Noncollagenous region of the streptococcal collagen-like protein is a trimerization domain that supports refolding of adjacent homologous and heterologous collagenous domains

Zhuoxin Yu, Oleg Mirochnitchenko, Chunying Xu, Ayumi Yoshizumi, Barbara Brodsky,* and Masayori Inouye*

Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854

Received 27 October 2009; Revised 27 January 2010; Accepted 28 January 2010

DOI: 10.1002/pro.356

Published online 16 February 2010 proteinscience.org

Abstract: Proper folding of the (Gly-Xaa-Yaa)_n sequence of animal collagens requires adjacent N- or C-terminal noncollagenous trimerization domains which often contain coiled-coil or beta sheet structure. Collagen-like proteins have been found recently in a number of bacteria, but little is known about their folding mechanism. The Scl2 collagen-like protein from *Streptococcus pyogenes* has an N-terminal globular domain, designated V_{sp}, adjacent to its triple-helix domain. The V_{sp} domain is required for proper refolding of the Scl2 protein *in vitro*. Here, recombinant V_{sp} domain alone is shown to form trimers with a significant α -helix content and to have a thermal stability of T_m = 45°C. Examination of a new construct shows that the V_{sp} domain facilitates efficient *in vitro* refolding only when it is located N-terminal to the triple-helix domain but not when C-terminal to the triple-helix domain. Fusion of the V_{sp} domain N-terminal to a heterologous (Gly-Xaa-Yaa)_n sequence from *Clostridium perfringens* led to correct folding and refolding of this triple-helix, which was unable to fold into a triple-helical, soluble protein on its own. These results suggest that placement of a functional trimerization module adjacent to a heterologous Gly-Xaa-Yaa repeating sequence can lead to proper folding in some cases but also shows specificity in the relative location of the trimerization and triple-helix domains. This information about their modular nature can be used in the production of novel types of bacterial collagen for biomaterial applications.

Keywords: collagen; *streptococcus*; trimerization; folding; triple-helix

Abbreviations: Scl2, streptococcal collagen-like protein 2; V_{sp}, globular domain of *S. pyogenes* Scl2 protein; CL_{sp}, collagen domain of *S. pyogenes* Scl2 protein; CL_{cp}, collagen domain of *C. perfringens* protein; Hyp, hydroxyproline; GAS, group A *streptococcus*; DLS, dynamic light scattering; Rh, hydrodynamic radius; CD, circular dichroism; DSC, differential scanning calorimetry; BclA, *Bacillus anthracis* collagen-like protein A; CTD, C-terminal domain of *B. anthracis* BclA protein.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH grant; Grant number: R21EB007198.

*Correspondence to: Masayori Inouye, Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, 679 Hoes Lane, Piscataway, NJ 08854. E-mail: inouye@umdnj.edu or Barbara Brodsky, Department of Biochemistry, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854. E-mail: brodsky@umdnj.edu

Introduction

Collagen is a ubiquitous feature of the extracellular matrix of multicellular animals. The characteristic conformation of collagen is a supercoiled triple-helix that requires the distinctive (Gly-Xaa-Yaa)_n repeating tripeptide sequence.^{1–3} Recently, collagen-like proteins with long stretches of (Gly-Xaa-Yaa)_n repeats and a high proline content have been found in many bacteria and bacteriophages.⁴ The length and amino acid composition of these putative triple-helix domains vary widely. They are distinguished from animal collagens by the absence of hydroxyproline (Hyp), as bacteria do not appear to have the enzyme prolyl hydroxylase.⁵ Despite the absence of Hyp, a number of bacterial proteins with a (Gly-Xaa-Yaa)_n sequence have been shown to form typical collagen triple-helix structures with stabilities close to that of mammalian collagens.^{6–9} Electrostatic interactions make major contributions to the stabilization of some of these bacterial collagen-like proteins,⁷ whereas others may be influenced by a very high content of proline or large amounts of polar residues such as Thr and Gln.⁴

Animal collagens are modular in nature. Triple-helix domains are flanked by noncollagenous domains shown to have distinct structural and biological functions.^{10,11} For instance, the globular C-propeptide of Type I collagen has been shown to be responsible for trimerization and chain selection,¹² and may be important in chain registration, while the noncollagenous C-terminal domain of type XVIII collagen is processed to yield the angiogenic peptide endostatin.¹³ Bacterial collagen-like proteins are also likely to have a similar organization, as the (Gly-Xaa-Yaa)_n triple-helix domain is flanked by variable lengths of sequence that could represent independent folding domains with distinct structural and functional roles. For example, the N-terminal globular domains of the *Streptococcus pyogenes* collagen-like proteins have been implicated in folding.^{6,8} In *Bacillus anthracis*, the highly stable beta-sheet-containing-terminal globular domain is likely to be important for folding and stability of the BclA triple-helix, whereas its N-terminal noncollagenous domain is essential for basal layer attachment.^{9,14,15}

Two collagen-like proteins of *S. pyogenes* Scl1 and Scl2 have been well characterized in terms of protein structure and their potential role as GAS virulence factors.^{6,16–20} The collagen-like domain of Scl1, but not Scl2 can interact with the $\alpha 2\beta 1$ integrin collagen receptor, promoting adherence to mammalian cells, internalization and re-emergence into the extracellular environment.^{21–24} The Scl1 and Scl2 proteins can bind to thrombin-activatable fibrinolysis inhibitor, which may counteract the host response.²⁵ Both Scl1 and Scl2 related proteins have been expressed in *E. coli*, and their conformation, stability, and folding have been studied.^{6–8} The Scl2

protein from GAS serotype M28 contains a signal peptide, an N-terminal globular domain (denoted as V_{sp}) of 73 residues, a collagen-like domain (CL_{sp}) (Gly-Xaa-Yaa)₇₉, and a cell wall binding domain which contains an LPXTG motif followed by a hydrophobic transmembrane sequence and a short charged tail. The lengths of the V_{sp} domain vary from 58 to 77 residues in different strains and the sequences are specific to the M serotypes.¹⁸ As expressed in *E. coli*, the Scl2 V_{sp}-CL_{sp} protein folds correctly into a stable triple-helix, and this construct is capable of refolding *in vitro* following heat denaturation.^{6,8}

