HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2024 March 08.

Published in final edited form as:

J Am Chem Soc. 2023 March 08; 145(9): 5285–5296. doi:10.1021/jacs.2c12931.

Hollow Octadecameric Self-Assembly of Collagen-like Peptides

Le Tracy Yu[1], **Maria C. Hancu**[1], **Mark A. B. Kreutzberger**[2], **Amy Henrickson**[3], **Borries Demeler**[3], **Edward H. Egelman**[2], **Jeffrey D. Hartgerink**[1,4,*]

[1]Department of Chemistry, Rice University, 6100 Main Street, Houston, TX 77005, United States

[2]Department of Biochemistry and Molecular Genetics, University of Virginia Box 800733, Charlottesville, VA 22908, United States

[3]Department of Chemistry & Biochemistry, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

[4]Department of Bioengineering, Rice University, 6100 Main Street, Houston, TX 77005, United States

Abstract

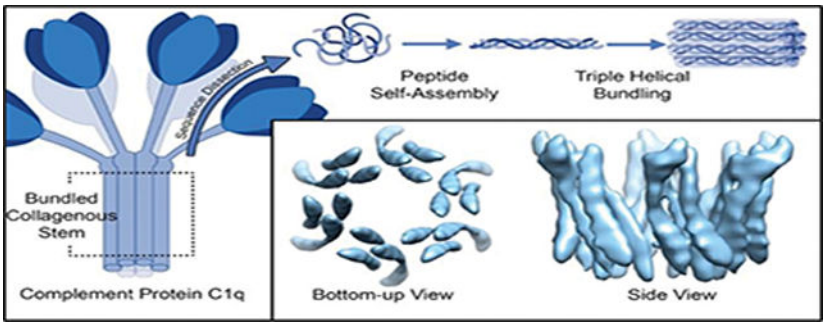
The folding of collagen is a hierarchical process which starts with three peptides associating into the characteristic triple helical fold. Depending on the specific collagen in question, these triple helices then assemble into bundles reminiscent of alpha-helical coiled-coils. Unlike alpha-helices, however, the bundling of collagen triple helices is very poorly understood with almost no direct experimental data available. In order to shed light on this critical step of collagen hierarchical assembly, we have examined the collagenous region of Complement Component 1q. Thirteen synthetic peptides were prepared to dissect the critical regions allowing for its octadecameric self-assembly. We find that short peptides (under 40 amino acids) are able to self-assemble into specific $(ABC)_6$ octadecamers. This requires the ABC heterotrimeric composition as the self-assembly subunit, but does not require disulfide bonds. Self-assembly into this octadecamer is aided by short non-collagenous sequences at the N-terminus, although they are not entirely required. The mechanism of self-assembly appears to begin with the very slow formation of the ABC heterotrimeric helix, followed by rapid bundling of triple helices into progressively larger oligomers, terminating in the formation of the $(ABC)_6$ octadecamer. Cryo-electron microscopy reveals the $(ABC)_6$ assembly as a remarkable, hollow, crown-like structure with an open channel approximately 18 Å at the narrow end and 30 Å at the wide end. This work helps to illuminate the structure and assembly mechanism of a critical protein in the innate immune system and lays the groundwork for the *de novo* design of higher order collagen mimetic peptide assemblers.

Graphical Abstract

[*]**Corresponding Author: Jeffrey D. Hartgerink**- to whom correspondence should be addressed: jdh@rice.edu.

ASSOCIATED CONTENT

Methods and further experimental details are available in the Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org."



Keywords

Collagen triple helix; C1q; fibril; interhelical interactions; higher-order assembly

INTRODUCTION

The structure of collagen and collagen-like proteins is defined by a triple helical folding motif in which three peptide strands, adopting a left-handed polyproline type II (PP2) helix, intertwine in a right-handed superhelix. Three chains associate with a one amino acid stagger as the natural register.[1,2] This triple helix is stabilized by a repetitive primary sequence in each peptide conforming to an $(Xaa\text{-}Yaa\text{-}Gly)_n$ repeat where Glycine is required due to steric constraints and also forms stabilizing interstrand hydrogen bonds. Positions Xaa and Yaa can be any amino acid, but are frequently Proline and Hydroxyproline respectively which help to stabilize the helix by preorganization.[3–8] Additionally, it has been found that pairwise interactions between peptide strands within a helix[9–14], synthetic amino acid side chains[15–17], or covalent linkages[18–21] can be either stabilizing or destabilizing. These interactions can be used to bias helix assembly into specific compositions as well as to control registration between peptide strands. (Figure 1a).

Collagen structural hierarchy, however, does not end at triple helix formation as most collagens undergo a multistep assembly in which triple helices associate to form higher order structures. For example, Type I Collagen triple helices bundle together to form fibrils.[22,23] These fibrils further associate and eventually create macroscopic fibers. Despite the nearly ubiquitous higher order assembly of triple helices in nature, synthetic Collagen Mimetic Peptides (CMPs) have very rarely been observed to assemble beyond a triple helix and the ones that do undergo such assembly form extremely large, polydisperse assemblies.[24–32] Therefore, unlike the bundling of α-helices into coiled coils for which interhelix interactions are reasonably well understood and good predictive models exist,[33–36] understanding the mechanism of triple helix bundling (Figure 1b), and the sequence constraints of this process for the collagen triple helix, is almost entirely unknown. The reasons for this gap in knowledge results from 1) the challenges associated with the study of suitable natural systems (extremely large size, extensive posttranslational modification and crosslinking, difficult expression and poor solubility) and 2) the lack of small, synthetically accessible model systems which undergo controlled self-assembly beyond the triple helix.

Complement Component 1q (C1q)[37,38] is a critical component of the innate immune system and is one example of a family of "defense collagens" which include MBL (mannose binding lectin)[39,40], SP-A (surfactant protein A)[41–43] and adiponectin[44–46] amongst others. C1q is a "bouquet" shaped structure self-assembled from 18 peptide strands. Peptides C1q-A, C1q-B and C1q-C associate to form an ABC heterotrimer and six of these heterotrimers further assemble to generate the full $(ABC)_6$ octadecamer.[47–49] The C-terminal half of each peptide is composed of a globular domain involved in antigen recognition.[50] The N-terminal half contains a long Collagen-Like Region (CLR) with the characteristic $(Xaa\text{-}Yaa\text{-}Gly)_n$ repeat. This repeat is disrupted once in the middle of this domain, creating a kink in the overall bouquet structure that allows the large globular "flowers" to spread out.[51–54] This kink also delineates two different portions of the CLR which we define here as the "stem" region (where all six triple helices are tightly packed) and "branch" region (where the helices splay out and away from one another) (Figure 1c). Finally, the last several amino acids of the N-terminus do not follow the Xaa-Yaa-Gly motif and also contain cysteine residues involved in inter-strand disulfide bonds connecting C1q-A and C1q-B into a heterodimer and two copies of C1q-C into a homodimer.[55,56]

Our hypothesis is that the stem region of C1q can self-assemble into a controlled octadecamer based entirely on interactions between triple helices and independent of any other portion of the molecule. Additionally, we hypothesize that this assembly does not require posttranslational modifications (other than hydroxyproline) and is not dependent on cysteine meditated covalent crosslinks. If true, this fragment of C1q would serve as a molecularly well-defined starting point to understand helix-helix interactions in collagen. This could then provide insights into the higher-level assembly and structure of this critical family of defense collagens as well as traditional fibrillar collagens in their hierarchical assembly. Additionally, we hope that advances provided here will lead to a new generation of *de novo* designed peptides similar to the robust design criteria available to alpha-helical coiled coils.

In the current study we describe a small (less than 40 amino acid) synthetic peptide derived from the collagen-like protein C1q which is able to self-assemble into an $(ABC)_6$ octadecamer with a remarkable hollow, crown-like sructure revealed by cryo-Electron Microscopy. To our knowledge, this is the first report of a collagen-like triple helical system which undergoes a second step of self-assembly to form a discrete (non-polydisperse) oligomer.

RESULTS AND DISCUSSION

C1q Dissection.

The Collagen Like Region (CLR) of C1q has previously been demonstrated to have an octadecameric structure based on enzymatic degradation of the globular domains off from the full native C1q structure.[53] More recently, an expression system was developed which contains the full CLR and a designed trimeric coiled coil at the C-termini, which also maintains the octadecameric assembly.[57] To dissect out the minimum domain necessary for oligomerization, we prepared thirteen peptides derived from human C1q[58] by standard Fmoc solid phase peptide synthesis[59] (Table 1). See supporting information for

details of synthesis, purification by high-performance liquid chromatography (HPLC), and characterization by mass spectrometry (Table S1 and Figure S1–16). Peptides were tested for two degrees of self-assembly: triple helix formation and oligomerization of triple helices into octadecamers. Triple helix formation and thermal stability (melting temperature, T_m) were determined by circular dichroism spectroscopy (CD) while octadecamer assembly was determined by size exclusion chromatography (SEC), analytical ultra-centrifugation (AUC) and / or cryo-EM. Methods and conditions used in all cases are described in the supporting information.

The fundamental question we wanted to answer was, what is the minimum structure required to allow self-assembly into an octadecamer? We started by preparing “Stem” region peptides A, B and C (see Table 1). These peptides contain the full amino acid sequence of C1q starting from the N-termini and extending to the Xaa-Yaa-Gly discontinuity. These peptides are believed to be tightly packed with one another in the full native C1q structure and also contain N-terminal cysteines for covalent crosslinking between peptides. While we include hydroxyproline post-translational modification, other known post-translational modifications such as hydroxylysine were not included. All peptides were synthesized as N-acetylated and C-amidated derivatives for self-assembly stabilization.[60–62] After solid phase synthesis, these peptides were crosslinked by iodine oxidation and HPLC purified to generate the C1q-A – C1q-B (A-B) disulfide dimer and C1q-C – C1qC (C-C) disulfide dimer. These dimers were then allowed to fold and assemble independently or mixed in a 2:1 ratio and similarly allowed to fold and assemble for over one month. Figure 2 shows the CD and SEC data from this study. Circular dichroism shows that while A-B and C-C dimers independently are only suggestive of weak triple helices (melting at or below 15 °C), the 2:1 mixture of A-B with C-C results in a much more robust triple helical signal with a melting temperature of approximately 42 °C. Furthermore, SEC demonstrated that the 2:1 (A-B):(C-C) mixture forms a species much higher in mass than a simple triple helix. Based on SEC calibration curves (see Figure S19A) this high mass peak is consistent with $(A\text{-}B)_6(C\text{-}C)_3$ assembly mimicking native C1q. In contrast, for C-C alone there was no trace of a higher mass species while for A-B alone a small high mass peak was observed. While in previous work, Brodsky and Kajava studied the thermal stability of the full CLR domain of C1q and found that the melting temperature was near 37 °C, here we observe a much shorter region, yet with higher thermal stability.[63,64] This is the first demonstration of C1q region assembling from such a short N-terminal fragment.

