Collagen triple helices fold slowly and inefficiently, often requiring adjacent globular domains to assist this process. In the *Streptococcus pyogenes* collagen-like protein Scl2, a V domain predicted to be largely α-helical, occurs N-terminal to the collagen triple helix (CL). Here, we replace this natural trimerization domain with a de novo designed, hyperstable, parallel, three-stranded, α-helical coiled coil (CC), either at the N terminus (CC-CL) or the C terminus (CL-CC) of the collagen domain. CD spectra of the constructs are consistent with additivity of independently and fully folded CC and CL domains, and the proteins retain their distinctive thermal stabilities, CL at −37 °C and CC at >90 °C. Heating the hybrid proteins to 50 °C unfolds CL, leaving CC intact, and upon cooling, the rate of CL refolding is somewhat faster for CC-CL than for CC-CL. A construct with coiled coils on both ends, CC-CL-CC, retains the −37 °C thermal stability for CL but shows less triple helix at low temperature and less denaturation at 50 °C. Most strikingly however, in CC-CL-CC, the CL refolds slower than in either CC-CL or CL-CC by almost two orders of magnitude. We propose that a single CC promotes folding of the CL domain via nucleation and in-register growth from one end, whereas initiation and growth from both ends in CC-CL-CC results in mismatched registers that frustrate folding. Bioinformatics analysis of natural collagens lends support to this because, where present, there is generally only one coiled-coil domain close to the triple helix, and it is nearly always N-terminal to the collagen repeat.

The collagen triple helix and the α-helical coiled coil (CC) are well characterized superhelical motifs in proteins (1–3). They form rod-like structures, which are directed by clear amino acid patterns in their sequences. Collagen triple helices require glycine as every third residue and often have a high imino acid (proline and hydroxyproline) content. These features lead to the formation of polyproline II helices, which trim-erize via interchain hydrogen bonding and close packing to form the collagen triple helix (see Fig. 1A). By contrast, most coiled coils are low in glycine and proline and have a so-called heptad, or related repeats in which hydrophobic residues alternate three and four residues apart. This pattern promotes the formation of amphipathic α-helices that combine via their hydrophobic faces to form rope-like helical bundles (Fig. 1, B and C). Some proteins contain both collagen and α-helical coiled-coil domains. For instance, three-stranded coiled coils occur immediately C-terminal to collagen triple helices in lung surfactant apoprotein D, lung surfactant apoprotein A, mannose-binding protein, and other collectins (4), whereas the macrophage scavenger receptor has a coiled coil N-terminal to a collagen triple helix (5). Previous sequence analyses suggest that putative coiled-coil domains are often found in members of the collagen superfamily, and these coiled coils may serve as oligomerization domains important for collagen assembly (6). In vitro triple helix formation from three (Gly-Xaa-Yaa)₇ polypeptide chains is an inherently slow process (7), and it appears that adjacent coiled coils facilitate proper registration, nucleation, and folding of the triple helix. Here, we explore the relationship between coiled-coil domains and an adjacent collagen triple helix using a designed coiled coil and a recombinant bacterial collagen.

Although collagens were originally thought to be restricted to multicellular animals, a number of collagen-like triple helix domains have been identified in bacteria (8). Of these, the cell-surface proteins from *Streptococcus pyogenes* are among the best characterized in terms of structure and function. Scl2 (*S. pyogenes* collagen-like protein 2) contains an N-terminal globular domain (denoted V), a (Gly-Xaa-Yaa)₇ polypeptide domain (CL), and linker and transmembrane regions (9). A recombinant construct with the globular V domain and the
adjacent triple helix, designated as V-CL, has been expressed in Escherichia coli and shown to form a triple helix structure with a conformation and thermal stability similar to that of mammalian collagens (see Fig. 2A) (9–11).

The N-terminal V domain has been shown to be a trimerization domain and is essential for in vitro refolding of the CL triple helix (9, 12). In addition, the S. pyogenes V domain can assist correct folding of a heterologous triple helix sequence from Clostridium perfringens, which is incapable of folding in its original context (12, 13). Although a three-dimensional structure of the V domain has yet to be determined, two α-helical regions are predicted from its sequence. The isolated V domain has an α-helical CD spectrum (9, 12), suggesting that the ScI2 bacterial collagen protein may represent another instance where an adjacent α-helical domain promotes triple helix formation. The two α-helical regions in the V domain were assigned previously as coiled coils (9, 12), but re-examination using the currently available prediction tools indicates that the coiled-coil nature of regions are predicted with ~25% confidence or less (supplemental Fig. S1).

