Article

Exploration of the hierarchical assembly space of collagen-like peptides beyond the triple helix

Le Tracy Yu [1], Mark A. B. Kreutzberger[2], Thi H. Bui[1], Maria C. Hancu[1], Adam C. Farsheed [3], Edward H. Egelman [2] & Jeffrey D. Hartgerink [1,3] ✉

The de novo design of self-assembling peptides has garnered significant attention in scientific research. While alpha-helical assemblies have been extensively studied, exploration of polyproline type II helices, such as those found in collagen, remains relatively limited. In this study, we focus on understanding the sequence-structure relationship in hierarchical assemblies of collagen-like peptides, using defense collagen Surfactant Protein A as a model. By dissecting the sequence derived from Surfactant Protein A and synthesizing short collagen-like peptides, we successfully construct a discrete bundle of hollow triple helices. Amino acid substitution studies pinpoint hydrophobic and charged residues that are critical for oligomer formation. These insights guide the de novo design of collagen-like peptides, resulting in the formation of diverse quaternary structures, including discrete and heterogenous bundled oligomers, two-dimensional nanosheets, and pH-responsive nanoribbons. Our study represents a significant advancement in the understanding and harnessing of collagen higher-order assemblies beyond the triple helix.

The de novo design of self-assembling peptide systems is of great interest to scientists. Among the various assemblies described to date, structures assembled from α-helices are the most extensively studied resulting in precise design capabilities through rational and computational approaches[1–13]. In addition to α-helices, researchers have explored the self-assembly of β sheets[14–17], hybrid assemblies[18,19], and the 3₁₀ helix[20]. The Polyproline type II (PPII) helix is a protein secondary structure frequently found in collagen, the most abundant extracellular protein[21,22]. However, compared to other secondary structures, PPII helices have been underexplored. In collagen, PPII helices are characterized by the Xaa-Yaa-Gly sequence repeat, with three PPII helices associating into a collagen triple helix[23]. In this super helix, the three peptide strands are offset from one another by $3n+1$ amino acids, where n is an integer. This offset defines a leading, middle and trailing peptide strand and is termed the helix's "register"[24]. Several structural studies of collagen triple helices have been conducted to

develop synthetic collagen-mimetic materials with the polymeric higher-order assemblies being the most well-developed as reported by the Conticello group[25–29], the Raines group[30], the Xu group[31] and our group[32]. However, achieving precise control over the self-assembled state to the degree which has been achieved in α-helical coiled coils remains elusive due to limited understanding of the sequence-structure relationship in collagen higher-order assembly. For instance, computational algorithms like Alphafold2 and Rosetta, while highly effective in predicting α-helical assemblies, encounter difficulties in accurately predicting collagen oligomer formation[6,33–35].

Previously we introduced an approach to creating a hollow bundled collagen oligomer assembly using peptides derived from the immune protein component C1q[36]. This family of proteins, which also includes SP-A, ficolins, and mannose binding lectins, is characterized by a bouquet-like structure, with the stem region consisting of bundled collagen triple helices, a branching region and a globular region[37].

[1]Department of Chemistry, Rice University, Houston, TX, USA. [2]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA. [3]Department of Bioengineering, Rice University, Houston, TX, USA. ✉e-mail: jdh@rice.edu

Specifically, the C1q stem forms a hexameric assembly of triple helices (octadecamer), with each subunit being a heterotrimer containing three different peptides: A, B, and C. Through the synthesis and characterization of short peptides derived from the stem region of C1q we explored the self-assembly mechanism of this region. We found that the formation of the hollow (ABC)$_6$ octadecameric bundle is independent of the C-terminus collagenous and globular fragments as well as disulfide bonding near the N-terminus. However, a short N-terminal non-collagenous domain is necessary for effective assembly of the (ABC)$_6$ complex[36]. This work demonstrated the structure of a discrete (non-polymeric, monodisperse) oligomeric collagen triple helical assembly, providing a valuable model for future investigation of inter-triple helical interactions. While some studies have suggested particular registrations of C1q's triple helical domains[38,39], challenges associated with the characterization and ambiguities introduced by discontinuities in the Gly-Xaa-Yaa repeat leave the register of the three peptides uncertain. This poses a challenge to the investigation of amino acid sequences critical for the formation of this quaternary structure. In this work, we turn our attention to another protein within the C1q family, Pulmonary Surfactant Protein A (SP-A). In contrast to C1q, SP-A features a homotrimeric collagen-like stem region, consisting of only one type of peptide. This characteristic dramatically simplifies the study of the sequence-structure relationship in hierarchical assemblies of collagen-like peptides, facilitating structural studies.

We dissected the sequence derived from the stem region of SP-A protein and synthesized these relatively short, collagen-like peptides. Upon self-assembly, we have successfully constructed a hollow octadecameric bundle similar to that observed for C1q. Subsequently, we conducted amino acid substitution studies to locate residues critical for oligomer formation. Amino acid sequences including hydrophobic Isoleucine residues and the charged Arginine and Aspartic acids, were found to favor the oligomer formation. Furthermore, we identified the N-terminal noncollagenous domain as critical for protein assembly, as peptide mutants lacking this domain failed to form discrete collagen oligomers. The sequence-structure relationship information was applied to guide the de novo design of collagen-like peptides to form bundled structures of triple helices. We used a (Gly-Pro-Hyp)$_{10}$ peptide as a template starting point in the design, which has been known to create a stable collagen triple helix, but without higher-order oligomerization[40–42]. Amino acid residues that facilitate interhelical charge-pair interactions and hydrophobic packing were introduced into the designed peptides. We have successfully obtained hierarchical assemblies of four different architectures including discrete oligomeric bundles, heterogeneous oligomeric bundles, polymeric ribbons, and 2D nanosheets, showcasing the potential for forming diverse quaternary structures with collagen-like peptides (Fig. 1). This study demonstrates peptide sequences conforming to a collagen-like sequence provide a rich space for hierarchical assembly, providing new approaches for future synthetic design as well as insight into natural collagen hierarchical assembly.

Collagen-like homotrimer and octadecamer formation

We started with amino acid sequence analysis of protein SP-A. We obtained the full-length amino acid sequence of natural SP-A and highlighted the collagen-like region by identifying the glycine residue in every third amino acid in the sequence, and recognized the discontinuity of the (Gly-Xaa-Yaa) sequence pattern starting from Gly45 (numbering starts after the signal sequence)[43]. The N-terminal 1-44 amino acid sequences ahead of the discontinuity are therefore assumed to be the stem collagen-like region (see Fig. S1 for sequence
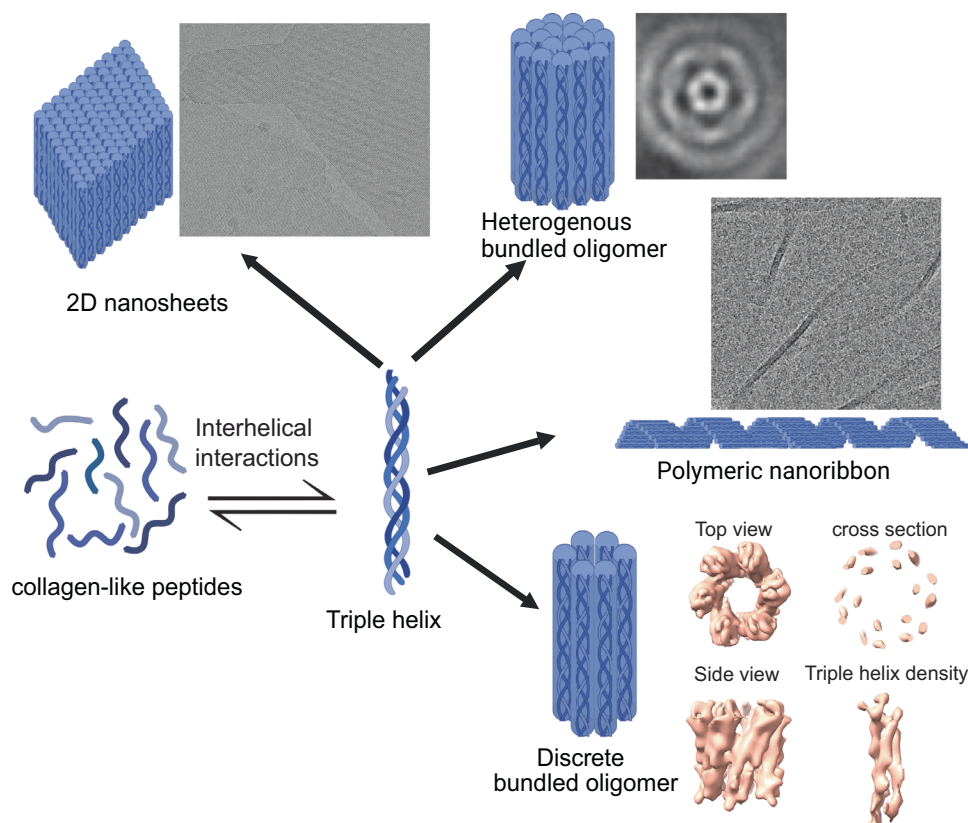

