nature portfolio

Peer Review File

Exploration of the Hierarchical Assembly Space of Collagen-like Peptides Beyond the Triple Helix

Corresponding Author: Professor Jeffrey Hartgerink

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authos made de novo design of collagen-like peptides, which were self-assembled into diverse quaternary structures, such as bundled oligomers, nanosheets, and nanoribbons. This highly novel study showed a significant advancement in the understanding higher-order assemblies of collagen-like peptides. I highly recommend publishing this manuscript after the authors address the questions and comments listed as below .
Comments:
(1) As the SPA-bcb peptide precipitated, it is not suitable for circular dichroism, please use Differential Scannining Calerometry to measure its stability (Karunakar Kar, et al Journal of Biological Chemistry 281(44): 33283-33290.)

(2) All the results are very interesting and show high novelty. However, the authors should address three questions explicitly in the conclusion:
(a) What is the role of DAA in DES-H, DES-1, DR1 and DR2 in the self-assembly?
Except for DES2-no DAA, the authors may provide experimental data of no DAA in DR1, DES-H, DES-1 peptides. Based on these data, the authors could make thorough discussion about the roles of DAA, hydrophobic and electrostatic interactions in the self-assembly.
(b) What is the role of hydrophobic interaction in the self-assembly?
DES-H with hydrophobic side chains cannot be self-assembled. However, similar hydrophobic interactions could induce self-assembly of some collagen-like peptides (Kenneth McGuinness et al, ACS Nano, 2014, 8(12) 12514–12523). DES-H has DAA while the peptides by McGuinness has no DAA. Does the DAA inhibit the self-assembly? Or maybe the locations and numbers of the hydrophobic interactions cause difference in the self-assembly?
(c) What is the role of electrostatic interactions in the self-assembly?
RD2, with the electrostatic interactions at the N- and C-termini, formed sheet-like assemblies. RD2-noDAA also formed sheet-like assemblies, which may have larger size than RD2 (Fig S29). In previous studies, peptides, with similar charged patterns but no DAA-region, also formed sheet-like assemblies (Avanish S. Parmar, et al J. Am. Chem. Soc. 2016, 138, 4362–4367; Tao Jiang et al, J. Am. Chem. Soc. 2015, 137, 7793–7802). RD1 and DES-1, with the electrostatic interactions at the center, formed oligomeric assemblies. The authors may discuss the roles of both DAA and location of the electrostatic interactions in the self-assembly.
Minor issues:
(1) Typo: Page 5 1st paragraph, 8th line: the pepetide aggregation

Reviewer #2

(Remarks to the Author)
1、In the sample preparation, lacey carbon 300 mesh copper grids were used. However, the molecular weight of these proteins are relatively small, resulting in a low signal-to-noise ratio of 2D classification. In the Figures S15 and S20, the quality of 2D classification is bad. Therefore, no high resolution details can be distinguished after 3D reconstruction. Have the authors tried to use holey copper grids instead?

2、Especially for the dataset of DES-1 peptide at pH 3.5, the protein complex seems to well match the helical symmetry. Maybe the authors could try helical reconstruction in cryoSPRAC to increase the final resolution. Considering the shape, preferred orientation would be quite severe if the authors used conventional single particle reconstruction.
3. A cryo-electron microscopy (cryo-EM) data processing workflow is required to demonstrate how the data was processed. The workflow should include additional information, such as the Gold Standard Fourier Shell Correlation (GSFSC) curve, a 3D Fourier Shell Correlation (FSC) histogram, and the local resolution for each cryo-EM map.
4. “For the SPA-human hexamers, DES-1 pentamers, and DES-1 hexamers 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.”
Given the relatively low resolution of the cryo-EM maps, the application of C5 or C6 symmetry might have led to artificial results.
5. In the legend of Figure S15b, “cry-EM” should be replaced with “cryo-EM”.

Reviewer #3

(Remarks to the Author)
The authors studied the sequence-structure relationship of assemblies formed by collagen-like peptides using the SP-A protein as a model system. Evaluation of the self-assembly of the collagen-like region of SP-A by cryo-EM and AFM identified the formation of different architectures, including oligomeric bundles, polymeric ribbons and 2D nanosheets.

Whereas the study did not provide general lessons for the design of hierarchical collagen-based structures, the insights into the underexplored self-assembly of PPII helical higher-order structures are valuable. These data will be of interest to the scientific community and specialists in the field of peptide-based supramolecular assemblies. However, several control experiments are missing, which could impact the results and their interpretation. In addition, based on a rather minimalistic data set, several conclusions are likely too general. Furthermore, one of the studied compounds is not sufficiently pure, with a possibly effect on its assembly properties.

Mutation Studies Reveal Sequence-Structure Relationship Section
1. “Even though the Ile to Pro mutant has enhanced thermal stability and favored triple helix formation, it did not promote oligomerization. This suggests that oligomerization is not correlated with triple helix formation, a critical finding."
This conclusion could be incorrect since two variables changed: the replacement of Ile by Pro and, as a consequence, the Tm. For instance, upon deleting a C-terminal POG repeat, the modified SPA-rat peptide would likely not form a stable triple helix and also not the higher-order assemblies. In contrast, an additional POG unit would stabilize the triple helix (while keeping the key Ile residue), and the system could have a higher propensity to oligomerize. Here, the conclusion would be the opposite: oligomerization correlates with triple helix formation and stability. Without such control experiments, the only valid conclusion is that the Ile residue is important for the oligomerization of SPA-rat.
Also, the term “mutation” refers to DNA, use “replacement of Ile by Pro.
Also, I would use “extended assembly” rather than “oligomerization” since oligomerization refers to covalent bonds.

2. “The point-mutation studies together revealed that charged-pair interactions involving arginine and aspartate and hydrophobic interactions may exist to favor collagen oligomerization.” This conclusion is likely too “big”, considering the rather minimalistic study.

a) The study focuses on one amino acid variations in the stem collagen region. Here, the authors replaced only Arg24 and Asp25, but there are two more Asp residues within the stem region. In addition, Asp25 is in the Yaa position, whereas Asp33 is in the Xaa position (per collagen canonical sequence). As Xaa and Yaa positions are not equal within the triple helix (Xaa position is, for example, more solvent exposed), substituting an Asp residue in the Xaa position (e.g. Asp33) could have different effect on the structure and its assembly than substituting an Asp in the Yaa position. Thus, the conclusion could change significantly, e.g., substituting Asp33 could have no impact on the triple helical or the further assembly, which would then contradict the current conclusion.

b) For the positively charged residues, the argument is the same. Lys31 is in the Yaa, compared to the mutated Xaa Arg24. Thus, to corroborate the conclusion Arg27 needs to be substituted.

Design of pH-responsive Nanoribbons
1. p.9: Initially, a physiological pH is mentioned for folding, and a few lines later, pH 6.1. On p10, the default annealing condition at pH 6.1 is mentioned. Which pH was used? Were all peptides assembled in the same buffer and at the same concentration? This is important for comparisons as the buffer and the ionic strength of the solutions can have a significant impact on self-assembling systems.

2. Fig. 4b: It would be better to show the raw melting curve rather than the CD spectrum as it carries more critical information. The CD spectrum can be moved into the SI.

3. “No higher-order assembly was observed with the DES-H peptide, indicating that the hydrophobic residues are insufficient to drive assembly alone and that the charged residues included in DES-1 are critical.”
A very nice step-wise study was pursued here, however, with a missing control. The authors show the parent peptide, then introduce only hydrophobic residues and then hydrophobic and charged residues. However, the peptide DAA(GPO)4(GRD)2(GPO)4 is missing. This control would show the influence of only the charged residues. If this peptide

formed higher-order assemblies, it would strengthen the claim above. If no assembly formed or the assemblies were rather amorphous, the hydrophobic residues may not be sufficient driving forces for the assembly but still be important for directing the assembly. I note that the subsequent exploration of peptides RD1 and RD2 does not compensate for this lack of control, as these have double the amount of charged residues.

Additionally, no controls are presented of how the assemblies would look without the DAA domain in both DES-H and DES-1. The authors claim, “We added a short noncollagenous domain of “DAA” at the N-terminus of the peptide as the presence of noncollagenous domain reduces unwanted aggregation and favors a more specific oligomer formation.” Peptides DES-H and DES-1 have, however, a completely different sequence and the findings from SPA-bcb cannot be generalized. The evaluation of a DES-1 analog lacking DAA would be an appropriate control.

4. “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.” – This statement is too general.
The rationale for choosing DAA as a non-collagenous domain within the new peptides is reasonable. However, appropriate control experiments are needed to corroborate the claim above. This DAA sequence has not been part of any of the previous peptides. The pI argument should, for example, be probed by substituting Asp in DAA of DES-1 with a more acidic amino acid, e.g. cysteic acid (Cya), that remains deprotonated at pH 3.5.

De novo Design of Heterogeneous Oligomeric Bundle and 2D Nanosheets Section
1. The melting curve (Fig 5c,d) is broad and shallow. The authors reason that this is due to multiple species of higher-order assemblies present, but there are alternative explanations:
a) Impurities. The UPLC trace (Fig S11) shows a shoulder at the head of the major peak and an additional peak at ~3.2 min, which indicates impurities. Based on the analytical data, this sample has a purity of less than 85%. Such an amount of impurities likely influences the results. The CMP needs to be purified and all relevant experiments repeated with the purified sample.

b) The very shallow transition in the melting curve could also arise from the co-existence of multiple triple helices with different thermal stabilities. The presence of consecutive GRD units could promote the formation of triple helices with non-canonical strand arrangements (other than 012 trimers). NMR studies with labeled RD1 could shed light on the presence of multiple triple helices. In addition, the authors measured the height of the assemblies derived from RD2, concluding that the height is consistent with that of a single triple helix. Fig. 5g (right) shows a cartoon suggesting that the clusters from RD1 also have a height of a canonical triple helix, but there is no AFM data to support this conclusion. AFM data could shed light on whether “longer” trimers with non-canonical strand arrangements form.

2. “RD1 and RD2 have identical amino acids compositions and therefore might be expected to have similar thermal stability.”
This sentence needs rewording since frame-shifted collagen triple helices that are composed of identical amino acids are known to have different Tm values (ref. 41).

Other comments
1. Figure S1 is very informative. I suggest moving it to the main text.
(also “collagen” not “collage”).

2. Please put the native SP-A sequence in all the figures where you make replacements, so that the reader can easily follow how the structure changed.

3. Figures 2, 3, 4 and 5: the basic experimental parameters of the CD experiments (buffer, concentration, heating rate) need to be listed in the caption, not only in the SI, since the Tm values strongly depend on these parameters.

