Supplementary Information

Peptide tessellation yields micron-scale collagen triple helices

I. Caglar Tanrikulu¹, Audrey Forticaux², Song Jin² & Ronald T. Raines¹,²

¹Department of Biochemistry and ²Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.

Experimental Procedures and Results

Peptide Purification and Characterization  page 2
Circular Dichroism Spectroscopy  2
Analytical Ultracentrifugation  2
Nanoimaging  3

Strand-Association Landscapes

Construction  4
Properties  4
F₂ and F₀ Association Landscapes  5
Heterotrimeric Infinite Strands  6

Supplementary Fig. S1  Effect of CMP size on symmetric assembly  7
Supplementary Fig. S2  Tessellation of peptides in two dimensions  8
Supplementary Fig. S3  AUC analysis of 4sb-like assemblies  9
Supplementary Fig. S4  Additional CD thermal denaturation experiments  9
Supplementary Fig. S5  Uninterrupted triple helices in 4sb-like assemblies  10
Supplementary Fig. S6  Salt-bridge propagation in 4sb-like assemblies  10
Supplementary Fig. S7  Alternative 4sb-like symmetric designs  10
Supplementary Fig. S8  HPLC chromatograms of purified CMPs  11
Supplementary Fig. S9  Additional TEM images of 4sb-like assemblies  12
Supplementary Fig. S10  Additional AFM images of OKD assemblies  16
Supplementary Fig. S11  F₂ and F₀ association landscapes  17
**Experimental Procedures and Results**

*Peptide Purification and Characterization*

Dried crude peptides were dissolved in 0.1% v/v TFA and filtered and purified by preparative HPLC at 45 °C using gradients of CH$_3$CN/water containing 0.1% v/v TFA on a Shimadzu Prominence instrument from Shimadzu (Kyoto, Japan) equipped with a VarioPrep 250/21 C18 column from Macherey–Nagel (Düren, Germany). All peptides were >90% pure according to analysis with analytical HPLC and MALDI–TOF mass spectrometry (MS). MALDI–TOF analysis was carried out with a Voyager DE-Pro mass spectrometer from Thermo Fischer Scientific (Waltham, MA) at the UW Biophysics Instrumentation Facility (www.biochem.wisc.edu/bif). For all 42-residue peptides, ($m/z$) [M + H]$^+$ calcd 3889.8, found 3889.5 for OKD; found 3890.0 for KDO; found 3889.3 for DOK; found 3890.1 for DOKctrl. For F0$^{19}$, calcd 3355.6, found 3355.4. Analytical HPLC chromatograms for purified peptides are displayed in Supplementary Fig. S8.

*Circular Dichroism Spectroscopy*

Thermal denaturation experiments were performed in both high ionic-strength (10 mM sodium phosphate buffer, pH 7.0, containing 180 mM NaCl; $I = 200$ mM) and low ionic-strength conditions (10 mM sodium phosphate buffer, pH 7.0; $I = 20$ mM). The results of these experiments are presented in Fig. 3b (DOK and DOKctrl) and Supplementary Fig. S4b (OKD and KDO). CMP assemblies that are not prone to homotrimer formation experience a 7–10 °C drop in $T_m$ upon an order-of-magnitude increase in ionic strength, consistent with electrostatic stabilization of assemblies.

Among self-assembling 36-mer CMPs designed by Hartgerink and coworkers, the F0 peptide$^{19}$ exhibits higher thermostability ($T_m = 40$ °C) than do others ($T_m = 15$–25 °C for A1, A2, F1 and F2)$^{21}$. To explore whether this surprising thermostability is due to sticky-ended triple-helix formation or to higher-order assembly, we monitored F0 denaturation at different concentrations by CD spectroscopy (Supplementary Fig. S4a). At low concentration, the F0 assemblies display a single melting-transition near 25 °C, while increasing F0 concentration promotes a second transition near 40 °C. Both F0 phases co-exist at 0.6 mg/mL, the room-temperature transition being prominent, while at 1.8 mg/mL the 40 °C transition dominates. We do not believe the two transitions correspond to different structures at the strand-association level, as the emergence of the 40 °C transition does not lead to a drastic change in CD signal. A more likely explanation is the concentration-dependent formation of higher-order F0 assemblies, which could boost the thermostability without inducing a significant change in triple-helical content. Thus, we believe a ~25 °C melting temperature better reflects the thermostability of sticky-ended F0 assemblies.