To further explore the role of the N-terminal globular domain in triple-helix folding, the biochemical and biophysical properties of the recombinant V_{sp} protein were characterized, and its location with respect to the triple-helix was modified. In addition, fusion of V_{sp} domain N-terminal to a (Gly-Xaa-Yaa)₆₃ heterologous collagen domain CL_{cp} from *Clostridium perfringens* facilitates its correct folding in *E. coli* and its refolding *in vitro*. The ability of the V_{sp} domain to fold a heterologous collagenous domain supports the modular nature of these proteins and raises the possibility that a bacterial trimerization domain may facilitate production of a range of prokaryotic collagen-like proteins for biomaterial applications.

Results

Characterization of V_{sp} domain

A construct of the V_{sp} domain from Scl 2.28 (*Streptococcal* collagen-like protein 2 from a serotype M28 GAS) with an N-terminal His-tag was constructed and was found to be expressed well in an *E. coli* cold shock expression system [Fig. 1(A)].²⁶ The protein was found in the soluble fraction [Fig. 1(B)], and purification of the V_{sp} domain yielded a single band on the SDS-PAGE gel with an apparent size similar to the expected molecular weight 10.2 kD [Fig. 1(C), Lane 1].

To investigate whether this isolated domain will oligomerize, glutaraldehyde was added to covalently cross-link the V_{sp} protein prior to SDS-PAGE analysis. The apparent sizes of the bands correspond to the predicted molecular weights of trimer, dimer, and monomer, indicating that the V_{sp} domain is capable of forming a trimer [Fig. 1(C), Lane 2]. When applied to a gel filtration column, the V_{sp} domain was eluted as a single peak with a predicted molecular weight of 31 kD, supporting its existence as a trimer in solution [Fig. 1(D) and Supporting Information Fig. 1]. Dynamic light scattering (DLS) studies of the V_{sp} domain at 4°C indicated a hydrodynamic radius (R_h) value of 3.3 ± 0.2 nm, which corresponds to a mass of 53 ± 10 kD for a globular

Figure 1. Expression and characterization of *V_{sp}* domain (A) Sequence of the His-tagged *V_{sp}* construct, beginning with a translation enhancing element MNHKV followed by a His-tag of 6 histidines (in italics) and the sequence of *V_{sp}* domain from Scl 2.28. Above the sequence is shown the coiled-coil structure prediction by COILS program. The cylinders indicate predicted coiled-coil regions. (B) SDS-PAGE of the expression of His-tagged *V_{sp}* in *E. coli*. Lane 1, cell lysate without IPTG induction; Lane 2, cell lysate with IPTG induction; Lane 3, insoluble fraction; Lane 4, soluble fraction. (C) SDS-PAGE of purified *V_{sp}* alone and with glutaraldehyde cross-linking. Lane 1: *V_{sp}* purified on a Ni-column, showing its apparent size; Lane 2, *V_{sp}* after cross-linking with glutaraldehyde, showing bands of putative monomers, dimers and trimers. (D) Gel filtration of the *V_{sp}*, with a single peak eluted at ~180 min, which corresponds to a molecular weight ~31 kD. See Supporting Information Figure 1 and Table for the gel filtration molecular weight standards.

protein (data not shown). This mass is larger than the 31 kD expected for a trimer, consistent with a shape that is somewhat elongated rather than spherical. The DLS experiments also detected a larger species with $R_h = 49.5 \pm 7.1$ nm, in addition to the $R_h = 3.3$ nm species, suggesting that small amounts of *V_{sp}* aggregates may be present.

The circular dichroism (CD) spectrum of the *V_{sp}* domain (PBS buffer, $c = 0.6$ mg/mL, 4°C) shows two minima at 208 nm and 222 nm and a positive peak at 198 nm, which are characteristic of α -helical structure [Fig. 2(A), solid line]. Deconvolution of the spectrum using the program CDNN (written by Gerald Böhm, Institut für Biotechnologie, Martin-Luther-Universität, Halle-Wittenberg, Germany) gives values of ~55% α -helical structure, ~10% β -strand, ~15% β -turn, and ~20% random coil structure. Analysis of the amino acid sequence using the COILS program (http://www.ch.embnet.org/software/COILS_form.html) predicts coiled-coil structure between residues 12–29 and between residues 56–71 [Fig. 1(A)].

The thermal stability of *V_{sp}* ($c = 0.6$ mg/mL) was determined by monitoring the CD signal at 220 nm with increasing temperature [Fig. 2(B), solid line, right arrow]. The observed thermal transition is cooperative with $T_m = 45^\circ\text{C}$, and the T_m value is concentration dependent [Fig. 2(C)], supporting a trimer to monomer transition as well as unfolding of the monomer chain. Refolding of the *V_{sp}* domain was monitored following $\text{MRE}_{220\text{nm}}$ as the temperature was decreased from 70°C to 0°C at the same rate [Fig. 2(B), solid line, left arrow], and the similarity to the unfolding curve suggests a transition that is close to equilibrium under these conditions. About 90% of the initial CD signal is recovered after refolding, indicating that the *V_{sp}* domain is able to refold on its own. After refolding, the *V_{sp}* domain again forms a trimer, as indicated by cross-linking studies showing an SDS-PAGE gel profile very similar to that seen in Figure 1(C) for the cross-linked *V_{sp}* domain before denaturation (data not shown).

The *V_{sp}* domain ($c = 1.0$ mg/mL) was also examined by differential scanning calorimetry (DSC) [Fig. 2(D)], giving a single transition with $T_m = 48.0^\circ\text{C}$ and $\Delta H_{\text{cal}} = 320$ kJ/mol. The DSC profile is completely reversible, again supporting the reversibility and equilibrium nature of the *V_{sp}* thermal transition.

