as it enables 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 neighbors. We name this infinite homotrimer, which bears 4 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 reestablish all contacts on the template upon assembly. We chose to synthesize and assess three 42-mers (DOK, KDO, and OKD; Fig. 1f), representing 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, allowing 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 for DOK, KDO, and OKD,

respectively. Thermal denaturation experiments revealed "melting" temperatures (T_m) above 37 °C for all assemblies. All 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 increasing size with increasing 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 negative and positively charged central sections form assemblies with higher T_m values (46 and 49 °C, respectively). The denaturation curves are increasingly cooperative for more stable peptides, displaying 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 36-mers reported previously (T_m = 15–25 °C; Supplementary Fig. 4) could be attributable to the $(POG)_2$ extension, this addition is expected to increase the T_m only by ~8 °C[22]. Thus, we believe that the additional interstrand salt bridges and uniform sticky-ends enabled by our design play a major role in improving assembly stability.

Another expected outcome of our strategy is the absence of higher-order association of nanofibers, which we confirmed by negative-stained transmission electron microscopy (TEM) imaging. The 4sb-like fibers are <10 nm in diameter, are oriented and distributed irregularly, and exhibit features that are as fine as 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 nanofibers (mean ± SD = 8 ± 2 nm). Comparison with previous TEM studies of single triple helices of natural collagen (~4 nm)[31,32] suggests that DOK fibers might be formed through the association of multiple triple helices. KDO and OKD nanofibers 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 nanofibers have a greater tendency for bundling, commonly remaining aligned for 30–50 nm before separating.

The structure and stability of OKD nanofibers are most impressive. Although individual OKD nanofibers are not isolable through dilutions, their bundling is rare, and triple-helical stretches can reach 0.3 μm in length. 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 nanofibers appear to be distributed randomly, and are intertwined extensively without forming bundles (Fig. 2c). These solvated nanofibers have a uniform diameter (3.8 ± 0.5 nm) that is slightly thicker than those recorded for dried samples, likely due to tightly associated water and buffer molecules. Individual OKD nanofibers observed 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). Fiber diameter is maintained even when two fibers cross on the mica surface, as overlaid fibers are detected at twice the height of single fibers. Overall, data on OKD assemblies are indicative of nanofibers composed of a single triple-helix. Hence, the surprising stability of OKD assemblies must be due to interstrand rather than extratriple helical interactions, providing 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 neighbors that interact most closely in DOK assemblies (Supplementary Fig. 6), which could destabilize sticky-ends in particular. A more general explanation becomes apparent upon consideration of 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, keeping those termini associated closely with the growing nanofiber. In contrast, termini that associate loosely with the growing assembly could compromise structure (translating to larger diameters) and charge-pairing (translating to increased bundling). As CMP termini are displayed at the edges of the assembly as sticky-ends, improved triple-helical preorganization at the termini could also enhance assembly kinetics. Interestingly, these benefits require sequences of low triple-helical propensity to be central on the peptide. This behavior is dichotomous to that of blunt-ended triple helices, in which destabilization to central residues is most damaging to 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 neighboring strands. The design of a CMP that leads to a sticky-ended assembly requires careful positioning of 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 yields an exceptional design readily (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 difference 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 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 never-before-seen 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 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, allowing CMP building blocks to add directly to the fiber. Moreover, because the symmetric-association state is unique, its stabilization does not lead to unintended competing states with matching performance (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 $n_T = 3\nu \pm 1$ through a unique strand-association state, which can be stabilized through axial salt bridges or other means. In this work, we have 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, though peptides violate the template individually, the template would be reinstated upon 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 (*e.g.*, 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. Nanofibers 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 DNA/RNA, the ease of their interfacing with both biology[9,38] and nanotechnology[16,39] encourage their development.

Methods

Peptide synthesis

All CMPs were synthesized on polyethylene glycol-based resins on a Prelude peptide synthesizer from Protein Technologies (Tucson, AZ) 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 (Bubendorf, Switzerland) 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 and NMM, were coupled to the free amine of the growing chain for 60 min. Addition onto Pro or Hyp residues were performed through 30-min double couplings. Peptides were cleaved from the resin and deprotected in 95:2.5:2.5 TFA/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 Text.

