Circular Permutation Directs Orthogonal Assembly in Complex Collagen Peptide Mixtures*

Fei Xu, Teresita Silva, Mihir Joshi, Sohail Zahid, and Vikas Nanda

From the Center for Advanced Biotechnology and Medicine, Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, New Jersey 08854

**Background:** Multiple types of collagens specifically assemble and co-exist as a major component in extracellular matrix. Although noncollagenous trimerization domains facilitate the folding of triple-helical regions, it is intriguing to ask whether collagen sequences are also capable of controlling heterospecific association. In this study, we designed a model system mimicking simultaneous specific assembly of two collagen heterotrimerers using a genetically inspired operation, circular permutation. Previously, surface charge-pair interactions were optimized on three collagen peptides to promote the formation of an abc-type heterotrimer. Circular permutation of these sequences retained networks of stabilizing interactions, preserving both triple-helical structure and heterospecificity of assembly. Combining original peptides A, B, and C and permuted peptides D, E, and F resulted primarily in formation of A:B:C and D:E:F, a heterospecificity of 2 of 56 possible stoichiometries. This degree of specificity in collagen molecular recognition is unprecedented in natural or synthetic collagens. Analysis of natural collagen sequences indicates low similarity between the neighboring exons. Combining the synthetic collagen model and bioinformatic analysis provides insight on how fibrillar collagens might have arisen from the duplication of smaller domains.

Multiple types of natural collagens specifically assemble and co-exist in the extracellular matrix. Although noncollagenous trimerization domains facilitate the folding of triple-helical regions, it is intriguing to ask whether collagen sequences are also capable of controlling heterospecific association. In this study, we designed a model system mimicking simultaneous specific assembly of two collagen heterotrimerers using a genetically inspired operation, circular permutation. Previously, surface charge-pair interactions were optimized on three collagen peptides to promote the formation of an abc-type heterotrimer. Circular permutation of these sequences retained networks of stabilizing interactions, preserving both triple-helical structure and heterospecificity of assembly. Combining original peptides A, B, and C and permuted peptides D, E, and F resulted primarily in formation of A:B:C and D:E:F, a heterospecificity of 2 of 56 possible stoichiometries. This degree of specificity in collagen molecular recognition is unprecedented in natural or synthetic collagens. Analysis of natural collagen sequences indicates low similarity between the neighboring exons. Combining the synthetic collagen model and bioinformatic analysis provides insight on how fibrillar collagens might have arisen from the duplication of smaller domains.

**Results:** Orthogonal assembly of two collagen heterotrimerers was induced by circularly permuting the sequences of a previously designed heterotrimer.

**Conclusion:** High collagen assembly specificity can be achieved with a simple sequence rearrangement.

**Significance:** This study provides insights on how fibrillar collagens might have arisen from gene duplication and domain shuffling.

Collagen is a family of fibril-forming proteins with a highly consistent sequence pattern at helical regions, where sequences are repeats of Xaa-Yaa-Gly tripletts. Genetic analyses suggest that collagen genes may originate from a common ancestor and expand by gene duplication and exon shuffling (1–3). However, different types of natural collagens discriminate each other well, specifically assembling in the extracellular matrix (4). For example, type I and III are co-localized in dermis (5). Type I and V co-assemble into heterotypic fiber in corneal stroma (6). Isoforms of type V/XI co-exist in cartilage, where the composition varies with ages (7). Although noncollagenous trimerization domains facilitate the folding of triple-helical regions (8–10), it is intriguing to ask whether collagen sequences also control this degree of heterospecific association. Design of synthetic collagens that recapitulate this natural specificity would be a powerful tool to investigate the relationship between primary sequence and folding heterospecificity.

Protein engineering strategies are often inspired by molecular processes driving natural protein evolution (11). In circular permutation, the order of amino acids is altered without significantly changing the three-dimensional fold. Conceptually, the termini of a protein are joined, resulting in a circular intermediate that is then broken at another location, giving new start and end positions. If successful, the folds of the original and circularly permuted proteins are similar with the exception of the new termini and the introduced loop that fuses the original termini (Fig. 1a). Circular permutation has been used in protein engineering to probe the location of structural intermediates in folding and to enhance protein stability (12–15). The success of circular permutation as an engineering strategy hinges on the close spatial proximity of the N and C termini in most globular proteins (16), making it possible to fuse termini without the introduction of long, unwieldy linkers.

