Dissecting a bacterial collagen domain from *Streptococcus pyogenes*:
Sequence and length dependent variations in triple-helix stability and folding

Zhuoxin Yu¹, Barbara Brodsky²,³ and Masayori Inouye¹,*

¹Department of Biochemistry, and Center for Advanced Biotechnology and Medicine, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854; ²Department of Biochemistry, Robert Wood Johnson Medical School; ³Department of Biomedical Engineering, University of Tufts

Running title: Triple-helix subdomains of a bacterial collagen

*To correspondence should be addressed: inouye@umdnj.edu, Barbara.Brodsky@tufts.edu

To better investigate the relationship between sequence, stability and folding, the *S. pyogenes* collagenous domain CL (Gly-Xaa-Yaa)ₙ was divided to create three recombinant triple-helix subdomains A, B, and C of almost equal size with distinctive amino acid features: an A domain high in polar residues, a B domain containing the highest concentration of Pro residues, and a very highly charged C domain. Each segment was expressed as a monomer, a linear dimer and a linear trimer fused with the trimerization domain (V domain) in *Escherichia coli*. All recombinant proteins studied formed stable triple-helical structures, but the stability varied depending on the amino acid sequence in the A, B and C segments and increased as the triple-helix got longer. When heat-denatured, the V domain enhanced refolding for all the constructs; however the folding rate was affected by their amino acid sequences and became reduced for longer constructs. V-AAA was found to melt at a much lower temperature (31.0°C) than V-A-B-C (V-CL), while V-BBB melted at the almost same temperature (around 36-37°C). The folding rates of all the other constructs were lower than that of the natural V-A-B-C protein. Amino acid substitution mutations at all Pro residues in the C fragment dramatically decreased stability but increased the folding rate. These results indicate that the thermostability of the bacterial collagen is dominated by the most stable domain in the same manner as found with eukaryotic collagens.

The amino acids occupying the Xaa and Yaa positions of the (Gly-Xaa-Yaa)ₙ sequence are biologically important, forming binding and degradation sites, as well as modulating global and local stability of the triple-helix. Host-guest model peptides have been used to determine the triple-helix propensity of individual residues and of pairs of residues (3). Here, the relation between amino acid sequence and triple-helix stability and folding is investigated through the use of a recombinant bacterial collagen system, where it is possible to manipulate triple-helix sequence and length and to make large amounts of protein for biophysical studies.

Collagen is an important structural protein in the extracellular matrix of animals, and collagen-like sequences have now been identified in prokaryotic genomes (5). Recombinant bacterial collagen-like proteins have been shown to adopt a triple-helix structure, with a thermal stability similar to that seen for human collagens (6-10). Bacteria lack the prolyl hydroxylase enzyme necessary for post-translational modification of Pro to Hyp, and in the absence of Hyp, alternative strategies to stabilize the triple-helix must be present, involving electrostatic interactions, polar residues, or high Pro content (5,11). One well characterized prokaryotic collagen is the *Streptococcus pyogenes* collagen-like protein 2 (Scl2), a cell surface protein which may help the pathogenic bacteria adhere to animal cells (12,13). The Scl2 protein contains an N-terminal globular V domain, a collagen-like CL domain and a transmembrane domain. The Scl2.28 (Scl2 protein from serotype M28) derived protein, containing only the V and CL domains, was successfully purified from a high yield expression system in *E.coli* (8,9). The *S. pyogenes* V-CL protein forms a stable triple-helix with *T_m* = ~36°C, and the collagen domain alone has the same thermal stability (8,9). The N-terminal globular V domain is an α-helix containing protein that forms trimers (14), and the V domain was shown to be essential for triple-helix refolding (7,9,14).
The S. pyogenes collagenous domain CL consists of 79 Gly-Xaa-Yaa tripeptides. The distribution of charged residues, Pro and polar residues is not even throughout the CL domain. The CL domain can be divided into three subdomains (A, B and C) of almost equal size with distinctive amino acid features. We hypothesize that each of these unique domains may play a critical role in the folding of the CL domain and its overall stability. The C-terminal third of CL contains the highest concentration of charges, (31 charges/78 residues, ~38.3 %), while the N-terminal region has the least charge, and the central region has the most Pro (Supplemental Table 1). These sub-domains were individually expressed in E. coli fused with the N-terminal V domain to promote trimerization and folding. In addition, each sub-domain was tandemly fused to construct a homo-dimer (A-A, B-B or C-C) or a homo-trimer (A-A-A, B-B-B or C-C-C) again fused with the N-terminal V domain. All recombinant proteins studied formed stable triple-helical structures; however, their stabilities were quite different dependent upon their amino acid compositions and lengths. The V domain promoted refolding in all segments, but the folding rate depended on sequence and length and was always less than the natural V-CL protein.