Sample Preparation

Peptide solutions were prepared in 10 mM sodium phosphate buffer, pH 7.0 (ionic strength, $I = 20$ 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 min at 65 °C, then annealed to 4 °C at a rate of -12 °C/h. Samples were left at 4 °C for at least 48 h before data acquisition.

Circular Dichroism Spectroscopy

All CD data were acquired on 0.6 mg/mL (~160 μM) samples with a 202SF or 420 CD spectrophotometer from Aviv Biomedical (Lakewood, NJ) 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 and an averaging time of 4 s, in 0.1-cm pathlength, quartz cuvettes. For thermal denaturation experiments, the CD signal was monitored at 226 nm while the sample was heated at a rate of 12 °C/h 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 ($\pm 1$ °C) of melting temperature (T_m) were determined assuming constant curvature in the immediate vicinity of the transition minimum.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed at the UW BIF with an XL-A analytical ultracentrifuge from Beckman Coulter (Brea, CA) equipped with an An-60 Ti rotor. Samples prepared at 0.6 mg/mL were diluted to 0.3 mg/mL prior to experiment.

Sample (100 μL) and matching buffer (110 μL) were placed in a cell with an Epon 12-mm double-sector charcoal-filled centerpiece from Beckman Coulter. Experiments were run at 4 °C for more than 7 days at speeds of 8.8, 15, 25, 36, and 60 k rpm, and gradients recorded at 231 nm were monitored until superimposable 4 h apart. Equilibrium gradients at 4 °C were modeled as single and multiple non-interacting species through nonlinear least-squares fits to gradient data, using a buffer density of 1.0011 g/mL and a partial specific volume of 0.6955 mL/g as calculated from the amino-acid content[40]. Non-sedimenting baselines between 0.04–0.07 OD were applied for all samples during analysis, which was performed with programs written by D. R. McCaslin (UW BIF) for IGOR PRO software from WaveMetrics (Lake Oswego, OR).

Transmission electron microscopy

All the imaging was performed at the UW Materials Science Center (msc.engr.wisc.edu) with a Tecnai TF-30 TEM instrument (300 kV) from FEI (Hillsboro, OR). Filtered samples were allowed to adhere to Quantifoil R1.2/1.3 holey-carbon mesh on copper grids at 4 °C for 1 min. Excess was blotted with filter paper and the grid was allowed to dry for 5 min. Negative stain was applied for 10 min 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. 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) was applied to a TEM grid, glow discharged for 0.5 min at 25 mA beforehand with a Pelco easiGlow unit from Ted Pella (Redding, CA). 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 was stored in liquid nitrogen until imaging. Acquisition was performed at −180 °C with low dose and long exposures.

Atomic force microscopy

Samples (0.03 mg/mL) were applied onto freshly cleaved mica at 4 °C. After 10 min, peptide solution was washed with cold water for 30 s, and excess water was removed with filter paper. 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 (Billerica, MA) and processed using Gwyddion ver. 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. Construction and properties of these landscapes are further discussed in the Supplementary Text.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Cohen SN, Chang ACY, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in-vitro. Proc Natl Acad Sci USA. 1973; 70:3240–3244. [PubMed: 4594039]

2. Gibson DG, et al. Enzymatic addembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6:343–345. [PubMed: 19363495]

3. Khalil AS, Collins JJ. Synthetic biology: Applications come of age. Nat Rev Genet. 2010; 11:367–379. [PubMed: 20395970]

4. Seeman NC. Nanomaterials based on DNA. Annu Rev Biochem. 2010; 79:65–87. [PubMed: 20222824]

5. Zimenkov Y, et al. Rational design of a reversible pH-responsive switch for peptide self-assembly. J Am Chem Soc. 2006; 128:6770–6771. [PubMed: 16719440]

6. Banwell EF, et al. Rational design and application of responsive α-helical peptide hydrogels. Nat Mat. 2009; 8:596–600.