Fig. 1 | Schematic illustration of the diverse hierarchical assemblies formed in this study using collagen-like peptides. Peptides self-assemble to form triple helices. These triple helices can then form a variety of higher-order assemblies including 2D nanosheets, heterogenous bundled oligomers, polymeric nanoribbons as well as discrete bundled oligomers similar to the organization of defense collagens such as SP-A and C1q. Created in BioRender. Yu, T. (2024) BioRender.com/b15h427.

dissection details)[43]. Similar to C1q, disulfide-bonded dimer formation has previously been observed at the Cys6 position in protein SP-A[44]. Our previous study of C1q protein suggests that the higher-order assembly is independent of the disulfide bond. We, therefore, designed peptides Cys6Ala mutants of protein SP-A for self-assembly study.

We conducted an analysis of the amino acid sequences of SP-A protein from various organisms, revealing differences in the amino acid sequences of the N-terminal noncollagenous domain. To

investigate the structural influence of the noncollagenous region, we made three collagen-like peptides for self-assembly studies. The first peptide, SPA-bcb, contains the collagen-like region exclusively (Fig. 2a). The other two peptides, SPA-Human and SPA-Rat are derived from the stem region of the human SP-A α2 chain and the rat SP-A protein, respectively (Fig. 2a). All these peptides were synthesized using solid-phase peptide synthesis, purified via reverse-phase HPLC, and characterized using mass spectrometry (refer to Supplementary Information Table S1 and Figs. S2–S18 for details).
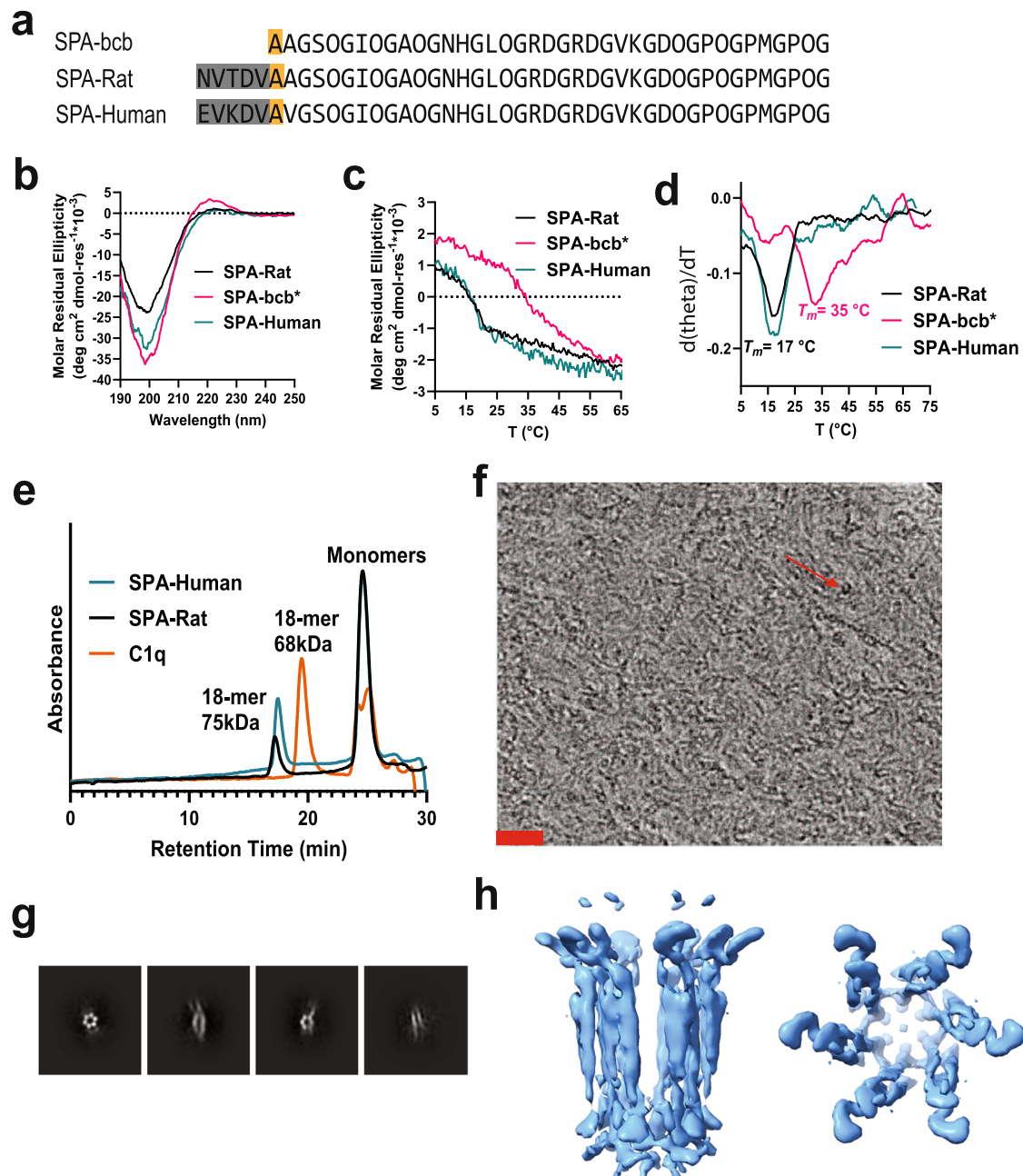

Fig. 2 | Characterization of SP-A derived peptide assemblies. a Amino acid sequences of SP-A derived collagen-like peptides SPA-bcb, SPA-Rat, and SPA-Human. Grey highlighted the non-collagen-like domain, and yellow highlighted the alanine residue, which is cysteine in the native pulmonary surfactant protein-A. Yellow highlights Cys6 to Ala substitution. Numbering starts after the signal sequence. **b** CD spectra of the folded triple helices. Black: SPA-Rat, green: SPA-human, red: SPA-bcb. **c** Thermal stability study of the assemblies. Signals were monitored at 224 nm. *SPA-bcb peptides formed a heterogeneous assembly that

contains precipitate. **d** The first-order derivatives of the melting curves presented in (**c**). See "Methods" for CD conditions. **e** Size exclusion chromatography of SPA-Rat (black), SPA-Human(green), and C1q 18-mer(orange). Source data for CD and SEC are provided as a Source Data file. **f** Representative cryo-EM image of SPA-Human peptide assembly. The scale bar is 20 nm. The red arrow highlights the hollow protein nanoparticles. **g** 2D class averaging processing of the cryo-EM images of SPA-Human peptide assembly. **h** Low-resolution density map of SPA-Human peptide oligomeric assembly.