**Experimental procedures**

**Construction and expression of recombinant collagen proteins**-

All the constructs were obtained by inserting the PCR products into pCold II vector (15) which has an N-terminal His tag, using the Ndel and BamHI restriction sites. The DNA fragment of V-CΔP was synthesized by Genscript Inc. (Piscataway, NJ) and cloned into pCold II vector using Ndel and BamHI sites.

All the constructs were confirmed by DNA sequencing and then transformed into E. coli BL21 strain. Cells were shaken in 1 ml M9-Casamino Acid medium with 50 µg/ml ampicillin at 37°C for 5 hr. 1 mM IPTG was added to induce protein expression and the cells were then shaken at room temperature. After overnight culture, 1 ml cultures were centrifuged at 4°C and the cell pellets thus obtained were resuspended in 100 µl lysis buffer. Both the soluble and insoluble fractions were analyzed by SDS-PAGE (data not shown). If the expressed protein was present mostly in the soluble fractions, further purification of the proteins was carried out as described below. Since V-CC and V-CCC were present only in the insoluble fractions, they were not further purified.

**Purification of His-tagged proteins**-

A single colony of E. coli BL21 cells containing the corresponding plasmid was inoculated into 5 ml M9-Casamino Acid medium with 50 µg/ml ampicillin. The culture was shaken at 37°C for 16 hr and transferred into a 250 ml M9-Casamino Acid medium, which was then shaken at 37 °C until OD₆₀₀ reached 0.8. 1 mM IPTG was added to induce protein expression and the cultures were shaken at room temperature. After overnight, the cells were collected by centrifugation at 11,500 rpm at 4 °C and resuspended in 25 ml lysis buffer, followed by cell disruption using a French press. After ultracentrifugation at 45,000 rpm at 4°C to remove insoluble materials, the supernatant was applied to a Ni-NTA agarose column (25 ml) at room temperature. After washing the column with wash buffer (20 mM NaPO₄ + 500 mM NaCl + 20 mM imidazole, pH 7.4), elution buffers with increasing concentrations of imidazole (50 mM, 100 mM, 125 mM and 400 mM) were used for stepwise elution of proteins. The protein purity was examined by SDS-PAGE and the concentration was determined by absorbance at 280 nm with ε = 9129 M⁻¹.cm⁻¹.

**CD (Circular dichroism) spectroscopy**-

CD data were obtained using an AVIV Model 62DS spectropolarimeter (Aviv Associates Inc., Lakewood, NJ). Before CD scanning, proteins were kept in 1 mm cuvettes at 4°C for at least 24 hr. CD spectra were collected from 260 nm to 195 nm at 0.5 nm interval with an averaging time of 5 sec and each scanning was repeated three times. Protein melting was examined at 220 nm with increasing temperature from 0°C to 60°C in 0.33°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. Tₘ was obtained by taking the peak of the 1st derivative of the melting curve. The thermal transition was observed at both neutral pH (PBS buffer, 20 mM NaPO₄ + 150 mM NaCL) and at pH2.8 (0.1 M acetic acid).
After incubation of the samples at 65°C for 20 min to denature the sample, protein refolding was observed at 220 nm at 0°C for 60,000 sec. The half time was defined as the time point when the fraction folded reaches 0.5. The percentage of refolding was defined as the ratio of the CD signal at 0°C after refolding to the initial CD signal before melting.

**DSC (Differential Scanning Calorimetry)** - DSC data were obtained using a NANO DSC II Model 6100 (Calorimetry Sciences Corp., Provo, UT). Before DSC scanning, each protein was dialyzed against PBS buffer (pH 7.0) and equilibrated at 4°C for 24 hr. The dialysis buffer was used for the baseline scan. Protein scanning was carried out from 0°C to 80°C with an increasing rate of 1°C/min.

**Trypsin digestion** - 20 mM purified samples in PBS buffers were incubated with 0.01 mg/ml trypsin at 15°C for 10 min. The reaction was stopped by adding 1 mM PMSF and then the samples were run on an SDS-PAGE. For the trypsin digestion on the cell lysate, 0.005 mg/ml trypsin was added to the cell lysate, 0.005 mg/ml trypsin was added to the cell lysate and a time course was recorded, with time equals 0, 1, 3, 7 and 24 hr.