For the present study, we constructed recombinant fusion proteins in which the natural N-terminal V domain of the ScI2 protein was replaced by a parallel, homotrimeric coiled coil of de novo design and high thermal stability. Specifically, a four-heptad sequence was placed either N- or C-terminal to, or at both ends of, the collagen domain. The effect of the position of the coiled-coil domains on folding, stability, and assembly of the triple helix was investigated by equilibrium, thermal unfolding, and kinetic CD spectroscopy experiments. These data indicate that each domain retained its independent stability in all cases, largely unaffected by the presence of the other domain in the same protein. Conditions were established where the collagen domain unfolded, whereas the coiled coil remained structured. In these cases, the refolding kinetics for the collagen triple helix depended markedly on the position of the coiled-coil domain. Intriguingly, the construct with coiled-coil domains at both ends of the collagen domain folded more slowly and less completely than those with a single coiled coil, which suggests that some structural frustration occurs if collagen triple helices are nucleated from both ends. The biological consequences of this hypothesis are explored through bioinformatics analysis of protein sequences that harbor both collagen and coiled-coil repeats.

**EXPERIMENTAL PROCEDURES**

**Peptide Synthesis**—Rink amide ChemMatrix™ resin was obtained from PCAS Biometry, Inc. (St.-Jean-sur-Richelieu, Canada); Fmoc-amino acids were obtained from AGTC Bioproducts (Hesse, UK); 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate was obtained from GL Biochem (Shanghai, China); all other reagents were of peptide synthesis grade and obtained from Thermo Fisher Scientific (Loughborough, UK). Peptide BS-3pCC4 was synthesized on a 0.1-mmol scale on Rink amide resin using a Liberty™ microwave peptide synthesizer (CEM; Matthews, NC) employing Fmoc solid-phase techniques (for review, see Ref. 14) and systematically repeated steps of coupling and deprotection interspersed with washings (5 × 7 ml dimethylformamide). Coupling was performed as follows: Fmoc amino acid (5 eq), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (4.5 eq), and diisopropylethylamine (10 eq) in dimethylformamide (7 ml) for 5 min with 20-watt microwave irradiation at 75 °C. Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 20-watt microwave irradiation at 75 °C. Following linear assembly, the peptide was acetylated (acetic anhydride (3 eq) and diisopropyl-ethylamine (4.5 eq) in dimethylformamide (7 ml) for 20 min) and then cleaved from the resin with concomitant removal of side chain-protecting groups by treatment with a cleavage mixture (10 ml) consisting of TFA (95%), trisopropylsilane (2.5%), and H2O (2.5%) for 3 h at room temperature. Suspended resin was removed by filtration, and the peptide was precipitated in ice-cold diethyl ether, centrifuged, and then the pellet was dissolved in 1:1 MeCN/H2O and freeze-dried. Purification was performed by RP-HPLC using a Kromatex C18 reverse phase column (semi-micro, 5 µm, 10 mm inner diameter × 150 mm long). Eluents were used as follows: 0.1% TFA in H2O (A) and 0.1% TFA in MeCN (B). The peptide was eluted by applying a linear gradient (at 3 ml/min) of 20% to 80% B over 40 min. Fractions collected were examined by MALDI-TOF mass spectrometry, and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%. Successful synthesis was confirmed by MALDI-TOF mass spectrometry (BS-3pCC4: m/z [M + H]⁺ observed: 3466.06; m/z [M + H]⁺ theoretical: 3465.95).

**Construction of pCold II-CC-CL, pCold II-CL-CC, and pCold II-CC-CL-CC and Protein Expression**—A DNA fragment encoding the BS-3pCC4 peptide sequence was synthesized by GenScript and inserted in place of the coding region for the variable globular domains of pCold III-V-CL and pCold III-CL-V (10, 11, 15) bearing a His₆ tag at the N terminus. For pCold II-CC-CL-CC, the coiled-coil gene sequence was inserted at both the 5’ and 3’ ends of the coding regions for CL in pCold II-CL. All resulting plasmids were confirmed by DNA sequencing and then transformed into E. coli BL21 strains. Cells were cultured in 5 ml of M9 casamino acid medium containing ampicillin (50 µg/ml) and incubated at 37 °C for 12 h. The cultures were then transferred into 250 ml of M9 casamino acid medium containing ampicillin (50 µg/ml) and incubated at 37 °C for approximately 4 h until the A₆₀₀ value reached 0.8 absorbance units. The culture was shifted to room temperature, and 1 mM isopropyl 1-thio-β-D-galactopyranoside was added to induce protein expression. After overnight expression, cells were harvested by centrifugation and disrupted by French pressing. Cellular debris was removed by centrifugation at 4 °C. All proteins were found in the soluble supernatant fraction. The pure proteins were obtained by elution with imidazole from a nickel–Sepharose resin column initially equilibrated with binding buffer (20 mM phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole) at room temperature, as described previously (11).

