Published in final edited form as:

Proteins. 2013 March ; 81(3): 386–393. doi:10.1002/prot.24194.

Sequence Recombination Improves Target Specificity in a Redesigned Collagen Peptide *abc*-type Heterotrimer

Sumana Giddu, Fei Xu, and Vikas Nanda

Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane West, Piscataway, New Jersey 08854

Abstract

Stability of the collagen triple helix is largely governed by its imino acid content, namely the occurrence of proline and 4R-hydroxyproline at the X and Y positions respectively of the periodic (Gly-X-Y)_n sequence. Although other amino acids at these positions reduce stability of the triple helix, this can be partially compensated by introducing intermolecular side chain salt bridges. This approach was previously used to design an *abc*-type heterotrimer composed of one basic, one acidic and one neutral imino acid rich chain (Gaub & Hartgerink, 2007). In this study, an *abc*-type heterotrimer was designed to be the most stable species using a sequence recombination strategy that preserved both the amino acid composition and the network of interchain salt-bridges of the original design. The target heterotrimer had the highest T_m of 50°C, 7°C greater than the next most stable species. Stability of the heterotrimer decreased with increasing ionic strength, consistent with the role of intermolecular salt bridges in promoting stability. Quantitative meta-analysis of these results and published stability measurements on closely related peptides was used to discriminate the contributions of backbone propensity and side chain electrostatics to collagen stability.

Keywords

backbone propensity; proline; hydroxyproline; electrostatics; salt bridge; folding; circular dichroism; triple-helix; energy landscape

Introduction

Protein design tests our fundamental understanding of the forces that govern molecular stability and structure. Collagen presents many unique challenges as a design target. In most proteins, an extensive hydrophobic core stabilizes the folded structure, but the collagen triple-helix lacks buried hydrophobic interactions. Instead a network of backbone hydrogen bonds promotes assembly. Amino acid side chains are extensively solvated and are flexible due to minimal tertiary structure constraints. This makes it challenging to model and calculate their contributions to protein energetics. Fibril forming regions of natural collagens can extend over a thousand amino acids, making them difficult to express and characterize. As such, much of our understanding of collagen folding has come from work on short collagen mimetic peptides. These have also served as a powerful platform for developing molecular design rules to manipulate stability and specificity of assembly.

Three polypeptide chains combine to form the collagen triple-helix, each comprised of a canonical (Gly-X-Y)_n repeat. Nearly all amino acids are tolerated at both X and Y positions,

but the glycine is essential for folding. Glycine mutations significantly reduce stability in model systems^{1,2} and result in collagen-mediated pathologies.³ Natural collagens are rich in imino acids with proline at position X and hydroxyproline at position Y, which serve to stabilize the triple-helix fold.^{4,5} The most stable collagen sequence using biogenic amino acids consist solely of repeating Gly-Pro-Hyp triplets (Hyp or O = 4R-hydroxyproline), and one can roughly estimate the stability of collagen based on the imino acid content of the sequence alone.⁶

Following Pro and Hyp, the next most over-represented amino acids in natural collagens are acidic and basic residues.^{7,8} Often, these form networks of favorable interchain charge pairs that contribute to triple-helix stability.^{9–11} Due to the regular structure of collagen, it is possible to infer the presence of side chain charge pairs from sequence. Amino acids at the Y position on one chain are in close proximity to X and X' positions on the adjacent chain. Side chain salt bridges can confer significant stability, in some cases rivaling the contributions of imino acids.¹²

These two features of collagen: networks of side chain electrostatic interactions, and high imino acid content, were designed into a series of collagen peptides to promote formation of an *abc*-type heterotrimer.¹³ Three peptides **R10**, **E10** and **POG** were studied with sequences (Pro-Arg-Gly)₁₀, (Glu-Hyp-Gly)₁₀ and (Pro-Hyp-Gly)₁₀ respectively. The charged amino acids formed a network of interchain salt-bridges; ideally, a maximum of nineteen ion pairs could exist in the heterotrimer (Figure 1A). The neutral **POG** chain did not contribute significant side chain interactions, but promoted stability through its rich imino acid content. The **E10:R10:POG** heterotrimer had a stability of 54°C, superseded by the **POG** homotrimer ($T_m = 67.5^\circ\text{C}$). Upon thermal denaturation, a single transition attributed to an *abc*-type heterotrimer was observed if the peptides were first heated and annealed.