Disulfide Bonds Formed through Covalent Capture.

As an alternative approach to disulfide bond formation, instead of mixing preformed and purified disulfide bonded peptides, we instead equilibrated the three Stem Region peptides A, B and C together under reducing conditions and subsequently oxidized the solution to form the disulfide bonds *in situ* only after oligomerization was confirmed by SEC. This allows us to test if the bonds form with specificity through covalent capture or would result in a statistical production of all possible disulfide bonds. Figure 2 (“ABC-cc”, green) shows the data for this covalently captured assembly which has nearly identical CD spectra and thermal unfolding characteristics as the pre-synthesized dimeric peptides. HPLC and Mass spectral analysis of the covalent captured solutions revealed A-B, A-C and C-C dimeric

species in addition to some remaining monomers, but no other dimers were observed (no A-A, B-B, nor B-C dimers Figure S11). This non-statistical distribution of disulfide bonds strongly suggests the oligomeric self-assembly preorganizes the formation of suitable disulfide bonds. Interestingly, however, A-C disulfide bonded pairs have not previously been shown in the literature for the native C1q assembly but were observed in our synthetic system and suggests that the cysteines in peptides A and C may be near enough to one another after assembly to covalently crosslink while this is apparently not possible for A-A, B-B nor B-C.

Disulfide Bonds are not Required for Stabilization or Self-Assembly.

Next, we wanted to determine the requirement of disulfide bonding in our short C1q mimic peptides. To do this we prepared peptides A-Ala, B-Ala, C-Ala (Table 1) in which the cysteine from each peptide was replaced with alanine. From these peptides and the previously described cysteine containing versions we prepared three additional series of self-assembling systems: One in which the A-B disulfide linkage was left intact but the C-C dimer was replaced with C-Ala, a second in which the C-C disulfide linkage was left intact but the A-B dimer was replaced with A-Ala and B-Ala and third a system in which all the disulfide bonds were eliminated using a mixture of A-Ala, B-Ala and C-Ala. The four possible self-assembling systems are shown schematically in Figure 3A. In all cases we observed triple helix formation by CD (Figure S18A & Figure 3B & C) as well as higher order assembly consistent with octadecamer formation by SEC (Figure 3D). Impressively, the thermal stability of the self-assembling system with no cysteine (A-Ala, B-Ala, C-Ala) showed the highest thermal stability (47 °C) and an SEC trace that showed a higher fraction of oligomer formation than any system containing disulfide linkages (Figure 3D). The C-C disulfide bond actually appears to moderately destabilize the triple helix as observed by thermal unfolding while the A-B disulfide bond does not seem to impact stability positively or negatively. These results are consistent with our study of the ABC assembly under reduced conditions (Figure S20) in which peptides A, B, and C assembled in the presence of DTT (dithiothreitol) formed a heterotrimer with a thermal stability of 46°C, comparable to (ABC)-Ala and (ABC)-cc (Fig. S20 A B). This sample under reduced conditions also assembled into an octadecameric structure according to the SEC characterization (Fig. S20 C). From this data, we can conclude that triple helix formation and oligomerization of C1q requires neither the C-terminal amino acids beyond the Xaa-Yaa-Gly disruption nor does it require covalent stabilization from disulfide bond formation. In fact, the thermal stability of this C1q mimic is improved when cysteine is entirely replaced with alanine.

An ABC-type Heterotrimer is Required.

To understand the requirements of self-assembly of the stem region of C1q in more detail we investigated the self-assembly of the supramolecular peptides C1q A-Ala, B-Ala and C-Ala individually and in pairs. Figure 4A and Figure S17 show that none of these peptides individually displays significant PP2 character. Pairwise mixtures also lack significant triple helical character as thermal analysis showed the stabilities of the weak helices that form to be below 10 °C. In contrast, the combination of all three peptides forms a strong PP2 helix with a melting temperature of 47 °C (Figure 4C). This demonstrates that triple helix formation in C1q is strongly dependent on the combination of all three peptides into a

unique heterotrimer as only the ABC combination shows strong PP2 character and good thermal stability.

Additionally, we carried out a detailed analytical ultracentrifugation (AUC) study of the combined ABC system as well as each peptide alone and in pairwise mixtures. The results for the sedimentation velocity experiments are shown in Figure 5. Measurements of peptides A-Ala, B-Ala and C-Ala alone, and in pairs, resulted in virtually identical, near-vertical sedimentation patterns, reflecting a homogeneous composition with a weight-average sedimentation coefficient of 0.54s and a weight average molar mass of 3,620 Da, indicative of a monomeric composition – neither triple helix nor oligomeric assembly is observed (Figure 5A & Figure S19C). On the other hand, mixtures containing all three peptides resulted in one additional species sedimenting at 4.12s (Figure 5B,C), with apparent molar mass of 71.4 kDa (Figure 5D). Molar masses are approximate since the precise partial specific volume is an estimate from the amino acid sequence. However, this mass is in good agreement with what is expected from an 18-peptide assembly composed of six units each of A-Ala, B-Ala and C-Ala, $(ABC\text{-}Ala)_6$, which has an expected mass of 67.9 kDa. The measurements of the three-peptide mixture at multiple concentrations did not produce a significant change in composition suggesting that the portion of the sample co-sedimenting at the same speed as the peptide monomeric controls is a result of non-stoichiometric mixing, with one or two of the peptides being present in excess. The remainder of the material is fully assembled into the species with a sedimentation coefficient of 4.12 s and does not change composition when the concentration is changed. This suggests that the K_d (the dissociation constant) is lower than the lowest concentration accessed in this experiment (14 μM), and the assembly has strong binding, but only when all three peptides are present in the mixture. This is consistent with CD and SEC data presented above. When the three-peptide mixture was diluted and heated at 65 °C for 20 minutes then cooled at room temperature for 2 hours before measurements, only monomeric species were observed which did not reassemble within 2 hours. Finally, we examined the effect of ionic strength on the 3-peptide mixture. The sample measured in pure water exhibited a slight heterogeneity of the larger species (Figure 5C). Genetic algorithm analysis detected a shift of the major peak to 4.38s and additional minor species sedimenting at 4.87s and 5.49s. A similar shift was observed at each peptide concentration (data not shown). Due to the significant charge of the peptide, a measurement without the concentration of screening salts present in buffered samples may result in non-ideal results, hence we did not investigate these changes in sedimentation further.

While our CD, SEC and AUC mixing experiments demonstrate that an ABC composition is required for assembly, we are unable to determine the register of the three peptides with respect to one another in a triple helix. Predictions of thermal stability from our previously published algorithm "SCEPTTr"[9,10], which are usually quite accurate, are extremely poor for this system of peptides (−2 °C T_m predicted vs. 47 °C T_m experimentally observed) which suggests that inter-helical contacts are a major contributor in the overall stability of the system and also mean that prediction of register cannot be made. This is a stumbling block for future designed systems based on C1q which we hope to overcome with future molecular level characterization.

Structure of the $(ABC\text{-}Ala)_6$ Assembly.

We used cryo electron microscopy (cryo-EM) to directly investigate the structure of the $(ABC\text{-}Ala)_6$ peptide assembly. Electron micrographs of the sample revealed small ring-like particles approximately 50 Å in diameter (Figure 6A). The ring-like structure was further evident from 2D class averages containing top views which revealed 6-fold symmetry (Figure 6B). This structure was reconstructed with C6 symmetry imposed and filtered to ~5 Å. Attempts to reconstruct with C1, C2, and C3 symmetries yielded lower quality density maps. In density maps viewed from the "bottom" (Figure 6C) or "top" (Figure 6D) six triple helical structures are evident which form a super-coil around a hollow center. When looking at the three-dimensional density map from a "side view" it is clear that the triple helices interact to form a crown-like structure (Figure 6E) which has a pore diameter of about 18 Å at the base and opens to approximately 30 Å at the top. While the strands in each triple helix are observed to be in close proximity with each other, adjacent triple helices only made close contacts at the narrower base of the crown and then moved further apart. The clear triple-helical density that we observed (Figure 6C–F) represent less than half of the amino acids from each strand. As a result of the relatively modest resolution of the structure, as well as the small sequence coverage in the map, we are unable to definitely determine the handedness observed in the images. The hand imposed on the images in Figure 6 is based on the assumption of the right-handed helicity of the component triple helices, which are well known. Assuming this chirality, the crown itself appears as a left-handed super helix of right handed triple helices. Strang et al. and Shelton et al. have previously performed electron microscopic studies of whole C1q that contains the full length of the collagenous and globular domains.[48,65] Their results suggest the complete C1q is a bouquet structure containing a stalk-like central region 3–6nm diameter and a length of 10–12.5 nm. Our results show a crown-like structure which is consistent with their low-resolution rotary shadowing and negative stain TEM results.

Rate and Mechanism of Folding.

The rate of assembly of the $(ABC\text{-}Ala)_6$ was monitored by CD over time (Figure S18B & Figure 7). The maximum near 225nm, which corresponds to PP2 character, slowly increased over a very long time - one month - where it reached a plateau demonstrating its very slow folding rate. Interestingly, the thermal unfolding profile of the system at different time points also changed with the T_m slowly increasing from just below 40 °C to its final value of 47 °C. SEC over a similar time frame initially show only very small quantities of assembled material. At no time do we observe assemblies corresponding to a simple triple helix by SEC (which elute near 15 minutes). Instead, the first observable oligomers correspond to at least hexamers with an elution time of 13 minutes. With extended equilibration time, both the quantity of oligomers increases and the size of the oligomers increase, ultimately eluting just after 11 minutes which matches expectations for the ultimate octadecamer assembly of $(ABC\text{-}Ala)_6$ (Figure 7c). This data suggests that the rate limiting step of assembly is triple helix formation. Once formed, triple helices are rapidly consumed into a higher order oligomer, which may be as small as $(ABC\text{-}Ala)_2$. This explains the lack of observable ABC-Ala triple helix at any time point by SEC.