The N and C termini for fibrous proteins such as coiled-coils or collagen are far apart. The effect of circular permutation on such proteins has not been studied to the same extent as globular proteins. Due to the linearity of fibrous protein structure, circular permutation would result in three-dimensional domain swapping, where a structural superposition of the fold would align different portions of the sequence. This may have functional implications, such as altering interactions with other proteins or self-assembly into higher order structures.

Highly specific abc-type collagen heterotrimerers have been previously designed using computational sequence patterning (17, 18). We examine the effect of circular permutation on a previous computationally designed collagen peptide abc-type heterotrimer (17). The sequences of peptides A, B, and C were patterned such that they self-assembled into a stable triple helix...
collagen heterotrimer with circularly permuted counterparts. The folded fraction was calculated as

\[ F(T) = \frac{\theta(T) - \theta_u(T)}{\theta(T) - \theta_u(T)} \]

where \( \theta(T) \) was the observed ellipticity at a temperature \( T \), and \( \theta_u(T) \) were ellipticity estimated from a linear fitting of folded and unfolded base lines. The apparent melting temperatures, \( T_m \), were estimated as \( F(T) \).

**Streptavidin-Biotin Binding Assay**—Two biotin-tagged peptides, \( C^* \) and \( F^* \), were used in this assay: \( C^* \) biotin, GGKOGPDGDPGPKGOKGKOGKOKGOG; \( F^* \) biotin, GGDOGDODGDPGPKGPKGPKGPKGPKGOGPDG. For synthesis of peptides, see supplemental Peptide Synthesis.

Before the assay mixtures A+B+C*+D+E+F and A+B+C+D+E+F*, of 0.4 mM total concentration were prepared in pH 7, 10 mM phosphate buffer with 100 mM NaCl. They were incubated at 50 °C for 15 min and then at 4 °C for 48 h to induce triple-helix formation. Samples were incubated with streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1; Invitrogen) at 4 °C for 4 h to bind the biotin-tagged peptides on the beads. Beads were washed with 1 ml of washing buffer (pH 7, 10 mM phosphate buffer with 100 mM NaCl) three times to eliminate nonspecific interactions. The beads were suspended in 75 μl of washing buffer and incubated at 60 °C for 15 min to denature triple helices, from which the supernatants containing the non-biotin-tagged peptides were collected. The beads were incubated in 6 M guanidine HCl at 96 °C for 15 min to release the biotin-tagged peptides. Supernatants from these two elution processes were combined and characterized by HPLC and mass spectrometry.

**HPLC and Mass Spectrometry**—For the mixtures before the streptavidin-biotin binding assay, each peptide of 250 nm, e.g., 3.75 μl of the mixtures with 0.4 mM total concentration, were prepared in 0.2% trifluoroacetic acid (TFA) and loaded on to a pre-packed Poroshell 120, EC-C18, 2.1 × 100 mm, 2.7-μm column (Agilent Technology) in 0.2% TFA/water as the mobile phase A and eluted with a linear gradient of 2–25% of 0.2% TFA.

---

Collagen Peptide Circular Permutation

only when all three were combined with equimolar stoichiometry. This is a specificity of one often possible trimetric states. The heterotrimer was stabilized by acidic and basic amino acids that formed an extensive network of surface charge-pair interactions. Due to the linear structure of collagen, these electrostatic interactions are local in both sequence and structure. We hypothesize that circular permutation of A, B, and C sequences will have minimal impact on these networks, resulting in a peptide system D, E, and F that maintains the structure and heterospecificity of the original design.

The sequence transformation of A:B:C to D:E:F swaps the N- and C-terminal domains, potentially affecting interactions with other proteins. It is interesting to consider the behavior of mixtures containing all six peptides. At least two outcomes may be considered (Fig. b): (i) domain-swapping drives fiber formation through assembly of sticky-ended precursors, or (ii) A:B:C and D:E:F assembling orthogonally, a stoichiometric specificity of 2 of 56 possible combinations of the six peptides. Circular permutation may either prevent protein-protein interactions between A, B, C and D, E, F, or it may promote supramolecular assembly. There are precedents for heteromeric peptide sticky-ended assembly; collagen fibril formation was promoted by engineering interchain disulfides that stabilized a sticky-ended precursor (20). Fiber-forming heterodimeric α-helix peptide coiled-coils were designed to assemble in a sticky-ended fashion, held together by internal hydrogen bonds and surface charge-pair interactions (21). In that example, circular permutation was used to generate a third peptide capable of forming a stable blunt-ended species that blocked fibril formation. Whether the designed collagen peptides proposed here form sticky-ended precursors en route to fibers, or orthogonally assemble as two abc-type heterotrimers will depend on the relative stability of target and competing triple-helical states.
Collagen Peptide Circular Permutation

in acetonitrile (phase B) in 40 min. HPLC profiles were monitored as UV absorbance at 214 nm. Due to identical molecular weights, original and permuted sequences had similar HPLC retention. To identify species in peak fractions from retention time from 16 to 22 min, MS and tandem MS at Reflector mode experiments were carried on an ABI-MDS SCIEX 4800. The peak area was integrated with Gilson Unipoint System software.