*Analytical Ultracentrifugation*

Assembly of OKD, KDO, and DOK into large structures was apparent in AUC experiments, as the shift of mass from meniscus to cell bottom was observed even at the start-up speed of 3 k rpm. Peptides in this work were not designed to assemble into structures of a specific size, and hence equilibrium gradients were expected to feature many different species with a wide range of molecular weights (MWs), requiring complex models featuring numerous interacting species for
an accurate description. The two-species models we employ consist of a low- and a high-MW species, which is a gross simplification of the self-assembly process. Nevertheless, these models describe the gradients at a reasonable level, especially when monomers are included as the low-MW species. The high-MW species represents the behavior of a complex mixture of peptide oligomer mixture, and is assigned a higher mass when steeper equilibrium gradients are modeled. This constitutes a lower-bound for the weight-average oligomer size, since the amount of depleted material at the cell bottom cannot be determined accurately. This value is highly sensitive to acquisition-speed and fit parameters, yet the trends revealed between different peptides are not sensitive to such factors. We list the oligomerization state of the high-MW species assigned at 15 k rpm to highlight differences in the extent of assembly between OKD, KDO and DOK (Table 1). Plots of gradients at 15 k rpm and their fits to multiple-species models are shown in Supplementary Fig. S3.

DOKctrl gradients are dominated by a trimer at low speeds, but the gradients cannot be explained by a trimer alone. Application of a two-species trimer + multimer model to low-speed data (8.8 and 15 k rpm) makes use of a 54-mer component to describe large masses at the cell bottom. In contrast to those of DOK and KDO, DOKctrl gradients harbor small amounts of this high-MW species (Supplementary Fig. S3). Correspondingly, DOKctrl exhibits minimal depletion when compared with peptides designed for self-assembly. Although the mechanism is not apparent, we suspect concentration-induced blunt-ended assembly (Fig. 1b) for the presence of such high-MW components. The absence of multiple DOKctrl phases in the thermal denaturation experiments, and their improved thermostability at high ionic strength supports this idea.

**Nanoimaging**

Additional TEM and cryo-TEM images collected on DOK, KDO, and OKD nanofibers at 3.5, 5.0, and 0.3 mg/mL peptide concentrations, respectively, are presented in Supplementary Fig. S9.

Additional AFM images collected on OKD nanofibers are provided in Supplementary Fig. S10. Although the softness of the material led to artifacts and necessitated frequent use of fresh probes, the images clearly and consistently present the self-assembled OKD nanofibers.
Strand-Association Landscapes

Construction

Association states for sticky ended assemblies were enumerated through a set of purpose-written scripts in Python (v.2.7) that construct all possible offsets between three infinite, periodic strands given a CMP sequence. An input sequence of the form \((XYG)_n\) is repeated without gaps to create a periodic, infinite strand, which is paired with two other identical strands to create an infinite homotrimer. The strands of the infinite homotrimer are shifted with respect to each other to produce all possible strand-association states. For each state, any instances of Lys···Asp pairs positioned to enable an axial contact are identified, and their counts are reported on the strand-association landscape (Fig. 3a).