Minimum Required Sequence.

Finally we set about reducing the length of the peptide to further isolate the minimum sequence necessary to form a stable triple helix and octadecamer. The N-terminal portion of peptide A-Ala and B-Ala contain a short domain that is not consistent with the Xaa-Yaa-Gly requirement of a triple helix (see Table 1, grey highlighted area) while C-Ala's sequence is entirely compatible with triple helix formation. In order to determine if this non-collagenous region is required for triple helix or oligomer assembly, two new peptides were prepared that eliminated the non-collagenous sequence named A-cx and B-cx (A & B "collagen exclusive"). Mixtures of these peptides that are fully consistent with the sequence of a triple helix and those that contain additional N-terminal amino acids were made thereby eliminating one or both N-terminal regions for a total of three new mixtures: A-cx/B-cx/C-Ala, A-cx/B-Ala/C-Ala, and A-Ala/B-cx/C-Ala. In all cases CD showed the formation of stable triple helices, however melting data demonstrate that elimination of either or both non-collagenous regions result in moderate to significant thermal destabilization (Figure 8A,B). Of the two N-terminal non-collagenous regions, the one associated with C1q-A appears to be substantially more important than that associated with C1q-B. Elimination of the non-collagenous domain of C1q-B left the triple helix with only moderately reduced melting temperature (40 °C vs. 47 °C). Furthermore, the same system retains strong oligomerization as demonstrated by SEC (Figure 8C). In contrast, eliminating the non-collagenous domain of A or both non-collagenous domains result in a melting temperature of 34 °C and 30 °C respectively and elimination of most, but not all, high mass oligomers. These results suggest that the non-collagenous domain of C1q-A plays an important, but not absolutely required, structural role in triple helix oligomerization.

To further investigate the importance of C1q-A's non-collagenous domain, we synthesized five additional mutants with different length of the non-collagenous domain (Table 1, N-terminal deletion series) and mixed them for self-assembly with peptide B-Ala and C-Ala and conducted similar CD and SEC experiments (Figure 9). Peptide mutant A-5 has the shortest length of the non-collagenous domain, including only one alanine residue. According to the CD melting results, the heterotrimer assembled from A-5, B-Ala and C-Ala has a melting temperature of 41 °C which is 7 °C more stable than the heterotrimer from A-cx B-Ala and C-Ala. The A-5 assembly also oligomerized more effectively than the A-cx B-Ala and C-Ala assembly (Figure 9 C), indicating that just one additional amino acid results in a substantial improvement of self-assembly. Further increasing the length of the non-collagenous domain moderately increases the thermal stability of the assembly.

CONCLUSION

In this work we have begun a dissection of the self-assembly mechanism of the defense collagen protein C1q. We found that well controlled assembly of C1q can be maintained despite elimination of the majority of the native protein. Specifically, the stem region of C1q stretching from the N-terminus to the Xaa-Yaa-Gly discontinuity is sufficient to allow well controlled self-assembly of an octadecamer. Self-assembly does not occur to an appreciable extent in the absence of any of the three peptides supporting the requirement of the formation of an ABC heterotrimeric helix. Disulfide bonds are not required for the

formation or stabilization of this system, nor are posttranslational modifications other than hydroxyproline. Cryo-EM reveals a beautiful crown-like, hollow, hexameric assembly that perfectly matches expectations from SEC and AUC. Individual peptide strands are able to be visualized that form the characteristic triple helix and six of these helices form a super coil around a large hollow pore. The overall structure resembles a crown in that the triple helices are tightly packed on one end and open up on the other changing the pore diameter from approximately 18 to 30 Å. In two of the three peptides there is a short sequence which does not conform to the Xaa-Yaa-Gly pattern at the N-terminus. While oligomerization is maintained when both of these regions are deleted, the stability and fraction of oligomer formed is substantially reduced without them. This appears to be primarily driven by the N-terminal sequence associated with C1q-A. Self-assembly of the C1q stem peptides into the final octadecamer is a very slow process requiring approximately one month of folding. The rate limiting step appears to be triple helix formation which then leads to a series of higher order oligomers, with increasing stability, that ultimately terminates with the final formation of the $(ABC)_6$ octadecamer.

This study helps to dissect the mechanisms of self-assembly of this critical protein family involved in innate immunity and is also the first example of a triple helical peptide which can self-assemble into a discrete higher order structure without significant polydispersity. Many of the previous structural studies looking at collagen and collagen-like higher order assemblies have been based upon very low resolution TEM studies[66] which provide few insights into the molecular packing of triple helices. Additionally, most of the structural models for collagen packing are based on X-ray fiber diffraction data combined with X-ray crystallography.[67,68] The crystallographic studies do not necessarily reflect the solution state of these macromolecules and the fiber diffraction data can be consistent with many different models.[69] The structure described in this paper is a critical first step in understanding the molecular interactions between collagen triple helices. We expect that these results will lay the groundwork for the design and synthesis of *de novo* designed triple helical bundles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

Le Tracy Yu thanks Caroline M. Peterson and Adam C. Farsheed for their help on instrument training. The cryo-EM data was collected at the Molecular Electron Microscopy Core (MEMC) at the University of Virginia. We would like to thank Dr. Michael Purdy at the MEMC for his assistance with the imaging,

Funding Sources

This work was supported by the NSF CHE grant number 2203937 (awarded to J.D.H.) and NIH GM122510 (to E.H.E.). AUC analysis was supported by the Canada 150 Research Chairs program (C150-2017-00015), the Canada Foundation for Innovation (CFI-37589), the National Institutes of Health (1R01GM120600) and the Canadian Natural Science and Engineering Research Council (DG-RGPIN-2019-05637). UltraScan supercomputer calculations were supported through NSF/XSEDE grant TG-MCB070039N, and University of Texas grant TG457201. (awarded to B.D.) The Canadian Natural Science and Engineering Research Council support AH through a scholarship grant.

Data Availability

Cryo-EM density map for the $(ABC\text{-}Ala)_6$ assembly is available from the Electron Microscopy Data Bank with accession code EMD-28960.

REFERENCES

(1). Boudko SP; Bächinger HP Structural Insight for Chain Selection and Stagger Control in Collagen. Sci Rep 2016, 6 (1), 37831. 10.1038/srep37831. [PubMed: 27897211]

(2). Kramer RZ; Venugopal MG; Bella J; Mayville P; Brodsky B; Berman HM Staggered Molecular Packing in Crystals of a Collagen-like Peptide with a Single Charged Pair11Edited by I. A. Wilson. Journal of Molecular Biology 2000, 301 (5), 1191–1205. 10.1006/jmbi.2000.4017. [PubMed: 10966815]

(3). Kramer RZ; Bella J; Brodsky B; Berman HM The Crystal and Molecular Structure of a Collagen-like Peptide with a Biologically Relevant Sequence. J Mol Biol 2001, 311 (1), 131–147. 10.1006/jmbi.2001.4849. [PubMed: 11469863]

(4). Shoulders MD; Raines RT Collagen Structure and Stability. Annual Review of Biochemistry 2009, 78 (1), 929–958. 10.1146/annurev.biochem.77.032207.120833.

(5). Fields B, G. Synthesis and Biological Applications of Collagen -Model Triple-Helical Peptides. Organic & Biomolecular Chemistry 2010, 8 (6), 1237–1258. 10.1039/B920670A. [PubMed: 20204190]

(6). Bretscher LE; Jenkins CL; Taylor KM; DeRider ML; Raines RT Conformational Stability of Collagen Relies on a Stereoelectronic Effect. J. Am. Chem. Soc. 2001, 123 (4), 777–778. 10.1021/ja005542v. [PubMed: 11456609]

(7). Erdmann RS; Wennemers H Importance of Ring Puckering versus Interstrand Hydrogen Bonds for the Conformational Stability of Collagen. Angewandte Chemie 2011, 123 (30), 6967–6970. 10.1002/ange.201008118.

(8). Egli J; Schnitzer T; Dietschreit JCB; Ochsenfeld C; Wennemers H Why Proline? Influence of Ring-Size on the Collagen Triple Helix. Org. Lett. 2020, 22 (2), 348–351. 10.1021/acs.orglett.9b03528. [PubMed: 31682124]

(9). Jalan AA; Sammon D; Hartgerink JD; Brear P; Stott K; Hamaia SW; Hunter EJ; Walker DR; Leitinger B; Farndale RW Chain Alignment of Collagen I Deciphered Using Computationally Designed Heterotrimers. Nat Chem Biol 2020, 16 (4), 423–429. 10.1038/s41589-019-0435-y. [PubMed: 31907373]

(10). Walker DR; Hulgan SAH; Peterson CM; Li I-C; Gonzalez KJ; Hartgerink JD Predicting the Stability of Homotrimeric and Heterotrimeric Collagen Helices. Nat. Chem. 2021, 13 (3), 260–269. 10.1038/s41557-020-00626-6. [PubMed: 33589786]

(11). Walker DR; Alizadehmojarad AA; Kolomeisky AB; Hartgerink JD Charge-Free, Stabilizing Amide−π Interactions Can Be Used to Control Collagen Triple-Helix Self-Assembly. Biomacromolecules 2021, 22 (5), 2137–2147. 10.1021/acs.biomac.1c00234. [PubMed: 33881314]

(12). Jalan AA; Hartgerink JD Pairwise Interactions in Collagen and the Design of Heterotrimeric Helices. Current Opinion in Chemical Biology 2013, 17 (6), 960–967. 10.1016/j.cbpa.2013.10.019. [PubMed: 24252327]

(13). Jalan AA; Demeler B; Hartgerink JD Hydroxyproline-Free Single Composition ABC Collagen Heterotrimer. J. Am. Chem. Soc. 2013, 135 (16), 6014–6017. 10.1021/ja402187t. [PubMed: 23574286]

(14). Persikov AV; Ramshaw JAM; Kirkpatrick A; Brodsky B Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability. Biochemistry 2005, 44 (5), 1414–1422. 10.1021/bi048216r. [PubMed: 15683226]