Stability Scores—The details of stability score calculation have been described previously (17, 22). Briefly, a discrete sequence-based model was adopted from an α-helical coiled coil design (23); the stability score counted electrostatic interactions between of proximal residues on adjacent chains, plus an imino acid (Pro + Hyp) content term. The best stability score among all the possible staggering registries was selected to represent a particular stoichiometry.

Sequence Similarity—Exon sequences of collagen types I, III, and V were downloaded from the ENSEMBL genome browser (sequence ID codes ENST00000225964, ENST00000304636, ENST00000371817, ENST00000374866, ENST00000264828, and ENST00000297268). A few large exons consisting of multiples of 15- or 18-residue-long fragments were subdivided into smaller domains. Amino acids were categorized by physiochemical properties, i.e. polar, hydrophobic, positively charged, negatively charged. The amino acids in the same category were considered to be identical. Multiple alignments were generated for fragment pairs of different lengths, among which the highest similarity score was selected.

RESULTS AND DISCUSSION

Sequence Design—To create a triple helix a few key design constraints must be considered. Collagen consists of repeated Xaa-Yaa-Gly triplets. In natural collagens these can extend for >1000 amino acids, but in the model peptides only 10 repeats are used resulting in 30-residue-long chains. A network of interchain backbone hydrogen bonds maintains the triple-helical fold, and additional stability and specificity are modulated by the identities of solvent exposed residues. Xaa and Yaa positions are often proline or hydroxproline, respectively, which stabilize the extended polyproline-II conformation of the triple-helical backbone. Charged amino acids may also be placed at these positions to promote electrostatic interactions that modulate stability and specificity (24–27). Previously, three peptidic sequences, A, B, and C were designed to form an abc-type heterotrimer A:B:C to the exclusion of competing heterotrimers and two-species heterotrimers were assayed by circular dichroism. Ten peptide mixtures at equivalent total peptide concentrations were prepared: A, B, C, A:2B, A:2C, 2A:C, B:2C, 2B:C, and A:B:C. Equivalent mixtures of D, E, and F were also tested. Both A:B:C and D:E:F mixtures formed a triple-helical structure with a strong positive ellipticity around 223 nm (Fig. 2, e and f). Other mixtures showed little or no triple-helical structure, indicating that the heterospecificity was preserved by circular permutation. Additionally, only A:B:C and D:E:F mixtures produced species with significant stabilities and cooperative unfolding transitions (Fig. 2, c and d). The melting temperature, Tm of A:B:C at 29 °C was slightly greater than D:E:F, Tm = 24 °C, consistent with the loss of two charge pairs upon circular permutation. Whereas B+C mixtures did form weak triple helices with borderline cooperative unfolding transitions of Tm < 5 °C, equivalent species were not found in E+F mixtures.

Assembling Formation of Sticky-ended Precursors—Combining all six peptides from the two groups could lead to the creation of sticky-ended precursors that further assemble into fibers. Two peptides from a group plus the circular permutation of the third peptide could form a sticky-ended precursor with an overlap of 15 residues as pictured in Fig. 1b. There are six such combinations: A:B:F, A:E:C, D:B:C, D:E:C, D:B:F, and A:E:F. When examined by circular dichroism spectroscopy, none had a positive ellipticity at 223-nm diagnostic of triple helix, nor did any exhibit cooperative loss of structure upon thermal denaturation (supplemental Table S1). Dynamic light scattering measurements showed no evidence of higher order assembly. Single particle size distribution peaks at approximately 2 nm were consistent with the observed hydrodynamic radius of triple helix and random-coil states in previous studies (supplemental Fig. S1) (29). Negligible signals were found for larger particle sizes. Solutions remained clear and did not form precipitates even after storage for long periods of time. In all, no evidence was found to support that the existence of sticky-ended precursors or higher order structures under the conditions studied. An insufficient number of backbone-backbone hydrogen bonding interactions in the offset species might account for the absence of such intermediates.