After folding under buffered conditions, the peptide assemblies were characterized for their secondary structure using circular dichroism (CD) and oligomeric state using size exclusion chromatography (SEC) and cryo-EM (Fig. 2). Peptides SPA-bcb, SPA-Rat, and SPA-Human all present characteristic features of collagen triple helices including a positive signal at approximately 224 nm (Fig. 2b) and a cooperative thermal transition. The thermal stability of the triple helices assembled by peptide SPA-Rat and SPA-Human are comparable, having a melting temperature (T_m) of 17 °C, which indicates that the sequence difference near the N-terminus of these two proteins did not alter the triple helix folding stability (Fig. 2c, d). The SPA-bcb peptide precipitated upon folding and remained insoluble in different solvent systems such as aqueous buffers, isopropanol, or 2,2,2-tri-fluoroethanol. Such turbidity impacted CD measurements and resulted in an increased signal-to-noise level. According to the CD melting experiments, the peptide aggregation derived from SPA-bcb has a T_m of 35 °C. The sample solution became clear above this indicated melting temperature. We hypothesize that the aggregation may enhance the thermal stability of individual triple helices. While we suspect that this precipitation may consist of polydisperse triple helical aggregates, Cryo-EM characterization of the SPA-bcb sample only showed amorphous features (Fig. S19).

Size exclusion chromatography (SEC) was used to characterize the oligomeric state of the soluble peptide assembles of SPA-Rat and SPA-Human. The octadecameric assembly previously derived from C1q peptides was used as a point of comparison[36] in addition to a collagen triple helix standard[45] and globular protein standards (Fig. S21). SEC of both human and rat SP-A peptides revealed peaks eluting at 25 min corresponding to monomeric peptides and at 17 min corresponding to an octadecameric assembly with a molecular weight near 75 kDa (Fig. 2e). The SEC results confirmed that SPA-Rat and SPA-Human peptides associate into homogenous collagen-like oligomers. The equilibrium of the 18-mer formation is favored in the SPA-Human compared to sample SPA-Rat, indicating that the difference in the N-terminal amino acid sequence of these two peptides affects oligomer formation (Fig. 2a). These results demonstrated that the oligomer formation in SP-A is independent of the C-terminal region of the protein and the N-terminal disulfide bonds. The presence of the short noncollagenous domain, however, prevents protein aggregation and favors the discrete oligomer assembly. These are similar to results we had observed for C1q[36].

We further characterized the octadecameric peptide assembly of SPA-Rat (Fig. S20) and SPA-Human using cryo-EM (Figs. 2f and S22). The raw cryo-EM images of SPA-Human peptide assembly show hollow protein particles of ~ 6 nm in diameter (Fig. 2f) as well as more filamentous assemblies (Sup. Fig. S22). The 2D class averages show side views and top views of the hollow oligomeric peptide assemblies (Fig. 2g). Similar cryo-EM results were observed for SPA-Rat (Fig. S20). The size of the particles in the SP-A human sample appeared rather heterogenous, especially when looking at the side views. This could be because some of the assemblies' side views were actually filaments containing multiple particles. The overall heterogeneity complicated three-dimensional structure determination. Despite this, a low-resolution electron density map was obtained after 3D reconstruction with C6 symmetry imposed (Fig. 2h). We observed features similar to the previously published structure of the C1q peptide assembly. Given the similarities to the C1q structure and the low resolution (~6 Å) triple helical features, we believe that this structure is representative of a portion of the SP-A peptide assembly.

Amino acid substitution studies reveal critical sequence-structure relationship

The supramolecular octadecameric assembly from the SP-A stem region provided a model for studying the interhelical interactions that drive the collagen triple helices to oligomerize. We next conducted amino acid substitution studies to investigate the residues critical for higher-order assembly. We focused our preliminary screening on the charged and hydrophobic amino acid residues. To investigate the effects of charge-pair interactions on the oligomer assembly, we made peptide mutants of Arg24 in the Xaa position to be proline and Asp25 in the Yaa position to be hydroxyproline residue since the (Gly-Pro-Hyp)$_n$ sequence is regarded as the canonical sequence for collagen assembly (Fig. 3a). The samples were prepared for self-assembly and then subject to CD and SEC characterizations. Figures 3b, c and S23a show the CD characterization of the SP-A mutant peptides. A homo-trimeric helix formed with a T_m of 19 °C for sample SPA-D-O but no triple helix formation for sample SPA-R-P was observed. Despite the fact that proline is known to be more stabilizing for triple helices than arginine[46,47], the Arg24 to proline substitution eliminated self-assembly, indicating that Arg24 is critical in some way for the folding of the octadecameric assembly. SEC results of SPA-R-P peptide only show peaks of monomer species while SPA-D-O peptide assembly suggests formation of trimer and oligomer (Fig. 3d). The oligomeric peak in sample SPA-D-O assembly is broad and shifted to a lower mass region compared to the 18-mer peak, indicating that the assembly decreased in size. This suggests that aspartate also plays an important role in helix oligomerization, possibly through inter-triple helical interactions.

We additionally investigated hydrophobic residues since hydrophobic packing is one of the main interactions driving folding and self-assembly as illustrated by α-helical coiled coils. To test the importance of hydrophobic packing in collagen oligomerization, we replaced the conserved isoleucine (Ile12) residue with the canonical proline (SPA-I-P) or hydrophilic serine (SPA-I-S). The CD spectra of both are presented in Fig. 3e, displaying the characteristic maximum near 224 nm associated with PP2 helicity in peptide SPA-I-P, but not I-S. CD melting curve analysis supports triple helicity for SPA-I-P. No thermal transition was observed for peptide SPA-I-S sample in the CD melting plot while a T_m of 22 °C was observed by peptide SPA-I-P (Figs. 3f and S23b). SEC characterization of sample SPA-I-P suggests triple helix formation but no higher-order oligomerization (Fig. 3g). Even though the peptide containing the Ile to Pro substitution has enhanced triple helical thermal stability (T_m = 22 °C vs. 19 °C), oligomerization was not observed by SEC (Fig. 3e, f and g). This demonstrates the importance of Ile in the oligomerization process and also that triple helix stability as assessed by T_m is not necessarily correlated with higher-order assembly. Replacement of this hydrophobic residue with a hydrophilic Serine ablated triple helix assembly completely: the CD spectrum lacked the PP2 signature, and only monomeric species were observed in the SEC (Fig. 3e, f, g). Together these amino acid substitution studies suggest the importance of charged-pair interactions involving arginine and aspartate and hydrophobic interactions for collagen oligomerization. This hypothesis was explored further in the following de novo-designed peptides.