Please correct spelling mistakes, typos and inaccuracies, examples:
• p.3: row 2 “relatively” is correct not “relative”, row 14: “sequence” not “sequences”
• p.4: “designed” peptides not “design peptides”
• p. 5: 2,2,2-triflouroethanol should be all lowercase.
• p. 9: “PP2” should be “PPII”.
• p. 16: "PyMol" not "pymol"
• p. 17: Fmoc not "FMOC"
• “The CD spectra give information about existing PPII helicity and triple helicity.”
This inaccurate statement occurs several times throughout the text (p5, p8, p9). On p8 "CD melts" is an incorrect term. Whereas CD spectra confirm PPII helicity, thermal denaturation provides the melting temperature of a collagen helix.

Further
1. Experimental details: only Fmoc chemistry is mentioned, but no details on which other protection groups on the side chain of the amino acids were used
2. In the SI, the scale bars of several cryo-EM images are missing, figures S22, S20, S17, S15 (scale bar present, but no corresponding length in caption).
3. Table S1: HRMS data of the peptides are missing.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I am satisfied with the authors' response and modification of the manuscript. I recommend for publication.

Reviewer #2

(Remarks to the Author)
My concerns have been addressed in the revisions.

Reviewer #3

(Remarks to the Author)
The authors have addressed major concerns by synthesizing and characterising four new peptides. These data have strengthened their claims. Additionally, some of the data was corrected, which put one of the compound's purity under question. However, some inaccuracies still persist, including missing data such as melting curves and CD spectra for the newly introduced peptide RD2-noDAA. See my responses to the authors' rebuttal in red.

Reviewer #3 (Remarks to the Author):
The authors studied the sequence-structure relationship of assemblies formed by collagen-like peptides using the SP-A protein as a model system. Evaluation of the self-assembly of the collagen-like region of SP-A by cryo-EM and AFM identified the formation of different architectures, including oligomeric bundles, polymeric ribbons and 2D nanosheets. Whereas the study did not provide general lessons for the design of hierarchical collagen-based structures, the insights into the underexplored self-assembly of PPII helical higher-order structures are valuable. These data will be of interest to the scientific community and specialists in the field of peptide-based supramolecular assemblies. However, several control experiments are missing, which could impact the results and their interpretation. In addition, based on a rather minimalistic data set, several conclusions are likely too general. Furthermore, one of the studied compounds is not sufficiently pure, with a possibly effect on its assembly properties.

Thank you for your detailed critique. Based on your suggestions, and those of the other reviewers, we have synthesized and characterized five new peptides that substantially help to elucidate the mechanisms of assembly and create an overall stronger manuscript. We have also provided new analysis of the one peptide with purity problems. Details of the changes made and response to specific criticisms follow in blue underneath each comment.

Mutation Studies Reveal Sequence-Structure Relationship Section
1. "Even though the Ile to Pro mutant has enhanced thermal stability and favored triple helix formation, it did not promote oligomerization. This suggests that oligomerization is not correlated with triple helix formation, a critical finding."

This conclusion could be incorrect since two variables changed: the replacement of Ile by Pro and, as a consequence, the Tm. For instance, upon deleting a C-terminal POG repeat, the modified SPA-rat peptide would likely not form a stable triple helix and also not the higher-order assemblies. In contrast, an additional POG unit would stabilize the triple helix (while keeping the key Ile residue), and the system could have a higher propensity to oligomerize. Here, the conclusion would be the opposite: oligomerization correlates with triple helix formation and stability. Without such control experiments, the only valid conclusion is that the Ile residue is important for the oligomerization of SPA-rat.

We have reworded the sentences in question as follows, "Even though the peptide containing the Ile to Pro substitution has enhanced triple helical thermal stability (Tm = 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 Tm is not necessarily correlated with higher order assembly."

OK

Also, the term "mutation" refers to DNA, use "replacement of Ile by Pro.
Throughout the manuscript we have changed this to "substitution" or "replacement" instead of mutation.

OK

Also, I would use "extended assembly" rather than "oligomerization" since oligomerization refers to covalent bonds.

The term "oligomerization" refers to more than covalent bonds and generally refers to anything of which there are several of something. It is commonly used in work related to protein association and self-assembly more generally. For example, see 10.1126/science.aad8865 or 10.1002/pro.5560040818. Based on this we left the term "oligomerization" in the manuscript. This sentence has been reworded as follows, "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."

OK

2. "The point-mutation studies together revealed that charged-pair interactions involving arginine and aspartate and hydrophobic interactions may exist to favor collagen oligomerization." This conclusion is likely too "big", considering the rather minimalistic study.
a) The study focuses on one amino acid variations in the stem collagen region. Here, the authors replaced only Arg24 and Asp25, but there are two more Asp residues within the stem region. In addition, Asp25 is in the Yaa position, whereas Asp33 is in the Xaa position (per collagen canonical sequence). As Xaa and Yaa positions are not equal within the triple helix (Xaa position is, for example, more solvent exposed), substituting an Asp residue in the Xaa position (e.g. Asp33) could have different effect on the structure and its assembly than substituting an Asp in the Yaa position. Thus, the conclusion could change significantly, e.g., substituting Asp33 could have no impact on the triple helical or the further assembly, which would then contradict the current conclusion.

See next answer below.

b) For the positively charged residues, the argument is the same. Lys31 is in the Yaa, compared to the mutated Xaa Arg24. Thus, to corroborate the conclusion Arg27 needs to be substituted.
The scope of the current work is a targeted substitution study on natural SPA in an attempt to identify amino acids important for its self-assembly into an octadecamer. Based on more than a decade of work with peptides that assemble into collagen triple helices, but show no sign of any oligomerization, the RDGRDG triplets and repetitious hydrophobic groups every three triplets in SPA (as well as similar patterns observed in other defense collagens) seemed unusual and worth a targeted study. Our studies indeed found that these sequences were important in the context of natural SPA, and also that these features could be implemented in de novo designed peptides (eg. DES-1).Without the use of some chemical intuition as to what all these amino acids are doing, every amino acid of the SPA sequence could conceivably play a role in self-assembly and one could easily suggest dozens, if not hundreds, of single and double substitution experiments that could potentially be enlightening. In the future we expect to carry out some of those experiments ourselves and hope to see work from other groups investigate this further as well.
Within the context of the current work we feel that the substitutions that were investigated were appropriate and sufficient since they led to successful criteria for de novo design. We have tried to capture this idea more effectively in the closing two sentences of this section which says, "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."

OK

Design of pH-responsive Nanoribbons
1. p.9: Initially, a physiological pH is mentioned for folding, and a few lines later, pH 6.1. On p10, the default annealing condition at pH 6.1 is mentioned. Which pH was used? Were all peptides assembled in the same buffer and at the same concentration? This is important for comparisons as the buffer and the ionic strength of the solutions can have a significant impact on self-assembling systems.
2. Fig. 4b: It would be better to show the raw melting curve rather than the CD spectrum as it carries more critical information. The CD spectrum can be moved into the SI.

To help clarify this we have changed the wording to "Peptides were prepared at a concentration of 3mM at a pH of 6.1 ± 0.2 in 20mM MES buffer, see methods for details." The methods have been clarified to state,
"Purified peptide solutions were freeze-dried with a FreeZone 4.5L 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 1M 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 1M HCl solution and the sample at pH 10.0 was adjusted with 1M NaOH. Peptide solutions were confirmed by pH meter to be the expected pH values before characterizations."
We have updated figure 4 as suggested.

OK

3. "No higher-order assembly was observed with the DES-H peptide, indicating that the hydrophobic residues are insufficient to drive assembly alone and that the charged residues included in DES-1 are critical."

A very nice step-wise study was pursued here, however, with a missing control. The authors show the parent peptide, then introduce only hydrophobic residues and then hydrophobic and charged residues. However, the peptide DAA(GPO)4(GRD)2(GPO)4 is missing. This control would show the influence of only the charged residues. If this peptide formed higher-order assemblies, it would strengthen the claim above. If no assembly formed or the assemblies were rather amorphous, the hydrophobic residues may not be sufficient driving forces for the assembly but still be important for directing the assembly. I note that the subsequent exploration of peptides RD1 and RD2 does not compensate for this lack of control, as these have double the amount of charged residues.

Thank you for this suggestion. We have now synthesized and characterized this control peptide (called DES-P) as well as the variant that lacks the "DAA" triplet (called DES-P-noDAA). While both of these peptides self-assemble into triple helices,

neither self-assemble beyond the triple helix to higher order aggregates, as expected. Data for DES-P is incorporated into Figure 4. DES-P-noDAA is shown in supporting information figures S24 and S25. These two additional peptides help to clarify the role of each of the three designed regions of the "DES" series of peptides. A new / updated paragraph has been added to the manuscript reproduced here:
"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."

OK

Additionally, no controls are presented of how the assemblies would look without the DAA domain in both DES-H and DES-1. The authors claim, "We added a short noncollagenous domain of "DAA" at the N-terminus of the peptide as the presence of noncollagenous domain reduces unwanted aggregation and favors a more specific oligomer formation." Peptides DES-H and DES-1 have, however, a completely different sequence and the findings from SPA-bcb cannot be generalized. The evaluation of a DES-1 analog lacking DAA would be an appropriate control.

As mentioned above, and in response to similar critique from reviewer 1, we have prepared and characterized a complete series of peptides in which the "DAA" triplet has been removed:
Peptide Name DES-1-no DAA DES-H-no DAA DES-P-no DAA RD1-no DAA RD2-no DAA
Sequence
GlOGPOGPOGLOGRDGRDGVOGPOGPOGPM GlOGPOGPOGLOGPOGPOGVOGPOGPOGPM
GPOGPOGPOGRDGRDGPOGPOGPOGPOGPO GPOGPOGPOGRDGRDGRDGRDGPOGPOGPO
GPOGRDGRDGPOGPOGPOGPOGRDGRDGPO
RD1-noDAA was in the original manuscript, but all the others are new for this revision. In summary, we find that for systems that the DAA triplet is extremely important for the control of self-assembly in higher order systems. For example, DES-1 undergoes higher order assemblies including 15- and 18-mers. Removal of the DAA removes a degree of electrostatic repulsion and because of this we find that DES-1-noDAA aggregates and precipitates. In contrast systems that only form simple triple helices such as DES-H and DES-P, the DAA triplet can be removed with only very small changes in assembly state (they still form simple triple helices). RD2 and RD2-no DAA displayed similar behavior in which the "no DAA" variant generated much larger aggregates. RD1 and RD1-noDAA behaved somewhat differently in that elimination of the DAA triplet completely eliminated triple helical folding and the peptide appeared to remain soluble suggesting that the DAA triplet was involved in some way in the stabilization of the ring-like higher order aggregates observed. While this is a bit different than the other cases, in every case observed well controlled higher order assembly required the DAA triplet.
We have updated our manuscript with a discussion of these results which can be found on pages 9-10 and 15-16 of the updated manuscript. Complete characterization is found in the SI in particular figures S24 and S25.

I do not see any CD or melting curves for RD2-no DAA. Please add them. Also fig S25c is missing the Tm values and neither are the Tm values mentioned in the caption. Please add the corresponding values to the figure (as you already have in all such figures).