Assessing Orthogonal Assembly—The term “orthogonal assembly” is used to describe the spontaneous formation of more than one abc-type heterotrimer where each is composed of a unique set of peptides. As such, orthogonal assembly would entail the folding of the six peptides A+B+C+D+E+F into A:B:C and D:E:F. This requires that these species are signifi-
significantly more stable than competing mixtures. There are 216 (6 \times 3) possible triple-helical states that can be constructed from combinations of the six peptides. This corresponds to 56 stoichiometric combinations compatible with a collagen trimer.

Given the subset of amino acids used in these designs, it is possible to score the stability of each state a weighted sum of the number of imino acids (Pro/Hyp) and the total predicted charge-pair interactions between structurally adjacent residues. Despite the simplicity of this metric, it is reasonably predictive of observed stabilities (22, 30). Computed stabilities from all 216 states were assigned to the 56 stoichiometries based on the best stability score among the set of triple-helical states associated with that mixture (supplemental Table S2). Target heterotrimers A:B:C and D:E:F had the highest stabilities both computationally and experimentally (Figs. 3b and 4a).

The next best scoring stoichiometries were C:E:F and the corresponding circularly permuted state, p5(C:E:F) = F:B:C. Both mixtures showed weaker triple-helical structures and stabilities with T_m of 15 °C and 11 °C, respectively (Fig. 3b). Other states and their corresponding circular permutation with similar energies, A:B:F and D:E:C or A:B:D and D:E:A, did not form cooperative triple-helical structures.

FIGURE 2. Circular permutation of A, B, and C maintains heterospecificity. a, all except two predicted charge-pair interactions in ABC are preserved upon circular permutation to DEF. b, computed stability scores for all mixtures of A+B+C and for D+E+F (A, B, C, A:2B, 2A:B, B:2C, 2B:C, A:2C, 2A:C, A:B:C, D, E, F, D:F, 2D:E, E:2F, 2E:F, D:2F, D:E:F) predict that target stability and energy gap separating competing states are maintained in both peptide systems. c–f, thermal denaturation profiles (c and d) and circular dichroism spectra (e and f) at 4 °C of the 10 combinations of A,B,C and D,E,F show a cooperative unfolding transition only in the case of 1:1:1 mixtures. Samples are in phosphate buffer, pH 7.0, 100 mM NaCl. MRE is in 10^2 deg cm^2 dmol res^{-1}.  

Collagen Peptide Circular Permutation
FIGURE 3. Nonspecific assembly across the original and permuted sets. CD spectra at 4 °C (a) and temperature denaturation monitored at 223 nm of the 10 cross-assembly mixtures (b) i.e. A:B:F, A:E:C, A:E:F, D:B:C, D:B:F, D:E:C, A:D:B, A:D:E, B:C:F, and C:E:F. In addition to A:B:C and D:E:F, All of the samples of 0.2 mM total concentration were prepared in 10 mM phosphate buffer, pH 7, with 100 mM NaCl.

Although it is reasonable to expect that A:B:C and D:E:F orthogonally assemble in mixture given their observed stabilities, spectroscopic measurements alone are insufficient to demonstrate this. A pulldown assay was developed to observe the composition of triple-helical species formed in complex mixtures. Biotin separated by a Gly-Gly linker was conjugated to N termini of peptides C and F, referred to as C* and F*. Triple-helical structure and stability of A:B:C* were very similar to the unmodified A:B:C mixture, indicating that the modification did not perturb folding (supplemental Fig. S2). The same was found for D:E:F* compared with D:E:F. A:B:C*, D:E:F, and A:B:C; D:E:F* mixtures were allowed to equilibrate overnight using a standard folding protocol. Streptavidin-coated magnetic beads were subsequently added to extract the biotinylated constructs and associated peptides. Peptides released from streptavidin-coated beads were identified by mass spectrometry and quantified from peak areas in the HPLC chromatograms.

The six peptides were initially combined in equimolar concentrations (Fig. 4, b and c). In both C* and F* mixtures, three major peptide peaks remained after the streptavidin bead pull-down. The post-C* pulldown peptides identified by tandem MS were A, B, and C* (supplemental Fig. S3). The ratios of peak areas of peptide A, B, and C* normalized to C* were 0.5:0:5:1 (supplemental Table S3), indicating C* predominantly bound to A and B, and half of the C* peptides were participating in heterotrimers. D, E, and F* were found in the same relative ratios in the post-F* pulldown, indicating that F* interacted specifically with D and E in the six-peptide mixture (Fig. 4c). These observations support orthogonal assembly of A:B:C and D:E:F in solution.