We hypothesized a thermodynamically optimal *abc*-type species could be obtained starting from the original peptide sequences. The **E10**, **R10** and **POG** peptides sequences were recombined as to maintain key physiochemical elements. The recombination would specifically destabilize competing states (*aaa*, *aab*, etc), leaving the *abc*-type heterotrimer as the most stable species. This strategy is referred to as *negative* design. Positive design features the introduction of amino acid substitutions that favor stability of the target conformation, whereas negative design promotes target formation by disfavoring competing states. The redesign was successful, allowing us to develop a quantitative model relating the contributions of backbone propensity and side chain electrostatics to triple-helix stability.

Methods

Peptides synthesis

Peptides **R**, **Z**, and **E** were synthesized at the Tufts University Core Facility (<http://tufts.org>) using solid-phase synthesis Fmoc chemistry.

R: Ac – PRGPRGPRGPRGPRGPOGPOGPOGPOG – NH₂

E: Ac – POGPOGPOGPOGPOGEEOGEEOGEEOG – NH₂

Z: Ac – EOGEEOGEEOGEEOGPRGPRGPRGPRG – NH₂

The peptides were purified by reverse-phase HPLC to greater than 95% purity. The identities of products were verified by mass spectrometry (Figure S1). Peptide concentrations were determined by measuring absorbance at 214 nm, using an extinction coefficient of 2200 M⁻¹ c⁻¹ per peptide bond.

Measuring Structure and Stability by Circular Dichroism

Circular dichroism (CD) experiments were performed on AVIV Model 400 CD Spectrophotometer (AVIV Biomedical Inc., Lakewood, NJ), with optically matched 0.1 cm path length quartz cuvettes (model 110-OS; Hellma USA). Peptide concentrations of 0.2 mM were prepared in 10 mM phosphate buffer at pH 7.0. Two annealing protocols were used to assess the role of assembly kinetics in modulating the distribution of species. For slow cooling, samples were incubated at 80°C for 45 minutes, gradually cooled to 50°C for 30 min, then 25°C for 30 minutes, and finally 4°C for 48 hours. For rapid cooling, samples were immediately moved into 4°C conditions after incubating at 80°C for 45 minutes.

Wavelength scans were conducted from 190 to 260 nm at 0 °C (averaging time 1.0 s). Values were reported as molar ellipticity corrected for concentration, sequence length and path length. Thermal denaturation was monitored using ellipticity at 225 nm. Temperature was raised from 0 to 80 °C in 0.3 °C steps with an equilibration time of 2 min at each step. Savitzky-Golay smoothing of the denaturation profiles was carried out over a span of eleven points using a third-order polynomial.¹⁴ The T_m 's of melting transitions were assigned to extrema of the first derivative of the denaturation profile. Reported values and standard error are from triplicate thermal denaturation experiments. To assess the effect of ionic strength, peptides were allowed to fold in the presence of salt at a series of concentrations: 0.0, 0.01, 0.1, 0.5 and 1 M NaCl.

Results and Discussion

Design Strategy

The original sequences¹³ were bisected into 15-residue blocks and recombined so that charged amino acids and imino acids were redistributed among the three peptides:

R (basic): (Pro-Arg-Gly)₅ (Pro-Hyp-Gly)₅ = $\frac{1}{2}$ **R10** + $\frac{1}{2}$ **POG**

E (acidic): (Pro-Hyp-Gly)₅ (Glu-Hyp-Gly)₅ = $\frac{1}{2}$ **POG** + $\frac{1}{2}$ **E10**

Z (zwitterionic): (Glu-Hyp-Gly)₅ (Pro-Arg-Gly)₅ = $\frac{1}{2}$ **E10** + $\frac{1}{2}$ **R10**

The total amino acid composition across the three peptides was preserved and eighteen salt bridge side chain pairs were possible, one less than the original design (Figure 1B). By eliminating the high imino-acid content **POG** peptide, it was expected that all competing species would be less stable than an **R:Z:E** heterotrimer.