Design of assemblies beyond the triple helix

Directed by the sequence-structure relationship information revealed above, we prepared collagen-like peptides designed to assemble beyond the level of a triple helix. The (GPO)$_{10}$ peptide sequence was used as a template for the design (Fig. 4a)[41,42]. Three features were imprinted onto this template: First, we added a short noncollagenous domain of "DAA" at the N-terminus of the peptide as the presence of noncollagenous domain in SP-A appears to reduce unwanted aggregation and favors a more specific oligomer formation. The choice of a charged residue "Asp" in this region decreased the pI point of the peptide to approximately 4, lower than the physiological pH that peptides were prepared to fold. This was anticipated to have a similar effect as the natural non-collagenous sequences which also have a net negative charge. Second, in the SP-A peptides, the hydrophobic residues Ile, Leu, Val, and Met were spaced out by approximately three triplets. We therefore introduced the same types of hydrophobic
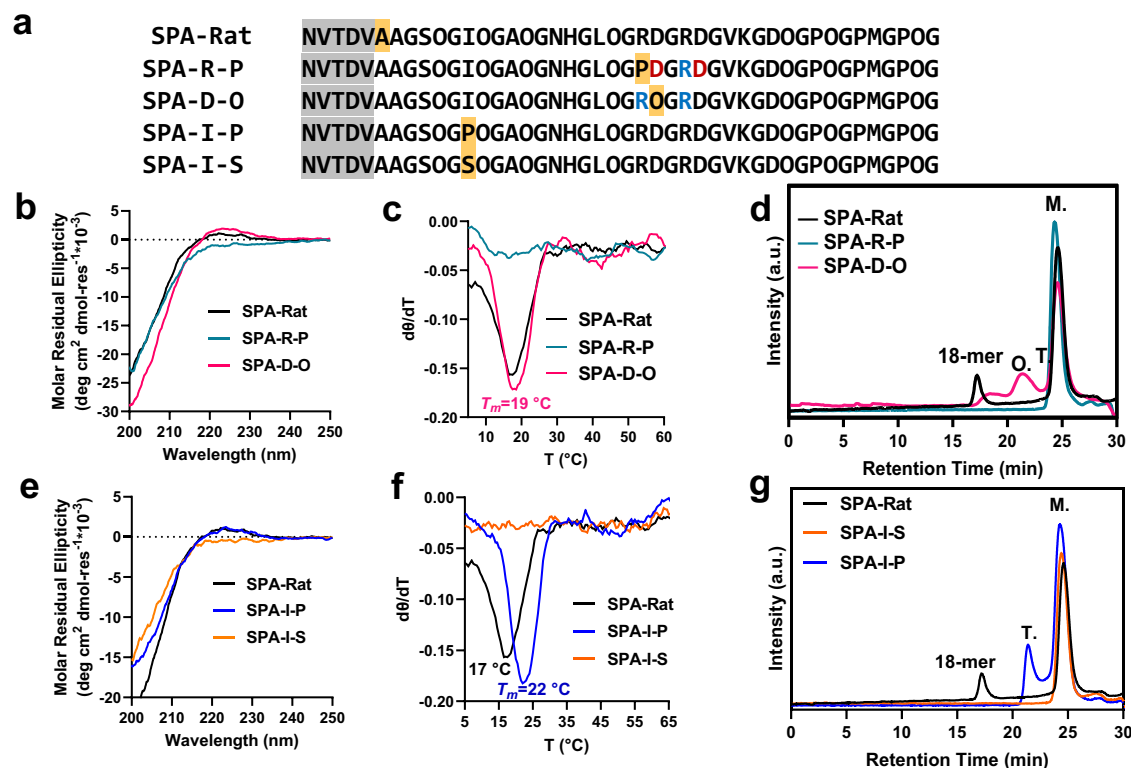
Fig. 3 | Amino acid substitution study of SP-A derived peptides. a Amino acid sequence of the mutant collagen-like peptides derived from SPA-Rat peptide. **b** CD spectra of peptide mutants involving Arg24Pro and Asp25Hyp substitutions. The amino acid sequence numbering starts after the cleavage of the signal peptide[43]. Black: SPA-Rat, green: SPA-R-P, red: SPA-D-O. **c** Thermal stability study of the mutant peptides SPA-R-P and S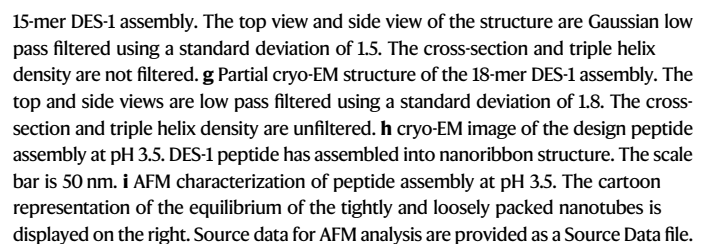PA-D-O. **d** SEC results of peptide variants SPA-R-P and SPA-D-O. **e** CD spectra of the SPA peptides involving Ile12Pro and Ile12Ser substitutions. Black: SPA-Rat, blue: SPA-I-P, orange: SPA-I-S. **f** Thermal stability study of the mutant peptides SPA-I-P and SPA-I-S. **g** SEC results of peptide mutants SPA-I-P and SPA-I-S. "M." represents monomer and "T." represents trimer. Source data for CD and SEC are provided as a Source Data file.

amino acid residues into the designed peptide in the i, $i+9$, $i+18$, and $i+28$ positions. Finally, "RDGRDG" charged sequences were introduced into the middle of the peptide for electrostatic interactions, similar to the SP-A peptides. DES-1 contains all three of these features. DES-P contains the "DAA" N-termini and polar middle while DES-H contains the "DAA" N-termini and hydrophobic groups. Three additional peptides (DES-1-noDAA, DES-P-noDAA, and DES-H-noDAA were prepared in which the N-terminal triplet was removed. All these peptides were synthesized and characterized in the same fashion as previous peptides.

Peptides were prepared at a concentration of 3 mM at a pH of 6.1 ± 0.2 in 20 mM MES buffer, see "Methods" for details. The CD spectra and melting curve analysis of all six peptides are indicative of triple helix formation based on a positive peak near 225 nm which undergoes a cooperative transition upon heating (Figs. 4b, c, S24a, and S25). The thermal stability of the formed triple helices was found to be 37 °C, 49 °C, and 51 °C for the DES-1, DES-H, and DES-P triple helices, respectively (Fig. 4b, c). Of these peptides, only DES-1 was found to assemble into structures beyond the triple helix as observed in both SEC and cryo-EM. In fact, DES-1 assembled into hollow protein nanoparticles as demonstrated in cryo-EM images (Fig. S29). Analysis of the cryo-EM data suggests that these hollow particles are a mixture of three different assemblies: a relatively rare doublet ring structure in addition to much more common bundles of five and six collagen triple helices (Fig. 4d–g). (The resolution information of the structure is listed in Table S2.) This indicates that the predicted sequence-structure relationship investigated in the amino acid replacement study correctly identified key elements controlling higher-order assembly. Whether the doublet structure was made up of octadecamers, pentadecamers, or a combination of each was impossible to determine at

the low resolution of images obtained. No higher-order assembly was observed with the DES-H or DES-P peptides, only trimer species were observed from the SEC chromatography (Fig. S24b), indicating that the hydrophobic residues and charged residues are insufficient to drive assembly alone and that the combined effects of both charged and hydrophobic residues in DES-1 are critical. Removal of the "DAA" N-terminal triplet led to the precipitation of DES-1-noDAA suggesting that the charge repulsion it provides is critical for preventing uncontrolled aggregation in a fashion similar to the natural SP-A derived sequences. As DES-P and DES-H did not assemble past simple triple helices, the presence or absence of DAA had little impact on their assembly.