**Results**

**Sub-fragmentation of the CL triple-helix domain** - The collagenous domain (CL) of *S. pyogenes* Scl2 protein is highly charged (30% residues), with 12% Pro, 13% polar residues, 6% Ala, and 6% larger hydrophobic residues, but without Cys, Tyr and Phe residues. It is an acidic protein with the pI value of 5.19. As shown in Figure 1A, this CL domain can be dissected into three unique fragments of approximately equal lengths namely on the basis of their amino acid compositions; the N-terminal 78 residues (fragment A consisting of 27 tripeptides, the central 78 residues (fragment B, 26 tripeptides) and the C-terminal 81 residues (fragment C, 26 tripeptides). Fragment A has the highest content of polar residues, fragment B the highest Pro content, and fragment C the highest charged residue content (Fig. 1B; Supplemental Table 1). Each fragment contains more acidic than basic residues so that the pI value of each fragment is between 5.1 and 5.3, similar to that of the complete CL domain. Notably fragment C contains no Arg or Glu, but only Lys and Asp with a repetitive sequence of GKGQNGKGDGLPGKD. To study the length effect of the triple helices, constructs with 1, 2 and 3 tandem repeats of each fragment were designed (Fig. 1A). For all constructs the V trimerization domain was added at the N-terminal end.

All the proteins were expressed in *E.coli* at room temperature using a cold-shock vector, pCold II, as previously described (15). The V-CC and V-CCC proteins were expressed well but were not soluble. Therefore, they were not further studied. All the soluble proteins were purified on a Ni-agarose column and analyzed by SDS-PAGE gel electrophoresis (Supplemental Fig. 3).

**Conformation of collagen constructs** - Circular Dichroism (CD) spectra of the A, B and C constructs, recorded from 260 nm to 195 nm, show typical triple-helical spectra with a positive peak at 220 nm (MRE\(_{220}\) ~ +8000 deg.cm.dmol\(^{-1}\)) and a minimum at 198 nm (MRE\(_{198}\) = -60,000 deg.cm.dmol\(^{-1}\)) (Fig. 2A). The Rpn values (ratio of the value of positive peak to that of the negative maximum) of the A, B and C collagen fragments (0.12-0.13) are comparable to those seen for mammalian collagens and model peptides, suggesting that they form complete triple-helical molecules (16).

Previous studies showed that the V domain has a typical \(\alpha\)-helical CD spectrum, with \(T_m = 45^\circ C\) (14). The CD spectra of the A, B and C constructs with the V domain include opposing contributions from the \(\alpha\)-helix structure of the V domain (MRE\(_{222}\) = -20100 deg.cm.dmol\(^{-1}\); MRE\(_{208}\) = -21700 deg.cm.dmol\(^{-1}\); MRE\(_{197}\) = +24300 deg.cm.dmol\(^{-1}\)) and the triple-helix structure (MRE\(_{223}\) = +8000 deg.cm.dmol\(^{-1}\); MRE\(_{199}\) = -60,000 deg.cm.dmol\(^{-1}\)). Note that the net spectrum depends on the relative fraction of residues in each domain; there are 74 residues in V domain, while there are 78 residues in A and B and 81 residues in C. For the V-B spectrum, there is a net negative MRE\(_{220}\) = -5000 deg.cm.dmol\(^{-1}\) and MRE\(_{208}\) = -27000 deg.cm.dmol\(^{-1}\) (Fig. 2A). As the proportion of triple-helical structures relative to \(\alpha\)-helix structures increases in the longer tandem constructs, the peak at 220 nm gets higher. The MRE value of V-BBB at 220 nm is almost the same as that of the full-length V-CL molecule (9) (Fig. 2A).

The CD spectra of the molecules derived from fragment A (A, V-A, V-AA, and V-AAA) or fragment C (C, V-C) showed patterns similar to
those described above for fragment B (data not shown), suggesting that all three fragments are capable of forming a triple-helical structure connected to the α-helical V domain, and that the V and triple-helix domains contribute additively to their CD spectra.