The *V_{sp}* structure is not stable under acidic conditions. The CD spectrum of *V_{sp}* ($c = 0.6$ mg/mL) at pH 2.9 [Fig. 2(A), dash line] shows a large decrease in magnitude at 222 nm and a change in shape, indicating partial loss of α -helical structure. Deconvolution of the spectrum at pH 2.9 suggests the presence of ~20% α -helical structure, ~25% β -strand, ~20% β -turn, and ~35% random coil structure. No thermal transition is seen at pH 2.9 [Fig. 2(B), dash line].

In vitro refolding of the *S. pyogenes* Scl2 protein

The recombinant collagenous domain of the Scl1 protein was previously shown to fold into a stable triple-helix when expressed in *E. coli* without its N-terminal globular domain.⁶ Similar correct triple-helix folding is also observed for the recombinant (Gly-Xaa-Yaa)₇₉ CL_{sp} domain from the Scl2 protein expressed as His-tagged CL_{sp} in *E. coli* [Fig. 3(A), top construct]. The recombinant CL_{sp} domain shows a characteristic collagen CD spectrum and a thermal stability of 36°C [Fig. 3(B), right arrow]. The stability of this recombinant triple-helix domain is the same as seen for the triple-helix domain obtained after trypsin digestion of *V_{sp}*-CL_{sp}.⁸ But after heat denaturation, the collagenous domain of the Scl2 protein shows no ability to refold *in vitro* on this time scale [Fig. 3(B), left arrow], consistent with results reported previously for the Scl1 triple-helix domain.⁶

Figure 2. Circular dichroism spectrum, thermal stability, and refolding of the V_{sp} domain. (A) CD spectra of V_{sp} domain at pH 7.0 (solid line) and pH 2.9 (dash line). (B) Melting and refolding of the V_{sp} domain at pH 7.0 (solid line) as monitored by CD. →, melting with increasing temperature; ←, refolding with decreasing temperature. The dash line shows the MRE_{220nm} as a function of increasing temperature at pH 2.9. (C) Concentration dependence of the equilibrium melting temperatures. C_M indicates the concentration of V_{sp} monomers, which are 36 μM, 90 μM, 180 μM, and 360 μM in this study, and T_m is the melting temperature at which the fraction folded is 50% in the curve fitted to thermal transition. (D) DSC of the V_{sp} domain. Heat capacity (C_p) is plotted against increasing temperature.

To investigate whether the position of V_{sp} domain with respect to the triple-helix domain is important for folding in *E. coli* or refolding *in vitro*, a construct was designed to make His-tagged CL_{sp}-V_{sp} protein, where the globular domain is now C-terminal to the triple-helix [Fig. 3(A), bottom construct]. Recombinant CL_{sp}-V_{sp} forms a soluble triple-helix containing molecule with a CD spectrum (MRE₂₂₀~1470 deg cm² dmol⁻¹) very similar to that of the original V_{sp}-CL_{sp} construct. The thermal denaturation of CL_{sp}-V_{sp} protein purified from *E. coli* is similar to that seen for the V_{sp}-CL_{sp} protein and the CL_{sp} domain alone [Fig. 3(B,C,D), right arrow]. However, there is some indication of a second transition for CL_{sp}-V_{sp}, where the CD signal at 220nm starts to increase after the temperature reaches ~40°C, which could reflect independent denaturation of the V_{sp} domain.

In vitro refolding experiments were carried out after heat denaturation on CL_{sp}, V_{sp}-CL_{sp}, and CL_{sp}-V_{sp} proteins. No indication of refolding is seen for His-CL_{sp} [Fig. 3(B), left arrow], while V_{sp}-CL_{sp} protein successfully achieves a substantial amount of

refolding [Fig. 3(C), left arrow] as reported previously.^{6,8} The inability of V_{sp}-CL_{sp} protein to fully refold could be related to aggregation, which removes some of the protein from solution, or to the conditions of refolding. Experiments with other conditions have resulted in much higher percentages of refolding (data not shown). In contrast to V_{sp}-CL_{sp}, CL_{sp}-V_{sp} does not appear to be capable of significant *in vitro* refolding [Fig. 3(D), left arrow]. As the temperature drops below 45°C, the CD signal starts to decrease, suggesting the formation of α-helix and refolding of the V_{sp} domain. When the temperature falls below 25°C, there is a small increase in MRE_{220nm}, but it never reaches a value expected for formation of a significant amount of triple-helix.

To investigate the relationship between refolding of these proteins and trimerization, cross-link experiments were carried out to determine whether the CL_{sp}, V_{sp}-CL_{sp}, and CL_{sp}-V_{sp} proteins reform trimers after cooling the denatured protein [Fig. 4(A)]. The presence of cross-linked oligomeric species suggests that V_{sp}-CL_{sp} and CL_{sp}-V_{sp} proteins reform trimers, whereas CL_{sp} does not. The drop in MRE₂₂₀

Figure 3. Melting and refolding when the V_{sp} domain is located N-terminal vs. C-terminal to the triple-helix domain. (A) Schematic drawing of the constructs of CL_{sp} , V_{sp} - CL_{sp} , and CL_{sp} - V_{sp} . (B) Melting and refolding curves of CL_{sp} as monitored by MRE_{220nm} . (C) Melting and refolding curves of V_{sp} - CL_{sp} , as monitored by MRE_{220nm} . (D) Melting and refolding curves of CL_{sp} - V_{sp} , as monitored by MRE_{220nm} .

upon refolding and the presence of cross-linked oligomers suggest the V_{sp} domain of the CL_{sp} - V_{sp} construct can refold to form trimers, but this is not followed by effective folding of the adjacent CL_{sp} domain. These experiments demonstrate that the V_{sp} domain can trimerize on its own, whether it is N-terminal or C-terminal to the triple-helix but it can only promote efficient *in vitro* refolding when on the N-terminal side of the CL_{sp} domain.