To test our hypothesis that charged residues are critical for the protein oligomer assembly, we tested DES-1 at different pH. We made a sample at pH 3.5, lower than the pKa of Asp, a sample at pH 6.1 which is the default pH we made our samples, and a sample at the mildly basic pH of 8.0. After 3 weeks of equilibration, the sample prepared at pH 3.5 became turbid. CD spectra and melting curve analysis in Fig. S26 suggest peptide DES-1 folded into collagen triple helices at all tested pH values. CD melting results show melting temperatures of DES-1 peptide folded in pH 3.5, pH 6.1 and pH 8.0 are 31, 38, and 37 °C respectively. The T_m differences indicate that the assemblies in pH 6.1 and 8.0 are likely similar to each other but different from the assembly in pH 3.5. Cryo-EM characterization demonstrated that DES-1 peptide had undergone further assembly into nanoribbons with a width of approximately 50 nm and varied length at pH 3.5 (Figs. 4h and S27). The 2D class average of the cryo-EM results of the nanotube region in this DES-1 peptide assembly at pH 3.5 shows a uniform diameter of 50 Å with a space between striations of 10 Å (Fig. S27c). Scanning AFM of this peptide at pH 3.5 exhibited similar nanoribbons (Figs. 4i and S28).

Fig. 4 | Structural characterization of the pH-responsive DES-1 peptide assembly. a Amino acid sequence of template peptide $(GPO)_{10}$ and the design peptides. **b** CD melting curves of the design peptide (left panel). **c** The first-order derivative curve of the CD melting curves of the designed peptides. (right panel) Black: DES-1, green: DES-P, red: DES-H. Source data for CD and SEC are provided as a Source Data file. **d** 2D class averages of the particles showing a range of classes including pentadecameric (15-mer) assemblies, assemblies with a doublet, and octadecameric (18-mer) assemblies. The peptide assembles in pH 6.1. **e** Low-resolution partial cryo-EM structure of the doublet DES-1 particle assembly. **f** Partial cryo-EM structure of the 15-mer DES-1 assembly. The top view and side view of the structure are Gaussian low pass filtered using a standard deviation of 1.5. The cross-section and triple helix density are not filtered. **g** Partial cryo-EM structure of the 18-mer DES-1 assembly. The top and side views are low pass filtered using a standard deviation of 1.8. The cross-section and triple helix density are unfiltered. **h** cryo-EM image of the design peptide assembly at pH 3.5. DES-1 peptide has assembled into nanoribbon structure. The scale bar is 50 nm. **i** AFM characterization of peptide assembly at pH 3.5. The cartoon representation of the equilibrium of the tightly and loosely packed nanotubes is displayed on the right. Source data for AFM analysis are provided as a Source Data file.

The nano-ribbon assembly is heterogeneous with two general architectures observed: tightly packed nanotubes and more loosely packed nanoribbons, indicating the possibility of equilibrium between the two assemblies. The nanoribbon structure has a height of ~9 nm and the double-layer region has a height of approximately 18 nm (Fig. 4i). The nanoribbon formation has slow kinetics and nanostructure assembly was observed only after at least 2 weeks of equilibration at room temperature (Fig. S30). These supramolecular polymers help to illustrate the role of the DAA non-collagenous domain and control over the isoelectric point of the assembly as the low pH employed here allows for neutralization of long-range ion-ion repulsion and subsequent extensive assembly. The cryo-EM image of the sample equilibrated at pH 8.0 showed similar hollow protein particles as seen in the peptide assembled at pH 6.1 (Fig. S29). This is reasonable as there is no expected net charge difference of the peptide at pH 6.1 vs. pH 8.0. These results suggest that the DES-1 peptide assembly is pH-responsive and the aspartate side chain is involved in the hierarchical assembly process.

De novo design of heterogeneous oligomeric bundle and 2D nanosheets

The pH-responsiveness of the DES-1 peptide indicates that charged residues play a significant role in driving the higher-order assembly. Motivated by this information, we designed collagen-like peptides containing only charged amino acid residues, but with no other deviations from the standard (Gly-Pro-Hyp) repeat (Fig. 5a). RD1 and RD2 peptides contain a short non-collagenous DAA triplet to prevent unwanted aggregation and each peptide contains four triplets of the "RDG" sequence (in contrast to the DES-1 and DES-P which included only two). The "RDG" sequence in the RD1 peptide is centered in the amino acid sequences whereas that of the RD2 is spaced out by four triplets of "GPO" sequences. Both peptides assembled into collagen triple helices as indicated by the CD spectrum and melting curve analysis presented in Fig. 5b–d. However, the thermal stability of the two peptide assemblies varied significantly. The CD melting results of RD1 and RD2 show T_m values of 50 °C and 34 °C (Fig. 5c, d), respectively. It has been suggested that the thermal stability of collagen triple helices is related with the amino acid propensity[23,46,48]. RD1 and RD2 have identical amino acid compositions and therefore might be expected to have similar thermal stability. By contrast, the melting temperatures vary significantly, suggesting that different higher-order assemblies are likely formed by these two peptides. Additionally, the melting curve of RD1 peptide has broad thermal transition, suggesting the higher-order assemblies formed by this peptide is heterogenous.

Dynamic light scattering (DLS) of RD1 showed the formation of nanoparticles of approximately 19 nm (Fig. S31). Cryo-EM also demonstrated nanoparticle formation (Fig. S32). The 2D class averages of the cryo-EM data show hollow protein particles with multiple layers and a concentric pore (Fig. 5f). Most of the particles observed form either pentadecameric (15-mer) or octadecameric (18-mer) structures in the innermost layer (Fig. 5g). The innermost layer of the hexameric structures has a hollow pore size of 17 Å, similar to the C6 assembly formed by C1q peptides[36]. The width of each layer of the ring is approximately 10 Å, matching the diameter of a single triple helix (Fig. 5g). Based on this we assume that triple helices oligomerized into hollow 15-mer or 18-mer assemblies via charge-pair interactions between the side chain of arginine and aspartate in the RD1 peptides. Subsequently, more triple helices bundled onto the oligomer, forming the second layer of triple helices. The unsatisfied charge-pair interactions likely initiate the formation of multiple layers of triple helices.

Different from RD1 peptide, the RD2 peptide formed two-dimensional nanosheets with varied sizes as observed from the cryo-EM images (Figs. 5h and S33). The width of each assembly unit is approximately 10 Å, corresponding to the width of a triple helix and the space between the striation of the nanosheets is approximately

14 Å, which may correspond to the distance between individual triple helices. AFM measurements of the 2D nanosheets suggest that the height of the nanosheets is ~10 nm (Figs. 5i and S34), which corresponds to the height of a triple helix. We hypothesize that the triple helices pack parallely into 2D nanosheets via charge-pair interactions. Similar collagen triple helical 2D nanosheets, which like these lack patterned hydrophobic amino acids, have been reported by the Conticello group[26,28,29].
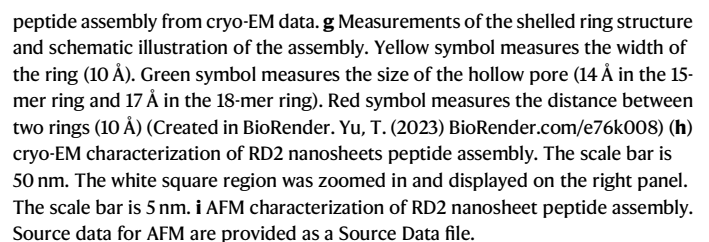
In our SP-A peptides and DES series of peptides, we found the short noncollagenous DAA triplet to be important in controlling higher-order association and eliminating precipitation. To examine its role in the RD peptides we prepared RD1-no DAA and RD2-noDAA. Interestingly RD1-noDAA did not fold into stable triple helices according to CD and SEC characterizations (Fig. S35) nor did it precipitate. We reason that the formation of heterogeneous rings of RD1 peptide observed must be stabilized in part through interaction of this triplet. In contrast, RD2-no DAA to was found to have a notable difference in assembly behavior. The RD2-no DAA peptide formed 2D microcrystals exceeding 3 micrometers in size (refer to Figs. S36 and S37 for TEM and AFM), significantly larger than those formed by the RD2 peptide. This observation lends additional credence to our hypothesis, suggesting that the short noncollagenous domain may serve to deter uncontrolled aggregation.