4. "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." – This statement is too general. The rationale for choosing DAA as a non-collagenous domain within the new peptides is reasonable. However, appropriate control experiments are needed to corroborate the claim above. This DAA sequence has not been part of any of the previous peptides. The pI argument should, for example, be probed by substituting Asp in DAA of DES-1 with a more acidic amino acid, e.g. cysteic acid (Cya), that remains deprotonated at pH 3.5.

We feel that the new peptides and the results from them described above now more effectively justify our initial claim.

OK

De novo Design of Heterogeneous Oligomeric Bundle and 2D Nanosheets Section
1. The melting curve (Fig 5c,d) is broad and shallow. The authors reason that this is due to multiple species of higher-order assemblies present, but there are alternative explanations:
a) Impurities. The UPLC trace (Fig S11) shows a shoulder at the head of the major peak and an additional peak at ~3.2 min, which indicates impurities. Based on the analytical data, this sample has a purity of less than 85%. Such an amount of impurities likely influences the results. The CMP needs to be purified and all relevant experiments repeated with the purified sample.

Thank you for this criticism. After significant reevaluation of this data, we uncovered a rather embarrassing mistake on our part. The UPLC trace reported in the initial SI was actually of the crude material, not the HPLC purified material. We performed new UPLC and MS experiments on samples that had been purified, and were the same materials that were studied elsewhere in the manuscript for SEC and electron microscopy. The UPLC analysis of the purified material looks clean, unlike what we originally reported. Mass spec results also support the purity of these samples. Supporting Information Figure S15

OK

b) The very shallow transition in the melting curve could also arise from the co-existence of multiple triple helices with different thermal stabilities. The presence of consecutive GRD units could promote the formation of triple helices with non-canonical strand arrangements (other than 012 trimers). NMR studies with labeled RD1 could shed light on the presence of multiple triple helices. In addition, the authors measured the height of the assemblies derived from RD2, concluding that the height is consistent with that of a single triple helix. Fig. 5g (right) shows a cartoon suggesting that the clusters from RD1 also have a height of a canonical triple helix, but there is no AFM data to support this conclusion. AFM data could shed light on whether “longer” trimers with non-canonical strand arrangements form.

RD1 and RD2 assemblies are markedly different thus the use of different methods for characterization here. While RD2 assembles into sheets whose heights can be assessed in a straightforward fashion by AFM, RD1 does not. Instead RD1’s assembly stops after some moderate number of helices have assembled. These types of aggregates are not suitable for AFM analysis. Instead cryo-EM is able to reveal a low-resolution structure with the suggested packing shown in cartoon form.
The authors did not address the comment regarding the possible formation of triple helices with non-canonical strand arrangements (other than 012 trimers) for RD1. They claim that “the melting curve of RD1 peptide has broad thermal transition, suggesting the higher-order assemblies formed by this peptide is heterogenous”. I still think that non-canonical strand arrangements in the triple helix rather than heterogeneity in higher-order assemblies are an equally plausible explanation and the previously suggested NMR studies should be performed to shed light on the origins of the shallow transition in the melting curve.

2. “RD1 and RD2 have identical amino acids compositions and therefore might be expected to have similar thermal stability.” This sentence needs rewording since frame-shifted collagen triple helices that are composed of identical amino acids are known to have different Tm values (ref. 41).
While this is a known phenomenon to those who study collagen, this may not be immediately apparent to all readers which is why we used the phrase “might be expected”.

OK

Other comments
1.Figure S1 is very informative. I suggest moving it to the main text.
We are not opposed to moving this figure to the main manuscript. However, we are already at 5 figures and are concerned about the overall length of the work. We will happily move this figure to the main text if the editorial staff support this decision.

OK

(also “collagen” not “collage”).
This typo has been corrected.

OK

2. Please put the native SP-A sequence in all the figures where you make replacements, so that the reader can easily follow how the structure changed.
We have updated Figures 2-5 with this suggestion.

OK

3. Figures 2, 3, 4 and 5: the basic experimental parameters of the CD experiments (buffer, concentration, heating rate) need to be listed in the caption, not only in the SI, since the Tm values strongly depend on these parameters.
We have updated our figure captions with this suggestion.

OK

Please correct spelling mistakes, typos and inaccuracies, examples:
• p.3: row 2 “relatively” is correct not “relative”, row 14: “sequence” not “sequences” Fixed OK
• p.4: “designed” peptides not “design peptides” Fixed OK
• p. 5: 2,2,2-triflouroethanol should be all lowercase. Fixed OK
• p. 9: “PP2” should be “PPII”. Both PP2 and PPII are commonly used in the literature. OK
• p. 16: "PyMol" not "pymol" Fixed, not fixed on p17
• p. 17: Fmoc not "FMOC" Fixed OK
• “The CD spectra give information about existing PPII helicity and triple helicity.”
This inaccurate statement occurs several times throughout the text (p5, p8, p9). On p8 "CD melts" is an incorrect term. Whereas CD spectra confirm PPII helicity, thermal denaturation provides the melting temperature of a collagen helix.
We have made our language more precise. In one instance we use the following: “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.” And later, “The CD spectra of both are presented in Fig. 3e, displaying the characteristic maximum near 224nm associated with PP2 helicity in peptide SPA-I-P, but not I-S. CD melting curve analysis

supports triple helicity for SPA-I-P."
p9: "the CD spectra of all six peptides show a characteristic signal of 225 nm, indicative of triple helix formation (S24a and S25a)" Here once again inaccurate language, please rephrase: CD spectra of all six peptides show a characteristic signal of 225 nm, but it is the is the melting curves that are indicative of triple helical formation, not the maxima of 225 nm signal on the CD curve. Same inaccuracy still stands in other places throughout the manuscript: p10, p 13.
Further
1. Experimental details: only Fmoc chemistry is mentioned, but no details on which other protection groups on the side chain of the amino acids were used
We have updated the method section with additional information.
OK
2. In the SI, the scale bars of several cryo-EM images are missing, figures S22, S20, S17, S15 (scale bar present, but no corresponding length in caption).
In all places where scale bars are shown, the corresponding length is now mentioned in the caption.
OK
3. Table S1: HRMS data of the peptides are missing.
The MS data has been added to Table S1.
OK

Additional small mistakes:
• Typo Fmoc-ASP(OtBu)-OH, should be "Asp".
• p9 bottom: typo and need of rephrasing: "Three additional peptides (DES-1-noDAA, DES-P-noDAA and DES-H-noDAA remove the N-terminal triplet." – sentence seems incomplete

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
The authors have addressed the concerns in the revised manuscript. I support publication as is.

Reviewer #1 (Remarks to the Author):

The authos made de novo design of collagen-like peptides, which were self-assembled into diverse quaternary structures, such as bundled oligomers, nanosheets, and nanoribbons. This highly novel study showed a significant advancement in the understanding higher-order assemblies of collagen-like peptides. I highly recommend publishing this manuscript after the authors address the questions and comments listed as below .

Thank you for your detailed critique of our work. Based on your suggestions we have synthesized and characterized five new peptides that substantially help to elucidate the mechanism of assembly and create an overall stronger manuscript. Details of the changes made and response to specific criticisms follow in blue underneath each comment.

Comments:

(1) As the SPA-bcb peptide precipitated, it is not suitable for circular dichroism, please use Differential Scannining Calorimetry to measure its stability (Karunakar Kar, et al Journal of Biological Chemistry 281(44): 33283-33290.)

As a precipitate, DSC would also report data on SPA-bcb that does not directly correspond to the stability of the peptide's fold and would instead report on changes in solubility (although these are likely correlated). These are similar limitations that exist with CD. Because of this we do not think this data would meaningfully improve the current manuscript. Our current CD data suggest that the assembly unfolds and dissolves at approximately the same temperature (35 °C), but the poor solubility results in a heterogeneous unfolding curve and limits our ability to interpret this value with high confidence as we mention in the manuscript.

(2) All the results are very interesting and show high novelty. However, the authors should address three questions explicitly in the conclusion:

(a) What is the role of DAA in DES-H, DES-1, DR1 and DR2 in the self-assembly? Except for DES2-no DAA, the authors may provide experimental data of no DAA in DR1, DES-H, DES-1 peptides. Based on these data, the authors could make thorough discussion about the roles of DAA, hydrophobic and electrostatic interactions in the self-assembly.

This is a very good suggestion. In this revision, **we have prepared four new peptides** that lack the "DAA" sequence. These complement the "RD2-no DAA" peptide included in the original manuscript and can be directly compared to their counterparts that do include the DAA N-terminal triplet. These are now included in Supplementary Table S1. In total we now include five "no DAA" peptides:

Peptide Name	Sequence
DES-1-no DAA	GIOGPOGPOGLOGRDGRDGVOGPOGPOGPM
DES-H-no DAA	GIOGPOGPOGLOGPOGPOGVOGPOGPOGPM
DES-P-no DAA	GPOGPOGPOGRDGRDGPOGPOGPOGPOGPO
RD1-no DAA	GPOGPOGPOGRDGRDGRDGRDGPOGPOGPO
RD2-no DAA	GPOGRDGRDGPOGPOGPOGPOGRDGRDGPO

In summary, we find that the DAA triplet is particularly important for the control of self-assembly in systems with the propensity to assemble beyond a triple helix. For example, DES-1 undergoes higher order assemblies including 15- and 18-mers. Removal of the DAA removes a degree of electrostatic repulsion and because of this we find that DES-1-noDAA aggregates and precipitates. In contrast systems that only form simple triple helices such as DES-H and DES-P, the DAA triplet can be removed with only very small changes in stability (they still form simple triple helices). RD2 and RD2-no DAA displayed similar behavior, but the "no DAA" variant generated much larger aggregates again suggesting the role of electrostatic repulsion limiting assembly size. RD1 vs. RD1-noDAA behaved somewhat differently in that elimination of the DAA triplet eliminated triple helical folding and the peptide appeared to remain soluble. This suggests that the DAA triplet may be involved in some way in the stabilization of the ring-like higher order aggregates observed. While this is somewhat different than the other cases, in all cases well controlled higher order assembly required the DAA triplet.

We have updated our manuscript with a discussion of these results which can be found on **pages 9-10** and **15-16** of the updated manuscript. Complete characterization is found in the SI in particular figures **S24** and **S25**.

(b) What is the role of hydrophobic interaction in the self-assembly?

For this revision we have synthesized a new peptide "DES-P" (and as mentioned above DES-P-noDAA) which complement the existing peptides DES-1 and DES-H. DES-1 contains both the hydrophobic and charged hydrophilic amino acids, DES-H contains only the hydrophobic amino acids while DES-P contain only the charged polar amino acids. In all cases the peptides conform to the (POG)n template overall. This allows us to much more effectively dissect out the role of each set of amino acids. We find that assembly into well controlled 15-mer and 18-mer assemblies requires both sets of amino acids as neither DES-P nor DES-H is found to assemble beyond a simple triple helix.