Stability and Structure Determination

Peptides **E**, **R** and **Z** were synthesized, purified and reconstituted in 10 mM phosphate buffer, pH 7.0 as described in the Methods. **E** and **R** formed structured homotrimers in solution, showing the characteristic circular dichroism (CD) spectrum of a triple-helix with a positive band at 225 nm (Figure 2). The structure of **Z** was difficult to characterize due to its tendency to aggregate in the absence of **R** or **E**. Aggregation of zwitterionic collagen peptides has been observed for many systems^{15–18}, purportedly due to complementary inter-helical electrostatic interactions driving higher order assembly. Consistent with this, transmission electron microscopy of **Z** aggregates revealed nanometer scale, fibrillar structures (Figure S2). In the presence of **R** and/or **E**, **Z** no longer precipitated and was incorporated into soluble triple-helical heterotrimers.

A stable triple-helix was formed by combining **R**, **Z** and **E**. Stability was assessed by thermal denaturation monitored with CD for nine mixture stoichiometries of the three peptides: 3**R**, 3**E**, 2**E**:**R**, **E**:2**R**, 2**E**:**Z**, **E**:2**Z**, 2**R**:**Z**, **R**:2**Z** and **R**:**Z**:**E** (Figure 3, Table I). Melting temperatures were determined by finding extrema of $d\theta/dT$ in the first derivative

plot of the denaturation profile (θ = ellipticity). Combining **R**, **Z** and **E** in a 1:1:1 ratio produced a major transition at 50°C, the highest observed in any of the mixtures. Only minor differences in the stability of **R:Z:E** mixtures were noted when using a rapid or slow annealing schedule, suggesting heterotrimer formation was not sensitive to the annealing protocol (Figure 4). A mixture of species was present as evidenced by the broad unfolding transition at lower temperatures. CD cannot discriminate between single and mixed-composition species or specify the register of the triple-helix (RZE versus REZ, ERZ, EZR, ZER, ZRE)^{15,19} – but we can conclude that the optimally stable species requires the presence of all three peptides.

Binary composition heterotrimers were also formed, but these were less stable than **R:Z:E**. Combining **R** and **Z** resulted in two discernable transitions, one with the same stability as an **R** homotrimer and a less stable species that was more prevalent at higher concentrations of **Z**. Combining **E** and **Z** also resulted in two transitions, one with the same stability as an **E** homotrimer. The two transitions in both **R:Z** and **E:Z** mixtures may be explained by the presence of **R:2Z + 2R:Z** and **E:2Z + 2E:Z**, although it is not possible to assign transitions to a specific species.

The relative populations of the two species in **R:Z** or **E:Z** mixtures depended on the annealing protocol. Surprisingly, rapid annealing of the **R:Z** mixture favored a more uniform, stable transition than the slow annealing protocol. This is counter to what is commonly observed in nucleic acid annealing, where rapid annealing can result in a heterogeneous mixture of species. Although the mechanism for this effect is not clear, it is often observed that the folding/unfolding of collagen peptides can be slow – on the order of minutes to hours – and variations on the annealing protocol in at this time scale might be expected to have dramatic effects such as this.

The **R:Z:E** heterotrimer was designed such that an extensive network of salt bridges would stabilize the target oligomer. To assess the salt bridge contributions to stability, ionic strength was varied using a series of salt concentrations from 0.0 to 1.0 M NaCl. Triple helical structure was observed under all conditions, but the T_m decreased by 9°C between the lowest and highest ionic strength conditions (Figure 5, Table II).

Stability of homotrimers also varied based ionic strength. **E** homotrimers increased in stability by 10°C upon addition of salt, suggesting that repulsive interactions between adjacent amino acids of like charge disfavored triple-helix formation. The effect was less pronounced for **R** homotrimers, where stability first decreased and then increased with the addition of salt. This may have been due to an initial loss of favorable guanidyl side chain to backbone carbonyl interactions,²⁰ followed by screening of side chain repulsions at higher ionic strengths.