Detectable peaks corresponding to E and F were also pulled down with C*. The relative ratios of E and F to C* were 0.06:0.05:1, suggesting the presence of a minor C:E:F species, consistent with the observed stability of C:E:F by circular dichroism spectroscopy. In a focused pulldown experiment where only three peptides were included in the initial mixture E, F, and C*, 5-fold greater amounts of both E and F were associated with C* compared with the six-peptide mixture (supplemental Fig. S4). This change in C:E:F fraction in the presence of the remaining three peptides is presumably due to competing formation of the more stable A:B:C and D:E:F species. In contrast, the B:C:F species which was shown to have marginal structure, and stability by circular dichroism was not found in a focused pulldown experiment consisting initially of B, C, and F*. The low stability of these competing unstable complexes would preclude them from forming at higher temperatures where A:B:C and D:E:F are present.

**Limits of Circular Permutation**—The orthogonal assembly of A:B:C and D:E:F was the result of the stability of target species and the instability of competing species, including sticky-ended precursors of higher order structure. Circular permutation serves to shift the network of electrostatic interactions so that they are out of phase with the optimal stagger of the triple helix. The five-triplet circular permutation is the largest perturbation possible in a 10-triplet peptide. We wanted to investigate the effect of smaller shifts, ranging from one to four triplets on triple helix formation. Would it be plausible to create a larger interactome where, for example, a three-triplet circular permutation would assemble orthogonally to the zero and five-triplet circular permutation sequences?

Four additional permutations of A were synthesized: p1(A), p2(A), p3(A), and p4(A) and studied along with p0(A) = A and p5(A) = D:

- p1(A), PKGPKPGPKGKOGPDGDOGDOGPGPK;
- p2(A), PKGPKPGPKGKOGPDGDOGDOGPGPK;
- p3(A), DOGPKPGPKPGKOGPDGDOGDOGPGPK;
- p4(A), DOGDOGPKPGPKPGKOGPDGDOGDOGPGPK;

Computed stabilities of p(x)A peptides were evaluated in both a B+C and a E+F background. The computed heterospecific association p(x)A:B+C was most favorable for x = 0 and progressively diminished with increasing x (supplemental Table S4). Conversely, the predicted stability of heterotrimers p(x)A:E+F increased with increasing index x. p(x)A:E+F and D:E+F had similar computed energies.
The stability and structure of each frame-shifted peptide was assessed by circular dichroism in the B+C and E+F backgrounds. No triple helix was detected in p^x(A) in any of the three-component mixtures of p^x(A):B and p^x(A):C for any value of x (supplemental Table S5). Consistent with the predicted stability scores, p^x(A):B:C and p^x(A):B:C formed stable heterotrimers with the stabilities lower than A:B:C (Fig. 5). Due to the inability to assign folded base lines, the melting temperatures of mixtures p^x(A):B:C and p^x(A):B:C were estimated to be <10 and 5 °C, respectively. It was similarly challenging to assign a specific T_m to p^x(A):E:F. Experimental and calculated stabilities correlate in both the B+C and E+F backgrounds.

In these designs, a five-triplet permutation is optimal for preventing cross-assembly. Smaller permutations result in marginally stable species where stability correlates with the extent of permutation and where triple-helical structures are formed in both backgrounds. In the 10-triplet systems commonly studied, we would not expect a p^x(A,B,C) + p^x(A,B,C) mixture to orthogonally assemble for |x-y| ≠ 5. This places a practical limit on the extent to which circular permutation can be utilized to direct orthogonal assembly in a multispecies collagen interactome. A nine-peptide orthogonally assembling interactome would not be generated by simply combining x = 0,3,6 permutations of p^x(A,B,C). Other approaches for diversifying the sequence would be needed to achieve specificity of molecular recognition in mixtures greater than six components.