DES-H with hydrophobic side chains cannot be self-assembled. However, similar hydrophobic interactions could induce self-assembly of some collagen-like peptides (Kenneth McGuinness et al, ACS Nano, 2014, 8(12) 12514–12523). DES-H has DAA while the peptides by McGuinness has no DAA. Does the DAA inhibit the self-assembly? Or maybe the locations and numbers of the hydrophobic interactions cause difference in the self-assembly?

As mentioned above, new to this revision, we prepared DES-H-noDAA and found that removal of the DAA sequence had very little impact on it: it folds into a triple helix with similar stability to DES-H, but neither DES-H nor DES-H-noDAA are observed to form higher order assemblies.

While both our system and the system studied by McGuinness contain hydrophobic amino acids, the sequences examined and assemblies observed by McGuinness are both different from what we have observed in the present study. The McGuinness study observed a wide range of nanostructures and what appear to be very dense assemblies. It is possible that their system's assembly might become more specific upon addition of a DAA triplet. A new study to examine this might be fruitful.

(c) What is the role of electrostatic interactions in the self-assembly?

As mentioned above, for this revision we have synthesized the new peptide "DES-P" which contains the polar / charged amino acids but lacks the hydrophobic residues. The role of the electrostatic interactions appear to be synergistic with the hydrophobic residues as neither alone can drive higher order assembly.

RD2, with the electrostatic interactions at the N- and C-termini, formed sheet-like assemblies. RD2-noDAA also formed sheet-like assemblies, which may have larger size than RD2 (Fig S29). In previous studies, peptides, with similar charged patterns but no DAA-region, also formed sheet-like assemblies (Avanish S. Parmar, et al J. Am. Chem. Soc. 2016, 138, 4362−4367; Tao Jiang et al, J. Am. Chem. Soc. 2015, 137, 7793−7802). RD1 and DES-1, with the electrostatic interactions at the center, formed oligomeric assemblies. The authors may discuss the roles of both DAA and location of the electrostatic interactions in the self-assembly.

As mentioned above we have prepared new peptides that do not include the DAA sequence to help illustrate the importance of this triplet for the control of assembly. The expanded discussion of this effect is included on pages **pages 9-10** and **15-16** of the updated manuscript.

Minor issues:

(1) Typo: Page 5 1st paragraph, 8th line: the pepetide aggregation

Thank you - we have made this correction.

Reviewer #2 (Remarks to the Author):

1、In the sample preparation, lacey carbon 300 mesh copper grids were used. However, the molecular weight of these proteins are relatively small, resulting in a low signal-to-noise ratio of 2D classification. In the Figures S15 and S20, the quality of 2D classification is bad. Therefore, no high resolution details can be distinguished after 3D reconstruction. Have the authors tried to use holey copper grids instead?

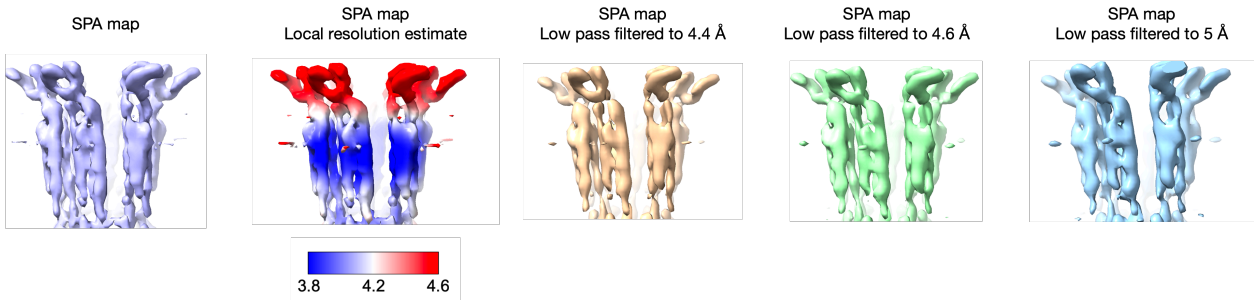
The issue with reconstructing particles such as this does not appear to be grid type. We have tried nearly every type of grid and they all yield similar results. As we did not see differences between grid types, we went with Lacey because we have an abundance of them in supply in our lab and they are relatively cheap compared to other grids. We have also tried glow discharging the grids with ethyl amine, coating grids with graphene, and much more. Instead, we believe the poor 2D-classification results in **Fig S20** are the result of the inherent heterogeneity of the SP-A-rat assembly.

2、Especially for the dataset of DES-1 peptide at pH 3.5, the protein complex seems to well match the helical symmetry. Maybe the authors could try helical reconstruction in cryoSPARC to increase the final resolution. Considering the shape, preferred orientation would be quite severe if the authors used conventional single particle reconstruction.

We have attempted helical reconstruction on the pH 3.5 DES-1 peptide assembly and encountered several problems due to sample heterogeneity (diameter heterogeneity as well as tubes unravelling into sheets) which could not simply be solved by collecting more data. Therefore, no cryo-EM structure of the filaments is reported. All other samples we looked at using cryo-EM were of the single particle nature or 2D-crysalline sheets.

3. A cryo-electron microscopy (cryo-EM) data processing workflow is required to demonstrate how the data was processed. The workflow should include additional information, such as the Gold Standard Fourier Shell Correlation (GSFSC) curve, a 3D Fourier Shell Correlation (FSC) histogram, and the local resolution for each cryo-EM map.

The requested workflows and the FSC curves have been added to the **supplemental figures S38 & S39**. In our experience we have found that, with these assemblies, over-reliance on the map:map FSC is problematic. Therefore, we would prefer to not include 3D histograms and the local resolution maps. For the SPA assembly, when we examine the map by eye, the overall quality of the map is at best around 4.4 to 5 Å. In the image below the SPA map and its local resolution map are shown. The maps of the SPA assembly filtered to 4.4, 4.6, and 5 Å resolution are also shown. The local resolution map indicates a strip of density that is around 3.8 Å (dark blue). When we low-pass filter the SPA map to 4.4 Å or 4.6 Å essentially all of same information that was present in the original map is retained except for noise. When we low-pass filter the map to 5 Å the map looks a little lower in resolution but one can draw the same conclusions about triple helix positioning.



Thus, in the case of these small particle assemblies we find that reliance on the GFSC is problematic because it overestimates the resolution. Seeing separation between the individual polypeptides of the triple helices in our SPA, and DES-1 18-mer and 15-mer density maps is a good indicator that the resolution of these structures is around 4.5 to 5 Å. In general, we prefer to use the map:model FSC curve to estimate the resolution of our structures. However, the structures in this study are too low resolution to allow for accurate model building. Thus, we think that the best metric for resolution estimation of each of the assemblies in this study is the "eye test". That being said, the resolutions reported in **Table S2** are calculated using the map:map FSC curves because it's the most feasible method available in the absence of a model. In the case of the SPA assembly, we used the 0.5 map:map FSC to estimate the

resolution because it reflected the overall quality of the map better. The pitfalls of using the map:map FSC to estimate resolution, especially in cases of symmetry imposition, have been shown before by Subramaniam et al. (doi: 10.1016/j.sbi.2016.07.009).

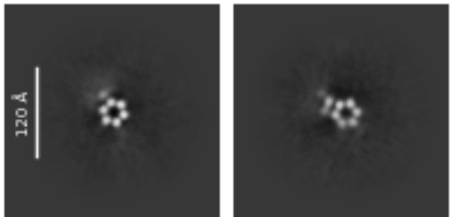
4. “For the SPA-human hexamers, DES-1 pentamers, and DES-1 hexamers ab initio reconstruction with multiple classes was performed with either C5 or C6 symmetry and the resulting “good” classes were then subjected to homogenous refinement with the same point group symmetry.”

Given the relatively low resolution of the cryo-EM maps, the application of C5 or C6 symmetry might have led to artificial results.

(For clarity, the previously titled hexamers and pentamers are now referred to as octadecamer (18-mers) and pentadecamers (15-mers) respectively. This is to be consistent with the literature regarding C1q which is defined as an octadecamer.)

We to shared your concerns with the possibility of artificial results given the relatively low-resolution of the dataset. For the SPA we have little doubt that the sample predominantly features hexamer ring structures (18-mer) based on top views we see in the 2D class averages.

SP-A Human 2D class averages

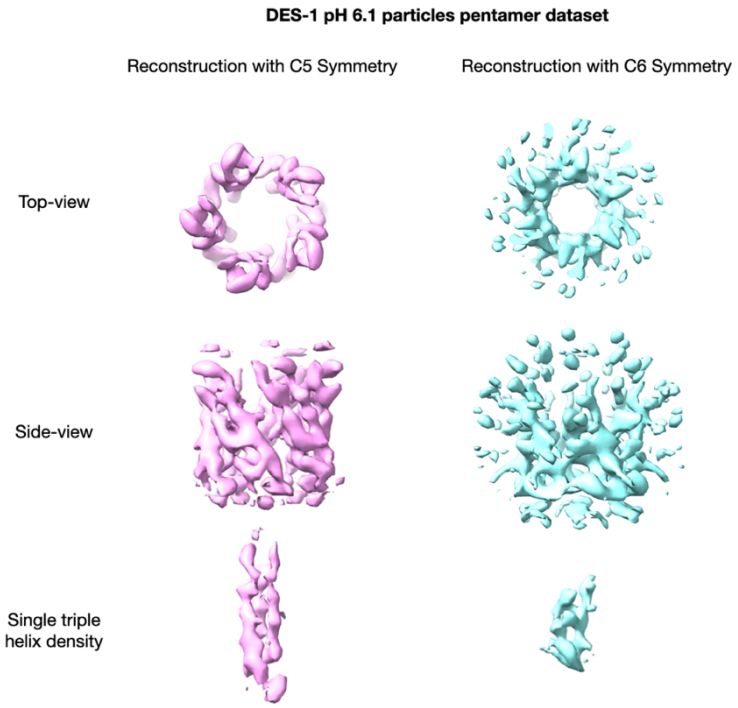


For the DES-1, we think that our results strongly support the presence of pentameric (15-mer), hexameric rings (18-mer), and doublets. Here is our expanded reasoning:

In the 2D class averages below we find evidence for the presence of both 15-mers (red circle) and 18-mers (yellow circle). Additionally, if one looks at some of the doublet classes (purple circle) hexameric packing is apparent as well.

DES-1 pH 6.1 2D class averages:

Additionally, when we reconstructed the 15-mer dataset using C6 symmetry we found that the resulting structure was poorer than the C5 symmetry structure as no triple helical strands were observed (see below). We believe this is strong evidence that these results are not artifactual.



5. In the legend of Figure S15b, “cry-EM” should be replaced with “cryo-EM”.

Thank you - this has been corrected.