V_{sp} domain facilitates the folding and refolding of a heterologous collagen-like protein

The ability of V_{sp} to promote folding of a heterologous triple-helix, which cannot fold on its own, was examined. The bacterium *C. perfringens* was found to contain a gene encoding a collagen-like protein,²⁷ composed of a 53 residue N-terminal non-collagenous domain (N_{cp}), a collagen-like domain (Gly-Xaa-Yaa)₆₃ (denoted as CL_{cp}) and a 161 residue C-terminal non-collagenous domain (C_{cp}) [Fig. 5(A)]. Four constructs N_{cp} - CL_{cp} - C_{cp} , N_{cp} - CL_{cp} , CL_{cp} - C_{cp} , and CL_{cp} [Fig. 5(A)] were previously expressed in *E. coli* using the cold shock vector system and all were found in inclusion bodies, suggesting an inability to correctly fold to form stable triple-helices.²⁷ In an attempt to promote proper folding, the V_{sp} domain from *S. pyogenes* was fused N-terminal to the CL_{cp}

domain, creating the chimeric protein V_{sp} - CL_{cp} [Fig. 5(A)]. The V_{sp} - CL_{cp} protein was fully soluble [Fig. 5(B)], suggesting that the V_{sp} domain facilitates the correct folding of this heterologous collagen-like domain as expressed in *E. coli*.

To further determine whether V_{sp} - CL_{cp} forms a proper triple helix, the physicochemical properties of V_{sp} - CL_{cp} and the CL_{cp} domain, obtained by trypsin treatment of V_{sp} - CL_{cp} , were determined. The CD spectra of both V_{sp} - CL_{cp} and the CL_{cp} domain show characteristic features of the collagen triple-helix (data not shown). The magnitude of the CD peak at 220 nm is lower for V_{sp} - CL_{cp} than for the CL_{cp} domain, as expected when the V_{sp} domain contributes a negative α -helical peak around 222 nm that partly cancels out the positive 220 nm peak of the collagen triple-helix signal.^{6,8} The thermal transitions of both V_{sp} - CL_{cp} and CL_{cp} domains are not as sharp as seen for V_{sp} - CL_{sp} and CL_{sp} domains and show some sign of multiple transitions [Fig. 5(C,D), right arrow]. The melting temperatures were $T_m = 40.6^\circ\text{C}$ for V_{sp} - CL_{cp} and $T_m = 38.8^\circ\text{C}$ for CL_{cp} , suggesting that the presence of the V_{sp} domain provides a small degree of stabilization. The *in vitro* refolding of both proteins was also studied [Fig. 5(C,D), left arrow]. For V_{sp} - CL_{cp} , cooling results in an initial drop in the MRE_{220nm} CD signal, consistent with the formation

Figure 4. SDS-PAGE of the cross-linking studies of the bacterial collagen-like proteins and their collagenous domains, to define the trimerization state of the proteins initially and after *in vitro* refolding. (A) Cross-linking studies of CL_{sp}, V_{sp}-CL_{sp}, and CL_{sp}-V_{sp}. Lane 1, purified CL_{sp} domain, showing a single band at a slower migration position expected for the rod-like triple-helix; Lane 2, cross-linked CL_{sp}, showing bands of putative monomers, dimers and trimers, with a somewhat weaker band of undefined source between the dimer and trimer positions; Lane 3, CL_{sp} cross-linked after denaturation and refolding, showing only one band at a position slightly larger than the monomer; Lane 4, purified V_{sp}-CL_{sp}, showing a very strong band at the expected monomer position and a weak band near the dimer position; Lane 5, cross-linked V_{sp}-CL_{sp}, showing bands at positions expected for monomers, dimers and trimers; Lane 6, V_{sp}-CL_{sp} cross-linked after denaturation and refolding, showing monomer, dimer, and trimer bands; Lane 7, purified CL_{sp}-V_{sp}; Lane 8, cross-linked CL_{sp}-V_{sp}, showing bands expected for monomers, dimers and trimers; Lane 9, CL_{sp}-V_{sp} cross-linked after denaturation and refolding, showing monomers, dimers, and trimers again. (B) Cross-linking studies of V_{sp}-CL_{cp} and the purified CL_{cp} domain obtained by trypsin digestion. Lane 1, purified V_{sp}-CL_{cp}; Lane 2, cross-linked V_{sp}-CL_{cp}, showing bands of putative monomers, dimers, trimers, as well as some higher molecular weight oligomers; Lane 3, V_{sp}-CL_{cp} cross-linked after denaturation and refolding, showing monomers, dimers, and trimers; Lane 4, purified CL_{cp} showing one strong monomer band and a weaker faster migrating band which could reflect partial degradation; Lane 5, cross-linked CL_{cp}, showing bands of putative monomers, dimers, and trimers; Lane 6, CL_{cp} cross-linked after denaturation and refolding, showing only monomers.

of α -helical structure, followed by an increase in MRE_{220nm} and the recovery of about 80% of the initial CD signal. This suggests refolding of the N-terminal V_{sp} domain must occur before refolding of the triple-helix structure takes place. Cross-link experimental evidence indicates that V_{sp}-CL_{cp} trimers are reformed [Fig. 4(B)]. In contrast, after the isolated CL_{cp} domain alone is heat denatured and then cooled down, no CD signal is recovered and no cross-linked trimers are formed [Figs. 4(B) and 5(D)], implying that V_{sp} domain is required for efficient and proper refolding of the heterologous collagen-like CL_{cp} domain.