**Thermal stability**

Thermal stabilities of all the constructs were determined by monitoring the intensity at the triple-helix maximum, MRE$_{220\text{nm}}$, with increasing temperature at a rate of 0.1°C/min. The thermal transition of fragment B occurs at $T_m = 29.5°C$, a value lower than the 36.0°C observed for the full-length CL domain, and the B transition is broader than seen for CL (9) (Fig. 2B). In addition, the CD signal of B has a sloping baseline before the major transition. For V-B, in which both V and B domains consist of similar residue numbers (74 and 78 residues, respectively), the observed melting profile shows a combination of decreasing MRE$_{220\text{nm}}$ values due to triple-helix unfolding and increasing MRE$_{222\text{nm}}$ values due to the unfolding of α-helix containing V domain (Fig. 2C). The first derivative of the V-B thermal transition shows a single peak at 31.5°C, a value 2°C higher than the $T_m$ of B alone and lower than the value observed for the V domain alone ($T_m = 45.0°C$). The $T_m$ values increase to 36.0°C and 37.6°C for V-BB and V-BBB, respectively, as the length of the collagen domain increases (Fig. 2D and E). It is noteworthy that the melting temperatures of V-BB and V-BBB reach the value reported for the full length molecule V-CL (9). The melting of the V domain becomes less visible as the residue proportion of the α-helix in the V domain to the triple-helix in the collagen domains decreases.

The major thermal transition of fragment C is ~25.3°C, followed by a second transition near 36.5°C (Supplemental Fig. 1E); the higher transition may be due to the aggregation since visible aggregates were observed after incubation for a few days. The melting temperature of V-C is 28.5°C, similar to that of V-A and lower than that of V-B (Supplemental Fig. 1F). The melting profiles for V-CC and V-CCC were not studied due to insolubility. The melting temperatures of V-A, V-AA, V-AAA, V-B, V-BB, V-BBB, V-C and full-length V-CL were plotted against the numbers of (Gly-Xaa-Yaa)$_n$ tri-peptides (Fig. 2F), and the values were shown in Table 1.

Thermal transitions of V-A, V-B and V-C were also examined by differential scanning calorimetry (DSC) (Supplemental Fig. 2). Only one major transition is seen for each sample, with $T_m = 29.4°C$, 32.4°C and 29.4°C, for V-A, V-B, and V-C, respectively. These values are slightly higher than the $T_m$ values obtained by CD likely due to the faster DSC heating rate and the non-equilibrium conditions (17). The DSC profile of V-A also shows a small peak at 12.5°C, consistent with the small transition observed on CD (Supplemental Fig. 1B), which may reflect a small amount of a misfolded, less stable species in the protein preparations.

Since native triple-helices are resistant to trypsin cleavage, the protein constructs were digested with trypsin to assess the integrity and stability of their triple-helices (Supplemental Fig. 3). All fragments have many potential trypsin sensitive sites (Supplemental Table 1). Trypsin digestion of both purified fragments and cell lysates suggests that fragment A is the most susceptible to trypsin (Supplemental Fig. 3A, E), although it has the least number of trypsin-sensitive sites (8 sites, Supplemental Table 1), suggesting this may form a looser triple-helical structure. Fragments B and C are more resistant to trypsin, although there are even more trypsin-sensitive sites in the sequence, indicating a tighter triple-helical structure. The proteins with 2 and 3 tandem repeating fragments (V-AA, V-AAA, V-BB, and V-BBB) were more resistant to trypsin than their corresponding smaller constructs, but all were still more susceptible to trypsin digestion than the natural V-CL protein (Supplemental Fig. 3B-D). The results indicate that the shorter constructs are more susceptible to trypsin, but intriguingly that the V-AAA and
V-BBB constructs, having the same length as V-CL, are still more trypsin susceptible than V-CL.

**Stability of proteins in acid pH**

The stability of all the constructs was examined by CD at low pH (0.1 M acetic acid, pH 2.8) to evaluate how electrostatic interactions play roles in stabilization of the highly charged triple-helix domains: V-A (21.8 % charged residues), V-B (28.2 %), and V-C (38.3 %). It was previously reported that the V domain is denatured in acid (14) and that the triple-helix CL domain shows a decreased stability with two thermal transitions (T_m at 24 and 27°C) compared with PBS (8). The melting temperatures of all of the A, B, C, and V-A, V-B and V-C proteins showed one thermal transition at low pH, which were lower than their respective values at neutral pH (Table 2). Fragment A having the least charges (21.8 %) showed a small decrease in T_m (ΔT_m = -0.6°C) than seen for B (28.2 % charged residues; ΔT_m = -10.6°C) and C (38.3 % charged residues; ΔT_m = -12.8°C). The V-A protein showed a smaller decrease, ΔT_m = -4.9°C in comparison with V-B (ΔT_m = -9.1°C) and V-C (ΔT_m = -14.1°C). The decrease in thermal stability appears to correlate with the percentage of charged residues in the A, B and C fragments, suggesting that at neutral pH, electrostatic interactions play a significant role in protein stability.