(15). Egli J; Siebler C; Köhler M; Zenobi R; Wennemers H Hydrophobic Moieties Bestow Fast-Folding and Hyperstability on Collagen Triple Helices. J. Am. Chem. Soc. 2019, 141 (14), 5607–5611. 10.1021/jacs.8b13871. [PubMed: 30920819]

(16). Erdmann RS; Wennemers H Functionalizable Collagen Model Peptides. J. Am. Chem. Soc. 2010, 132 (40), 13957–13959. 10.1021/ja103392t. [PubMed: 20849115]

(17). Holmgren SK; Bretscher LE; Taylor KM; Raines RT A Hyperstable Collagen Mimic. Chemistry & Biology 1999, 6 (2), 63–70. 10.1016/S1074-5521(99)80003-9. [PubMed: 10021421]

(18). Ottl J; Battistuta R; Pieper M; Tschesche H; Bode W; Kühn K; Moroder L Design and Synthesis of Heterotrimeric Collagen Peptides with a Built-in Cystine-Knot Models for Collagen Catabolism by Matrix-Metalloproteases. FEBS Letters 1996, 398 (1), 31–36. 10.1016/S0014-5793(96)01212-4. [PubMed: 8946948]

(19). Tanrikulu IC; Westler WM; Ellison AJ; Markley JL; Raines RT Templated Collagen “Double Helices” Maintain Their Structure. J. Am. Chem. Soc. 2020, 142 (3), 1137–1141. 10.1021/jacs.9b07583. [PubMed: 31895554]

(20). Saccá B; Barth D; Musiol H-J; Moroder L Conformation-Dependent Side Reactions in Interstrand-Disulfide Bridging of Trimeric Collagenous Peptides by Regioselective Cysteine Chemistry. Journal of Peptide Science 2002, 8 (5), 205–210. 10.1002/psc.386. [PubMed: 12043995]

(21). Saccà B; Renner C; Moroder L The Chain Register in Heterotrimeric Collagen Peptides Affects Triple Helix Stability and Folding Kinetics. Journal of Molecular Biology 2002, 324 (2), 309–318. 10.1016/S0022-2836(02)01065-3. [PubMed: 12441109]

(22). Gelman RA; Williams BR; Piez KA Collagen Fibril Formation. Evidence for a Multistep Process. Journal of Biological Chemistry 1979, 254 (1), 180–186. 10.1016/S0021-9258(17)30289-2. [PubMed: 758319]

(23). Orgel JPRO; Irving TC; Miller A; Wess TJ Microfibrillar Structure of Type I Collagen in Situ. Proceedings of the National Academy of Sciences 2006, 103 (24), 9001–9005. 10.1073/pnas.0502718103.

(24). 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. Nature Chem 2011, 3 (10), 821–828. 10.1038/nchem.1123. [PubMed: 21941256]

(25). Rele S; Song Y; Apkarian RP; Qu Z; Conticello VP; Chaikof EL D-Periodic Collagen-Mimetic Microfibers. J. Am. Chem. Soc. 2007, 129 (47), 14780–14787. 10.1021/ja0758990. [PubMed: 17985903]

(26). Tanrikulu IC; Forticaux A; Jin S; Raines RT Peptide Tessellation Yields Micrometre-Scale Collagen Triple Helices. Nature Chem 2016, 8 (11), 1008–1014. 10.1038/nchem.2556. [PubMed: 27768103]

(27). Merg AD; Touponse G; Genderen E. van; Blum TB; Zuo X; Bazrafshan A; Siaw HMH; McCanna A; Brian Dyer R; Salaita K; Abrahams JP; Conticello VP Shape-Shifting Peptide Nanomaterials: Surface Asymmetry Enables PH-Dependent Formation and Interconversion of Collagen Tubes and Sheets. J. Am. Chem. Soc. 2020, 142 (47), 19956–19968. 10.1021/jacs.0c08174. [PubMed: 33170675]

(28). Przybyla DE; Chmielewski J Higher-Order Assembly of Collagen Peptides into Nano- and Microscale Materials. Biochemistry 2010, 49 (21), 4411–4419. 10.1021/bi902129p. [PubMed: 20415447]

(29). Pires MM; Chmielewski J Self-Assembly of Collagen Peptides into Microflorettes via Metal Coordination. J. Am. Chem. Soc. 2009, 131 (7), 2706–2712. 10.1021/ja8088845. [PubMed: 19182901]

(30). Hernandez-Gordillo V; Chmielewski J Mimicking the Extracellular Matrix with Functionalized, Metal-Assembled Collagen Peptide Scaffolds. Biomaterials 2014, 35 (26), 7363–7373. 10.1016/j.biomaterials.2014.05.019. [PubMed: 24933513]

(31). Przybyla DE; Rubert Pérez CM; Gleaton J; Nandwana V; Chmielewski J Hierarchical Assembly of Collagen Peptide Triple Helices into Curved Disks and Metal Ion-Promoted Hollow Spheres. J. Am. Chem. Soc. 2013, 135 (9), 3418–3422. 10.1021/ja307651e. [PubMed: 23402552]

(32). Knapinska AM; Tokmina-Roszyk D; Amar S; Tokmina-Roszyk M; Mochalin VN; Gogotsi Y; Cosme P; Terentis AC; Fields GB Solid-Phase Synthesis, Characterization, and Cellular Activities of Collagen-Model Nanodiamond-Peptide Conjugates. Peptide Science 2015, 104 (3), 186–195. 10.1002/bip.22636. [PubMed: 25753561]

(33). Wood CW; Woolfson DN CCBuilder 2.0: Powerful and Accessible Coiled-Coil Modeling. Protein Science 2018, 27 (1), 103–111. 10.1002/pro.3279. [PubMed: 28836317]

(34). Thomas F; Dawson WM; Lang EJM; Burton AJ; Bartlett GJ; Rhys GG; Mulholland AJ; Woolfson DN De Novo-Designed α-Helical Barrels as Receptors for Small Molecules. ACS Synth. Biol. 2018, 7 (7), 1808–1816. 10.1021/acssynbio.8b00225. [PubMed: 29944338]

(35). Rhys GG; Wood CW; Beesley JL; Zaccai NR; Burton AJ; Brady RL; Thomson AR; Woolfson DN Navigating the Structural Landscape of De Novo α-Helical Bundles. J. Am. Chem. Soc. 2019, 141 (22), 8787–8797. 10.1021/jacs.8b13354. [PubMed: 31066556]

(36). Beesley JL; Woolfson DN The de Novo Design of α-Helical Peptides for Supramolecular Self-Assembly. Current Opinion in Biotechnology 2019, 58, 175–182. 10.1016/j.copbio.2019.03.017. [PubMed: 31039508]

(37). Thielens NM; Tedesco F; Bohlson SS; Gaboriaud C; Tenner AJ C1q: A Fresh Look upon an Old Molecule. Molecular Immunology 2017, 89, 73–83. 10.1016/j.molimm.2017.05.025. [PubMed: 28601358]

(38). Reid KBM Complement Component C1q: Historical Perspective of a Functionally Versatile, and Structurally Unusual, Serum Protein. Frontiers in Immunology 2018, 9, 764. 10.3389/fimmu.2018.00764. [PubMed: 29692784]

(39). Ip WKE; Takahashi K; Ezekowitz RA; Stuart LM Mannose-Binding Lectin and Innate Immunity. Immunological Reviews 2009, 230 (1), 9–21. 10.1111/j.1600-065X.2009.00789.x. [PubMed: 19594626]

(40). Jensenius H; Klein DCG; van Hecke M; Oosterkamp TH; Schmidt T; Jensenius JC Mannan-Binding Lectin: Structure, Oligomerization, and Flexibility Studied by Atomic Force Microscopy. Journal of Molecular Biology 2009, 391 (1), 246–259. 10.1016/j.jmb.2009.05.083. [PubMed: 19501100]

(41). Haagsman HP; White RT; Schilling J; Lau K; Benson BJ; Golden J; Hawgood S; Clements JA Studies of the Structure of Lung Surfactant Protein SP-A. American Journal of Physiology-Lung Cellular and Molecular Physiology 1989, 257 (6), L421–L429. 10.1152/ajplung.1989.257.6.L421.

(42). Tenner AJ; Robinson SL; Borchelt J; Wright JR Human Pulmonary Surfactant Protein (SP-A), a Protein Structurally Homologous to C1q, Can Enhance FcR- and CR1-Mediated Phagocytosis. Journal of Biological Chemistry 1989, 264 (23), 13923–13928. 10.1016/S0021-9258(18)80088-6. [PubMed: 2788165]

(43). McCormack FX; Damodarasamy M; Elhalwagi BM Deletion Mapping of N-Terminal Domains of Surfactant Protein A: THE N-TERMINAL SEGMENT IS REQUIRED FOR PHOSPHOLIPID AGGREGATION AND SPECIFIC INHIBITION OF SURFACTANT SECRETION *. Journal of Biological Chemistry 1999, 274 (5), 3173–3181. 10.1074/jbc.274.5.3173. [PubMed: 9915857]

(44). Tsao T-S; Tomas E; Murrey HE; Hug C; Lee DH; Ruderman NB; Heuser JE; Lodish HF Role of Disulfide Bonds in Acrp30/Adiponectin Structure and Signaling Specificity: DIFFERENT OLIGOMERS ACTIVATE DIFFERENT SIGNAL TRANSDUCTION PATHWAYS *. Journal of Biological Chemistry 2003, 278 (50), 50810–50817. 10.1074/jbc.M309469200. [PubMed: 14522956]

(45). Briggs DB; Jones CM; Mashalidis EH; Nuñez M; Hausrath AC; Wysocki VH; Tsao T-S Disulfide-Dependent Self-Assembly of Adiponectin Octadecamers from Trimers and Presence of Stable Octadecameric Adiponectin Lacking Disulfide Bonds In Vitro. Biochemistry 2009, 48 (51), 12345–12357. 10.1021/bi9015555. [PubMed: 19943704]

(46). Suzuki S; Wilson-Kubalek EM; Wert D; Tsao T-S; Lee DH The Oligomeric Structure of High Molecular Weight Adiponectin. FEBS Letters 2007, 581 (5), 809–814. 10.1016/j.febslet.2007.01.046. [PubMed: 17292892]