**CD Spectroscopy**—CD spectra of BS-3pCC4 were recorded using a JASCO J815 spectropolarimeter, whereas recombinant collagen/coiled coil constructs were examined using an AVIV Model 62DS spectropolarimeter (Aviv Associates, Inc., Lake-
Fusing Coiled-coil and Collagen Domains

wood, NJ). Peptide concentrations were determined by UV absorption at 280 nm (ε(Trp) = 5690 mol⁻¹ cm⁻¹; ε(Tyr) = 1280 mol⁻¹ cm⁻¹). Ellipticities (millidegrees) were converted to mean residue ellipticity (MRE; deg cm² dmol⁻¹) by normalizing foramide bond concentration and cuvette path length. In the absence of aromatic residues, the concentration of the collagen domain (CL) was estimated by weighing and by absorbance at 214 nm and also by subtracting the CD spectrum of the isolated V domain from the CD spectrum of the intact V-CL molecule, giving MRE222 ~ 7000 deg·cm²·dmol⁻¹ and MRE198 ~ -60,000 deg·cm²·dmol⁻¹ for CL.

Secondary Structure and Thermal Stability—The four-heptad trimeric coiled coil, BS-3pCC4, the collagen triple helix domain (CL), the coiled coil/collagen constructs (CC-CL and CL-CC), and CC-CL-CC were prepared for analysis at 20 μM in PBS at pH 7.0. Complete spectra were obtained at 0 °C and 50 °C from 260 nm to 196 nm recording points at every 0.5 nm for 4 s using a bandwidth of 1 nm, averaging three scans for each sample. Spectra obtained for BS-3pCC4 and CL were also used as a basis for determining theoretical spectra for the coiled coil/collagen constructs to which the experimentally observed spectra were compared. Theoretical spectra were calculated by taking into account the relative sizes of both the BS-3pCC4 (32 residues) and CL (238 residues) domains, e.g. MRE(CC-CL)theoretical = (32/(238 + 32)) × MRE(3pCC4) + (238/(238 + 32)) × MRE(CL).

Experiments examining the thermal denaturation of coiled coil/collagen constructs were also performed at 20 μM concentration in PBS (pH 7.0). Molar ellipticity was recorded at 220 nm from 0 to 100 °C at an average ramp rate of 0.1 °C/min. The Tm was determined as the temperature at which the fraction folded was equal to 0.5 in the curve fitted to the trimer-to-trimeronomer transition. Owing to the high thermal stability of the latter showing the positions and juxtapositions of residues in the coiled-coil heptad repeat abcd efgh.

Kinetic Analysis of Coiled Coil/Collagen Refolding—V-CL and fused coiled-coil and collagen constructs (20 μM) were denatured at 50 °C for 30 min in PBS (pH 7.0), a temperature that is sufficient to denature the collagen triple helix but leaves the coiled-coil domain intact. After denaturation, the sample was immediately transferred to CD cells pre-equilibrated at 0 or 25 °C, and the ellipticity at 220 nm was monitored as a function of time (time constant, 2 s; time interval, 10 s) to follow refolding of the collagen triple helix. Because the folding curves did not fit any simple kinetics, the half-time of refolding was defined as the time for the fraction folded to reach 50% of the original MRE220 nm value at low temperature.

Analytical Ultracentrifugation—Sedimentation equilibrium experiments for the BS-3pCC4 peptide were conducted in a Beckman-Optima XL-I analytical ultracentrifuge at 20 °C using an An-60 Ti rotor. Solutions were prepared in PBS (pH 7.4) at peptide concentrations in the range of 25–400 μM. Centrifugation speeds were between 23,000–54,000 rpm. Data sets were fitted to a single, ideal species model using Ultrascan (16). The partial specific volumes (0.7716 ml/g) of the peptide and the solvent density (1.0054 g/ml) were calculated using Sednterp (17).