**Implications for Fibrous Collagen Evolution**—It is interesting to consider whether the same sequence constraints promoting orthogonal assembly in synthetic collagen peptides might prevent misfolding in much longer fibrillar collagens. The specificity of A:B:C was achieved by intentionally incorporating an energy gap in the design protocol. This gap was preserved in D:E:F. However, gaps between targets and all other competing states in the six-peptide mixture were not explicitly modeled at the design stage. This unintentional gap arose from sequence diversity, the disparate charge patterns in the N- versus C-terminal halves of the designed sequences. This diversity comes from the stochastic nature of the simulated-evolution computational design protocol. Low sequence complexity or short periodic sequence motifs would be expected to increase the probability of undesired interactions between the N- and C-terminal domains, resulting in a mixture of misfolded states. An equivalent phenomenon is found in naturally occurring multidomain proteins, where sequence diversity across individual domains is an important factor in preventing misfolding (31, 32). In particular, adjacent subunits tend to have lower sequence identities than more distant pairs, as their spatial proximity makes misfolding more likely (33).

To examine the sequence diversity within collagen domains, it was first necessary to define the domains themselves. Unlike multidomain proteins such as fibronectin or titin, there are no clear structural boundaries in fibrillar collagens, which consist of a single, extended triple helix. In this case, a genetic definition of the domain based on exons is more useful. In fibrillar collagens, exons are predominantly 45 or 54 nucleotides in length, corresponding to 15 or 18 amino acid domains (34). This makes the fundamental domain in fibrillar collagen a five- or six-triplet sequence; the synthetic peptides would consist of the equivalent of two domains.

An autocorrelation plot of sequence similarity versus the number of intervening domains was computed for several human fibrillar collagens. A peak in similarity is observed for domains separated by one intervening domain, i.e. domain n versus n+2 (Fig. 6). This supports a domain repeat expansion model where gene duplication occurred for pairs of exons, rather than one exon at a time (35). If a constraint on multidomain protein folding is the low sequence similarity of adjacent domains, then the duplication of pairs of domains would be a mechanism to accomplish this. The difference in sequence similarity between domain n to itself (100%) and n+1 (~37%) is large. With reference to domain n, the difference in similarity of n+2
Collagen Peptide Circular Permutation

and \(n+3\) suggests that sequence diversity is maintained, perhaps for the reasons outlined above, to mitigate misfolding. Despite the absence of structural domains within fibrillar collagen, sequence diversity between adjacent genetically specified domains is present and may be a mechanism to prevent misfolding. Although collagen registry is specified during procollagen folding by N- and C-terminal globular domains, the matrix is a dynamic structure that is constantly being remodeled (36), and such sequence diversity may minimize frustration caused by misannealed states during transient unfolding/refolding.

CONCLUSIONS

Circular permutation preserves three-dimensional structure in both globular and fibrous proteins. In our case, both the triple-helical fold and abc-type heterospecificity are maintained. Whereas circular permutations of globular proteins generally do not perturb intermolecular interactions (37, 38), fibrous protein circular permutation swaps domains around the point of permutation, modifying intermolecular interactions and in this case promoting orthogonal assembly. Success of circular permutation as a strategy for engineering orthogonal assembly hinges on the sequence properties of each peptide, particularly the presence of diverse, aperiodic charge patterns. Computationally designed collagen heterotrimers differ in this respect to rationally designed sequences, which often use repeating charge patterns through the entire peptide length (39, 40). Our results suggest that it is not essential to design heterotrimers using 10-triplet sequences. Instead, libraries of five-triplet fragments can be combined to yield longer designs that preserve heterospecificity as long as sequence similarity of adjacent fragments is low. This reduces potential sequence space by 10 orders of magnitude, from \(5^{30} \times 10^{21}\) to \(5^{15} \times 10^{10}\) trimeric states, making it computationally feasible to pursue full sampling, expand the types of amino acids used, or to increase the level of theory used to evaluate designs.

The observed low sequence diversity of adjacent domains in natural collagens has implications for the design of long engineered collagens. The majority of biophysical analyses on collagen have focused on short peptides due to their synthetic accessibility. Recent studies on recombinant bacterial collagens demonstrate that it is possible to engineer much longer sequences and characterize their stability, folding kinetics, and other biophysical properties (41–43). A strategy for the engineering of long collagen sequences would be to use computational design to specify individual domains on the order of five to six triplets and then concatenate them in such a way as to minimize sequence identity of adjacent domains. This strategy may be extended to the design of other oligomeric fibrous proteins such as coiled-coils.

Six-component orthogonal assembly represents an unprecedented degree of molecular recognition specificity for collagen. Complex orthogonally assembling systems have been designed using \(\alpha\)-helical folds (44–46) and observed in natural protein-protein interactions (19, 47). One advantage of a triple-helical interaction network is the lack of structural overlap with existing interatomces. Due to the high proline content, collagenous peptides are unlikely to form other secondary structures.