(47). Svehag S-E; Manhem L; Bloth B Ultrastructure of Human C1q Protein. Nature New Biology 1972, 238 (82), 117–118. 10.1038/newbio238117a0. [PubMed: 4625706]

(48). Strang CJ; Siegel RC; Phillips ML; Poon PH; Schumaker VN Ultrastructure of the First Component of Human Complement: Electron Microscopy of the Crosslinked Complex. Proc Natl Acad Sci U S A 1982, 79 (2), 586–590. [PubMed: 6952210]

(49). 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. Journal of Molecular Biology 1985, 186 (2), 403–415. 10.1016/0022-2836(85)90114-7. [PubMed: 3878889]

(50). Gaboriaud C; Frachet P; Thielens N; Arlaud G The Human C1q Globular Domain: Structure and Recognition of Non-Immune Self Ligands. Frontiers in Immunology 2012, 2, 92. 10.3389/fimmu.2011.00092. [PubMed: 22566881]

(51). Reid KB; Porter RR Subunit Composition and Structure of Subcomponent C1q of the First Component of Human Complement. Biochem J 1976, 155 (1), 19–23. [PubMed: 938474]

(52). Yonemasu K; Stroud RM Structural Studies on Human Clq: Non-Covalent and Covalent Subunits. Immunochemistry 1972, 9 (5), 545–554. 10.1016/0019-2791(72)90064-X. [PubMed: 4113858]

(53). Reid KB Isolation, by Partial Pepsin Digestion, of the Three Collagen-like Regions Present in Subcomponent Clq of the First Component of Human Complement. Biochem J 1976, 155 (1), 5–17. [PubMed: 7240]

(54). Reid KB Complete Amino Acid Sequences of the Three Collagen-like Regions Present in Subcomponent C1q of the First Component of Human Complement. Biochem J 1979, 179 (2), 367–371. [PubMed: 486087]

(55). Tissot B; Gonnet F; Iborra A; Berthou C; Thielens N; Arlaud GJ; Daniel R Mass Spectrometry Analysis of the Oligomeric C1q Protein Reveals the B Chain as the Target of Trypsin Cleavage and Interaction with Fucoidan. Biochemistry 2005, 44 (7), 2602–2609. 10.1021/bi047802h. [PubMed: 15709773]

(56). Pflieger D; Przybylski C; Gonnet F; Caer J-PL; Lunardi T; Arlaud GJ; Daniel R Analysis of Human C1q by Combined Bottom-up and Top-down Mass Spectrometry: DETAILED MAPPING OF POST-TRANSLATIONAL MODIFICATIONS AND INSIGHTS INTO THE C1R/C1S BINDING SITES *. Molecular & Cellular Proteomics 2010, 9 (4), 593–610. 10.1074/mcp.M900350-MCP200. [PubMed: 20008834]

(57). Fouët G; Bally I; Signor L; Häußermann K; Thielens NM; Rossi V; Gaboriaud C Headless C1q: A New Molecular Tool to Decipher Its Collagen-like Functions. The FEBS Journal 2021, 288 (6), 2030–2041. 10.1111/febs.15543. [PubMed: 32869492]

(58). The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. Nucleic Acids Research 2023, 51 (D1), D523–D531. 10.1093/nar/gkac1052. [PubMed: 36408920]

(59). Merrifield RB Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. J. Am. Chem. Soc. 1963, 85 (14), 2149–2154. 10.1021/ja00897a025.

(60). Fiala T; Barros EP; Heeb R; Riniker S; Wennemers H Predicting Collagen Triple Helix Stability through Additive Effects of Terminal Residues and Caps. Angewandte Chemie International Edition n/a (n/a), e202214728. 10.1002/anie.202214728.

(61). Fiala T; Barros EP; Ebert M-O; Ruijsenaars E; Riniker S; Wennemers H Frame Shifts Affect the Stability of Collagen Triple Helices. J. Am. Chem. Soc. 2022, 144 (40), 18642–18649. 10.1021/jacs.2c08727. [PubMed: 36179150]

(62). Qi Y; Zhou D; Kessler JL; Qiu R; Yu SM; Li G; Qin Z; Li Y Terminal Repeats Impact Collagen Triple-Helix Stability through Hydrogen Bonding. Chem. Sci. 2022, 13 (42), 12567–12576. 10.1039/D2SC03666E. [PubMed: 36382282]

(63). Brodsky-Doyle B; Leonard KR; Reid KBM Circular-Dichroism and Electron-Microscopy Studies of Human Subcomponent C1q before and after Limited Proteolysis by Pepsin. Biochemical Journal 1976, 159 (2), 279–286. 10.1042/bj1590279. [PubMed: 187173]

(64). Tischenko VM; Ichtchenko AM; Andreyev CV; Kajava AV Thermodynamic Studies of the Collagen-like Region of Human Subcomponent C1q: A Water-Containing Structural Model. Journal of Molecular Biology 1993, 234 (3), 654–660. 10.1006/jmbi.1993.1618. [PubMed: 8254665]

(65). Shelton E; Yonemasu K; Stroud RM Ultrastructure of the Human Complement Component, Clq. Proceedings of the National Academy of Sciences 1972, 69 (1), 65–68. 10.1073/pnas.69.1.65.

(66). Meadows RS; Holmes DF; Gilpin CJ; Kadler KE Electron Cryomicroscopy of Fibrillar Collagens. In Extracellular Matrix Protocols; Streuli CH, Grant ME., Eds.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2000; pp 95–109. 10.1385/1-59259-063-2:95.

(67). Kramer RZ; Bella J; Brodsky B; Berman HM The Crystal and Molecular Structure of a Collagen-like Peptide with A Biologically Relevant Sequence12Edited by I. A. Wilson. Journal of Molecular Biology 2001, 311 (1), 131–147. 10.1006/jmbi.2001.4849. [PubMed: 11469863]

(68). Bella J; Eaton M; Brodsky B; Berman HM Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. Science 1994, 266 (5182), 75–81. 10.1126/science.7695699. [PubMed: 7695699]

(69). Pieri L; Wang F; Arteni A-A; Vos M; Winter J-M; Le Du M-H; Artzner F; Gobeaux F; Legrand P; Boulard Y; Bressanelli S; Egelman EH; Paternostre M Atomic Structure of Lanreotide Nanotubes Revealed by Cryo-EM. Proceedings of the National Academy of Sciences 2022, 119 (4), e2120346119. 10.1073/pnas.2120346119.

Synopsis

Collagen undergoes a hierarchical assembly starting from peptides to form a triple helix. These triple helices can pack into fibril bundles and subsequently into even larger structures. Despite the ubiquity of this multi-step assembly in nature, very few synthetic systems have ever demonstrated self-assembly beyond the triple helix. Those that have, generate large polydisperse structures. Here for the first time, we describe the self-assembly of a collagen mimetic system, based on complement protein C1q, which self-assembles into a discrete $(ABC)_6$ octadecamer of collagen-like peptides which are arranged in a hollow, crown-like structure. This work is important for enhancing our understanding of this ubiquitous class of proteins and also sets the stage for future generations of de novo designed triple helical bundles.

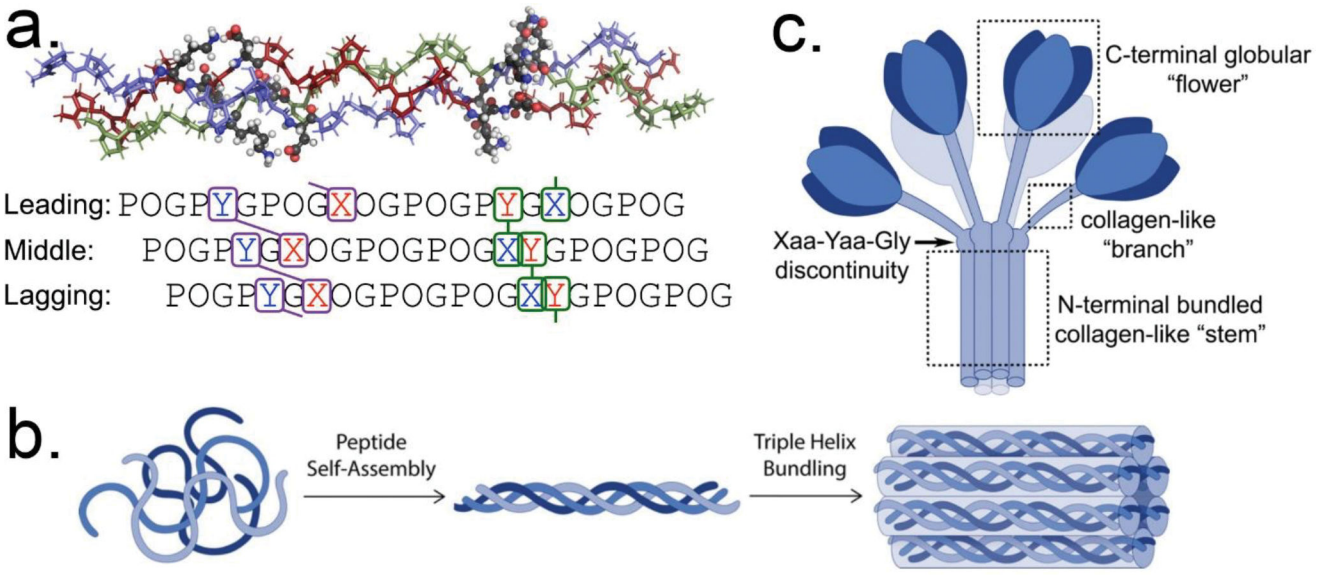


Figure 1.

a) Molecular model and sequence of a collagen triple helix highlighting possible axial (purple lassos) and lateral (green lassos) pairwise interactions. “X” and “Y” indicate amino acids in the Xaa or Yaa position of the Xaa-Yaa-Gly collagen repeat. Letters shown in blue are frequently positively charged amino acids while letters in red are frequently negatively charged amino acids. b) Depiction of peptide self-assembly to triple helices followed by triple helical bundling. c) Cartoon of the protein C1q which is one of several structurally related “defense collagens” that play important roles in innate immunity. The stem of C1q is composed of sequences which closely match the requirement of a collagen triple helix.