Discussion

Although the folding of the rod-like collagen triple-helix molecule appears to be straightforward compared with the folding of a globular protein, a number of factors make triple-helix folding a complex, multistep, slow process that requires a noncollagenous domain to promote. The repetitive nature of its (Gly-Xaa-Yaa)_n sequence and the interchain nature of its peptide hydrogen bonds complicate trimerization and nucleation in the correct 1-residue stagger register, and linear propagation of the triple-helix is limited by the slow *cis-trans* isomerization of the

many imino acids in collagen.²⁸ To facilitate proper folding of animal collagens and other triple-helix containing proteins, noncollagenous domains adjacent to the (Gly-Xaa-Yaa)_n sequence independently form stable trimers and act to promote triple-helix nucleation. The structures of many of these folding domains have been shown to contain three-stranded α -helical coiled-coil motifs, such as those in collectins and transmembrane collagens, and analysis shows that they can be located C-terminal as well as N-terminal to the triple-helix domain.^{29–31} Meanwhile, high resolution structures have shown beta-sheet structures in the NC1 domains of network-forming collagens Type IV, Type VIII, and Type X,^{32–35} and the C-terminal noncollagenous region of Type XVIII collagen.³⁶

Examination of bacterial collagens extends the categories of collagen folding domains. The globular domain C-terminal to the triple-helix (denoted as CTD for C-terminal domain) in the *B. anthracis* collagen-like protein BclA forms trimers containing a beta-sheet structure¹⁴ with an extremely high stability near 95°C in PBS buffer.⁹ In the recombinant BclA protein, the triple helix shows $T_m = 37^\circ\text{C}$, while the trimers do not dissociate until 95°C, consistent with trimerization determined by the CTD.

Figure 5. V_{sp} domain facilitates the folding and refolding of the heterologous CL_{cp} domain. (A) Schematic drawing of constructs derived from the *C. perfringens* collagen-like protein. The protein is divided into three domains, an N-terminal domain (N_{cp}), a collagen-like domain (CL_{cp}) with (Gly-Xaa-Yaa)₆₃ repeats and a C-terminal domain (C_{cp}). V_{sp}-CL_{cp} is a chimera protein composed of the N-terminal globular domain from *S. pyogenes* and the collagen-like domain from *C. perfringens*. (B) The solubility analysis of N_{cp}-CL_{cp} and V_{sp}-CL_{cp} proteins on SDS-PAGE. Lane 1, N_{cp}-CL_{cp} total cell lysate; Lane 2, N_{cp}-CL_{cp} insoluble fraction; Lane 3, N_{cp}-CL_{cp} soluble fraction; Lane 4, V_{sp}-CL_{cp} total cell lysate; Lane 5, V_{sp}-CL_{cp} insoluble fraction; and Lane 6, V_{sp}-CL_{cp} soluble fraction. (C) Melting and refolding curves of V_{sp}-CL_{cp} as monitored by MRE_{220nm}. (D) MRE_{220nm} monitoring of melting and refolding curves of the CL_{cp} domain obtained by trypsin digestion of V_{sp}-CL_{cp}.

And the conservation of the two predicted coiled coil regions in the N-terminal noncollagenous V domains of Scl proteins of different *S. pyogenes* serotypes, despite differences in primary structure,^{6,37} suggests the V domains may be members of the group of triple-helix folding domains which requires coiled coil regions. Here, the N-terminal globular V_{sp} domain of the Scl2 collagen-like protein is shown to be independently capable of trimerization. The characteristic α -helix CD spectrum observed for V_{sp} is consistent with the prediction of two coiled coil regions within its sequence, and the absence of Cys residues indicates no disulfide bonds are involved in its self-association. It is interesting to note that V_{sp} denatures at a higher temperature (45°C) when it is alone than when it is conjugated to the collagen-like domain (36.8°C) and that the melting temperature of V_{sp} is higher than that of the collagen domain CL_{sp} alone (35.2°C). This indicates that in the V_{sp}-CL_{sp} protein, the V_{sp} domain denatures coopera-

tively with the triple-helix domain. The observation that the stability of the less stable triple-helix domain determines the thermal transition of V_{sp}-CL_{sp} may be related to the relatively small ~7–8°C increased stability of V_{sp} domain relative to the CL_{sp} domain, in contrast to the much larger 60°C difference between the thermal stabilities of CTD and the BclA triple-helix.

The requirement for the noncollagenous domains in successful folding of Scl2 protein differs when synthesized *de novo* in *E. coli* compared with *in vitro* refolding of the heat denatured protein. For instance, the *S. pyogenes* (Gly-Xaa-Yaa)₇₉ collagenous domain from Scl2 protein is purified from an *E. coli* expression system in a stable trypsin resistant triple-helical form, but after heat denaturation, the protein can not refold *in vitro* to a native triple-helix. The collagen domain alone does not have the information to properly refold the triple-helix, suggesting that the initial folding of the (Gly-Xaa-Yaa)₇₉

sequence is facilitated by some factor in *E. coli*, which could include chaperones, membrane binding, ribosomal disposition, or solvent environment. These observations are reminiscent of the stable triple-helical molecule formed when the recombinant (Gly-Xaa-Yaa)₃₃₈ domain of human Type I collagen without the C-propeptide is expressed in a yeast *pichia* system,³⁸ even though it is well known that the C-propeptide initiates trimerization, cross-linking and folding *in vivo* and *in vitro*³⁹ and that the purified triple-helix domain alone after heat denaturation can only form segments of misaligned triple-helices.⁴⁰