An infinite homotrimer approximation overlooks the impact of peptide termini placement, but captures all side-chain–side-chain interactions for any CMP that follows the sticky-ended assembly route. In this way, the strand-association landscape provides a comprehensive overview of all sticky-ended associations, and allows identification of productive and/or competing states. The infinite-strand approximation also ensures that circular permutations to the CMP sequence does not alter the strand-association landscape. Thus, all peptides constructed based on the 4sb-template (Fig. 1f) share the same strand-association states and landscapes (Fig. 3a and 4a), irrespective of the 42-residue section under consideration.

The number of XYG-shifts that offset any two strands of a sticky-ended trimer is determined relative to a reference state with zero offset. In this work, this offset is defined to be zero when two strands have a single-residue stagger, and the \(i^{th}\) and \(i^{th} + 1\) strands are leading and lagging, respectively, ensuring that the coordinate \([0, 0]\) on the strand-association landscape refers to the “blunt-ended” association state.

Properties

Strand-association states (SASs) are uniquely defined by the number of XYG-offsets between any two pairs of strands. In this work, we have used offsets between the 1st and 2nd, and 2nd and 3rd strands for this purpose. The assignment of the 1st strand is arbitrary for an infinite homotrimer. Once assigned, the 1st strand is set as the “leading strand”, allowing the assignment of “middle” and “lagging” strands as 2nd and 3rd. For strands \(i\) and \(j\), \(\Delta(i-j)\) represents the XYG-offset between \(i^{th}\) and \(j^{th}\) strands.

The strand-association landscape presents the number of axial contacts for each SAS (Fig. 3a; Supplementary Fig. S11b), comprehensively describing all possible states for a given sequence cataloged by \(\Delta(1–2)\) and \(\Delta(2–3)\) offsets.

Since the assignment of the “1st strand” is arbitrary, most SASs are represented three times on this landscape, at coordinates \([\Delta(1–2), \Delta(2–3)], [\Delta(2–3), \Delta(3–1)],\ and [\Delta(3–1), \Delta(1–2)]\). The symmetric association state is the only exception to this rule. Symmetric association ensures that all strand offsets are identical (\(i.e., \Delta(1–2) = \Delta(2–3) = \Delta(3–1)\)), and is the only state to appear once on the strand-association landscape. Consequently, the number of non-redundant strand-association states \(n_{SAS}\) for a single-CMP assembly system is
\[ n_{\text{SAS}} = \left\lceil \frac{n_T^2}{3} \right\rceil \]

where \( n_T \) is the number of XYG-trimers in the CMP sequence, and the ceiling function (\( \lceil \rceil \)) accounts for the symmetric association state when present.

By definition, states that neighbor each other vertically or horizontally on the strand-association landscape are separated by a single-XYG shift between their strands. Horizontal movements are equivalent to moving the 1\(^{\text{st}}\) strand with respect to the 2\(^{\text{nd}}\), and vertical movements are equivalent to moving the 3\(^{\text{rd}}\) strand with respect to the 2\(^{\text{nd}}\). In addition, movements in the direction of the bottom-left–top-right diagonal are equivalent to shifting the 2\(^{\text{nd}}\) strand with respect to the 1\(^{\text{st}}\) and 3\(^{\text{rd}}\), and also constitute single-XYG shifts. An example of this shifting can be seen on the “DOK” panel in Fig. 3a, where a single-XYG shift away from the symmetric association state by any strand reduces the number of possible axial interactions from 12 to 10.

**F2 and F0 Association Landscapes**

Small changes in the distribution of PKG and DOG units on the peptide sequence can have dramatic effects on the strand-association landscape (Fig. 3a). Interestingly, despite different ordering of their charged blocks, self-assembling peptides F2\(^{21}\) and F0\(^{19}\) (Supplementary Fig. S11a) generate similar strand-association landscapes (Supplementary Fig. S11b). Both peptides produce 5 top-performing SASs producing the maximum, 8 axial salt bridges per sticky-ended homotrimer. In both cases, all 5 top-performing states are grouped together on the landscape, and can access each other through only one or two moves, where a move is a single XYG-shift of a strand.