Reviewer #3 (Remarks to the Author):

The authors studied the sequence-structure relationship of assemblies formed by collagen-like peptides using the SP-A protein as a model system. Evaluation of the self-assembly of the collagen-like region of SP-A by cryo-EM and AFM identified the formation of different architectures, including oligomeric bundles, polymeric ribbons and 2D nanosheets.

Whereas the study did not provide general lessons for the design of hierarchical collagen-based structures, the insights into the underexplored self-assembly of PPII helical higher-order structures are valuable. These data will be of interest to the scientific community and specialists in the field of peptide-based supramolecular assemblies. However, several control experiments are missing, which could impact the results and their interpretation. In addition, based on a rather minimalistic data set, several conclusions are likely too general. Furthermore, one of the studied compounds is not sufficiently pure, with a possibly effect on its assembly properties.

Thank you for your detailed critique. Based on your suggestions, and those of the other reviewers, we have synthesized and characterized five new peptides that substantially help to elucidate the mechanisms of assembly and create an overall stronger manuscript. We have also provided new analysis of the one peptide with purity problems. Details of the changes made and response to specific criticisms follow in blue underneath each comment.

Mutation Studies Reveal Sequence-Structure Relationship Section

1. “Even though the Ile to Pro mutant has enhanced thermal stability and favored triple helix formation, it did not promote oligomerization. This suggests that oligomerization is not correlated with triple helix formation, a critical finding."

This conclusion could be incorrect since two variables changed: the replacement of Ile by Pro and, as a consequence, the Tm. For instance, upon deleting a C-terminal POG repeat, the modified SPA-rat peptide would likely not form a stable triple helix and also not the higher-order assemblies. In contrast, an additional POG unit would stabilize the triple helix (while keeping the key Ile residue), and the system could have a higher propensity to oligomerize. Here, the conclusion would be the opposite: oligomerization correlates with triple helix formation and stability. Without such control experiments, the only valid conclusion is that the Ile residue is important for the oligomerization of SPA-rat.

We have reworded the sentences in question as follows, “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.”

Also, the term “mutation” refers to DNA, use “replacement of Ile by Pro.

Throughout the manuscript we have changed this to “substitution” or “replacement” instead of mutation.

Also, I would use “extended assembly” rather than “oligomerization” since oligomerization refers to covalent bonds.

The term “oligomerization” refers to more than covalent bonds and generally refers to anything of which there are several of something. It is commonly used in work related to protein association and self-assembly more generally. For example, see 10.1126/science.aad8865 or 10.1002/pro.5560040818. Based on this we left the term “oligomerization” in the manuscript.

2. “The point-mutation studies together revealed that charged-pair interactions involving arginine and aspartate and hydrophobic interactions may exist to favor collagen oligomerization.” This conclusion is likely too “big”, considering the rather minimalistic study.

This sentence has been reworded as follows, “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.”

a) The study focuses on one amino acid variations in the stem collagen region. Here, the authors replaced only Arg24 and Asp25, but there are two more Asp residues within the stem region. In addition, Asp25 is in the Yaa position, whereas Asp33 is in the Xaa position (per collagen canonical sequence). As Xaa and Yaa positions are not equal within the triple helix (Xaa position is, for example, more solvent exposed), substituting an Asp residue in the Xaa position (e.g. Asp33) could have different effect on the structure and its assembly than substituting an Asp in the Yaa position. Thus, the conclusion could change significantly, e.g., substituting Asp33 could have no impact on the triple helical or the further assembly, which would then contradict the current conclusion.

See next answer below.

b) For the positively charged residues, the argument is the same. Lys31 is in the Yaa, compared to the mutated Xaa Arg24. Thus, to corroborate the conclusion Arg27 needs to be substituted.

The scope of the current work is a targeted substitution study on natural SPA in an attempt to identify amino acids important for its self-assembly into an octadecamer. Based on more than a decade of work with peptides that assemble into collagen triple helices, but show no sign of any oligomerization, the RDGRDG triplets and repetitious hydrophobic groups every three triplets in SPA (as well as similar patterns observed in other defense collagens) seemed unusual and worth a targeted study. Our studies indeed found that these sequences were important in the context of natural SPA, and also that these features could be implemented in *de novo* designed peptides (*eg.* DES-1). Without the use of some chemical intuition as to what all these amino acids are doing, every amino acid of the SPA sequence could conceivably play a role in self-assembly and one could easily suggest dozens, if not hundreds, of single and double substitution experiments that could potentially be enlightening. In the future we expect to carry out some of those experiments ourselves and hope to see work from other groups investigate this further as well. Within the context of the current work we feel that the substitutions that were investigated were appropriate and sufficient since they led to successful criteria for *de novo* design. We have tried to capture this idea more effectively in the closing two sentences of this section which says, "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 pH-responsive Nanoribbons

1. p.9: Initially, a physiological pH is mentioned for folding, and a few lines later, pH 6.1. On p10, the default annealing condition at pH 6.1 is mentioned. Which pH was used? Were all peptides assembled in the same buffer and at the same concentration? This is important for comparisons as the buffer and the ionic strength of the solutions can have a significant impact on self-assembling systems.

To help clarify this we have changed the wording to "Peptides were prepared at a concentration of 3mM at a pH of 6.1 ± 0.2 in 20mM MES buffer, see methods for details." The methods have been clarified to state,

> "Purified peptide solutions were freeze-dried with a FreeZone 4.5L 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 1M 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 1M HCl solution and the sample at pH 10.0 was adjusted with 1M NaOH. Peptide solutions were confirmed by pH meter to be the expected pH values before characterizations."

2. Fig. 4b: It would be better to show the raw melting curve rather than the CD spectrum as it carries more critical information. The CD spectrum can be moved into the SI.

We have updated figure 4 as suggested.

3. "No higher-order assembly was observed with the DES-H peptide, indicating that the hydrophobic residues are insufficient to drive assembly alone and that the charged residues included in DES-1 are critical."

A very nice step-wise study was pursued here, however, with a missing control. The authors show the parent peptide, then introduce only hydrophobic residues and then hydrophobic and charged residues. However, the peptide DAA(GPO)4(GRD)2(GPO)4 is missing. This control would show the influence of only the charged residues. If this peptide formed higher-order assemblies, it would strengthen the claim above. If no assembly formed or the assemblies were rather amorphous, the hydrophobic residues may not be sufficient driving forces for the assembly but still be important for directing the assembly. I note that the subsequent exploration of peptides RD1 and RD2 does not compensate for this lack of control, as these have double the amount of charged residues.

Thank you for this suggestion. We have now synthesized and characterized this control peptide (called DES-P) as well as the variant that lacks the "DAA" triplet (called DES-P-noDAA). While both of these peptides self-assemble into triple helices, neither self-assemble beyond the triple helix to higher order aggregates, as expected. Data for DES-P is incorporated into **Figure 4**. DES-P-noDAA is shown in **supporting information figures S24** and **S25**. These two additional peptides help to clarify the role of each of the three designed regions of the "DES" series of peptides. A new / updated paragraph has been added to the manuscript reproduced here:

> "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."

Additionally, no controls are presented of how the assemblies would look without the DAA domain in both DES-H and DES-1. The authors claim, "We added a short noncollagenous domain of "DAA" at the N-terminus of the peptide as the presence of noncollagenous domain reduces unwanted aggregation and favors a more specific oligomer formation." Peptides DES-H and DES-1 have, however, a completely different sequence and the findings from SPA-bcb cannot be generalized. The evaluation of a DES-1 analog lacking DAA would be an appropriate control.

As mentioned above, and in response to similar critique from reviewer 1, we have prepared and characterized a complete series of peptides in which the "DAA" triplet has been removed:

Peptide Name	Sequence
DES-1-no DAA	GIOGPOGPOGLOGRDGRDGVOGPOGPOGPM
DES-H-no DAA	GIOGPOGPOGLOGPOGPOGVOGPOGPOGPM
DES-P-no DAA	GPOGPOGPOGRDGRDGPOGPOGPOGPOGPO
RD1-no DAA	GPOGPOGPOGRDGRDGRDGRDGPOGPOGPO
RD2-no DAA	GPOGRDGRDGPOGPOGPOGPOGRDGRDGPO

RD1-noDAA was in the original manuscript, but all the others are new for this revision. In summary, we find that for systems that the DAA triplet is extremely important for the control of self-assembly in higher order systems. For example, DES-1 undergoes higher order assemblies including 15- and 18-mers. Removal of the DAA removes a degree of electrostatic repulsion and because of this we find that DES-1-noDAA aggregates and precipitates. In contrast systems that only form simple triple helices such as DES-H and DES-P, the DAA triplet can be removed with only very small changes in assembly state (they still form simple triple helices). RD2 and RD2-no DAA displayed similar behavior in which the "no DAA" variant generated much larger aggregates. RD1 and RD1-noDAA behaved somewhat differently in that elimination of the DAA triplet completely eliminated triple helical folding and the peptide appeared to remain soluble suggesting that the DAA triplet was involved in some way in the stabilization of the ring-like higher order aggregates observed. While this is a bit different than the other cases, in every case observed well controlled higher order assembly required the DAA triplet.

We have updated our manuscript with a discussion of these results which can be found on **pages 9-10** and **15-16** of the updated manuscript. Complete characterization is found in the SI in particular figures **S24** and **S25**.

4. “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.” – This statement is too general.
The rationale for choosing DAA as a non-collagenous domain within the new peptides is reasonable. However, appropriate control experiments are needed to corroborate the claim above. This DAA sequence has not been part of any of the previous peptides. The pI argument should, for example, be probed by substituting Asp in DAA of DES-1 with a more acidic amino acid, e.g. cysteic acid (Cya), that remains deprotonated at pH 3.5.

We feel that the new peptides and the results from them described above now more effectively justify our initial claim.

De novo Design of Heterogeneous Oligomeric Bundle and 2D Nanosheets Section
1. The melting curve (Fig 5c,d) is broad and shallow. The authors reason that this is due to multiple species of higher-order assemblies present, but there are alternative explanations:

a) Impurities. The UPLC trace (Fig S11) shows a shoulder at the head of the major peak and an additional peak at ~3.2 min, which indicates impurities. Based on the analytical data, this sample has a purity of less than 85%. Such an amount of impurities likely influences the results. The CMP needs to be purified and all relevant experiments repeated with the purified sample.