Bioinformatics—The list of representative architectures for the protein family PF01391 (collagen triple helix repeat) was retrieved from the Pfam database (18). Positions of coiled-coil domains and collagen repeats were identified from these sequences and augmented with data from Marcoil (19), a hidden Markov model-based coiled-coil sequence predictor, using a confidence level of 0.9. For those sequences where the coiled-coil domain and the collagen repeats were within 15 residues, oligomer state predictions were carried out using a combination of SCORER (20) and LOGICOIL. Given a coiled-coil sequence, both algorithms predict whether that sequence forms a dimer or a trimer, based on evidence garnered from training sets of sequences of known oligomeric state. From these, conservative definitions were taken; if SCORER and LOGICOIL concurred in their oligomer-state prediction, that state was assigned; if they disagreed, no state was assigned. Two of the sequences had structurally characterized coiled-coil domains identified from Pfam and did not require oligomeric state prediction.

RESULTS

Design and Characterization of Synthetic α-Helical Coiled Coil—BS-3pCC4—A short peptide sequence predicted to form a highly stable parallel trimeric coiled-coil was designed to be amenable to characterization via solution phase biophysical methods, as well as subsequent gene synthesis and cloning into ScI2 constructs for gene expression and protein production. The starting point for the design was the coiled-coil heptad repeat, designated abcd efgh and visualized on the helical wheel of Fig. 1C. Different amino acid combinations at the predominately hydrophobic, a and d positions direct different oligomer states (21, 22). For the required design, the combination of a = d = Ile was used to prescribe parallel trimers. The interface was cemented further with oppositely charged residues at

5 T. L. Vincent, P. J. Green, and D. N. Woolfson, unpublished results.
Fusing Coiled-coil and Collagen Domains

Addition of GdnHCl, however, resulted in sigmoidal thermal unfolding transitions, e.g. in 3 M GdnHCl, the peptide unfolded with a midpoint temperature ($T_{m}$) of 55 °C. Finally, we have recently solved an x-ray crystal structure for the peptide, which confirms it as an in-register parallel, three-stranded $\alpha$-helical coiled-coil.

Design, Expression, and Purification of a Bacterial Collagen with Appended Coiled-coil Domains—Previously, we have used a cold-shock vector system, pCold II, to express a portion of the gene for the Scl2.28 collagen-like protein of S. pyogenes. This portion covers the N-terminal globular domain (V) and the triple helical domain of sequence (Gly-Xaa-Yaa)79 (CL) (10–12). The construct has an N-terminal His$_6$ tag for purification and a protease-susceptible sequence, LVPRGSP, between the V and CL domains; this construct is referred to as V-CL (Fig. 2). We also have reported a permuted construct, CL-V, (12). For the constructs presented here, the V domain was replaced by the designed three-stranded coiled-coil domain to yield CC-CL (Fig. 2). The resulting peptide was shown to be expressed in bacteria and purified by reverse-phase HPLC, and successful synthesis was confirmed by MALDI-TOF mass spectrometry (m/z [M + H]$^+$ observed, 3466.06; m/z [M + H]$^+$ theoretical 3465.95).

The CD spectrum of the BS-3pCC4 peptide in PBS at pH 7.0 (Fig. 3A), showed minima near 222 nm (MRE 222 nm = 3466.06 deg cm$^2$/dmol$^{-1}$) and 208 nm (MRE 208 nm = 27,806 deg cm$^2$/dmol$^{-1}$) as expected for a predominantly $\alpha$-helical structure. Sedimentation equilibrium data from analytical ultracentrifugation (supplemental Fig. S2) modeled well as a single ideal species with a mass of 10,520 Da, close to three times the molecular weight of BS-3pCC4 (i.e. 3 $\times$ 3466 = 10,398). The peptide was extremely stable under thermal unfolding, and a complete thermal transition was not seen using CD spectroscopy even when heated to 90 °C (Fig. 4C).

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.