Discussion

In this study, we have investigated three peptides derived from protein SP-A. Folding studies suggested the short non-collagenous region prevents aggregation that happened in SPA-bcb peptide. This precipitation/aggregation might also be caused by the fact that the peptides were assembled at a pH close to the isoelectric point (pI) as decreased solubility of proteins was observed when pH is close to the pI[49]. However, we assume that the nocollagenous domain has additional underexplored structural functions that affect the bundled collagen-like oligomer folding. For example, we have observed that the oligomer formation has an enhanced equilibrium fraction in SPA-Human which differs from SPA-Rat only in the noncollagenous domain. The amino acids in SPA-Rat and SPA-Human at the 7th position are alanine and valine, respectively. Valine is more hydrophobic than alanine, and we suspect that additional hydrophobic interaction may exist in the assembly of SPA-Human peptide to enhance oligomer formation. Our design peptides that contain the short "DAA" non-collagenous domain displayed higher-order self-assembly that was more effectively controlled and avoided uncontrolled aggregation and precipitation.

From the characterization of the SPA-mutants, we observed no triple helix folding of SPA-R-P peptide but SPA-D-O mutant still form trimer and an oligomer of a different oligomeric status. This suggests that the Arginine side chain may form H-bonding with the carbonyl oxygen of the neighboring triple helix where the Asp residue was replaced. Therefore, oligomer formation was partially maintained in SPA-D-O but not SPA-R-P. Peptide mutant SPA-I-P gained enhanced thermal stability of the triple helix, but the oligomer formation was ablated, indicating that a stable triple helix will likely be more reluctant to form higher-order oligomers. We, therefore, do not think the formation of a triple helix is the prerequisite for oligomerization, rather, we believe that the oligomer formation happens simultaneously with the trimer formation. The fact that we did not observe trimer intermediate in SEC supported this hypothesis.

Interestingly, the DES-1 oligomers at pH 6.1 formed discrete structural populations of pentadecamers, octadecamers, and doublets. The SPA-human cryo-EM structure also revealed octadecameric structures in addition to heterogeneity. The discrete nature of these structures is likely a combination of the presence of both charged and hydrophobic interactions. The DES-1 structure with its two "back-to-back" repeating RDG motifs assembled into these discrete

a

RD1 DAAGPOGPOGPOGRDGRDGRDGRDGPOGPOGPO
RD2 DAAGPOGRDGRDGPOGPOGPOGPOGRDGRDGPO

Fig. 5 | Characterization of RDG-designed peptides. a Amino acid sequence of designed peptides containing different "RDG" sequence patterns. b CD spectra of design peptides. Black: RD1, green: RD2. c CD melting curves of design peptides. The signals were monitored at 225 nm. d The first-order derivative curves of the melting curves displayed in (c). Source data for CD and SEC are provided as a Source Data file. e Cartoon illustration of RD1 and RD2 triple helices. The model is created by amino acid substitution of PDB file 3bOs using PyMol. The charged side chains occupy one-third of the arc of the RD1 triple helix and the side chains of charged residues nearly align along the axis in the RD2 triple helix. For clarity, only charged residues of the leading chain are shown. f 2D class averages of the RD1

peptide assembly from cryo-EM data. g Measurements of the shelled ring structure and schematic illustration of the assembly. Yellow symbol measures the width of the ring (10 Å). Green symbol measures the size of the hollow pore (14 Å in the 15-mer ring and 17 Å in the 18-mer ring). Red symbol measures the distance between two rings (10 Å) (Created in BioRender. Yu, T. (2023) BioRender.com/e76k008) (h) cryo-EM characterization of RD2 nanosheets peptide assembly. The scale bar is 50 nm. The white square region was zoomed in and displayed on the right panel. The scale bar is 5 nm. i AFM characterization of RD2 nanosheet peptide assembly. Source data for AFM are provided as a Source Data file.

structural states while the addition of more RDG residues into the peptide sequence in RD1 resulted in the loss of discrete oligomeric states as additional "layers" of triple helical interactions appeared to form. Because our cryo-EM reconstructions were not of the entire complexes, as well as due to the limited resolution, we were unable to build atomic models of these assemblies. This limits our interpretation of the structural results. Despite this, it seems reasonable that the addition of more arginine and aspartate residues increases the charged-charged interaction between triple helices allowing for ordered layers outside of a octadecameric, pentadecameric, or doublet assembly.

Serendipitously, the DES-1 peptide is a pH-responsive assembly. Previously, a pH responsive collagen assembly was developed by the Conticello group, using charged amino acid side chains of arginine and glutamate residues[27]. The findings suggesting that charged residues can be included for future development of pH-sensitive materials.

RD1 and RD2 have formed different quaternary structures. Even though we do not have atomic level characterization of these assemblies, we believe that the difference between the sequence patterns is expected to result in the different three-dimensional geometry of the side chains in the triple helices. We have modeled a single triple helix of RD1 and RD2 peptides in PyMol. The side chains of RDG sequences in

RD1 peptide spiral up the helical axis from C to N term and occupy approximately 1/3 of the helical arc (Fig. 5e) versus the side chains of the RDG residues are on the same side of the triple helix (Fig. 5e). For clarity, only charged residues of the leading chain is shown. We hypothesize that in the folding process, the leading chain of one triple helix communicates with the leading chain of the neighboring triple helix, similarly for middle and lagging chains. Consequently, the differences in the tertiary structure of a triple helix will result in that the RD1 triple helices pack parallelly but arced whereas the RD2 triple helices pack parallelly and almost linear, forming hollow heterogeneous bundled oligomers and 2D nanosheets, respectively.

In this study, we obtained discrete octadecameric assembly using collagen-like peptides derived from protein SP-A. Our exploration into the sequence-structure relationship of this oligomer formation involved a detailed examination of a variety of peptide mutants, particularly emphasizing the role of charged and hydrophobic residues in protein folding. We discovered that the charged residues of arginine and aspartate side chains, along with hydrophobic residue isoleucine, are key factors of facilitating oligomer assembly. Using this information, we designed collagen-like peptides that formed a pH-responsive collagen bundle-nanoribbon assemblies, heterogenous bundled oligomers and 2D nanosheets. This finding highlights the significance of amino acid sequence patterns in determining the 3D structure of the assembly, combined with previously published research on collagen-like nanofibers, underscore the versatility and complexity of collagen-like peptide assembly[25,30–32] While our study does not fully exhaust the space of assembly possibilities, we anticipate that our findings will offer valuable insights for future efforts in protein design, especially concerning collagen-like mimetic peptides.

Methods

Peptide synthesis
Amino acid reagents Fmoc-Leu-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Hyp(tBu)-OH, Fmoc-Arg(pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Pro-OH, Fmoc-Gly-OH, Fmoc-Ala-OH, Fmoc-Pro-OH, Fmoc-Lys(Boc)-OH, Fmoc-Asn(Trt)-OH, Fmoc-His(Trt)-OH, Fmoc-Val-OH, Fmoc-Met-OH and all the other reagents were obtained from Sigma Aldrich. Peptides were synthesized manually using solid-phase peptide synthesis with Fmoc chemistry. In detail, a low-loading rink-amide resin was used to generate C-terminal amidated peptides. The resin was swelled twice with dichloromethane (DCM) and twice with N,N-dimethyl formaldehyde (DMF). 25 % v.v. piperidine in DMF was added into the resin to deprotect the Fmoc protecting group for 5 min. The resin was washed five times with DMF and tested with chloranil test or Kaser ninhydrin test to confirm presence of free amines. The amino acid, activating reagent hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) and a weak base diisopropylethylamine (DiEA) were dissolved in DMF in a molar ratio of 1:4:4:6 (resin/amino acid/HATU/DiEA) and mixed for 1 min for pre-activation. The activated amino acid solution was added into the resin for coupling of 20 min. After coupling, the reaction solution was filtered, and the resin was washed twice with DMF, one time with DCM and one time with DMF. Negative amine test was observed for each coupling. The deprotection, coupling, washing and testing procedures were repeated for each cycle of amino acid attachment.