Thank you for this criticism. After significant reevaluation of this data, we uncovered a rather embarrassing mistake on our part. The UPLC trace reported in the initial SI was actually of the crude material, not the HPLC purified material. We performed new UPLC and MS experiments on samples that had been purified, and were the same materials that were studied elsewhere in the manuscript for SEC and electron microcopy. The UPLC analysis of the purified material looks clean, unlike what we originally reported. Mass spec results also support the purity of these samples. **Supporting Information Figure S15** is reproduced here:

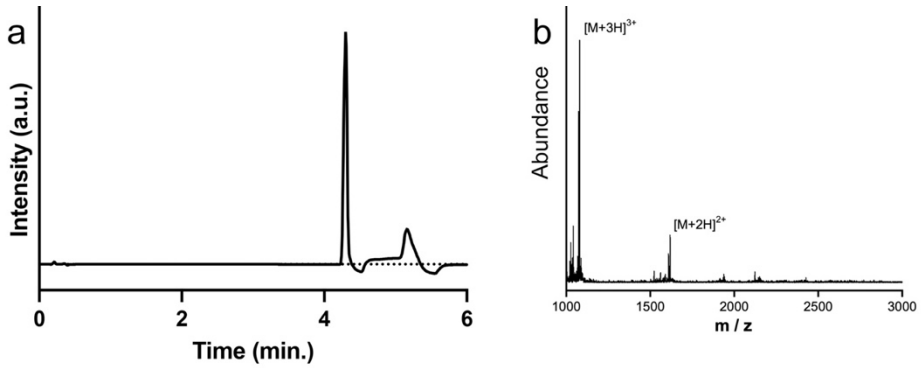


Figure S15. Peptide characterizations. a) Analytical HPLC (UPLC) of pure peptide RD1.b) ESI mass spectrometry of peptide RD1 showing the $[M+3H]^{3+}$ (Expected 1078.8, observed 1078.7) and $[M+2H]^{2+}$ (Expected $[M+2H]^{2+}$ 1617.8, observed 1617.9) ions.

b) The very shallow transition in the melting curve could also arise from the co-existence of multiple triple helices with different thermal stabilities. The presence of consecutive GRD units could promote the formation of triple helices with non-canonical strand arrangements (other than 012 trimers). NMR studies with labeled RD1 could shed light on the presence of multiple triple helices. In addition, the authors measured the height of the assemblies derived from RD2, concluding that the height is consistent with that of a single triple helix. Fig. 5g (right) shows a cartoon suggesting that the clusters from RD1 also have a height of a canonical triple helix, but there is no AFM data to support this conclusion. AFM data could shed light on whether “longer” trimers with non-canonical strand arrangements form.

RD1 and RD2 assemblies are markedly different thus the use of different methods for characterization here. While RD2 assembles into sheets whose heights can be assessed in a straightforward fashion by AFM, RD1 does not. Instead RD1’s assembly stops after some moderate number of helices have assembled. These types of aggregates are

not suitable for AFM analysis. Instead cryo-EM is able to reveal a low-resolution structure with the suggested packing shown in cartoon form.

2. “RD1 and RD2 have identical amino acids compositions and therefore might be expected to have similar thermal stability.” This sentence needs rewording since frame-shifted collagen triple helices that are composed of identical amino acids are known to have different Tm values (ref. 41).

While this is a known phenomenon to those who study collagen, this may not be immediately apparent to all readers which is why we used the phrase “might be expected”.

Other comments

1. Figure S1 is very informative. I suggest moving it to the main text.

We are not opposed to moving this figure to the main manuscript. However, we are already at 5 figures and are concerned about the overall length of the work. We will happily move this figure to the main text if the editorial staff support this decision.

(also “collagen” not “collage”).

This typo has been corrected.

2. Please put the native SP-A sequence in all the figures where you make replacements, so that the reader can easily follow how the structure changed.

We have updated Figures 2-5 with this suggestion.

3. Figures 2, 3, 4 and 5: the basic experimental parameters of the CD experiments (buffer, concentration, heating rate) need to be listed in the caption, not only in the SI, since the Tm values strongly depend on these parameters.

We have updated our figure captions with this suggestion.

Please correct spelling mistakes, typos and inaccuracies, examples:

- p.3: row 2 “relatively” is correct not “relative”, row 14: “sequence” not “sequences” Fixed
- p.4: “designed” peptides not “design peptides” Fixed
- p. 5: 2,2,2-triflouroethanol should be all lowercase. Fixed
- p. 9: “PP2” should be “PPII”. Both PP2 and PPII are commonly used in the literature.
- p. 16: "PyMol" not "pymol" Fixed
- p. 17: Fmoc not "FMOC" Fixed
- “The CD spectra give information about existing PPII helicity and triple helicity.”

This inaccurate statement occurs several times throughout the text (p5, p8, p9). On p8 "CD melts" is an incorrect term. Whereas CD spectra confirm PPII helicity, thermal denaturation provides the melting temperature of a collagen helix.

We have made our language more precise. In one instance we use the following: “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.” And later, “The CD spectra of both are presented in Fig. 3e, displaying the characteristic maximum near 224nm associated with PP2 helicity in peptide SPA-I-P, but not I-S. CD melting curve analysis supports triple helicity for SPA-I-P.”

Further

1. Experimental details: only Fmoc chemistry is mentioned, but no details on which other protection groups on the side chain of the amino acids were used

We have updated the method section with additional information.

2. In the SI, the scale bars of several cryo-EM images are missing, figures S22, S20, S17, S15 (scale bar present, but no corresponding length in caption).

In all places where scale bars are shown, the corresponding length is now mentioned in the caption.

3. Table S1: HRMS data of the peptides are missing.

The MS data has been added to Table S1.

Reviewer #1 (Remarks to the Author):

I am satisfied with the authors' response and modification of the manuscript. I recommend for publication.

We thank reviewer #1 for their assistance improving our manuscript.

Reviewer #2 (Remarks to the Author):

My concerns have been addressed in the revisions.

We thank reviewer #2 for their assistance improving our manuscript.

Reviewer #3 (Remarks to the Author):

The authors have addressed major concerns by synthesizing and characterising four new peptides. These data have strengthened their claims. Additionally, some of the data was corrected, which put one of the compound's purity under question. However, some inaccuracies still persist, including missing data such as melting curves and CD spectra for the newly introduced peptide RD2-noDAA. See my responses to the authors' rebuttal in red.

Thank you for your previous comments which have helped improve this manuscript and your continued attention to detail. For clarity we are only reproducing comments below in which the reviewer requested additional information or changes.

I do not see any CD or melting curves for RD2-no DAA. Please add them. Also fig S25c is missing the *T*m values and neither are the *T*m values mentioned in the caption. Please add the corresponding values to the figure (as you already have in all such figures).

We have updated Supporting Information **Figure S25** as requested and have added the melting curve analysis of RD2-no DAA as a new **Figure S41**. Both figures are reproduced below:

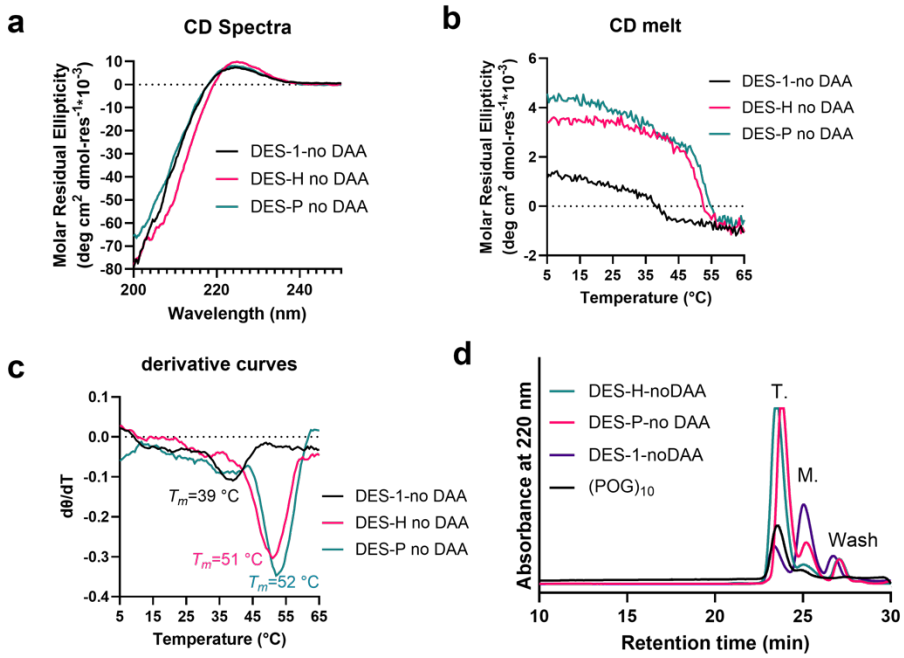


Figure S1 *CD characterization of design peptides with no "DAA" noncollagenous domain. a) CD spectra, b) CD melting curves, c) CD derivative curves. d) SEC chromatography.*

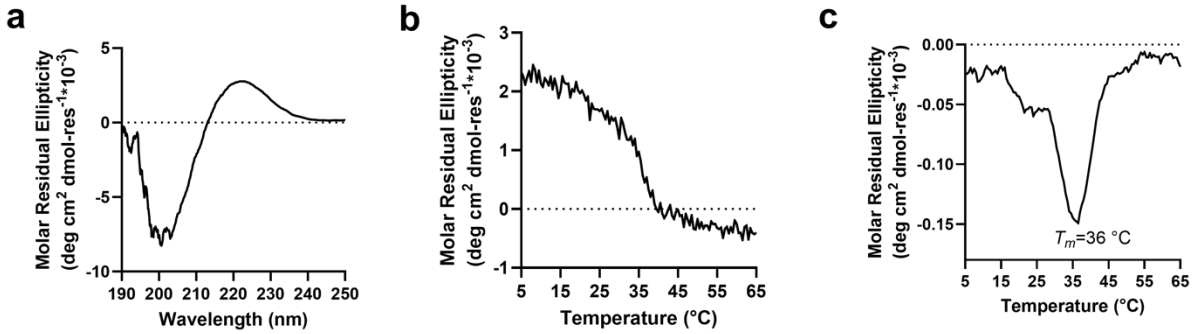


Figure S41. *CD characterization of RD2-noDAA assembly. A) CD spectrum b) thermal denaturation c) First derivative of the thermal denaturation presented in (b).*

The authors did not address the comment regarding the possible formation of triple helices with non-canonical strand arrangements (other than 012 trimers) for RD1. They claim that "the melting curve of RD1 peptide has broad thermal transition, suggesting the higher-order assemblies formed by this peptide is heterogenous". I still think that non- canonical strand arrangements in the triple helix rather than heterogeneity in higher-order assemblies are an equally plausible explanation and the previously suggested NMR studies should be performed to shed light on the origins of the shallow transition in the melting curve.