Using these principles, a four-heptad peptide, expected to form a stable, three-stranded, parallel coiled coil was designed with short Gly-based linkers at both ends to allow flexibility of the $g$ and following $e$ positions, as these flank the hydrophobic core and often form salt bridges (Fig. 1C). The remaining $b$, $c$, and $f$ sites, which lie away from the helix-helix interface and have less influence on oligomerization, were made helix-favoring and polar residues Ala and Gln, respectively. The $d$ positions were used to add charge (via a lysine residue) for solubility, and a chromophore (Trp) was used to aid concentration determination.
Fusing Coiled-coil and Collagen Domains

The recombinant CC-CL and CL-CC proteins both gave CD spectra at 0 °C with a maximum at 220 and a minimum at 198 nm as well, but the magnitudes of these peaks were lower than for CL alone. From the sequences of these constructs, the four-heptad coiled coil and collagen (Gly-Xaa-Yaa)79 domains account for 10 and 90% of the residues, respectively. On the basis of these results, we calculated predicted spectra for the two hybrid sequences, i.e., using the observed CD spectra for CC domain (0.1 fraction) and of the CL domain (0.9 fraction). These calculated spectra were very similar to those observed (Fig. 3, C–E). This excellent additivity indicates that the coiled-coil and collagen triple helix domains retain their original structures without perturbation when in the same molecule. The low temperature CD spectrum of the larger construct, CC-CL-CC (Fig. 3G), had a reduced $M_{\text{RE}_{220}}$ of $-1237 \text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$, which is consistent with the additional coiled-coil domain contributing negative ellipticity at this wavelength. The observed and calculated spectra for this construct are in general agreement; although there is a larger discrepancy than found for the single coiled-coil constructs.

Monitoring the $M_{\text{RE}_{220}}$ with increasing temperature showed that all three fusion proteins had sharp thermal unfolding transitions centered $\sim 36–37^\circ \text{C}$ (Fig. 4A). Interestingly, this $T_m$ value is indistinguishable from that observed for V-CL and CL proteins. Because the BS-3pCC4 peptide does not unfold until very high temperatures, these melting transitions likely reflect denaturation of only the collagen triple helix. To confirm this, CD spectra for BS-3pCC4, CC-CL, CL-CC, and CC-CL-CC were recorded at 50 °C, (Fig. 3, right panels) where the collagen triple helix portion is expected to be unfolded, whereas the coiled-coil region is expected to be intact. The observed CD spectra for CC-CL and CL-CC showed typical $\alpha$-helix minima at 208 and 222 nm, with magnitudes in excellent agreement with those calculated from a weighted average of spectra for isolated folded BS-3pCC4 and unfolded CL at 50 °C (Fig. 3B). The observed CD spectrum for CC-CL-CC at 50 °C showed more pronounced minima at 208 and 222 nm, consistent with the expected increased $\alpha$-helix content. However, in this case, the observed magnitudes were less than calculated. It is possible that there is incomplete unfolding of the central triple helix domain in this highly constrained context, or that the coiled-coil and collagen domains perturb each other in some other way.

The very high stability of the coiled coil made it difficult to observe directly the melting of both domains of the hybrid molecule. Therefore, increasing molarities of GdnHCl were introduced to lower the stability of the coiled coil (Fig. 4C). Addition of 1–3 M GdnHCl allowed observation of the independent thermal melting of the collagen triple helix domain (decreasing $M_{\text{RE}_{220}}$), followed by melting of the $\alpha$-helical coiled coil at higher temperatures (increasing $M_{\text{RE}_{220}}$) (Fig. 4B). Clearly, there is a region of intermediate temperature where the triple helix is fully melted, whereas the coiled-coil structure remains intact. This facilitated the following refolding experiments.

Refolding Kinetics of Fused Coiled-coil and Collagen Domains—Previous studies show that recombinant V-CL refolds in vitro following heat denaturation, whereas the collagen triple helix domain CL alone does not refold significantly even after weeks.

---

**FIGURE 4. Thermal unfolding behavior of the coiled-coil/collagen constructs.** A, V-CL (○), CL (●), CC-CL (■); CL-CC (△), and CC-CL-CC (◇) at 20 μM peptide/protein in PBS, pH 7.0. CC-CL (△) and BS-3pCC4, i.e., CC alone (●), in the presence of guanidine HCl at 0 m (□), 1.0 m (△), 2.0 m (◇), and 3.0 m (○). $T_m$ of coiled coil denaturation in 3 M guanidine HCl, CC-CL (69 °C, B (◇)) and BS-3pCC4 (55 °C, C (○)).
Parallel, three-stranded coiled coil adjacent to a proline-rich (Gly-Xaa-Yaa)_{79} domain that forms a stable collagen-like triple helix. As we discuss further below, there are natural precedents for such arrangements. Therefore, the juxtaposition of these domains in relatively "pure" forms within the same polypeptide makes an interesting study, particularly given the structural similarities and differences between the two motifs and the likely different mechanisms of folding.