7. Brinckmann J. Collagens at a glance. Top Curr Chem. 2005; 247:1–6.

8. Meyers MA, Chen PY, Lin AYM, Seki Y. Biological materials: Structure and mechanical properties. Prog Mat Sci. 2008; 53:1–206.

9. Chattopadhyay S, Raines RT. Collagen-based biomaterials for wound healing. Biopolymers. 2014; 101:821–833. [PubMed: 24633807]

10. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol. 2011; 3:a004978. [PubMed: 21421911]

11. Fields GB. Synthesis and biological applications of collagen-model triple-helical peptides. Org Biomol Chem. 2010; 8:1237–1258. [PubMed: 20204190]

12. Shoulders MD, Raines RT. Collagen structure and stability. Annu Rev Biochem. 2009; 78:929–958. [PubMed: 19344236]

13. Siebler C, Erdmann RS, Wennemers H. From azidoproline to functionalizable collagen. Chimia. 2013; 67:891–895. [PubMed: 24594333]

14. Kotch FW, Raines RT. Self-assembly of synthetic collagen triple helices. Proc Natl Acad Sci USA. 2006; 103:3028–3033. [PubMed: 16488977]

15. Rele S, et al. D-Periodic collagen-mimetic microfibers. J Am Chem Soc. 2007; 129:14780–14787. [PubMed: 17985903]

16. Gottlieb D, Morin SA, Jin S, Raines RT. Self-assembled collagen-like peptide fibers as templates for metallic nanowires. J Mat Chem. 2008; 18:3865–3870.

17. Cejas MA, et al. Thrombogenic collagen-mimetic peptides: Self-assembly of triple helix-based fibrils driven by hydrophobic interactions. Proc Natl Acad Sci USA. 2008; 105:8513–8518. [PubMed: 18559857]

18. Yamazaki CM, et al. A collagen-mimetic triple helical supramolecule that evokes integrin-dependent cell responses. Biomaterials. 2010; 31:1925–1934. [PubMed: 19853297]

19. O'Leary LER, Fallas JA, Bakota EL, Kang MK, Hartgerink JD. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. Nat Chem. 2011; 3:821–828. [PubMed: 21941256]

20. Xu F, et al. Compositional control of higher order assembly using synthetic collagen peptides. J Am Chem Soc. 2012; 134:47–50. [PubMed: 22171825]

21. Sarkar B, O'Leary LER, Hartgerink JD. Self-assembly of fiber-forming collagen mimetic peptides controlled by triple-helical nucleation. J Am Chem Soc. 2014; 136:14417–14424. [PubMed: 25494829]

22. Persikov AV, Ramshaw JAM, Brodsky B. Prediction of collagen stability from amino acid sequence. J Biol Chem. 2005; 280:19343–19349. [PubMed: 15753081]

23. Fallas JA, Gauba V, Hartgerink JD. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. J Biol Chem. 2009; 284:26851–26859. [PubMed: 19625247]

24. Coxeter HSM. Crystal symmetry and its generalizations. Trans Roy Soc Canada. 1957; 51:1–13.

25. Emmer, M.; Schattschneider, D., editors. MC Escher's Legacy: A Centennial Celebration. Springer; New York, NY: 2003.

26. Crick FHC, Watson JD. Structure of small viruses. Nature. 1956; 177:473–475. [PubMed: 13309339]

27. Padilla JE, Colovos C, Yeates TO. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci USA. 2001; 98:2217–2221. [PubMed: 11226219]

28. Tanrikulu IC, Raines RT. Optimal interstrand bridges for collagen-like biomaterials. J Am Chem Soc. 2014; 136:13490–13493. [PubMed: 25211141]

29. Gauba V, Hartgerink JD. Surprisingly high stability of collagen ABC heterotrimer: Evaluation of side chain charge pairs. J Am Chem Soc. 2007; 129:15034–15041. [PubMed: 17988128]