Yet, the two peptides reveal very different patterns when axial interactions and strand overlap are examined together (Supplementary Fig. S11c). The F2 peptide features a blunt-ended state stabilized by 4 axial interactions, and favors short overhangs for sticky-ended assembly. In contrast, the F0 peptide produces a pattern similar to that of OKD (Fig 4a). F0 does not allow axial interactions in the blunt-ended state, but enables increasingly more axial interactions for SASs with smaller strand overlap, and thereby supports longer overhangs.

Among peptides of similar architecture\(^{21}\), F0 appears better suited for sticky-ended self-assembly, and does indeed produce the most striking triple-helical nanofibers and hydrogels\(^{19}\). Although F0 shares its strand-association landscape with its circular permutations, A1 and A2, triple-helical structures formed by these permutations (\( T_m = -15 \, ^{\circ}\text{C} \)) were not amendable to visualization by TEM. This result contradicts our expectations. F0 and A1/A2 could, however, use different “top-performing” SASs on their shared landscape, producing assemblies with different structures, altering their potential for interhelical association, leading to the different observed morphologies. In contrast, F2 and its circular permutation, F1, assemble into fibers as thick as 0.2 \( \mu \text{m} \) (\( T_m = -20–25 \, ^{\circ}\text{C} \)) despite being supported with only short overhangs. It is possible that F2 assembles in a way that resembles blunt-ended assembly, stabilized by interhelical in addition to interstrand interactions during growth. Further work on this interesting system could elucidate the effect of sticky ends on self-assembly and hydrogel formation, both critical to the development of functional CMP-based materials.
Heterotrimeric Infinite Strands

Strand-association landscapes can also be constructed for sticky-ended assemblies wherein peptides of different composition occupy different strands. Such cases require the construction of AAB- or ABC-type infinite heterotrimers. With distinct sequences on each strand, explicit assignment of strand numbers becomes necessary. As a result, the coordinates $[\Delta(1–2), \Delta(2–3)]$, $[\Delta(2–3), \Delta(3–1)]$, and $[\Delta(3–1), \Delta(1–2)]$, redundant for an infinite homotrimer, point to distinct association states in a heterotrimeric system.

For ABC-type heterotrimers, two different strand orders (ABC and ACB) are possible. For such systems, different strand-association landscapes are necessary for each strand ordering, as a single coordinate, $[\Delta(1–2), \Delta(2–3)]$, cannot represent both $[\Delta(A–B), \Delta(B–C)]$ and $[\Delta(A–C), \Delta(C–B)]$ on a single landscape. Hence, a single homotrimeric SAS with redundancies evolves into 6 distinct states on an infinite ABC heterotrimer, analogous to 6 registers being available to a blunt-ended heterotrimer.

A set of strand offsets (i.e., $\Delta(i–j)$ for all strand pairs) for a sticky-ended heterotrimer with known strand ordering produces three distinct SASs for both AAB- and ABC-type heterotrimers. This relationship is true for all SASs but the symmetric association state, where this information uniquely specifies a single SAS. This property could provide an additional advantage for symmetrically designed CMPs, as the field moves on to more complex, multi-component or multi-stranded assembly systems.
Supplementary Fig. S1 | Effect of CMP size on symmetric assembly. Collagen structure requires a single-residue stagger between XYG-units of all strands of a triple helix (top). This requirement prevents (XYG)$_{12}$ strands from associating symmetrically, while such assemblies are accessible through (XYG)$_{11}$ and (XYG)$_{13}$ strands (bottom).
Supplementary Fig. S2 | Tessellation of peptides in two dimensions. a, The symmetric design of OKD self-assemblies ensures identical interactions for every peptide and strand in the assembly. The resulting network of inter- and intrastrand interactions encompasses the triple helix completely, yielding a tessellation in two dimensions. This tessellation can be visualized through “peptide tiles” that encode the sequence and interactions of a peptide. Upon OKD self-assembly, these tiles tessellate around the central axis of the triple helix, satisfying all possible inter-peptide contacts. “Unrolling” this cylinder to cover a plane presents a tessellation of peptide tiles belonging to the plane symmetry group, p1. b, The F2 peptide cannot produce symmetric assemblies, and thus cannot form such tessellations.
Supplementary Fig. S3 | AUC analysis of 4sb-like assemblies. Equilibrium gradients at 15 k rpm and 4 °C are shown (gray circles) with models that best explain the data. Here, the slope at any radial position is proportional to the weight-averaged molecular weight at that position in the centrifuge cell. A two-species model (red line) that includes a monomer and a multimer component (black lines) can describe patterns observed for OKD, KDO and DOK assemblies. Nonetheless, this model performs poorly for DOKctrl, necessitating a trimer + multimer model. Model (red) and components (black) shown are the 15 k rpm results of a fit to data at 8.8 and 15 k rpm.