We have explored non-canonical peptide associations in several of our past publications.**[See non canonical registration references below]** Non-canonical registrations are offset from one another by more than the standard single amino acid and are typically motivated by improving amino acid pairing. For example, one strand might be offset by four amino acids if this configuration provides a more energetically favorable interaction compared to what is lost at the termini of the triple helix, where only one or two strands are present. While such an arrangement is not outside the realm of possibility, there is nothing in RD1's sequence to suggest that this is likely. See comparisons below:

Comparisons of RD1 registrations:

```
012:        DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
(canonical)  DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
              DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
Reverse lateral RD pairs: 11

015:        DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
             DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                 DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
Reverse lateral RD pairs: 9

042:        DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
              DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
Reverse lateral RD pairs: 9

045:        DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                 DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
Reverse lateral RD pairs: 9

048:        DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
                    DAAPOGPOGPOGRDGRDGRDGRDGPOGPO
Reverse lateral RD pairs: 7
```

There are three reasons we believe these non-canonical registrations are not likely: First, no important new pairwise interactions are created by these offsets. In fact, all four non-canonical registers have fewer good pairwise interactions. More extreme examples of non-canonical registrations (not shown) would have even fewer. Second, as the offset increases, important backbone hydrogen bonding interactions are lost at both termini. Third, the DAA sequence lacks a glycine residue critical for effective steric packing within a triple helix. At the termini this has minimal impact on stability overall, but when this is drawn in closer to the center of the helix (as in all non-canonical registrations) it plays a larger, destabilizing role.

With respect to the suggested NMR experiment, we have published NMR determination of collagen triple helix registration data many times previously **[See NMR references below]** and this data would be very nice data to have had here. However, our experience with these higher order aggregates is that the data is extremely difficult, if not impossible to get. The mass of RD-1 is 2,909. A triple helix would be a manageable 8727 daltons, while an octadecamer of this peptide is 52,362 daltons. At this high size and mass, the tumbling rate results in very poor line width. This exasperates the already extremely challenging problem of registration determination in a molecule with very high degeneracy. In related systems we spent a great deal of effort trying to collect exactly this data to no avail. It was the failure of those experimental methods that led us to our current collaborative effort with the Egelman laboratory and electron microscopy characterization which has generally been found to be much more fruitful.

Finally, we believe that the statement of "heterogeneity" we use includes the possibility of non-canonical registers. Since we don't have definitive results to confirm the specifics of the registration, we would prefer to avoid making such a hypothesis in the manuscript.

[Non-Canonical Registers]

A. A. Jalan, K. A. Jochim, J. D. Hartgerink "Rational Design of a Non-canonical "Sticky-ended" ABC-Collagen Triple Helix" *J. Am. Chem. Soc.*, **2014**, 136, 7535-7538. DOI: 10.1021/ja5001246.

D. R. Walker, S. A. Hulgan, C. M. Peterson, I. Li, K. J. Gonzalez, J. D. Hartgerink "Predicting the stability of homotrimeric and heterotrimeric collagen helices" *Nature Chem.* **2021**, 13, 260-269. DOI: 10.1038/s41557-020-00626-6.

[NMR of collagen triple helices]

J. A. Fallas and J. D. Hartgerink "Computational Design of Self-Assembling Register-Specific Collagen Heterotrimers" *Nat. Comm.* **2012**, 3, 1087. DOI:10.1038/ncomms2084.

A. A. Jalan, B. Demeler, J. D. Hartgerink "Hydroxyproline-Free Single Composition ABC Collagen Heterotrimer" *J. Am. Chem. Soc.* **2013**, 135, 6014-6017. DOI: 10.1021/ja402187t.

D. R. Walker, A. A. Alizadehmojarad, A. Kolomeisky, J. D. Hartgerink "Charge-free, stabilizing amide-π interactions can be used to control collagen triple helix self-assembly" *Biomacromolecules* **2021**, 22, 5, 2137-2147. DOI: 10.1021/acs.biomac.1c00234.

C.C. Cole, D.R. Walker, S.A.H. Hulgan, B.H. Pogostin, J.W.R. Swain, M.D. Miller, W. Xu, R. Duella, M. Misiura, X. Wang, A.B. Kolomeisky, G.N. Phillips Jr., and J.D. Hartgerink "Heterotrimeric Collagen Helix with High Specificity of Assembly Results in a Rapid Rate of Folding" *Nature Chemistry* **2024**, available online. DOI: 10.1038/s41557-024-01573-2.

• p. 16: "PyMol" not "pymol" on p17

This has now been corrected.

p9: "the CD spectra of all six peptides show a characteristic signal of 225 nm, indicative of triple helix formation (S24a and S25a)" Here once again inaccurate language, please rephrase: CD spectra of all six peptides show a characteristic signal of 225 nm, but it is the is the melting curves that are indicative of

triple helical formation, not the maxima of 225 nm signal on the CD curve. Same inaccuracy still stands in other places throughout the manuscript: p10, p 13.

On **page 9** this text has been changed to, “The CD spectra and melting curve analysis of all six peptides are indicative of triple helix formation based on a positive peak near 225nm which undergoes a cooperative transition upon heating (Fig 4bc, Fig. S24a, and S25).”

On **page 10** this text has been changed to, “CD spectra and melting curve analysis in Fig. S26 suggests peptide DES-1 folded into collagen triple helices at all tested pH values.”

On **page 13** this text has been changed to, “Both peptides assembled into collagen triple helices as indicated by the CD spectrum and melting curve analysis presented in Fig. 5bcd.”

Additional small mistakes:

- Typo Fmoc-ASP(OtBu)-OH, should be “Asp”.

Capitalization has now been corrected.

- p9 bottom: typo and need of rephrasing: “Three additional peptides (DES-1-noDAA, DES-P-noDAA and DES-H-noDAA remove the N-terminal triplet.” – sentence seems incomplete

This sentence was corrected to read, “Three additional peptides (DES-1-noDAA, DES-P-noDAA and DES-H-noDAA were prepared in which the N-terminal triplet was removed.”

The authors have addressed major concerns by synthesizing and characterising four new peptides. These data have strengthened their claims. Additionally, some of the data was corrected, which put one of the compound's purity under question. However, some inaccuracies still persist, including missing data such as melting curves and CD spectra for the newly introduced peptide RD2-noDAA. See my responses to the authors' rebuttal in red.

Reviewer #3 (Remarks to the Author):

The authors studied the sequence-structure relationship of assemblies formed by collagen-like peptides using the SP-A protein as a model system. Evaluation of the self-assembly of the collagen-like region of SP-A by cryo-EM and AFM identified the formation of different architectures, including oligomeric bundles, polymeric ribbons and 2D nanosheets. Whereas the study did not provide general lessons for the design of hierarchical collagen-based structures, the insights into the underexplored self-assembly of PPII helical higher-order structures are valuable. These data will be of interest to the scientific community and specialists in the field of peptide-based supramolecular assemblies. However, several control experiments are missing, which could impact the results and their interpretation. In addition, based on a rather minimalistic data set, several conclusions are likely too general. Furthermore, one of the studied compounds is not sufficiently pure, with a possibly effect on its assembly properties.

Thank you for your detailed critique. Based on your suggestions, and those of the other reviewers, we have synthesized and characterized five new peptides that substantially help to elucidate the mechanisms of assembly and create an overall stronger manuscript. We have also provided new analysis of the one peptide with purity problems. Details of the changes made and response to specific criticisms follow in blue underneath each comment.

Mutation Studies Reveal Sequence-Structure Relationship Section

1. "Even though the Ile to Pro mutant has enhanced thermal stability and favored triple helix formation, it did not promote oligomerization. This suggests that oligomerization is not correlated with triple helix formation, a critical finding."

This conclusion could be incorrect since two variables changed: the replacement of Ile by Pro and, as a consequence, the Tm. For instance, upon deleting a C-terminal POG repeat, the modified SPA-rat peptide would likely not form a stable triple helix and also not the higher-order assemblies. In contrast, an additional POG unit would stabilize the triple helix (while keeping the key Ile residue), and the system could have a higher propensity to oligomerize. Here, the conclusion would be the opposite: oligomerization correlates with triple helix formation and stability. Without such control experiments, the only valid conclusion is that the Ile residue is important for the oligomerization of SPA-rat.

We have reworded the sentences in question as follows, “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.”

OK

Also, the term “mutation” refers to DNA, use “replacement of Ile by Pro.

Throughout the manuscript we have changed this to “substitution” or “replacement” instead of mutation.

OK

Also, I would use “extended assembly” rather than “oligomerization” since oligomerization refers to covalent bonds.

The term “oligomerization” refers to more than covalent bonds and generally refers to anything of which there are several of something. It is commonly used in work related to protein association and self-assembly more generally. For example, see 10.1126/science.aad8865 or 10.1002/pro.5560040818. Based on this we left the term “oligomerization” in the manuscript.

This sentence has been reworded as follows, “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.”

OK

2. “The point-mutation studies together revealed that charged-pair interactions involving arginine and aspartate and hydrophobic interactions may exist to favor collagen oligomerization.” This conclusion is likely too “big”, considering the rather minimalistic study.

a) The study focuses on one amino acid variations in the stem collagen region. Here, the authors replaced only Arg24 and Asp25, but there are two more Asp residues within the stem region. In addition, Asp25 is in the Yaa position, whereas Asp33 is in the Xaa position (per collagen canonical sequence). As Xaa and Yaa positions are not equal within the triple helix (Xaa position is, for example, more solvent exposed), substituting an Asp residue in the Xaa position (e.g. Asp33) could have different effect on the structure and its assembly than substituting an Asp in the Yaa position. Thus, the conclusion could change significantly, e.g., substituting Asp33 could have no impact on

the triple helical or the further assembly, which would then contradict the current conclusion.

See next answer below.

b) For the positively charged residues, the argument is the same. Lys31 is in the Yaa, compared to the mutated Xaa Arg24. Thus, to corroborate the conclusion Arg27 needs to be substituted.

The scope of the current work is a targeted substitution study on natural SPA in an attempt to identify amino acids important for its self-assembly into an octadecamer. Based on more than a decade of work with peptides that assemble into collagen triple helices, but show no sign of any oligomerization, the RDGRDG triplets and repetitious hydrophobic groups every three triplets in SPA (as well as similar patterns observed in other defense collagens) seemed unusual and worth a targeted study. Our studies indeed found that these sequences were important in the context of natural SPA, and also that these features could be implemented in *de novo* designed peptides (*eg.* DES-1).Without the use of some chemical intuition as to what all these amino acids are doing, every amino acid of the SPA sequence could conceivably play a role in self-assembly and one could easily suggest dozens, if not hundreds, of single and double substitution experiments that could potentially be enlightening. In the future we expect to carry out some of those experiments ourselves and hope to see work from other groups investigate this further as well.

Within the context of the current work we feel that the substitutions that were investigated were appropriate and sufficient since they led to successful criteria for *de novo* design. We have tried to capture this idea more effectively in the closing two sentences of this section which says, “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.”

OK

Design of pH-responsive Nanoribbons

1. p.9: Initially, a physiological pH is mentioned for folding, and a few lines later, pH 6.1. On p10, the default annealing condition at pH 6.1 is mentioned. Which pH was used? Were all peptides assembled in the same buffer and at the same concentration? This is important for comparisons as the buffer and the ionic strength of the solutions can have a significant impact on self-assembling systems.