Therefore, collagen peptides such as these expand the synthetic biology toolkit for directing protein-protein interaction networks or promoting self-assembly of complex structures.

Acknowledgment—We thank Dr. Awanish Parmar for assistance with collection and analysis of DLS.

REFERENCES

1. Pathy, L. (1999) Genome evolution and the evolution of the exon-shuffling: a review. Gene 238, 103–114
2. Boot-Handford, R. P., and Tuckwell, D. S. (2003) Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. Bioessays 25, 142–151
3. Heino, J., Huhtala, M., Käpylä, J., and Johnson, M. S. (2009) Evolution of collagen-based adhesion systems. Int. J. Biochem. Cell Biol. 41, 341–348
4. Birk, D. E. (2001) Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. Micron 32, 223–237
5. Vitellaro-Zaccarellò, L., Garbelli, R., and Rossi, V. D. (1992) Immunocytochemical localization of collagen types I, III, IV, and fibronectin in the human dermis. Cell Tissue Res. 268, 505–511
6. Birk, D. E., Fitch, J. M., Babiaz, J. P., and Linsenmayer, T. F. (1988) Collagen type I and type V are present in the same fibril in the avian corneal stroma. J. Cell Biol. 106, 999–1008
7. Wu, J. J., Weis, M. A., Kim, L. S., Carter, B. G., and Eyre, D. R. (2009) Differences in chain usage and cross-linking specificities of cartilage type V/XI collagen isoforms with age and tissue. J. Biol. Chem. 284, 5539–5545
8. McLaughlin, S. H., and Bulleid, N. J. (1998) Molecular recognition in procollagen chain assembly. Matrix Biol. 16, 369–377
9. Boudko, S. P., Engel, J., and Bächinger, H. P. (2012) The crucial role of trimerization domains in collagen folding. Int. J. Biochem. Cell Biol. 44, 21–32
10. Bourhis, J. M., Mariano, N., Zhao, Y., Harlos, K., Exposito, J. Y., Jones, E. Y., Moali, C., Aghajari, N., and Hulmes, D. J. (2012) Structural basis of fibrillar collagen trimerization and related genetic disorders. Nat. Struct. Mol. Biol. 19, 1031–1036
11. Eisenbeis, S., and Höcker, B. (2010) Evolutionary mechanisms as templates for protein engineering. J. Pept. Sci. 16, 538–544
12. Haglund, E., Lindberg, M. O., and Oliveberg, M. (2008) Changes of protein folding pathways by circular permutation: overlapping nuclei promote global cooperativity. J. Biol. Chem. 283, 27904–27915
13. Lindberg, M., Tängrot, J., and Oliveberg, M. (2002) Complete change of the protein folding transition state upon circular permutation. Nat. Struct. Biol. 9, 818–822
14. Iwakura, M., Nakamura, T., Yamane, C., and Maki, K. (2000) Systematic circular permutation of an entire protein reveals essential folding elements. Nat. Struct. Biol. 7, 580–585
15. Hennecke, J., Seppel, P., and Glocshuber, R. (1999) Random circular permutation of DsBA reveals segments that are essential for protein folding and stability. J. Mol. Biol. 286, 1197–1215
16. Krishna, M. M., and Englander, S. W. (2005) The N-terminal to C-terminal motif in protein folding and function. Proc. Natl. Acad. Sci. U.S.A. 102, 1053–1058
17. Xu, F., Zahid, S., Silva, T., and Nanda, V. (2011) Computational design of a collagen A:B:C-type heterotrimer. J. Am. Chem. Soc. 133, 15260–15263
18. Fallas, J. A., and Hartgerink, J. D. (2012) Computational design of self-assembling register-specific collagen heterotrimers. Nat. Commun. 3, 1087
19. Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D., and Sidhu, S. S. (2008) A specificity map for the PDZ domain family. PLoS Biol. 6, e239
20. Kotch, F. W., and Raines, R. T. (2006) Self-assembly of synthetic collagen triple helices. Proc. Natl. Acad. Sci. U.S.A. 103, 3028–3033
21. Pandya, M. J., Spooner, G. M., Sunde, M., Thorpe, J. R., Rodger, A., and Woolfson, D. N. (2000) Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. Biochemistry 39, 8728–8734
22. Xu, F., Zhang, L., Koder, R. L., and Nanda, V. (2010) De novo self-assembling collagen heterotrimers using explicit positive and negative design. Biochemistry 49, 2307–2316