The location of the noncollagenous folding domain with respect to the triple-helix domain is shown here to influence *in vitro* triple-helix refolding. Fibrillar collagens, basement membrane Type IV collagen, and collectins (e.g., lung surfactant protein D, mannose binding lectin) have folding domains C-terminal to the triple-helix domain, whereas transmembrane collagens and the macrophage scavenger receptor have an N-terminal folding domain. The ability of the peptide (Pro-Pro-Gly)₁₀ triple-helix to be nucleated at either end was shown by situating a short highly stable trimerization sequence (foldon) at either the N- or C-terminus.⁴¹ In this study, moving the V_{sp} domain to the C-terminal of the triple-helix led to molecules with similar conformation and stability as the original protein V_{sp}-CL_{sp}, but the molecules were incapable of efficient *in vitro* refolding after heat denaturation. This suggests that V_{sp} can only effectively play its role in trimerization and nucleation of triple-helix folding *in vitro* when it is N-terminal to the triple-helix. Experiments suggest that the V_{sp} domain has retained its trimerization ability when C-terminal to the triple-helix, but it is possible that at this location, it cannot promote initiation of the triple-helix or the proper registration of the three chains. This specificity could be due to the relationship of the V_{sp} domain with the ends of the triple-helix, for example, conformational constraints in the sequence linking the two domains or electrostatic repulsion between charged residues at the N-terminus of the V_{sp} domain and the highly charged C-terminus of the CL_{sp} domain. In bacterial collagens characterized thus far, the BclA protein is trimerized by a C-terminal domain CTD, while the *S. pyogenes* collagen-like proteins are trimerized by the N-terminal V domains, and recent characterization of collagen-like proteins from other bacteria show the locations of the noncollagenous folding domain can vary.²⁷ Thus in bacterial collagens, as well as in animal collagens and collagen containing proteins, trimerization domain is likely to be optimized with respect to location as well as sequence fit to get proper initiation of triple-helix and proper registration. The cell-wall binding domain C-terminal to the Scl2 triple-helix

may also be important in native folding and could be considered in future designs.

The successful folding and *in vitro* refolding of a chimeric molecule containing the *S. pyogenes* Scl2 V_{sp} domain and the *C. perfringens* triple-helix CL_{cp} domain support a modular nature, where the domain containing information for trimerization and registration is independent of an adjacent triple-helix domain. The inability of the *C. perfringens* triple-helix domain to fold in *E. coli*, either with or without its adjacent original noncollagenous domains, indicates not all (Gly-Xaa-Yaa)_n containing proteins can fold in *E. coli*. The ability of the V_{sp} domain to “rescue” this misfolded protein shows a functional trimerization module from one collagen-like protein can function as a folding domain for a different collagen-like protein. A modular model is also supported by the replacement of C-propeptides by the foldon in Type I and Type III collagen assembly,⁴² the ability of the trimeric neck domain of surfactant protein D to fold Type II procollagen,⁴³ and the ability of the chimeric protein with a Type III C-propeptide to assemble $\alpha 2(I)$ homotrimers.⁴⁴ The modular and interchangeable nature of the trimerization unit and triple-helix domain from bacterial collagens shown here is promising for practical applications, such as large scale production of these collagen-like proteins as biomaterials.

Material and Methods

Protein expression

All the proteins were expressed using the cold-shock vector system.²⁶ Except for V_{sp}-CL_{sp} and V_{sp}-CL_{cp}, all the plasmids used in this manuscript were constructed by inserting the PCR-generated DNA fragments into pColdII vector, which contains an N-terminal His₆ tag sequence. All the Scl2 protein derivatives were from *S. pyogenes* serotype M28. As described previously, V_{sp}-CL_{sp} was recloned from p163 plasmid¹⁸ into pColdIII vector which contains no His₆ tag, so the His₆ tag was introduced at the N-terminal end by PCR.⁸ V_{sp}, CL_{sp} and CL_{sp}-V_{sp} sequences were generated by PCR using V_{sp}-CL_{sp} as the template. The collagen-like sequences from *C. perfringens* were amplified by PCR using genomic DNA from strain SM101 as the template. V_{sp}-CL_{cp} sequence was constructed by inserting PCR-generated CL_{cp} fragment into pColdIII vector already containing His-tagged V_{sp} domain. The amino acid sequences of all the constructs are shown in the Supporting Information Figure 2. All the constructs were confirmed by DNA sequencing and then transformed into *E. coli* BL21 strain. Cells were cultured in 1 mL M9-casamino acid medium with 50 μ g/mL ampicillin at 37°C for 5 h. A total of 1 mM IPTG was added to induce protein expression, and the cells were then cultured at room temperature. After

overnight culture, cells from 100 μ L of the culture were harvested by centrifugation in a MICRO 240A centrifuge (Denville Scientific, Metuchen, NJ) at speed 14,000 rpm for 2 min. The cell pellets were then resuspended in 20 μ L H₂O and analyzed by SDS-PAGE.

Solubility analysis

Cells were cultured as described above. After overnight culture, cells from 1 mL cultures were collected by centrifugation at 4°C and resuspended in 100 μ L lysis buffer (20 mM NaPO₄ and 500 mM NaCl, pH 7.4). Then, the cells were sonicated using a Heat Systems sonicator (W-220F, Heat System Ultrasonics Inc.) for 4 cycles (80 Hz for 30 s and on ice for 30 s). And then the cell lysate was centrifuged at speed 14,000 rpm for 10 min at 4°C. The supernatant was collected as the soluble fraction. The pellet was resuspended in 100 μ L lysis buffer and this was the insoluble fraction. Both the soluble and insoluble fractions were analyzed by SDS-PAGE.

Protein purification

For V_{sp}, CL_{sp}, V_{sp}-CL_{sp}, CL_{sp}-V_{sp} and V_{sp}-CL_{cp}, a single colony of *E. coli* BL21 cells containing the corresponding plasmid was inoculated into 5 mL M9-casamino acid medium with 50 μ g/mL ampicillin. The culture was shaken at 37°C for 16 h and transferred into 250 mL M9-casamino acid medium at 37°C. When OD₆₀₀ reached 0.8, IPTG was added to a final concentration of 1 mM to induce protein expression, and the cells were cultured at room temperature. After overnight culture, the cells were collected by centrifugation in a Sorvall RC-5B refrigerated superspeed centrifuge (Du Pont Instruments) at speed 3000 rpm for 20 min at 4°C and then resuspended in 25 mL lysis buffer, followed by cell disruption by a French press. And then the cell lysate was centrifuged at speed 11,500 rpm for 20 min at 4°C. The supernatant was centrifuged in an Optima™ L-90 Ultracentrifuge (Beckman Coulter, Fullerton, CA) at speed 45,000 rpm for 1 h at 4°C. Then, the supernatant was applied to Ni-NTA agarose column (25 mL) at room temperature. After washing the column with wash buffer (20 mM NaPO₄, 500 mM NaCl and 20 mM imidazole, pH 7.4), elution buffers with stepwisely increasing concentrations of imidazole (50 mM, 100 mM, 125 mM and 400 mM) were used to elute proteins. Protein purity was analyzed by SDS-PAGE and the concentration was determined by absorbance at 280 nm.