After synthesis, the peptides were subject to the final deprotection cycle and washed three times with DCM. The peptides were acetylated with acetic anhydride. In detail, 1:8:4 resin/acetic anhydride/DiEA were mixed and reacted for 45 mins. The acetylation was repeated once. The resin was then washed three times with DCM.

After acetylation, the peptides on the resin were mixed with the cleavage cocktail (90% TFA/2.5% Anisole/2.5% Milli Q H$_2$O/2.5% TIPS/2.5% EDT by volume) and reacted for 3 h. The cleaved peptide solution was drained into a clean round bottom flask. The TFA was blown off with nitrogen gas. Ice-cold diethyl ether was added into the peptide solution to triturate the peptide. The mixture was then centrifuged to obtain the white pellet of the crude peptide. The white pellet was washed once more with cold ether to remove the cleavage cocktail further. After that, the crude peptide was air-dried in the fume hood.

Peptide purification and mass spectrometry
The peptides were purified with the reverse phase high-performance liquid chromatography (HPLC) in a Waters instrument equipped with a C18 column (Waters). Milli-Q H$_2$O with 0.05% w./v. TFA was mobile phase A and acetonitrile (Sigma Aldrich) with 0.05% w./v. TFA was mobile phase B. A general gradient of 5-50% B, 1.5%B/min, was used for peptide purification. The purified peptides were characterized of the molecular weight with a Bruker AutoFlex Speed MALDI ToF instrument or an Agilent qToF LC-MS instrument. Peptide purity post HPLC was checked with ultra performance liquid chromatography (UPLC, Waters).

Peptide sample preparation for folding
Purified peptide solutions were freeze-dried with a FreeZone 4.5 L Freeze Dry System. Peptides solutions of 3 mM peptide in 20 mM MES buffer (Sigma Aldrich), pH 6.1 ± 0.2 were made for self-assembly. The pH was adjusted with 1 M NaOH and the pH value was confirmed with a pH meter (Mettler Toledo). The pH of the DES-1 peptide sample at pH 3.5 was adjusted with 1 M HCl solution and the sample at pH 10.0 was adjusted with 1 M NaOH. Peptide solutions were confirmed by pH meter to be the expected pH values before characterizations. Peptide solutions were annealed at 65 °C for 15 min and then equilibrated at 4 °C for >72 h before characterizations. The nanoribbon assembly from DES-1 peptide and the nanosheets structure from RD2 peptide require three weeks to form at room temperature.

Circular dichroism
The circular dichroism (CD) data were collected on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller using a cuvette with 1 mm path length. Peptide solutions were diluted into 0.3 mM concentration with 20 mM MES buffer for data collection. The melting curves were collected from 5 °C to 65 °C with a heating rate of 10 °C/h at 225 nm of each sample. The first-order derivatives of the melting curves were calculated with a Savitzky-Golay smoothing algorithm, and temperature where the minimum derivative value appears was regarded as the melting temperature. The molar residue ellipticity (MRE) value was calculated with the equation, MRE = $(\theta \times m)/(c \times l \times nr \times 10)$ where θ represents the experimental ellipticity in millidegrees, m is the molecular weight of the peptide (g/mol), c is the peptide concentration (milligrams/milliliter), l is the path length of the cuvette (cm), and nr is the number of amino acid residues in the peptide.

Size exclusion chromatography
The SEC data were collected in a Varian Star instrument equipped with a Superdex 200 Increase 10/300 GL column. The mobile phase was 100 mM MES buffer, and the flow rate was 0.75 ml/min. The peptide stock solution was diluted with 20 mM MES buffer and 100 uL of 0.6 mM peptide solution was injected for each measurement. The signal was monitored at 220 nm wavelength. The system was cooled and equilibrated at 10 °C for 45 min before data collection to prevent denaturation of the supramolecular assemblies that are unstable at room temperature.

Cryo-EM characterization and structure reconstruction
Aliquots of the peptides were diluted with 20 mM MES buffer into 0.1 wt% for grid freezing. For sample dilution of DES-1 nanoribbon sample, the buffer was adjusted to be pH 3.5 before usage. All grids used for this study were on Lacey carbon 300 mesh copper grids. The grids were frozen with a Vitrobot™ Mark IV plunge freezer. Prior to the addition of the sample, the grids were cleaned using a Gloqube® Plus

glow discharge system. Grids were imaged on a Glacios 200 kV electron microscope with a cryo Autoloader, XFEG electron source and Falcon4 direct electron detector (1.2 Å/pixel). The basic reconstructions scheme for single particle cryo-EM is shown in Fig. S38 for the SPA octadecamer and Fig. S39 for the DES-1 doublets, octadecamer, and pentadecamers. Processing of raw movies and subsequent image processing and reconstruction steps of images from cryo-EM were handled in cryoSPARC[50]. Images were motion corrected and CTFs were estimated using the in house Patch Motion Correction and Patch CTF estimation software. For each dataset-500 or so "particles" were selected using the manual particle picker function and these were then used to generate initial 2D class averages using the 2D classification program in cryoSPARC. For many samples multiple rounds of sorting based on 2D classification were performed to sort through obvious "junk" particles. For the SPA-human octadecamers, DES-1 pentadecamers, and DES-1 octadecamers ab initio reconstruction with multiple classes was performed with either C5 or C6 symmetry and the resulting "good" classes were then subjected to homogenous refinement with the same point group symmetry. For the DES-1 doublets the same basic protocol was followed except that C1 symmetry was imposed. Any density maps that were determined were then analyzed in UCSF Chimera[51] and ChimeraX[52]. The map:map Fourier shell correlation curves for each structure are shown in Fig. S40.

AFM characterization

The atomic force microscope (AFM) characterization was performed in a Park NX20 AFM instrument using tapping mode, scanning rate 1.0 Hz. Measurements were performed using AFM tips with a force constant of 42 N/m and a resonance frequency of 320 kHz (Nano-World). The peptide solution was diluted using 20 mM MES buffer to a final concentration of 0.3 mg/mL. For sample dilution of the DES-1 nanoribbon sample, the buffer was adjusted to be pH 3.5 before usage.10 uL of diluted peptide solution was drop cast to a freshly cleaved mica slide. After 5 mins of bonding, the solution was wicked away, and the mica slide was rinsed with 200 uL of fresh MilliQ H_2O. The mica slide was air dry for 1 h at room temperature before imaging.

Dynamic light scattering

The samples were diluted tenfold with 20 mM MES buffer for measurement. The data were collected in a Malvern Zen 3600 Zetasizer instrument using Malvern ZEN0040 disposable cuvettes for size measurement. All measurements were done at 25 °C.

Data availability

The data generated in this study are provided in the Supplementary Information/Source Data file, and are available from the corresponding author(s) upon request. No PDBs were built for the structures in this study. The DES-1 doublet (EMD-44405), octadecamer (EMD-44406), and pentadecamer (EMD-44409) assemblies' cryo-EM density maps were deposited with the EMDB. The SPA-human octadecamer was deposited with the EMDB as EMD-44418. Source data are provided with this paper.