2. Fig. 4b: It would be better to show the raw melting curve rather than the CD spectrum as it carries more critical information. The CD spectrum can be moved into the SI.

To help clarify this we have changed the wording to "Peptides were prepared at a concentration of 3mM at a pH of 6.1 ± 0.2 in 20mM MES buffer, see methods for details." The methods have been clarified to state,

"Purified peptide solutions were freeze-dried with a FreeZone 4.5L 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 1M 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 1M HCl solution and the sample at pH 10.0 was adjusted with 1M NaOH. Peptide solutions were confirmed by pH meter to be the expected pH values before characterizations."

We have updated figure 4 as suggested.

OK

3. "No higher-order assembly was observed with the DES-H peptide, indicating that the hydrophobic residues are insufficient to drive assembly alone and that the charged residues included in DES-1 are critical."

A very nice step-wise study was pursued here, however, with a missing control. The authors show the parent peptide, then introduce only hydrophobic residues and then hydrophobic and charged residues. However, the peptide DAA(GPO)4(GRD)2(GPO)4 is missing. This control would show the influence of only the charged residues. If this peptide formed higher-order assemblies, it would strengthen the claim above. If no assembly formed or the assemblies were rather amorphous, the hydrophobic residues may not be sufficient driving forces for the assembly but still be important for directing the assembly. I note that the subsequent exploration of peptides RD1 and RD2 does not compensate for this lack of control, as these have double the amount of charged residues.

Thank you for this suggestion. We have now synthesized and characterized this control peptide (called DES-P) as well as the variant that lacks the "DAA" triplet (called DES-P-noDAA). While both of these peptides self-assemble into triple helices, neither self-assemble beyond the triple helix to higher order aggregates, as expected. Data for DES-P is incorporated into **Figure 4**. DES-P-noDAA is shown in **supporting information figures S24** and **S25**. These two additional peptides help to clarify the role of each of the three designed regions of the "DES" series of peptides. A new / updated paragraph has been added to the manuscript reproduced here:

"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."

OK

Additionally, no controls are presented of how the assemblies would look without the DAA domain in both DES-H and DES-1. The authors claim, "We added a short noncollagenous domain of "DAA" at the N-terminus of the peptide as the presence of noncollagenous domain reduces unwanted aggregation and favors a more specific oligomer formation." Peptides DES-H and DES-1 have, however, a completely different sequence and the findings from SPA-bcb cannot be generalized. The evaluation of a DES-1 analog lacking DAA would be an appropriate control.

As mentioned above, and in response to similar critique from reviewer 1, we have prepared and characterized a complete series of peptides in which the "DAA" triplet has been removed:

Peptide Name DES-1-no DAA DES-H-no DAA DES-P-no DAA RD1-no DAA RD2-no DAA

Sequence

GIOGPOGPOGLOGRDGRDGVOGPOGPOGPM
GIOGPOGPOGLOGPOGPOGVOGPOGPOGPM
GPOGPOGPOGRDGRDGPOGPOGPOGPOGPO
GPOGPOGPOGRDGRDGRDGRDGPOGPOGPO
GPOGRDGRDGPOGPOGPOGPOGRDGRDGPO

RD1-noDAA was in the original manuscript, but all the others are new for this revision. In summary, we find that for systems that the DAA triplet is extremely important for the control of self-assembly in higher order systems. For example, DES-1 undergoes higher order assemblies including 15- and 18-mers. Removal of the DAA removes a degree of electrostatic repulsion and because of this we find that DES-1-noDAA aggregates and precipitates. In contrast systems that only form simple triple helices such as DES-H and DES-P, the DAA triplet can be removed with only very small changes in assembly state (they still form simple triple helices). RD2 and RD2-no DAA displayed similar behavior in which the "no DAA" variant generated much larger aggregates. RD1 and RD1-noDAA behaved somewhat differently in that elimination of the DAA triplet completely eliminated triple helical folding and the peptide appeared to remain soluble suggesting that the DAA triplet was involved in some way in the stabilization of the ring-like higher order aggregates observed. While this is a bit different than the other cases, in every case observed well controlled higher order assembly required the DAA triplet.

We have updated our manuscript with a discussion of these results which can be found on **pages 9-10** and **15-16** of the updated manuscript. Complete characterization is found in the SI in particular figures **S24** and **S25**.

I do not see any CD or melting curves for RD2-no DAA. Please add them. Also fig S25c is missing the T_m values and neither are the T_m values mentioned in the caption. Please add the corresponding values to the figure (as you already have in all such figures).

4. "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." – This statement is too general. The rationale for choosing DAA as a non-collagenous domain within the new peptides is reasonable. However, appropriate control experiments are needed to corroborate the claim above. This DAA sequence has not been part of any of the previous peptides. The pI argument should, for example, be probed by substituting Asp in DAA of DES-1 with a more acidic amino acid, e.g. cysteic acid (Cya), that remains deprotonated at pH 3.5.

We feel that the new peptides and the results from them described above now more effectively justify our initial claim.

OK

De novo Design of Heterogeneous Oligomeric Bundle and 2D Nanosheets Section

1. The melting curve (Fig 5c,d) is broad and shallow. The authors reason that this is due to multiple species of higher-order assemblies present, but there are alternative explanations:

a) Impurities. The UPLC trace (Fig S11) shows a shoulder at the head of the major peak and an additional peak at ~3.2 min, which indicates impurities. Based on the analytical data, this sample has a purity of less than 85%. Such an amount of impurities likely influences the results. The CMP needs to be purified and all relevant experiments repeated with the purified sample.

Thank you for this criticism. After significant reevaluation of this data, we uncovered a rather embarrassing mistake on our part. The UPLC trace reported in the initial SI was actually of the crude material, not the HPLC purified material. We performed new UPLC and MS experiments on samples that had been purified, and were the same materials that were studied elsewhere in the manuscript for SEC and electron microscopy. The UPLC analysis of the purified material looks clean, unlike what we originally reported. Mass spec results also support the purity of these samples. **Supporting Information Figure S15**

OK

b) The very shallow transition in the melting curve could also arise from the co-existence of multiple triple helices with different thermal stabilities. The presence of consecutive GRD units could promote the formation of triple helices with non-canonical strand arrangements (other than 012 trimers). NMR studies with labeled RD1 could shed light on the presence of multiple triple helices. In addition, the authors measured the height of the assemblies derived from RD2, concluding that the height is consistent with that of a single triple helix. Fig. 5g (right) shows a cartoon suggesting that the clusters from RD1 also have a height of a canonical triple helix, but there is no AFM data to support this conclusion. AFM data could shed light on whether "longer" trimers with non-canonical strand arrangements form.

RD1 and RD2 assemblies are markedly different thus the use of different methods for characterization here. While RD2 assembles into sheets whose heights can be assessed in a straightforward fashion by AFM, RD1 does not. Instead RD1's assembly stops after some moderate number of helices have assembled. These types of aggregates are not suitable for AFM analysis. Instead cryo-EM is able to reveal a low-resolution structure with the suggested packing shown in cartoon form.

The authors did not address the comment regarding the possible formation of triple helices with non-canonical strand arrangements (other than 012 trimers) for RD1. They claim that "the melting curve of RD1 peptide has broad thermal transition, suggesting the higher-order assemblies formed by this peptide is heterogenous". I still think that non-canonical strand arrangements in the triple helix rather than heterogeneity in higher-order assemblies are an equally plausible explanation and the previously suggested NMR studies should be performed to shed light on the origins of the shallow transition in the melting curve.

2. "RD1 and RD2 have identical amino acids compositions and therefore might be expected to have similar thermal stability." This sentence needs rewording since frame-shifted collagen triple helices that are composed of identical amino acids are known to have different Tm values (ref. 41).

While this is a known phenomenon to those who study collagen, this may not be immediately apparent to all readers which is why we used the phrase "might be expected".

OK

Other comments

1.Figure S1 is very informative. I suggest moving it to the main text.

We are not opposed to moving this figure to the main manuscript. However, we are already at 5 figures and are concerned about the overall length of the work. We will happily move this figure to the main text if the editorial staff support this decision.

OK

(also “collagen” not “collage”).

This typo has been corrected.

OK

2. Please put the native SP-A sequence in all the figures where you make replacements, so that the reader can easily follow how the structure changed.

We have updated Figures 2-5 with this suggestion.

OK

3. Figures 2, 3, 4 and 5: the basic experimental parameters of the CD experiments (buffer, concentration, heating rate) need to be listed in the caption, not only in the SI, since the Tm values strongly depend on these parameters.

We have updated our figure captions with this suggestion.

OK

Please correct spelling mistakes, typos and inaccuracies, examples:

• p.3: row 2 “relatively” is correct not “relative”, row 14: “sequence” not “sequences” Fixed OK

• p.4: “designed” peptides not “design peptides” Fixed OK

• p. 5: 2,2,2-triflouroethanol should be all lowercase. Fixed OK

• p. 9: “PP2” should be “PPII”. Both PP2 and PPII are commonly used in the literature. OK

• p. 16: "PyMol" not "pymol" Fixed, not fixed on p17

• p. 17: Fmoc not "FMOC" Fixed OK

• “The CD spectra give information about existing PPII helicity and triple helicity.”

This inaccurate statement occurs several times throughout the text (p5, p8, p9). On p8 "CD melts" is an incorrect term. Whereas CD spectra confirm PPII helicity, thermal denaturation provides the melting temperature of a collagen helix.

We have made our language more precise. In one instance we use the following: “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." And later, "The CD spectra of both are presented in Fig. 3e, displaying the characteristic maximum near 224nm associated with PP2 helicity in peptide SPA-I-P, but not I-S. CD melting curve analysis supports triple helicity for SPA-I-P."

p9: "the CD spectra of all six peptides show a characteristic signal of 225 nm, indicative of triple helix formation (S24a and S25a)" Here once again inaccurate language, please rephrase: CD spectra of all six peptides show a characteristic signal of 225 nm, but it is the is the melting curves that are indicative of triple helical formation, not the maxima of 225 nm signal on the CD curve. Same inaccuracy still stands in other places throughout the manuscript: p10, p 13.

Further

1. Experimental details: only Fmoc chemistry is mentioned, but no details on which other protection groups on the side chain of the amino acids were used

We have updated the method section with additional information.

OK

2. In the SI, the scale bars of several cryo-EM images are missing, figures S22, S20, S17, S15 (scale bar present, but no corresponding length in caption).

In all places where scale bars are shown, the corresponding length is now mentioned in the caption.

OK

3. Table S1: HRMS data of the peptides are missing.

The MS data has been added to Table S1.

OK

Additional small mistakes:

- Typo Fmoc-ASP(OtBu)-OH, should be "Asp".
- p9 bottom: typo and need of rephrasing: "Three additional peptides (DES-1-noDAA, DES-P-noDAA and DES-H-noDAA remove the N-terminal triplet." – sentence seems incomplete