To obtain CL_{cp}, purified V_{sp}-CL_{cp} was dialyzed against 50 mM glycine-NaOH buffer (pH 8.6) and treated with 1:1000 (w:w) trypsin at room temperature for 4 h. Then, the digested protein was applied to a DEAE-Sephadex column and eluted with a NaCl gradient from 0 to 400 mM. CL_{cp} was eluted

between 150 and 250 mM NaCl, which was then applied to a Superdex™ 200 gel filtration column (GE Healthcare). After elution, CL_{cp} purity was analyzed by SDS-PAGE.

Cross-linking

For V_{sp}, CL_{sp}, V_{sp}-CL_{sp}, CL_{sp}-V_{sp}, V_{sp}-CL_{cp} and CL_{cp} in PBS buffer (20 mM NaPO₄ and 500 mM NaCl, pH 7.0), both the native proteins and the proteins after melting and refolding were incubated with 0.8% (v/v) glutaraldehyde for 2 min at 23°C. The cross-linking was stopped by adding 2 $\times$ loading buffer containing 0.1% SDS and 4% β -mercaptoethanol. The samples were boiled in a boiling water bath for 3 min. Then, the samples were analyzed by SDS-PAGE.

Gel filtration

Gel filtration standard (BIO-RAD) was applied to the Superdex™ 200 gel filtration column and eluted by PBS buffer. The standard components include thyroglobulin (MW = 670,000), γ -globulin (MW = 158,000), ovalbumin (MW = 44,000), myoglobin (MW = 17,000) and vitamin B₁₂ (MW = 1,350). The elution time of each component was plotted against the logarithm of its molecular weight, and a standard curve was generated. Purified V_{sp} was applied to the same column. Based on the standard curve, the molecular weight of V_{sp} was calculated according to the elution time.

Dynamic light scattering

DLS experiment was performed using a DynaPro Titan instrument (Wyatt Technology Corp., Santa Barbara, CA). V_{sp} sample of 1 mg/mL in PBS buffer was filtered through a 0.1 μ m Whatman Anotop filter then put in the quartz cuvette. The experiment was carried out at 25°C with the usage of a temperature controller. Dynamic software by Wyatt Technology Corp. was used to obtain the hydrodynamic radius (R_h) through the intensity autocorrelation function.

CD spectroscopy

CD data were obtained using an AVIV Model 62DS spectropolarimeter (Aviv Associates, Lakewood, NJ). Before CD scanning, proteins were kept in PBS buffer in 1 mm cuvettes at 4°C for at least 24 h. CD spectra were collected from 190 nm to 260 nm at 0.5 nm interval with an averaging time of 5 s and each scanning was repeated three times. Protein melting was observed at 220 nm with increasing temperature from 0 to 70°C in 0.3°C steps. Proteins were equilibrated at each temperature point for 2 min, and the temperature was increased with an average rate of 0.1°C/min. T_m was defined as the temperature at which the fraction folded was 50% in the curve fitted to thermal transition. Protein refolding

was observed at 220 nm with decreasing temperature from 70°C to 0°C with the same average rate as protein melting. The percentage of refolding was defined as the ratio of the CD signal resumed at 0°C after refolding to the initial CD signal before melting.

Differential scanning calorimetry

DSC data were obtained using a NANO DSC II Model 6100 (Calorimetry Sciences Corp., Provo, UT). Before DSC scanning, V_{sp} protein solution was dialyzed against PBS buffer and equilibrated at 4°C for 24 h. The dialysis buffer was used for the baseline scan. Protein scanning was carried out from 0 to 80°C with an increasing rate of 1°C/min and reversely. The enthalpy was calculated from the first scan.

Acknowledgments

The authors thank Dr. Slawomir Lukomski for supplying the original p163 plasmid. They also thank Dr. Takeshi Yoshida and Dr. Angela Mohs for the early design of V_{sp} -CL $_{sp}$ construct. The authors acknowledge Eileen Hwang for helpful discussions.