Additive Model for Backbone and Side chain Interactions

Success of the redesigned heterotrimer was predicated on balancing the distribution of backbone stabilizing amino acids and maintaining key side chain ion pairs specific to the target state. Straightforward sequence-structure mapping in the collagen fold has motivated sequence-derived stability prediction models based on the assumption that backbone and side chain forces additively contributed to overall stability.¹⁰ Together, the original and redesigned *abc*-type heterotrimer peptides provide a unique data set for developing a model that specifically addresses the relative importance of these interactions in this set of peptides:

$$T_m = \alpha_1 \bullet N_{\text{POG}} + \alpha_2 \bullet N_{\text{EOG}} + \alpha_3 \bullet N_{\text{PRG}} \quad \text{composition terms} \\ + \alpha_4 \bullet N_{\text{R-E}} + \alpha_5 \bullet N_{\text{R-R}} + \alpha_6 \bullet N_{\text{E-E}} \quad \text{pairwise interaction terms} \quad \text{Eq. 1}$$

N_{POG} , N_{EOG} , N_{PRG} were POG, EOG and PRG triplets counts, respectively. $N_{\text{R-E}}$ was the total number Arg-Glu interchain interactions at Y-X and Y-X' positions. The value used represented the maximal number of favorable interactions, assuming an optimal registry as depicted in Figure 1. $N_{\text{R-R}}$ and $N_{\text{E-E}}$ were the number of interchain and intrachain Arg-Arg or Glu-Glu unfavorable, spatially adjacent interactions using a model three-dimensional structure (Figure S3). Although the strength of electrostatic interactions for solvent exposed residues over 10–15 Å was expected to be small, it has been observed in globular proteins that multiple weak interactions can have significant effects on structure and stability.²¹ It was further assumed that under 1.0 M NaCl conditions, both favorable and unfavorable electrostatic interactions were completely screened.

Previous host-guest collagen peptide studies concluded T_m correlates linearly with free energy of folding over a wide temperature range, allowing us to evaluate the weights directly using experimental melting temperatures.¹¹ The model was fit to experimental stabilities for twelve peptide compositions (Table III). All peptides were the same length ($N_{\text{POG}} + N_{\text{EOG}} + N_{\text{PRG}} = 30$ triplets), with capped termini, composed solely of POG, EOG and PRG triplets. Only homotrimers and the *abc*-type heterotrimer species where stoichiometry could be reasonably assigned were used. With these assumptions a final model was determined using the nonlinear least-squares fitting module Solver in Microsoft Excel. Standard error of parameters was determined using a jackknife leave-one-out method:

$$T_m = 1.03 \pm 0.48 N_{\text{EOG}} + 1.34 \pm 0.54 N_{\text{PRG}} + 2.09 \pm 0.08 N_{\text{POG}} \\ + 0.62 \pm 0.39 N_{\text{R-E}} - 0.08 \pm 0.81 N_{\text{R-R}} - 0.34 \pm 0.60 N_{\text{E-E}} \quad \text{Eq. 2}$$

Weights describe the change in T_m per interaction – i.e. for a (POG)₁₀ homotrimer, the total stability is calculated as 2.1°C per triplet × 30 triplets ≈ 63°C. A reasonable correlation between observed and calculated stabilities was obtained with the full model ($R^2 = 0.81$) (Figure 6).

Removing the two repulsive terms from the model reduced the correlation ($R^2 = 0.67$), and if no pairwise terms were included, only a poor correlation was observed ($R^2 = 0.44$). Our previous scoring function that did not discriminate between the backbone propensities of Arg and Glu¹⁵ also gave a weak correlation of $R^2 = 0.44$.