30. Smulders MMJ, et al. How to distinguish isodesmic from cooperative supramolecular polymerisation. Chem—Eur J. 2010; 16:362–367. [PubMed: 19921721]

31. Gelman RA, Williams BR, Piez KA. Collagen fibril formation: Evidence for a multistep process. J Biol Chem. 1979; 254:180–186. [PubMed: 758319]

32. Bai HY, Xu K, Xu YJ, 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. 2007; 46:3319–3322.

33. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. Biochemistry. 2000; 39:14960–14967. [PubMed: 11101312]

34. Chen Y-S, Chen CC, Horng JC. Thermodynamic and kinetic consequences of substituting glycine at different positions in a Pro-Hyp-Gly repeat collagen model peptide. Biopolymers. 2011; 96:60–68. [PubMed: 20560144]

35. Jiang T, Xu C, Zuo X, Conticello VP. Structurally homogeneous nanosheets from self-assembly of a collagen-mimetic peptide. Angew Chem Int Ed. 2014; 53:8367–8371.

36. McGuinness K, Khan IJ, Nanda V. Morphological diversity and polymorphism of self-assembling collagen peptides controlled by length of hydrophobic domains. ACS Nano. 2014; 8:12514–12523. [PubMed: 25390880]

37. Wess TJ. Collagen fibril form and function. Adv Protein Chem. 2005; 70:341–374. [PubMed: 15837520]

38. Rad-Malekshahi M, Lempsink L, Amidi M, Hennink WE, Mastrobattista E. Biomedical applications of self-assembling peptides. Bioconjugate Chem. 2016; 27:3–18.

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. 2010; 49:8375–8378.