References

- Bella J, Eaton M, Brodsky B, Berman HM (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 266:75–81.
- Ramachandran GN, Chandrasekharan R (1968) Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers* 6:1649–1658.
- Rich A, Crick FH (1961) The molecular structure of collagen. *J Mol Biol* 3:483–506.
- Rasmussen M, Jacobsson M, Bjorck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* 278:32313–32316.
- Myllyharju J (2009) Recombinant collagen trimers from insect cells and yeast. *Methods Mol Biol* 522: 51–62.
- Xu Y, Keene DR, Bujnicki JM, Hook M, Lukomski S (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* 277:27312–27318.
- Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, Brodsky B (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J Biol Chem* 282:29757–29765.
- Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JA, Inouye M, Brodsky B (2009) Self-association of *streptococcus pyogenes* collagen-like constructs into higher order structures. *Protein Sci* 18:1241–1251.
- Boydston JA, Chen P, Steichen CT, Turnbough CL, Jr (2005) Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. *J Bacteriol* 187:5310–5317.
- Ortega N, Werb Z (2002) New functional roles for non-collagenous domains of basement membrane collagens. *J Cell Sci* 115:4201–4214.
- Engel J (1996) Domain organizations of modular extracellular matrix proteins and their evolution. *Matrix Biol* 15:295–299.
- Khoshnoodi J, Cartailier JP, Alvares K, Veis A, Hudson BG (2006) Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. *J Biol Chem* 281:38117–38121.
- O'reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR, Folkman J (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* 88:277–285.
- Rety S, Salamiou S, Garcia-Verdugo I, Hulmes DJ, Le Hegarat F, Chaby R, Lewit-Bentley A (2005) The crystal structure of the *Bacillus anthracis* spore surface protein BclA shows remarkable similarity to mammalian proteins. *J Biol Chem* 280:43073–43078.
- Tan L, Turnbough CL, Jr (2010) Sequence motifs and proteolytic cleavage of the collagen-like glycoprotein BclA required for its attachment to the exosporium of *Bacillus anthracis*. *J Bacteriol* 192(5):1259–68.
- Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Ireland RM, Reid SD, Adams GG, Musser JM (2000) Identification and characterization of the scl gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. *Infect Immun* 68:6542–6553.
- Rasmussen M, Eden A, Bjorck L (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* 68:6370–6377.
- Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Shelvin BJ, Graviss EA, Musser JM (2001) Identification and characterization of a second extracellular collagen-like protein made by group A *Streptococcus*: control of production at the level of translation. *Infect Immun* 69: 1729–1738.
- Rasmussen M, Bjorck L (2001) Unique regulation of SclB—a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol Microbiol* 40:1427–1438.
- Whatmore AM (2001) *Streptococcus pyogenes* sclB encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* 147: 419–429.
- Humtsoe JO, Kim JK, Xu Y, Keene DR, Hook M, Lukomski S, Wary KK (2005) A streptococcal collagen-like protein interacts with the $\alpha 2\beta 1$ integrin and induces intracellular signaling. *J Biol Chem* 280: 13848–13857.
- Caswell CC, Lukomska E, Seo NS, Hook M, Lukomski S (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the $\alpha 2\beta 1$ integrin enhances pathogen survival and re-emergence. *Mol Microbiol* 64:1319–1331.
- Caswell CC, Oliver-Kozup H, Han R, Lukomska E, Lukomski S (in press) Scl1, the multifunctional adhesin of group A *Streptococcus*, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol Lett*.
- Caswell CC, Barczyk M, Keene DR, Lukomska E, Gullberg DE, Lukomski S (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins $\alpha 2\beta 1$ and $\alpha 11\beta 1$. *J Biol Chem* 283: 36168–36175.
- Pahlman LI, Marx PF, Morgelin M, Lukomski S, Meijers JC, Herwald H (2007) Thrombin-activatable fibrinolysis inhibitor binds to *Streptococcus pyogenes* by interacting with collagen-like proteins A and B. *J Biol Chem* 282:24873–24881.
- Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T, Montelione GT, Ikura M, Inouye M (2004) Cold-shock

- induced high-yield protein production in *Escherichia coli*. *Nat Biotechnol* 22:877–882.
27. Xu C, Yu Z, Inouye M, Brodsky B, Mirochnitchenko O (2010) Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. *Biomacromolecules* 11(2): 348–56.
 28. Bachinger HP, Bruckner P, Timpl R, Prockop DJ, Engel J (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. *Eur J Biochem* 106:619–632.
 29. Hoppe HJ, Barlow PN, Reid KB (1994) A parallel three stranded alpha-helical bundle at the nucleation site of collagen triple-helix formation. *FEBS Lett* 344: 191–195.
 30. Latvanlehto A, Snellman A, Tu H, Pihlajaniemi T (2003) Type XIII collagen and some other transmembrane collagens contain two separate coiled-coil motifs, which may function as independent oligomerization domains. *J Biol Chem* 278:37590–37599.
 31. Mcalinden A, Smith TA, Sandell LJ, Ficheux D, Parry DA, Hulmes DJ (2003) Alpha-helical coiled-coil oligomerization domains are almost ubiquitous in the collagen superfamily. *J Biol Chem* 278:42200–42207.
 32. Sundaramoorthy M, Meiyappan M, Todd P, Hudson BG (2002) Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. *J Biol Chem* 277:31142–31153.
 33. Than ME, Henrich S, Huber R, Ries A, Mann K, Kuhn K, Timpl R, Bourenkov GP, Bartunik HD, Bode W (2002) The 1.9-Å crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of covalent Met-Lys cross-link. *Proc Natl Acad Sci USA* 99:6607–6612.
 34. Bogin O, Kvensakul M, Rom E, Singer J, Yayon A, Hohenester E (2002) Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* 10:165–173.
 35. Kvensakul M, Bogin O, Hohenester E, Yayon A (2003) Crystal structure of the collagen alpha1(VIII) NC1 trimer. *Matrix Biol* 22:145–152.
 36. Boudko SP, Sasaki T, Engel J, Lerch TF, Nix J, Chapman MS, Bachinger HP (2009) Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization fold. *J Mol Biol* 392:787–802.
 37. Han R, Caswell CC, Lukomska E, Keene DR, Pawlowski M, Bujnicki JM, Kim JK, Lukomski S (2006) Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of *Streptococcus pyogenes*. *Mol Microbiol* 61:351–367.
 38. Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA, Toman PD (2001) Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J Biol Chem* 276: 24038–24043.
 39. Engel J, Prockop DJ (1991) The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Biophys Chem* 20:137–152.
 40. Beier G, Engel J (1966) The renaturation of soluble collagen. Products formed at different temperatures. *Biochemistry* 5:2744–2755.
 41. Frank S, Boudko S, Mizuno K, Schulthess T, Engel J, Bachinger HP (2003) Collagen triple helix formation can be nucleated at either end. *J Biol Chem* 278: 7747–7750.
 42. Pakkanen O, Hamalainen ER, Kivirikko KI, Myllyharju J (2003) Assembly of stable human type I and III collagen molecules from hydroxylated recombinant chains in the yeast *Pichia pastoris*. Effect of an engineered C-terminal oligomerization domain foldon. *J Biol Chem* 278:32478–32483.
 43. Mcalinden A, Crouch EC, Bann JG, Zhang P, Sandell LJ (2002) Trimerization of the amino propeptide of type IIA procollagen using a 14-amino acid sequence derived from the coiled-coil neck domain of surfactant protein D. *J Biol Chem* 277:41274–41281.
 44. Lees JF, Tasab M, Bulleid NJ (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *EMBO J* 16: 908–916.