Both the collagen triple helix and the three-stranded coiled coil are rod-like assemblies in which three chains adopt defined secondary structures, combine, and intertwine to form supercoiled structures. However, this is where the similarities end. The collagen triple helix comprises three polyproline II-like helical chains supercoiled about a common axis, with hydrogen bonding between adjacent chains. The chains are staggered by one residue to allow close packing of the invariant glycine residues within the core of the triple helix (Fig. 1A). In contrast, in three-stranded a-helical coiled coils, each chain folds as a self-contained, internally hydrogen-bonded a-helix. Patterns of hydrophobic and polar residues on the surface of these structures drive helix association, which is usually perfectly in register (Fig. 1B).

These sequence and structural differences affect the mechanism of folding of the two domains. In vitro assembly of (Gly-Xaa-Yaa)_n repeats into triple helices is slow, with folding taking from minutes to days for collagens and model triple helical peptides (7, 25). Factors that limit triple helix folding include the following: 1) slow cis-trans isomerization of imino acids, which can limit nucleation and propagation of the secondary structure; 2) tertiary association; 3) (related to 1) the need for chains to adopt correct dihedral angles to allow close packing and hydrogen bonding between chains; and 4) the possibility of misfolding and misalignment of chains due to the repetitiveness of the sequences (25, 26). By contrast, the more modular assembly of a-helical coiled coils makes their folding generally rapid and efficient on the μs-ms time frame (27). However, we note for a-helical coiled coils that some require particular sequence motifs to trigger folding and assembly (28) and that the more general structure and mechanism of assembly can lead to promiscuity (21, 23) as many more coiled-coil architectures are possible in addition to the three-stranded parallel type (3, 29).

The study presented here for constructs containing both the trimeric a-helical coiled-coil and the collagen triple helix motif suggests little influence of one domain on the conformation of the other, which may not be surprising for these rod-like, linear motifs. At least in aqueous buffers without denaturants (see below), this independence extended to the thermal stabilities of the two domains; the ~37 °C thermal transition for the collagen triple helix remained unchanged regardless of the location of the coiled-coil domain(s), and the thermal stability of the coiled coil remained high and similar to that for an isolated, synthetic peptide of the same sequence. The use of a highly stable coiled coil allowed melting of the collagen triple helix domain, whereas the coiled-coil domain(s) remained intact, i.e. above 40 °C. This contrasts with the melting of a construct based on the full-length natural protein, V-CL, where the two domains are closer in thermal stability (T_m of the V domain ~ 46 °C (12))
Fusing Coiled-coil and Collagen Domains

and that of CL \( \sim 36 - 37 \, ^\circ C \) (Fig. 4A)), and the domains influence each other. The V-CL fusion protein melts at \( \sim 37 \, ^\circ C \), which is considerably lower than the isolated V domain. This indicates coupling in V-CL, whereby the V domain starts unfolding when the CL domain unfolds.

Replacement of the more complex natural V domain by the far more stable \( \alpha \)-helical coiled-coil sequence changed the nature of thermal transitions in the system. Thermal unfolding of the stable CC domain was not observed in the recombinant constructs until 3 \( m \) guanidine HCl was added (Fig. 4B). At this point, the CL domain appears to be completely unfolded from the outset of the thermal denaturation experiments (i.e. 0 \( ^\circ C \)), and the CC domain unfolds with a sigmoidal curve with a \( T_m \) of 69 \( ^\circ C \). This compares with 55 \( ^\circ C \) for the BS-3pCC4 peptide alone. Thus, the construct stabilizes the CC domain in some way. We suggest that any residual structure in or near the junction of the two domains (i.e. some remaining collagen or coiled-coil triple helix) will stabilize this domain through an entropic effect. Otherwise, we have no explanation for this effect at present, and further high resolution structural studies will be required to resolve this issue.