The rank stability of triplets: POG > PRG > EOG, was consistent with previously determined amino acid preferences for the triple helix.²² Based on this model, a single Pro→Glu substitution would change stability by $\alpha_1 - \alpha_3 = -1.1^\circ\text{C}$. In a homotrimer, this would amount to a ΔT_m of -4.3°C , comparable to prior estimates of -4.4°C .¹⁰ In contrast, a single Hyp→Arg substitution was predicted to reduce stability by 0.7°C , whereas the reported effect of single substitutions in homotrimers was negligible.¹⁰ Only in cases of multiple Hyp→Arg substitutions were ΔT_m values of up to $-0.8^\circ\text{C}/\text{substitution}$ observed.²³ Non-additive contributions of multiple substitutions for other amino acid types has also been observed in both homotrimers²⁴ and heterotrimers.²⁵

Inclusion of favorable pairwise interactions between charged side chains improves the correlation, and explains the effect of high ionic strength on the stability of the **R:Z:E** mixture. Although it is assumed here that Y-X and Y-X' interactions are energetically equivalent, recent structural studies indicate Y-X' interactions contribute more to stability,

depending on the residue identities.²⁶ As such, assuming an optimal number of favorable interactions in **R:Z:E** or **R10:E10:POG** may result in underestimating the contribution of individual salt bridges to overall stability.

The **R:Z:E** mixture can assume six registries assuming optimal interstrand hydrogen bonding: RZE, ZER, ERZ, REZ, EZR and ZRE. ZER and ERZ have the potential to form sixteen favorable charge pairs, two less than RZE, which corresponds to a modest 1.6°C difference in melting temperature. We do not expect the scoring function to reliably distinguish which of these registries is formed. In contrast, EZR ZRE and REZ are not able to form any charge pairs and their calculated T_m 's are 38°C, significantly lower than the observed transition. It is unlikely that these states are significantly populated in the R:Z:E mixture.

Conclusions

The collagen heterotrimer has become a useful target for developing positive and negative protein design methods. Our redesign of an *abc*-type heterotrimers by sequence recombination used a purely negative design strategy to enhance the specificity of the target state without significantly compromising its stability. Modulating stability by redistributing imino-acids among the three sequences was recently used to generate a single composition collagen ABC heterotrimer,²⁷ indicating that it may serve as a general strategy for tuning the oligomerization energy landscape.

Quantitative analysis and molecular modeling coupled with biophysical and structural characterization are improving our understanding of the molecular forces that drive collagen folding and stability. Although the model developed here is limited to specific amino acid compositions, peptide length and types of side chain interactions, it highlights the relative importance of backbone propensities and favorable and unfavorable side chain electrostatic interactions to collagen stability.

Understanding how sequence recombination alters stability and specificity should provide insight into the evolution of heterospecific natural collagens. Fibrous proteins with periodic structure presents unique challenges to traditional phylogenetic analysis methods due to the inherent repetition of low information content sequence motifs.²⁸ The exon organization of vertebrate collagens shows evidence of gene duplication and exon shuffling.^{29–31} Synthetic proteins such as these can be used to model how extant collagens may have evolved from the repetition and shuffling of simpler structural domains.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH DP2 OD006478-01 and NSF DMR-0907273.

References

1. Baum J, Brodsky B. Real-time NMR investigations of triple-helix folding and collagen folding diseases. *Fold Des.* 1997; 2(4):R53–60. [PubMed: 9269560]
2. Bhate M, Wang X, Baum J, Brodsky B. Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide. *Biochemistry.* 2002; 41(20): 6539–6547. [PubMed: 12009919]