40. Durchschlag H, Zipper P. Calculation of the partial volume of organic compounds and polymers. Ultracentrifugation. 1994; 94:20–39.
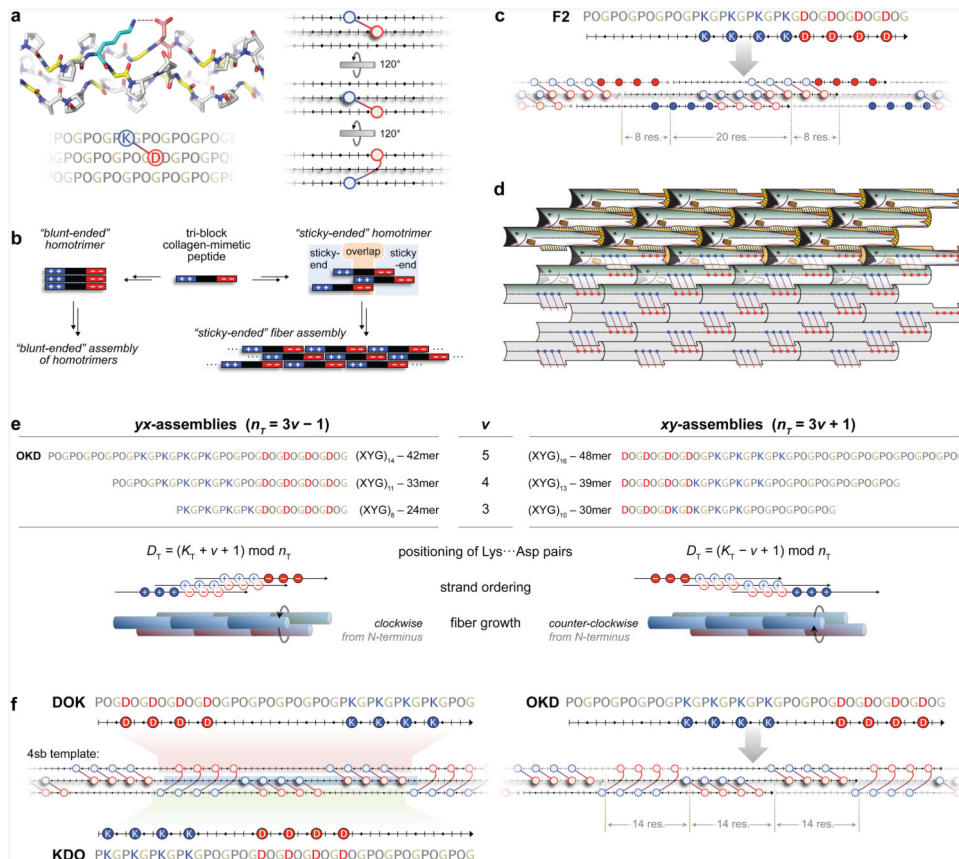
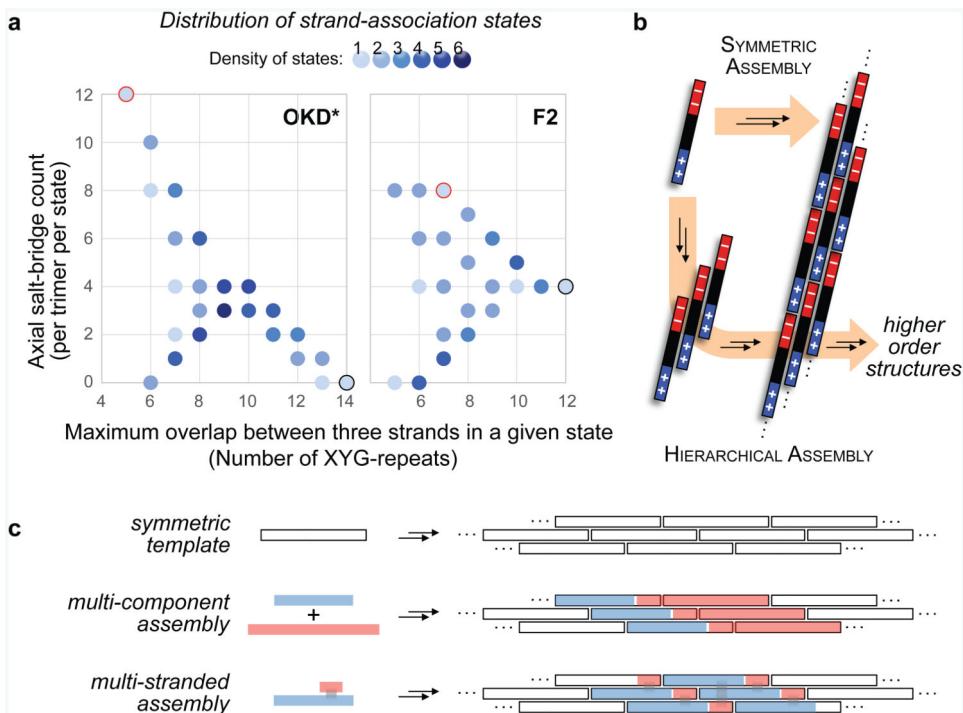**Figure 1. Design of symmetric CMP self-assemblies**