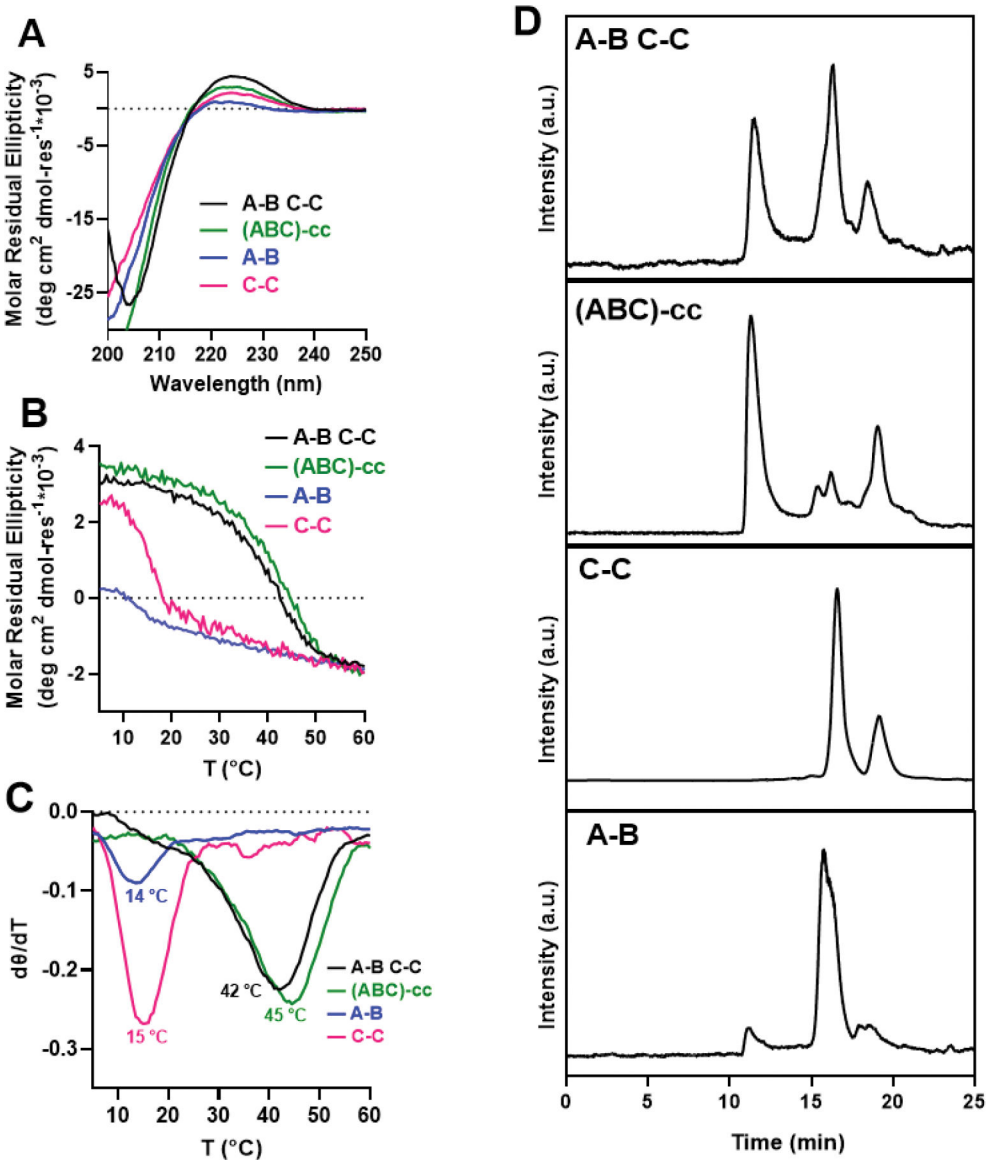


Figure 2.

Circular Dichroism (CD) and Size Exclusion Chromatography (SEC) of the stem region peptides. The CD spectra measurement was performed with 0.3 mM peptide in 10 mM tris·HCl buffer, with a wavelength scanning from 200–250 nm at 5 °C. The melting curves were collected from 5 °C to 60 °C with a heating rate of 10 °C/hour at the wavelength of the maximum MRE value of each sample. (A) CD spectra of disulfide bonded dimeric peptides and mixtures of these dimers. (B) CD melting curve of the same samples. (C) First-order derivative of the melting curves shown in B. (D) Size-exclusion chromatography of disulfide bonded peptide dimers and their mixtures.

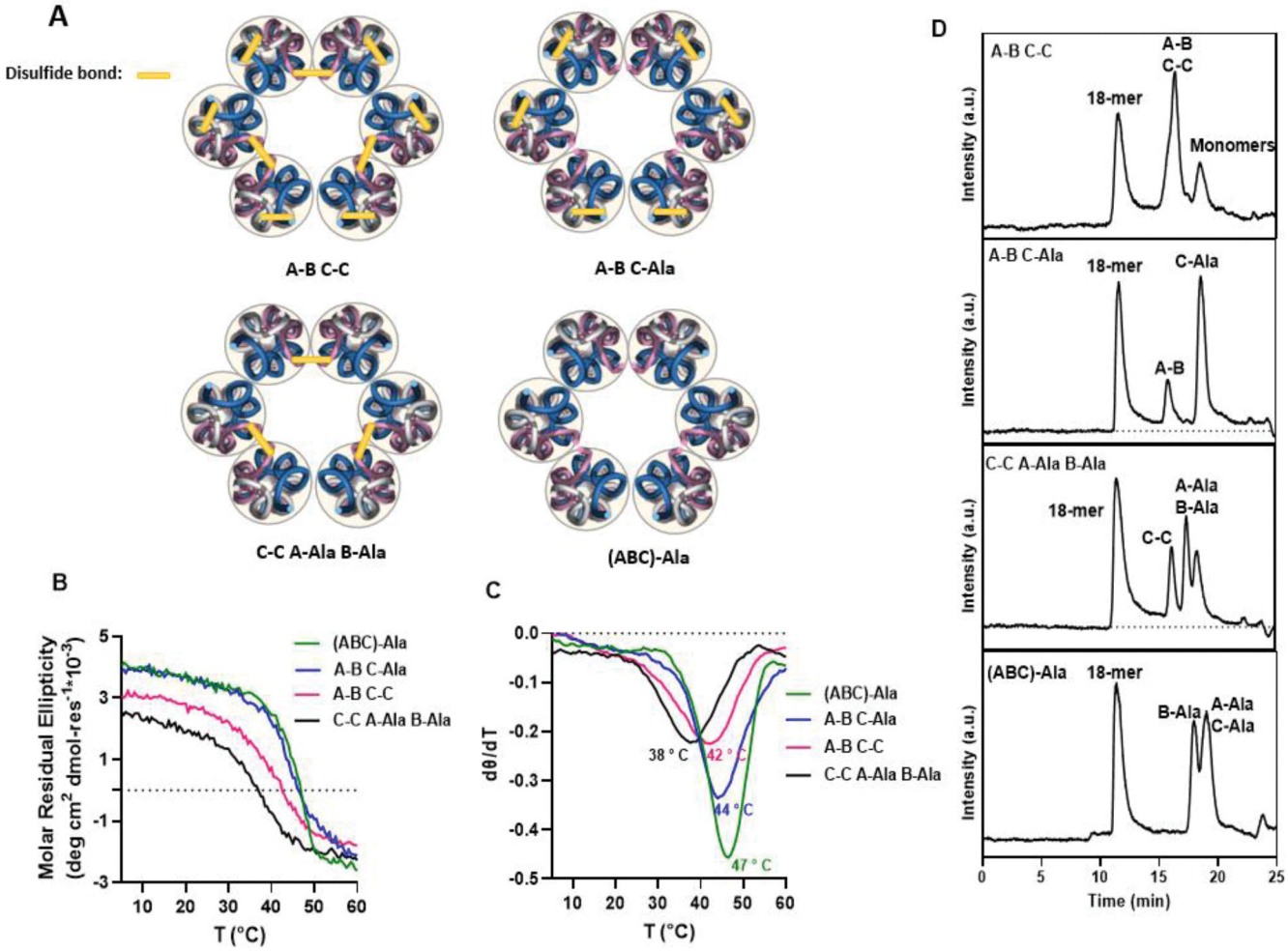


Figure 3.

(A) Schematic depiction of covalent and supramolecular assemblies of the octadecameric assembly of C1q stem peptides indicating the four different disulfide bonding patterns generated. (B) CD melting data. The melting curves were collected from 5 °C to 60 °C with a heating rate of 10 °C/hour at the wavelength that gives the maximum MRE value of each sample, sample concentration is 0.3 mM peptide in 10 mM tris·HCl buffer. (C) The first-order derivatives of the CD melting results. (D) SEC of assemblies with different disulfide bonding patterns. All CD and SEC data were collected after 60 days of equilibration. All SEC chromatographs were monitored at 220nm.

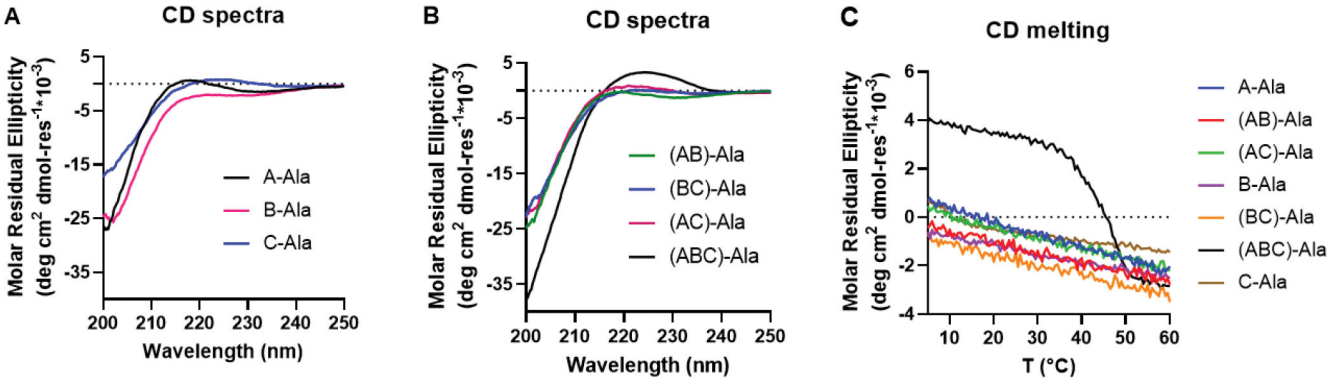


Figure 4.