3. Steinmann B, Rao VH, Vogel A, Bruckner P, Gitzelmann R, Byers PH. Cysteine in the triple-helical domain of one allelic product of the alpha 1(I) gene of type I collagen produces a lethal form of osteogenesis imperfecta. *The Journal of biological chemistry*. 1984; 259(17):11129–11138. [PubMed: 6469997]
4. Burjanadze T. Hydroxyproline content and location in relation to collagen thermal stability. *Biopolymers*. 1979; 18(4):931–938. [PubMed: 435609]
5. Berg RA, Prockop DJ. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochemical and biophysical research communications*. 1973; 52(1):115–120. [PubMed: 4712181]
6. Engel J, Chen HT, Prockop DJ, Klump H. The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)_n and (L-Pro-L-Hyp-Gly)_n. *Biopolymers*. 1977; 16(3):601–622. [PubMed: 843606]
7. Salem G, Traub W. Conformational Implications of Amino Acid Sequence Regularities in Collagen. *FEBS Lett*. 1975; 51(1):94–99. [PubMed: 1123071]
8. Ramshaw JA, Shah NK, Brodsky B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol*. 1998; 122(1–2):86–91. [PubMed: 9724608]
9. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry*. 2005; 44(5):1414–1422. [PubMed: 15683226]
10. Persikov AV, Ramshaw JA, Brodsky B. Prediction of collagen stability from amino acid sequence. *J Biol Chem*. 2005; 280(19):19343–19349. [PubMed: 15753081]
11. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Peptide investigations of pairwise interactions in the collagen triple-helix. *J Mol Biol*. 2002; 316(2):385–394. [PubMed: 11851346]
12. Gauba V, Hartgerink JD. Surprisingly high stability of collagen ABC heterotrimer: evaluation of side chain charge pairs. *J Am Chem Soc*. 2007; 129(48):15034–15041. [PubMed: 17988128]
13. Gauba V, Hartgerink JD. Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J Am Chem Soc*. 2007; 129(9):2683–2690. [PubMed: 17295489]
14. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*. 1964; 36(8):1627–1639.
15. Xu F, Zhang L, Koder RL, Nanda V. De novo self-assembling collagen heterotrimers using explicit positive and negative design. *Biochemistry*. 2010; 49(11):2307–2316. [PubMed: 20170197]
16. Rele S, Song Y, Apkarian RP, Qu Z, Conticello VP, Chaikof EL. D-periodic collagen-mimetic microfibers. *J Am Chem Soc*. 2007; 129(47):14780–14787. [PubMed: 17985903]
17. O'Leary LE, Fallas JA, Bakota EL, Kang MK, Hartgerink JD. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nature chemistry*. 2011; 3(10):821–828.
18. Xu F, Li J, Jain V, Tu RS, Huang Q, Nanda V. Compositional control of higher order assembly using synthetic collagen peptides. *Journal of the American Chemical Society*. 2012; 134(1):47–50. [PubMed: 22171825]
19. Fallas JA, Gauba V, Hartgerink JD. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J Biol Chem*. 2009; 284(39):26851–26859. [PubMed: 19625247]
20. Chan VC, Ramshaw JA, Kirkpatrick A, Beck K, Brodsky B. Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix. *J Biol Chem*. 1997; 272(50):31441–31446. [PubMed: 9395477]
21. Lee KK, Fitch CA, Garcia-Moreno EB. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci*. 2002; 11(5):1004–1016. [PubMed: 11967358]
22. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Amino acid propensities for the collagen triple-helix. *Biochemistry*. 2000; 39(48):14960–14967. [PubMed: 11101312]
23. Yang W, Chan VC, Kirkpatrick A, Ramshaw JA, Brodsky B. Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. *J Biol Chem*. 1997; 272(46):28837–28840. [PubMed: 9360948]

24. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Triple-helix propensity of hydroxyproline and fluoroproline: comparison of host-guest and repeating tripeptide collagen models. *J Am Chem Soc.* 2003; 125(38):11500–11501. [PubMed: 13129344]
25. Gauba V, Hartgerink JD. Synthetic collagen heterotrimers: structural mimics of wild-type and mutant collagen type I. *J Am Chem Soc.* 2008; 130(23):7509–7515. [PubMed: 18481852]
26. Fallas JA, Dong J, Tao YJ, Hartgerink JD. Structural insights into charge pair interactions in triple helical collagen-like proteins. *The Journal of biological chemistry.* 2012; 287(11):8039–8047. [PubMed: 22179819]
27. Fallas JA, Lee MA, Jalan AA, Hartgerink JD. Rational design of single-composition ABC collagen heterotrimers. *Journal of the American Chemical Society.* 2012; 134(3):1430–1433. [PubMed: 22239117]
28. Barua B, Pamula MC, Hitchcock-DeGregori SE. Evolutionarily conserved surface residues constitute actin binding sites of tropomyosin. *P Natl Acad Sci USA.* 2011; 108(25):10150–10155.
29. Patthy L. Genome evolution and the evolution of exon-shuffling--a review. *Gene.* 1999; 238(1): 103–114. [PubMed: 10570989]
30. Boot-Handford RP, Tuckwell DS. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *BioEssays: news and reviews in molecular, cellular and developmental biology.* 2003; 25(2):142–151.
31. Heino J, Huhtala M, Kapyla J, Johnson MS. Evolution of collagen-based adhesion systems. *Int J Biochem Cell Biol.* 2009; 41(2):341–348. [PubMed: 18790075]