23. Summa, C. M., Rosenblatt, M. M., Hong, I. K., Lear, J. D., and DeGrado, W. F. (2002) Computational de novo design, and characterization of an A2B2 diiron protein. J. Mol. Biol. 321, 923–938

24. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) Prediction of collagen stability from amino acid sequence. J. Biol. Chem. 280, 19343–19349

25. Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A., and Brodsky, B. (2005) Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry 44, 1414–1422

26. Chan, V. C., Ramshaw, J. A., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix. J. Biol. Chem. 272, 31441–31446

27. Yang, W., Chan, V. C., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (1997) Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. J. Biol. Chem. 272, 28837–28840

28. Hellinga, H. W., and Richards, F. M. (1994) Optimal sequence selection in proteins of known structure by simulated evolution. Proc. Natl. Acad. Sci. U.S.A. 91, 5803–5807

29. Kar, K., Amin, P., Bryan, M. A., Persikov, A. V., Mohs, A., Wang, Y. H., and Brodsky, B. (2006) Self-association of collagen triple-helix peptides into higher order structures. J. Biol. Chem. 281, 33283–33290

30. Giddu, S., Xu, F., and Nanda, V. (2013) Sequence recombination improves target specificity in a redesigned collagen peptide abc-type heterotrimer. Proteins 81, 386–393

31. Wright, C. F., Teichmann, S. A., Clarke, J., and Dobson, C. M. (2005) The importance of sequence diversity in the aggregation and evolution of proteins. Nature 438, 878–881

32. Lukatsky, D. B., Shakhnovich, B. E., Mintseris, J., and Shakhnovich, E. I. (2007) Structural similarity enhances interaction propensity of proteins. J. Mol. Biol. 365, 1596–1606

33. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007) The folding and evolution of multidomain proteins. Nat. Rev. Mol. Cell Biol. 8, 319–330

34. Buttice, G., Kayes, P., D’Armento, J., Vogeli, G., and Kurkinnen, M. (1990) Evolution of collagen IV genes from a 54-base pair exon: a role for introns in gene evolution. J. Mol. Evol. 30, 479–488

35. Björklund, A. K., Ekman, D., and Elofsson, A. (2006) Expansion of protein domain repeats. PLoS Comput. Biol. 2, e114

36. Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) Type I collagen is thermally unstable at body temperature. Proc. Natl. Acad. Sci. U.S.A. 99, 1314–1318

37. Bulaj, G., Koehn, R. E., and Goldenberg, D. P. (2004) Alteration of the disulfide-coupled folding pathway of BPTI by circular permutation. Protein Sci. 13, 1182–1196

38. Sanders, K. E., Lo, J., and Sligar, S. G. (2002) Intersubunit circular permutation of human hemoglobin. Blood 100, 299–305

39. Gauba, V., and Hartgerink, J. D. (2007) Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. J. Am. Chem. Soc. 129, 2683–2690

40. Mechling, D. E., and Bachinger, H. P. (2000) The collagen-like peptide (GER)15GPCCG forms pH-dependent covalently linked triple-helical trimers. J. Biol. Chem. 275, 14532–14536

41. Yoshizumi, A., Fletcher, J. M., Yu, Z., Persikov, A. V., Bartlett, G. J., Boyle, A. L., Vincent, T. L., Woolfson, D. N., and Brodsky, B. (2011) Designed coiled coils promote folding of a recombinant bacterial collagen. J. Biol. Chem. 286, 17512–17520

42. Yu, Z., Brodsky, B., and Inouye, M. (2011) Dissecting a bacterial collagen domain from Streptococcus pyogenes: sequence and length-dependent variations in triple-helix stability and folding. J. Biol. Chem. 286, 18960–18968

43. Xu, C., Yu, Z., Inouye, M., Brodsky, B., and Mirochnitchenko, O. (2010) Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. Biomacromolecules 11, 348–356

44. Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. ACS Synth. Biol. 1, 118–129

45. Reinke, A. W., Grant, R. A., and Keating, A. E. (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J. Am. Chem. Soc. 132, 6025–6031

46. Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature 458, 859–864

47. Tonikian, R., Xin, X., Toret, C. P., Gfeller, D., Landgraf, C., Panni, S., Paoluzi, S., Castagnoli, L., Currell, B., Seshagiri, S., Yu, H., Winsor, B., Vidal, M., Gerstein, M. B., Bader, G. D., Volkmer, R., Cesareni, G., Drubin, D. G., Kini, P. M., Sidhu, S. S., and Boone, C. (2009) Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. 7, e1000218