Characterization of the supramolecular (ABC)-Ala assembly composition. (A) CD spectra of monomeric peptide solutions. (B) CD spectra of peptide mixture solutions. (C) CD melting curves of monomeric peptides and peptide mixtures. The melting curves were collected from 5 °C to 60 °C with a heating rate of 10 °C/hour at the wavelength that gives the maximum MRE value of each sample, sample concentration is 0.3 mM peptide in 10 mM tris·HCl buffer. Only the complete ABC system shows significant triple helical character.

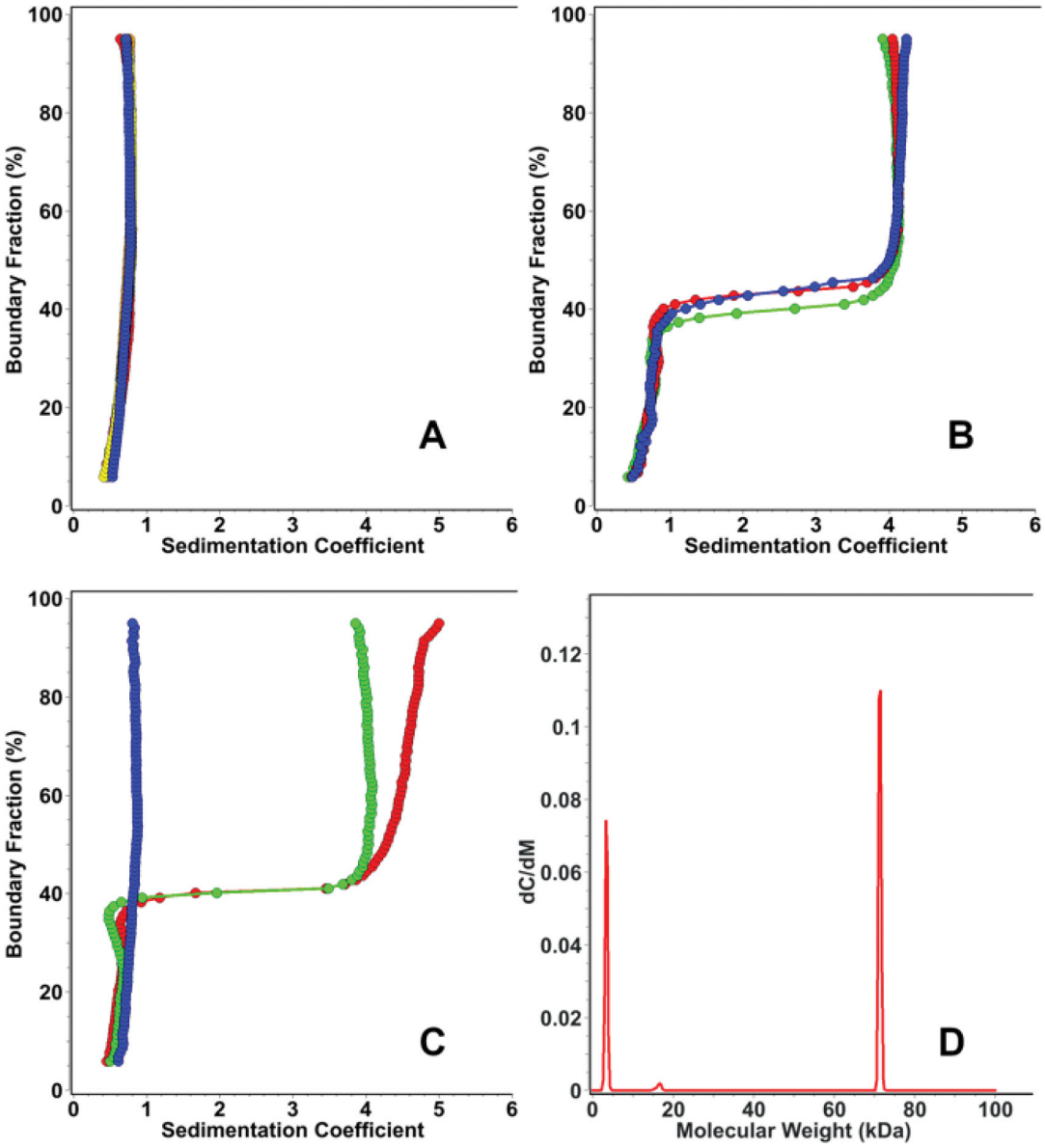


Figure 5.

Sedimentation velocity analysis of A-Ala, B-Ala, C-Ala peptides. (A) Individual peptides and mixtures of each pair of peptides in 150mM NaCl (red: peptide A-Ala alone, orange: peptide B-Ala alone, magenta: peptide C-Ala alone, cyan: peptide mixture of A-Ala and B-Ala, yellow: peptide mixture of A-Ala and C-Ala, blue: peptide mixture of B-Ala and C-Ala). All peptides have approximately the same molar mass and sedimentation behavior and therefore are indistinguishable. Peptide pairs do not associate to form higher order structures. (B) three different concentrations of all three peptides mixed together in 150mM NaCl (187μM at 276nm (green), 35μM at 230nm (red) and 14μM at 220nm (blue). (C) A mixture of all three peptides in 150mM NaCl (green), 0 mM NaCl (red), and 150mM NaCl after heat treatment (In detail, the three-peptide mixture was diluted and heated in a 65 °C water bath for 20 mins to fully denature the assembly and then cool to room temperature for two hours before measurements.) (blue)). (D) Representative molar mass distribution for

s-value distribution corresponding to the red trace (35 μM concentration) in panel B., molar mass data was determined by genetic algorithm – Monte Carlo analysis.

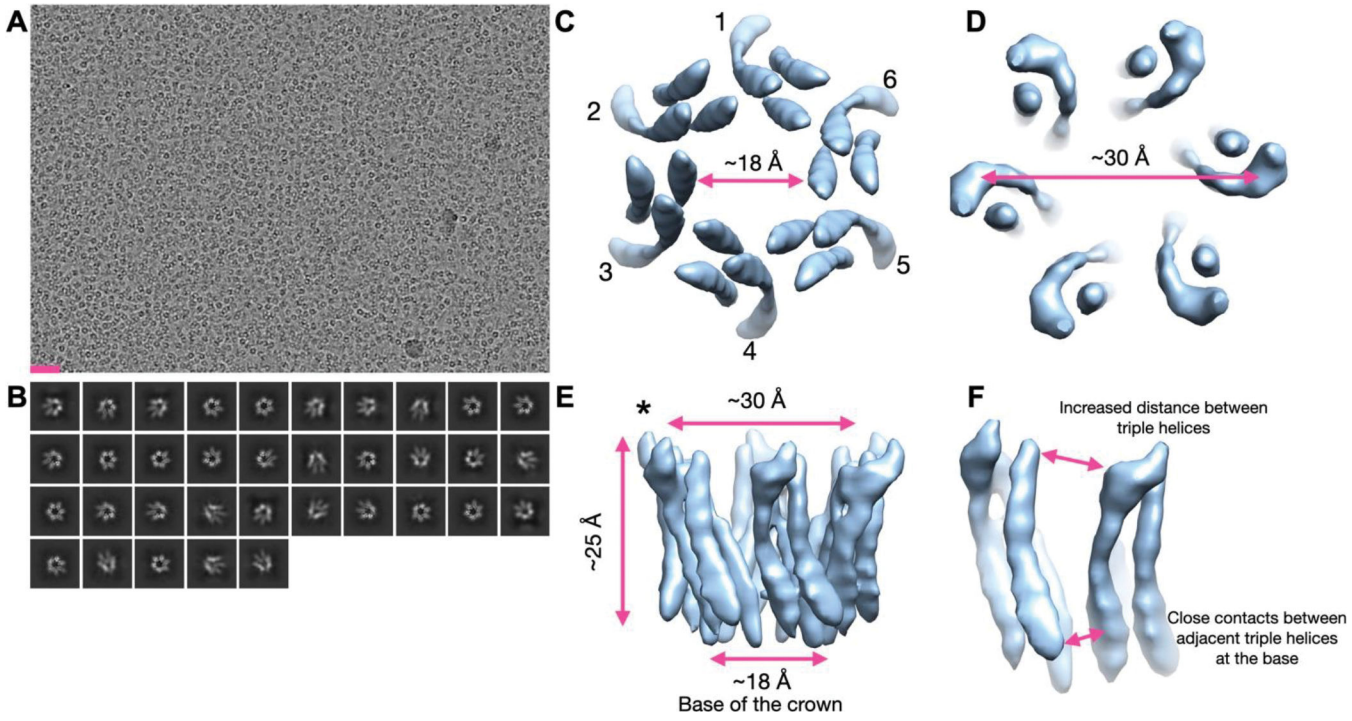


Figure 6.

Cryo-EM characterization of $(ABC\text{-}Ala)_6$. (A) Raw micrograph of $(ABC\text{-}Ala)_6$ particles. Scale bar is 20 nm. (B) 2D class averages of the $(ABC\text{-}Ala)_6$ particles. (C-F) Reconstructed images of the assembly. (B-C) Show the assembly from the "bottom" and "top" respectively illustrating the change in pore diameter which increases from approximately 18 to 30 Å. (E) Side view of the assembly with approximate dimensions indicated. (F) Zoomed side view showing the interface between adjacent triple helices. Strands in the same triple helix are always in close proximity to each other while adjacent triple helices make close contacts at the base of the crown and move further apart from each other near the top.