Figure 1. Recombination Design Strategy

(A) Original design (ref 13) modeled onto a triple-helix backbone. Interactions between Glu C δ and Arg C ζ within 8.0 Å are shown between the sequences of **R10**, **E10** and **POG**. (B) Structural cartoon of the current design with **R** (white), **Z** (light orange) and **E** (dark orange). The three-dimensional cartoons reflect sidechain orientations that form the maximal number of interchain charge pairs. Such conformations does not satisfy other known stabilizing interactions such as backbone carbonyl hydrogen bonding to the sidechain guanidyl group of the Arg sidechain (ref 20).

Figure 2. CD spectrum of triple-helix homotrimers

Wavelength spectrum of **E** (red) and **R**(black) peptide homotrimers at 4°C are consistent with formation of a triple-helix. Mean residue ellipticity (MRE) is in units of $10^3 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol} \cdot \text{residue}^{-1}$

Figure 3.

Thermal denaturation profiles (lines) and first-derivate plots (circles) for peptide mixtures. All samples are in 10 mM phosphate buffer, pH 7.

Figure 4.
Thermal denaturation of peptide mixtures following gradual annealing (black) or rapid annealing (red).

Figure 5.

Thermal denaturation profiles for the R:Z:E peptide mixture and R, E homotrimers at a series of NaCl concentrations: 0.0 M (bold black lines/circles), 0.01, 0.1, 0.5 and 1.0 M (red lines/circles). T_m values are in Table II.

Figure 6. Experimental versus model stabilities for the original design (ref 13) (black circles) and data from this study (red circles). Results of the fit are in Eq. 2.

Table I

Measured T_m for R, Z and E peptide mixtures.

peptide	T_m (°C)
R	42.5 ± 0.3
E	29.6 ± 0.6
R:2Z	$37.1 \pm 1.1, 43.0 \pm 0.6$
2R:Z	42.5 ± 0.7
R:2E	40.3 ± 0.7
2R:E	42.5 ± 0.6
E:2Z	$31.0 \pm 0.9, 40.5 \pm 0.3$
2E:Z	$30.3 \pm 0.6, 37.8 \pm 0.4$
R:Z:E	48.0 ± 0.7

Table II

Effect of ionic strength on the stability of R, E homotrimers and the R:Z:E mixture.

[NaCl] (M)	T _m (°C)		
	R:Z:E	R	E
0	50°C	43°C	31°C
0.01	50	43	34
0.1	47	42	37
0.5	43	43	40
1	41	46	42

Denaturation profiles are in Figure 5.

Table III

Parameters for the additive model (eq.2).

peptide	N _{POG}	N _{EOG}	N _{PRG}	N _{RE}	N _{RR}	N _{EE}	exp T _m	calc T _m
R	15	-	15	-	38	-	42.5°C	46.6°C
E	15	15	-	-	-	38	29.6	32.5
R:Z:E	10	10	10	18	8	8	48.0	52.6
E10:R10:POG	10	10	10	19	9	9	54 ^a	52.8
E3	21	9	-	-	-	9	51 ^a	51.5
R3	21	-	9	-	9	-	58 ^a	56.5
E5	15	15	-	-	-	15	46 ^a	41.5
R5	15	-	15	-	15	-	55 ^a	49.8
POG	30	-	-	-	-	-	62.5 ^b	66.4
R:Z:E + NaCl	10	10	10	-	-	-	41	44.0
E + NaCl	15	15	-	-	-	-	42	47.3
R + NaCl	15	-	15	-	-	-	46	51.9

^afrom ref 13.^a“+ NaCl” indicates expected conditions in 1.0 M NaCl.^bdenaturation profile in Figure S4.