References

1. Thomas, F. et al. De novo-designed α-helical barrels as receptors for small molecules. *ACS Synth. Biol.* **7**, 1808–1816 (2018).
2. Beesley, J. L. & Woolfson, D. N. The de novo design of α-helical peptides for supramolecular self-assembly. *Curr. Opin. Biotechnol.* **58**, 175–182 (2019).
3. Dawson, M. et al. Coiled coils 9-to-5: rational de novo design of α-helical barrels with tunable oligomeric states. *Chem. Sci.* **12**, 6923–6928 (2021).
4. Dawson, W. M. et al. Structural resolution of switchable states of a de novo peptide assembly. *Nat. Commun.* **12**, 1530 (2021).
5. Silva, D.-A. et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
6. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
7. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
8. Bermeo, S. et al. De novo design of obligate ABC-type heterotrimeric proteins. *Nat. Struct. Mol. Biol.* **29**, 1266–1276 (2022).
9. Chang, L. & Perez, A. Ranking peptide binders by affinity with AlphaFold. *Angew. Chem. Int. Ed.* Engl. **62**, e202213362 (2023).
10. Cao, L. et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
11. Boyken, S. E. et al. De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity. *Science* **352**, 680–687 (2016).
12. Joh, N. H. et al. De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
13. Huddy, T. F., et al. Blueprinting extendable nanomaterials with standardized protein blocks. *Nature* 1–7 (2024). https://doi.org/10.1038/s41586-024-07188-4
14. Vorobieva, A. A. et al. De novo design of transmembrane β barrels. *Science* **371**, eabc8182 (2021).
15. Shimizu, K. et al. De novo design of a nanopore for single-molecule detection that incorporates a β-hairpin peptide. *Nat. Nanotechnol.* **17**, 67–75 (2022).
16. Marcos, E. et al. Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017).
17. Schneider, J. P. et al. Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. *J. Am. Chem. Soc.* **124**, 15030–15037 (2002).
18. Huang, P.-S. et al. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
19. Sarkar, S., Anderson, C. F. & Schneider, J. P. The design of a participatory peptide nucleic acid duplex crosslinker to enhance the stiffness of self-assembled peptide gels. *Angew. Chem.* **136**, e202313507 (2024).
20. Kumar, P., Paterson, N. G., Clayden, J. & Woolfson, D. N. De novo design of discrete, stable 310-helix peptide assemblies. *Nature* **607**, 387–392 (2022).
21. Graham, L. A. & Davies, P. L. Glycine-rich antifreeze proteins from snow fleas. *Science* **310**, 461–461 (2005).
22. Rath, A., Davidson, A. R. & Deber, C. M. The structure of "unstructured" regions in peptides and proteins: Role of the polyproline II helix in protein folding and recognition*. *Pept. Sci.* **80**, 179–185 (2005).
23. Walker, D. R. et al. Predicting the stability of homotrimeric and heterotrimeric collagen helices. *Nat. Chem.* **13**, 260–269 (2021).
24. Fallas, J. A. & Hartgerink, J. D. Computational design of self-assembling register-specific collagen heterotrimers. *Nat. Commun.* **3**, 1087 (2012).
25. Rele, S. et al. D-periodic collagen-mimetic microfibers. *J. Am. Chem. Soc.* **129**, 14780–14787 (2007).
26. Merg, A. D. et al. 2D crystal engineering of nanosheets assembled from helical peptide building blocks. *Angew. Chem. Int. Ed.* **58**, 13507–13512 (2019).
27. Merg, A. D. et al. Shape-shifting peptide nanomaterials: surface asymmetry enables pH-dependent formation and interconversion of collagen tubes and sheets. *J. Am. Chem. Soc.* **142**, 19956–19968 (2020).
28. Merg, A. D. et al. Seeded heteroepitaxial growth of crystallizable collagen triple helices: engineering multifunctional two-dimensional core–shell nanostructures. *J. Am. Chem. Soc.* **141**, 20107–20117 (2019).
29. Jiang, T. et al. Structurally defined nanoscale sheets from self-assembly of collagen-mimetic peptides. *J. Am. Chem. Soc.* **136**, 4300–4308 (2014).

30. Tanrikulu, I. C., Forticaux, A., Jin, S. & Raines, R. T. Peptide tessellation yields micrometre-scale collagen triple helices. *Nat. Chem.* **8**, 1008–1014 (2016).

31. Hu, J. et al. Design of synthetic collagens that assemble into supramolecular banded fibers as a functional biomaterial testbed. *Nat. Commun.* **13**, 6761 (2022).

32. 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. *Nat. Chem.* **3**, 821–828 (2011).

33. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).

34. Davis, I. W. & Baker, D. Rosetta ligand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **385**, 381–392 (2009).

35. Park, J. et al. De novo design of a homo-trimeric amantadine-binding protein. *eLife* **8**, e47839 (2019).

36. Yu, L. T. et al. Hollow octadecameric self-assembly of collagen-like peptides. *J. Am. Chem. Soc.* **145**, 5285–5296 (2023).

37. Casals, C., García-Fojeda, B. & Minutti, C. M. Soluble defense collagens: sweeping up immune threats. *Mol. Immunol.* **112**, 291–304 (2019).

38. Kilchherr, E., Hofmann, H., Steigemann, W. & Engel, J. Structural model of the collagen-like region of C1q comprising the kink region and the fibre-like packing of the six triple helices. *J. Mol. Biol.* **186**, 403–415 (1985).

39. Fouët, G. et al. Headless C1q: a new molecular tool to decipher its collagen-like functions. *FEBS J.* **288**, 2030–2041 (2021).

40. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75–81 (1994).

41. Fiala, T. et al. Frame shifts affect the stability of collagen triple helices. *J. Am. Chem. Soc.* **144**, 18642–18649 (2022).

42. Qi, Y. et al. Terminal repeats impact collagen triple-helix stability through hydrogen bonding. *Chem. Sci.* **13**, 12567–12576 (2022).

43. UniProt Consortium, The UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

44. McCormack, F. X. et al. The Cys6 intermolecular disulfide bond and the collagen-like region of rat SP-A play critical roles in interactions with alveolar type II cells and surfactant lipids. *J. Biol. Chem.* **272**, 27971–27979 (1997).

45. Yu, L. T. & Hartgerink, J. D. Selective covalent capture of collagen triple helices with a minimal protecting group strategy. *Chem. Sci.* **13**, 2789–2796 (2022).

46. Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960–14967 (2000).

47. Erdmann, R. S. & Wennemers, H. Importance of ring puckering versus interstrand hydrogen bonds for the conformational stability of collagen. *Angew. Chem.* **123**, 6967–6970 (2011).

48. Persikov, A. V., Ramshaw, J. A. M. & Brodsky, B. Collagen model peptides: sequence dependence of triple-helix stability. *Pept. Sci.* **55**, 436–450 (2000).

49. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**, 1206–1215 (2001).

50. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

51. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

52. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

Acknowledgements

Author contributions

L.T. Yu: conceptualization, data curation, formal analysis, methodology, investigation, writing—Original draft. M.A.B. Kreutzberger: data curation, formal analysis, methodology, investigation, writing—review & editing. T.H. Bui: methodology, data curation, writing—review & editing. M.C. Hancu: data curation, investigation, writing—review & editing. A.C. Farsheed: methodology, writing—review & editing. E.H. Egelman: supervision, funding acquisition, writing—review & editing. J.D. Hartgerink: conceptualization, supervision, funding acquisition, writing—review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54560-z.

Correspondence and requests for materials should be addressed to Jeffrey D. Hartgerink.

Peer review information *Nature Communications* thanks Fei Xu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.