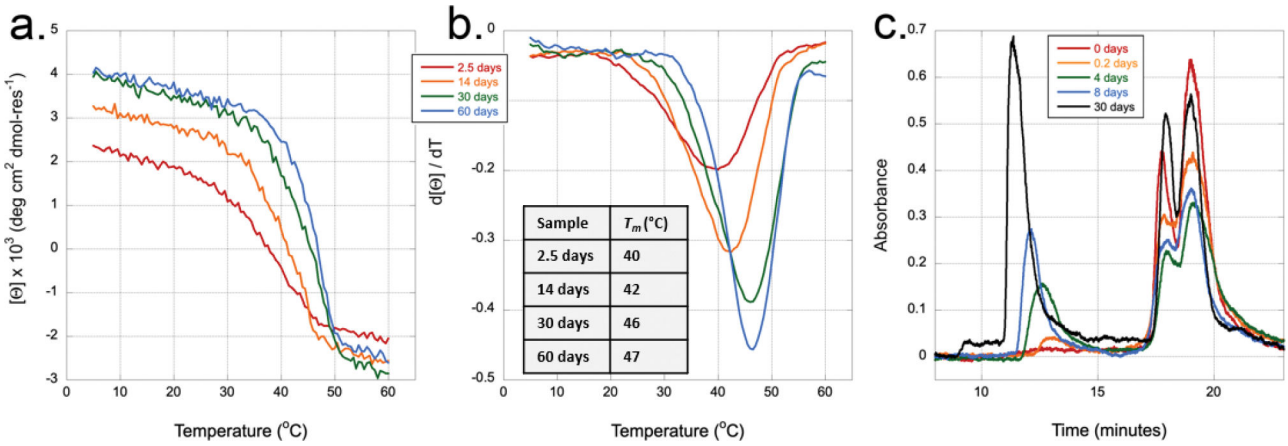


Figure 7.

Extended time course of $(ABC\text{-}Ala)_6$ folding. (a) CD melting data: signal was monitored of samples with a concentration of 0.3 mM peptide in 10 mM tris·HCl buffer at 225nm versus time with a heating rate of 10 °C/hour, and (b) the derivative of data presented in (a). Over time MRE and thermal stability increase reaching a plateau after 30 days. (c) SEC from zero to 30 days shows the slow increase in the proportion of oligomers as well as their increase in apparent mass over time.

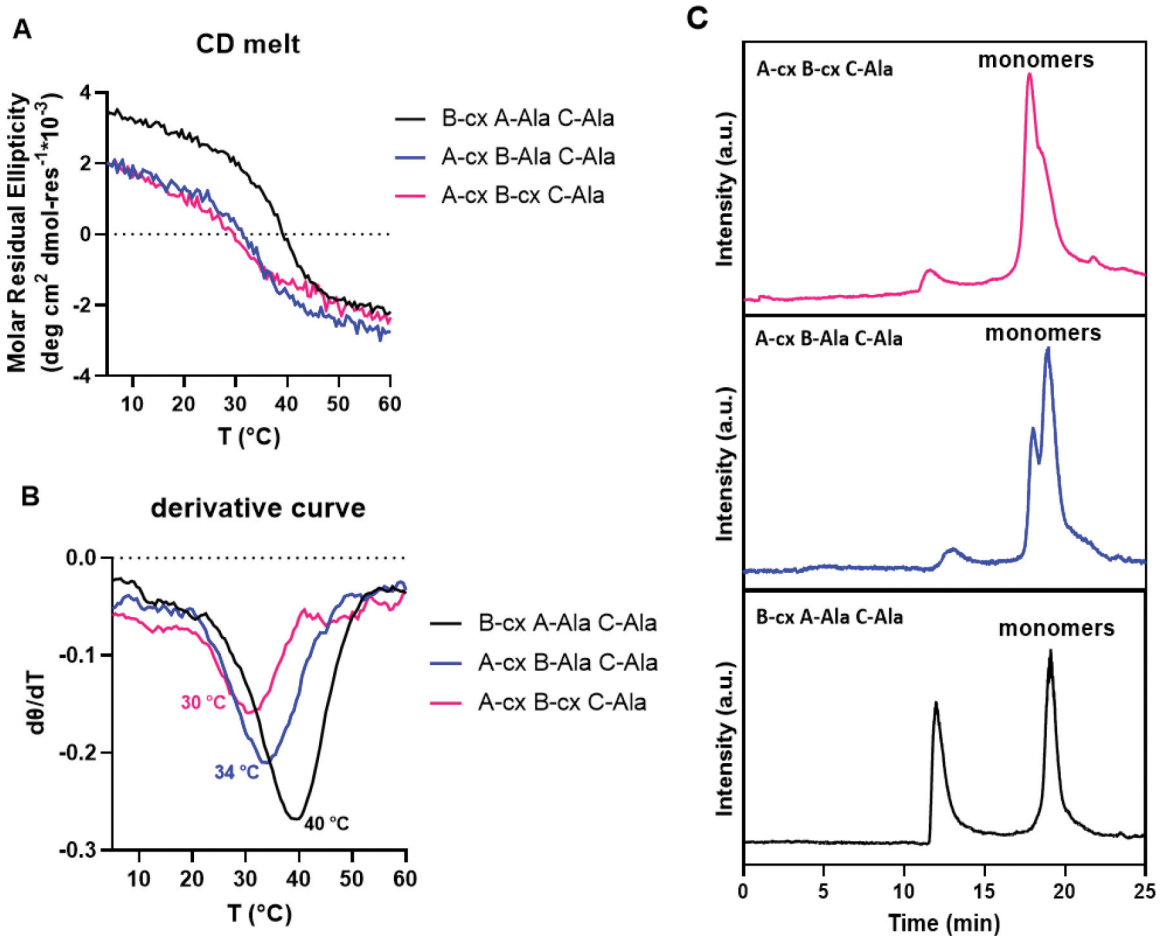


Figure 8.

CD and SEC characterizations of self-assembled samples with varied N-terminal domains. (A) CD melting curves of supramolecular heterotrimers with or without the N-terminal fragment of A-Ala and B-Ala peptides, the melting curves were collected from 5 °C to 60 °C with a heating rate of 10 °C/hour at the wavelength that gives the maximum MRE value of each sample, sample concentration is 0.3 mM in 10 mM. (B) First-order derivative of the melting curves in (A). (C) SEC of assemblies with or without the N-terminal domains of A-Ala and B-Ala peptides. (red): neither N-terminal domain, (blue): only B-Ala N-terminal domain, (black): only A-Ala N-terminal domain.

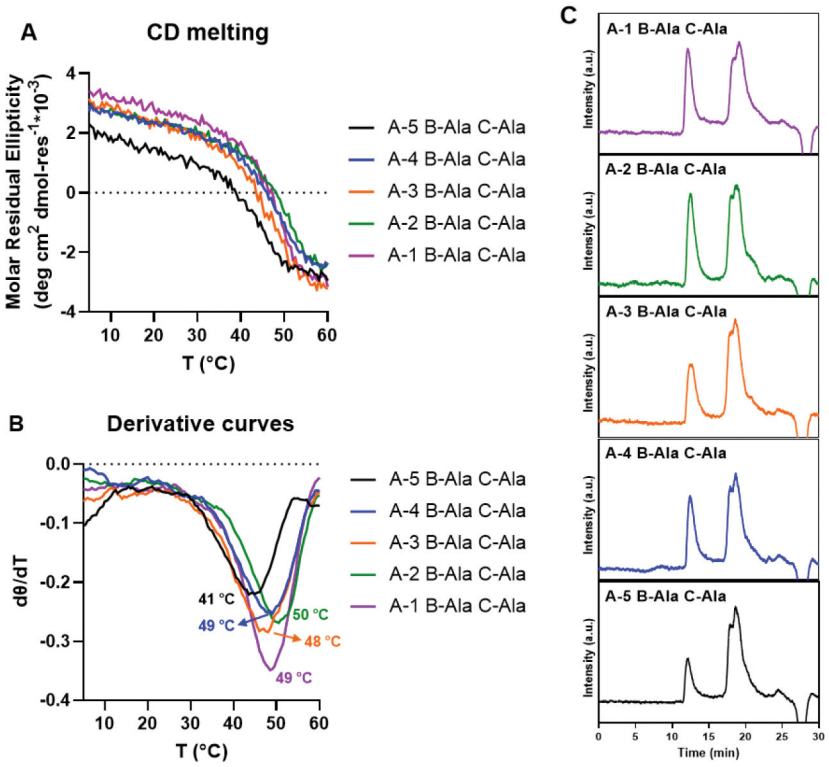


Figure 9.

CD and SEC of the self-assembled samples from the deletion mutants of peptide A and B-Ala and C-Ala. The deletion mutants have different length of the N-terminal NC domain compared to peptide A. (A) CD melting curves. The samples assembled from the deletion mutants and peptide B-Ala and C-Ala all assembled into collagen triple helices and show thermal transitions. (B) Derivative curves of the melting curves in (A). (C) SEC of the deletion mutants' assemblies. Oligomers shown in all the assemblies.

Table 1.

Names and sequences of peptides studied. “O” is (2*S*,4*R*)-Hydroxyproline. Glycines in the putative triple helical region are shown in bold. Cysteine (or Alanine mutant) is highlighted in orange and underlined. Positively charged amino acids K & R are shown in blue, negatively charged amino acids D & E are shown in red. Sequences not conforming to the canonical $(Xaa\text{-}Yaa\text{-}Gly)_n$ repeat are highlighted in grey.

“Stem” Region Peptides:	
A	EDLCRAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
B	QLSCTGPOAIO**G**IO**G**IO**G**TO**G**PD**G**QO**G**TO**G**IK**G**EK**G**LO**G**L
C	NT**G**CY**G**IO**G**MO**G**LO**G**AO**G**KD**G**YD**G**LO**G**PK**G**EP**G**IO
Cysteine to Alanine Replacement:	
A-Ala	EDLARAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
B-Ala	QLSATGPOAIO**G**IO**G**IO**G**TO**G**PD**G**QO**G**TO**G**IK**G**EK**G**LO**G**L
C-Ala	NT**G**AY**G**IO**G**MO**G**LO**G**AO**G**KD**G**YD**G**LO**G**PK**G**EP**G**IO
Collagen Exclusive Peptides:	
A-cx	PD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
B-cx	IO**G**IO**G**IO**G**TO**G**PD**G**QO**G**TO**G**IK**G**EK**G**LO**G**L
N-terminal deletion series:	
A-1	DLARAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
A-2	LARAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
A-3	ARAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
A-4	RAPD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR
A-5	APD**G**KK**G**EA**G**RO**G**RR**G**RO**G**LK**G**EQ**G**EP**G**AO**G**IR