Turning to the kinetics of refolding and assembly upon cooling from 50 \( ^\circ C \), the CL domain alone assembles slowly and incompletely over a period of weeks, whereas all of the engineered constructs combining the CC or V domains with a CL unit fold faster and to near completion. The ability of a CC domain on either the N or C terminus to promote triple helix formation in the CL domain is consistent with previous studies, showing that triple helix formation can be nucleated from either end (26). However, in contrast to the equilibrium CD spectra and thermal melting curves, the refolding rates do depend on the location of the CC domain. In the natural protein, the V domain is N-terminal to the CL domain, and moving it to the C-terminal slows down folding. For the designed fusions, however, the reverse is true; a C-terminal CC domain is more effective in promoting folding. Moreover, the V-CL protein is completely denatured at 50 \( ^\circ C \), with no remaining \( \alpha \)-helical structure; yet this protein refolds as quickly as the engineered CC-CL fusion where the coiled-coil structure is intact, and only the triple helix portion refolds. Very rapid refolding; i.e. within the 1 to 2 min it takes to cool the sample from 50 to 0 \( ^\circ C \), is observed for the isolated V domain,7 consistent with the V domain undergoing rapid trimerization prior to CL folding in V-CL. Taken together, the data support a relation between the V and CL domains that is highly evolved in the natural protein, such that the N-terminal original placement is optimized for folding, whereas the C-terminal is not. Such a relationship would most likely be disrupted when the V domain is placed C-terminal to CL. In our engineered constructs, the coupling between the domains seen in the natural protein appears to be lost. This could be related to the very large difference in stability between the CL and CC domains, compared with the closer \( T_m \) values of CL and V, or alternatively, to the inclusion of flexible Gly linkers between the CC and CL domains. Thus, in our engineered constructs, we propose that it is simply the proximity of the CC domains that promote CL assembly from either end rather than some specific interaction. A full understanding of this awaits structural resolutions of the V domain and the various V/CC-CL constructs, which are currently underway.

Incorporation of CC domains at both ends of the CL domain did not affect CL stability, although the triple helix content appeared a little lower than expected. However, this construct, CC-CL-CC, refolded from 50 \( ^\circ C \) with a half-time \( \sim 100 \) times longer than that of the fastest folding construct, CL-CC (Table 1). Thus, it appears that bracketing the collagen triple helix with two coiled coils hinders the assembly of the former. We posit that in the single coiled-coil constructs, folding of the collagen domain is initiated from the end proximal to the folded coiled coil and propagates toward the distal end. In this way, any misfolding or out-of-register alignment of the polypeptide chains is either avoided or readily corrected by unwinding. However, with the chains stabilized by trimeric coiled coils at both ends, folding and propagation can be initiated from both ends with the possibility of mismatched registers near the middle of the CL domain (Fig. 6). The results on the recombinant CC-CL-CC contrast with studies on bovine type III collagen, where the folding rate and folding directionality of the molecule with disulfide bonds at both ends is very similar to that with a disulfide bond only at the C terminus (26). The longer length of the type III collagen and the discontinuity between the N-terminal disulfide and the main triple helix may partially explain this difference. Examination of a simple model of three twisted ropes tethered at both ends revealed difficulty in unfolding and unwinding the central triple helix, whereas both ends are kept fixed, an observation that may be more relevant for the shorter triple helix.

Our data are consistent with and indeed shed further light on the domain structure of natural collagen proteins. In these, the presence of non-collagen domains, such as the V domain in ScI2, adjacent to the tripeptide repeats leads to rapid trimerization at one end and forces the three chains of the (Gly-Xaa-Yaa)\(_n\) sequence into close proximity at that end, promoting proper registration and nucleation of the triple helix. As suggested by Hoppe et al. (4), such results show that the non-collagen protein sequences, which have no inherent staggered arrangement of chains are sufficient to promote triple helix

\footnote{7 E. Hwang, Z. Yu, and B. Brodsky, unpublished data.}

\( A \) \( B \) \( C \)

\begin{center}
\includegraphics[width=0.8\textwidth]{figure6.png}
\end{center}

**FIGURE 6. Scheme of the possible mechanisms of collagen refolding in the three CL/CC constructs CC-CL (A), CL-CC (B), and CC-CL-CC (C).** In each case, step 1 depicts the CL domains undergoing thermal denaturation, and step 2 depicts their refolding. The constructs with a single CC domain (A and B) refold completely, whereas that with a CC at both ends folds incompletely. We posit that this is due to register mismatch when collagen assembly is nucleated from both ends.
Fusing Coiled-coil and Collagen Domains

nucleation in the correct one-residue stagger. In an extensive study (6), which examined the distribution of collagen and coiled-coil domains, McAlinden et al. (6) state that “coiled-coil domains are present in most members of the collagen superfamily, located either before, after or between collagen-like regions, suggesting a general role in triple helix assembly.” We sought to investigate this further with a focus, consistent with our study, on protein architectures containing coiled-coil and collagen domains proximal (i.e. <15 residues) to one another, as opposed to the more general cases examined by McAlinden et al. (6).