Supplementary Fig. S4 | Additional CD thermal denaturation experiments. a, CD spectra and temperature melts for F0 assemblies in 10 mM sodium phosphate buffer, pH 7.0, reveal concentration-dependent formation of presumed higher-order assemblies. b, Temperature melts for OKD and KDO assemblies (0.6 mg/mL) in high (dashed) and low ionic-strength sodium phosphate buffer (solid lines) at pH 7.0.
Supplementary Fig. S5 | Uninterrupted triple helices in 4sb-like assemblies. OKD, DOK and KDO assemblies form near-identical uninterrupted triple-helical sections (highlighted in gray). All termini appear within a single XYG-unit on the 4sb-template.

Supplementary Fig. S6 | Salt-bridge propagation in 4sb-like assemblies. When assembled, OKD and KDO peptides form salt bridges with neighbors that they overlap and interact with most closely, while DOK does not.

Supplementary Fig. S7 | Alternative 4sb-like symmetric designs. a, Examples of symmetrically designed peptides that allow perfect charge-pairing. Among 73 sequences that allow symmetric assembly for 4sb-like systems ($n_T = 14$; $yx$-class), all non-repetitive cases produce unique perfectly-paired states. b, Repeating sequences that support multiple symmetries are the only exceptions. This example is composed of two repeats of a shorter ($n_T = 7$; $xy$-class) "symmetric" peptide, which also satisfies $n_T = 14$ symmetry. In this way, perfectly-paired assemblies can be accessed through two distinct symmetric association states. This example is one of 3 from a possible 73.
Supplementary Fig. S8 | HPLC chromatograms of purified CMPs. Reverse-phase analytical HPLC chromatograms of peptides synthesized and purified for this study.
Supplementary Fig. S9 | Additional TEM images of 4sb-like assemblies. TEM images of dried DOK, KDO and OKD samples, and cryo-TEM images of OKD assemblies are provided to supplement Fig. 2 of the main text.
Supplementary Fig. S9 (continued) | Additional TEM images of 4sb-like assemblies.
Supplementary Fig. S9 (continued) | Additional TEM images of 4sb-like assemblies.
Supplementary Fig. S9 (continued) | Additional TEM images of 4sb-like assemblies.
**Supplementary Fig. S10 | Additional AFM images of OKD assemblies.** Probe amplitude data (right) is shown to facilitate identification of features in the height images (left).
**Supplementary Fig. S11 | Comparison of F2 and F0 association landscapes.**

**a.** Amino acid sequences for self-assembling CMPs, F2^{21} and F0^{19}. **b.** Strand-association landscapes for F2 and F0. The states accepted as dominant are indicated on the landscape with a black border. **c.** Distribution of strand-association states with respect to number of axial interactions and strand overlap for F2 and F0. The blunt-ended homotrimer (black outline) and the dominant SAS (red outline) are presented in comparison to all other states.