a, Molecular structure of the collagen triple helix and depictions of an axial salt bridge. The model, built based on the PDB entry 3u29 on PyMOL v1.3, 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 (gray 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 positions of XYG-units that host Lys and Asp residues of an interacting Lys···Asp pair to ensure 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.
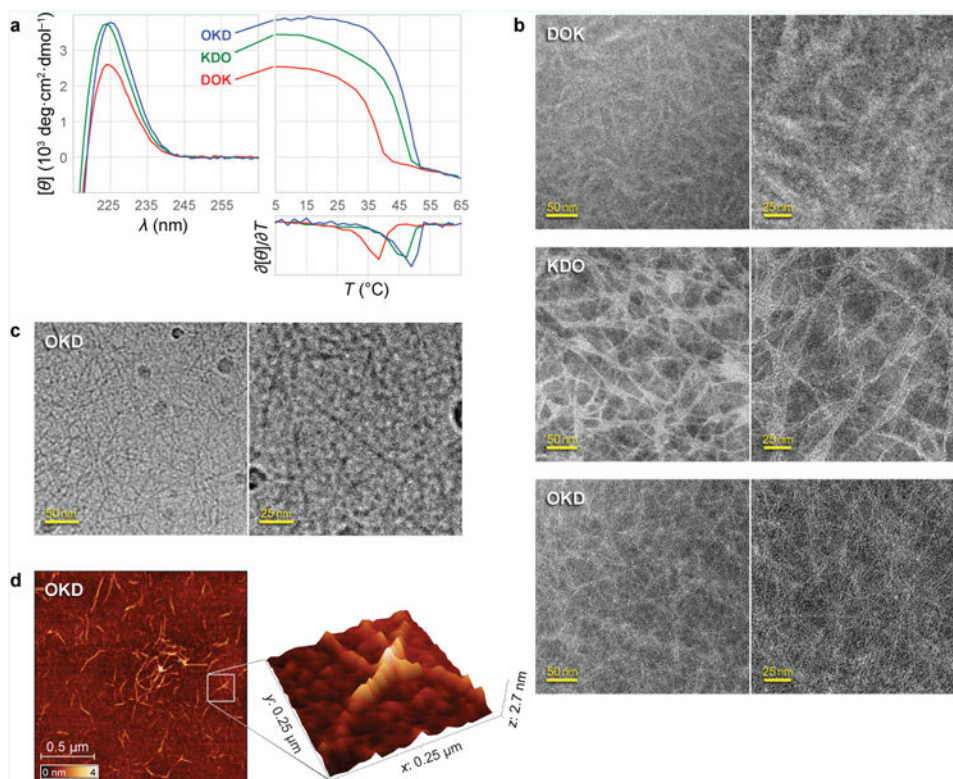
Figure 2. Characterization of 4sb-derived CMP self-assemblies

a, CD spectra and thermal denaturation curves at 0.6 mg/mL confirm triple-helical assemblies. All samples were prepared in 10 mM sodium phosphate buffer, pH 7.0, unless noted otherwise. **b**, TEM images of peptide assemblies reveal nanofibers for all permutations. **c**, Vitreous ice cryo-TEM images of OKD nanofibers (0.3 mg/mL). Triple-helices do not form bundles in solution. Circular spots are ethane artifacts introduced during flash-freezing. **d**, OKD nanofibers (0.03 mg/mL) imaged by AFM on freshly cleaved mica. Nanofibers retain their 1.2-nm diameter, even when overlaid.

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 lacking any defining features. Symmetric association states for DOK and DOKctrl are marked with black borders. *This landscape describes strand-association for not only DOK, but *all* 4sb-derived peptides. **b**, CD thermal denaturation curves for DOK and DOKctrl (0.6 mg/mL). Data acquired at low ionic strength (20 mM; solid line) and 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.

Figure 4. Benefits of symmetric assembly
a, The distribution of strand-association states available to OKD and F2 peptides. States leading 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 vs. 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 multi-component fibers 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.

Table 1

Assembly characteristics of 4sb-derived peptides

| Peptide | T_m (°C)[a] | | Average oligomerization by AUC[b] | TEM imaging of assemblies | |
	I = 20 mM	I = 200 mM		Nanostructure and Bundling	Feature dimensions (nm)
OKD	49	39	>188	single triple-helix (TH) nanofibers	2.3 ± 0.4
KDO	46	38	>61	single-TH nanofibers, intermittent bundling	3.8 ± 0.7
DOK	38	31	>47	multiple-TH nanofibers, no higher-level organization	7.9 ± 2.2
DOKctrl	27	31	3[c]	none	—

[a]Values of T_m were determined at both 20 and 200 mM ionic strength (I).

[b]Estimates for a lower-bound assembly size were obtained from a monomer + multimer model fit to AUC data at 15000 rpm, 4 °C.

[c]DOKctrl data does not agree with a monomer + multimer model, fitting best to a trimer + multimer model dominated by the trimeric species. See Supplemental Information and Supplemental Fig. 3 for discussion.