We used the Pfam database (18) to identify proteins that contain collagen and α-helical coiled-coil domains within 15 residues of each another. Each architecture is represented by one example protein sequence, shown as a line and with domains blocked as follows: white, collagen repeats; black, predicted trimeric coiled coils; cross-hatched, predicted dimeric coiled coils; and with domains blocked as follows: white, collagen repeats; black, predicted trimeric coiled coils; and with domains blocked as follows: white, collagen repeats; black, predicted dimeric coiled coils; and gray, those predicted as coiled coil, but with inconclusive oligomer state. Numbers of protein sequences in Pfam for each architecture are shown in gray, those predicted as coiled coil, but with inconclusive oligomer state. Predicted trimeric coiled coils; and with domains blocked as follows: white, collagen repeats; black, predicted trimeric coiled coils; and with domains blocked as follows: white, collagen repeats; black, predicted dimeric coiled coils; and gray, those predicted as coiled coil, but with inconclusive oligomer state.

FIGURE 7. Architectures of unique proteins that contain collagen and α-helical coiled-coil domains within 15 residues of each another. Each architecture is represented by one example protein sequence, shown as a line and with domains blocked as follows: white, collagen repeats; black, predicted trimeric coiled coils; cross-hatched, predicted dimeric coiled coils; and gray, those predicted as coiled coil, but with inconclusive oligomer state. Numbers of protein sequences in Pfam for each architecture are shown in parentheses after the example sequence identifier. Oligomeric states could be unambiguously assigned in the case of P and R, which had structurally characterized coiled-coil domains identified in Pfam. Image created using the Perl Bio::Graphics module.

protein sequence. Marcoil (19) was then used to predict all coiled-coil domains within these sequences. This returned 54 protein architectures that contained a coiled-coil region along with the collagen repeat. Only representative proteins with a coiled-coil domain within 15 residues of a collagen repeat are highlighted in Fig. 7. Two things are immediately apparent from this figure: first, all but two of these architectures have coiled-coil regions predicted on just one side of the collagen domain, fully consistent with our data and hypothesis; second, in the vast majority of cases, all except three sequences, the coiled-coil regions are N-terminal to the collagen region. Using the coiled-coil oligomer-state prediction algorithms SCORER (20) and LOGICOIL, 8/15 of the N-terminal coiled coils predicted as trimers, and 7/15 predictions were inconclusive. Two of the three C-terminal coiled coil domains have been structurally characterized and are trimeric; the third is predicted to be a weak dimer. The absence of any collagen domains flanked by coiled coils on both ends is consistent with the poorer folding that we observe with such an arrangement experimentally.

Perhaps the most important observation from the bioinformatics analysis, however, is that when coiled-coil and collagen regions are found within 15 residues of one another, they are almost exclusively arranged such that the coiled coil is N-terminal to the collagen domain; this extended to the set of 54 architectures, in which only six have coiled-coil domains C-terminal of a collagen domain. These results largely concur with those reported by McAlinden et al. (6) but highlight that when only proteins containing adjacent (<15 residues apart) coiled-coil and collagen domains are considered (as opposed to a more general case), there is a clear preference for architectures that place trimeric coiled coils N-terminal to their respective collagen domain. This suggests a possible role for coiled coils in the folding and assembly of collagen domains as they emerge from the protein synthesis machinery in the cell.

Acknowledgment—We thank Eileen Hwang for helpful discussions.

REFERENCES
Fusing Coiled-coil and Collagen Domains

Designed Coiled Coils Promote Folding of a Recombinant Bacterial Collagen
Ayumi Yoshizumi, Jordan M. Fletcher, Zhuoxin Yu, Anton V. Persikov, Gail J. Bartlett, Aimee L. Boyle, Thomas L. Vincent, Derek N. Woolfson and Barbara Brodsky


Access the most updated version of this article at doi: 10.1074/jbc.M110.217364

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/03/28/M110.217364.DC1

This article cites 27 references, 8 of which can be accessed free at http://www.jbc.org/content/286/20/17512.full.html#ref-list-1