**Refolding of triple-helix**

Refolding of all the fragments was examined by monitoring CD spectra at MRF_{222 nm} after heating samples at 50°C for 20 min followed by cooling to 0°C. Fragments A, B and C did not refold, consistent with previous observations that the CL domain cannot refold in the absence of the V domain. When the V domain was fused to these fragments, both V-domain α-helices and the CL triple-helix structure influence the CD signal at 220 nm. Studies on the isolated V domain showed that it trimerizes and is able to fold very quickly within 1 to 2 min required for cooling the sample from 50°C to 0°C (E. Hwang, pers. comm.). We assume that the V domain may be completely refolded within the first 1-min cooling time, and that the refolding to triple-helical structure starts at 0°C. Since the kinetics shown in Fig. 3 did not fit 1st, 2nd or 3rd order reactions for any samples, the half time for folding (t_{1/2}) was used to compare the folding rate for each sample. Protein V-B has the fastest folding rate (t_{1/2} ~ 330 sec), followed by V-A (t_{1/2} ~ 640 sec) and then V-C (t_{1/2} ~ 770 sec) (Fig. 3A and C). After 16 hr, the CD signals for V-B, V-A, and V-C reach 95.6 %, 90.4 % and 92.6 % of their individual original signals, respectively, and after 3 days incubation at 0°C, the CD signals of all constructs containing the V domain reached their original signal values.

Increasing length of the triple-helix led to slower folding (Table 3). For V-B, V-BB and V-BBB, the half time increased as the length of the collagen domain increased (330 sec, 410 sec and 700 sec, respectively) (Fig. 3B and D). The final refolding efficiency after 16 hr also dropped (90.0 % for V-BB and 89.2 % for V-BBB) as the number of collagen repeats increased, which might be due to more misfolding of proteins with increasing triple-helix length. Comparison of refolding of V-BBB with that of V-CL shows that the tandem repeating construct (V-BBB) folded more slowly than V-CL, even if both of them have the same length of the collagen domain (t_{1/2} ~ 351 sec for V-CL versus 700 sec for V-BBB). Similarly, for the V-A, V-AA, and V-AAA constructs, the folding rates and the duration times for completion of folding decreased as the lengths increased. Although V-A folded significantly slower than V-B, the folding of V-AAA was only slightly slower than that of V-BBB.

**Removal of Pro residues in the V-C construct**

Since imino acids, Pro and Hyp are known to play an important role in stabilizing mammalian collagens, the effect of eliminating Pro residues in fragment C (containing 7 Pro residues) on the thermal stability was investigated. Fragment C has the least number of Pro residues compared with fragments A and B (Supplemental Table 1). Furthermore, all of them are at the Yaa positions in Gly-Xaa-Yaa triplets. All these Pro residues are mutated to either Gln or Asp, creating V-CΔP (Fig. 4A). In this mutagenic strategy, all GLP tripeptides were altered to GLQ, because GLQ is the second most abundant tripeptide after GLP in the Scl2 protein, and the GQPGKP sequence was mutated to GQDGKD, because D is the most frequently used amino acid at the Yaa position if the Xaa position is occupied by Q or K in the CL triple-helix. The host-guest peptide set in a peptide context predicts that substitution of a Pro residue with Gln in the Y position will lower the stability by ~1.6°C if Leu is in the X position (3). On the other hand, a Pro to Asp substitution may decrease the stability by ~3.9°C if Lys is in the X position and by ~11.5°C if Gln is in the X position (3). By CD analysis, the V-CΔP protein showed a melting
temperature at 13.9°C, which is 14.6°C lower than V-C (Fig. 4B). A second thermal transition with \( T_m \) of \( \sim 45°C \) is also observed, consistent with independent melting of the V domain in this construct.

DSC of V-C\( \Delta P \) (Fig. 4C) (1°C/min) shows three peaks at 15.7°C, 32.8°C and 45.0°C. The 15.7°C peak represents the thermal transition of the triple-helix domain, which is slightly higher than the transition temperature measured by CD as expected (17). The 45.0°C peak is consistent with the thermal transition of V domain, which is the same value as obtained by CD for the isolated V domain. The origin of the 32.8°C peak is not clear, but it is not seen on a second DSC scan, suggesting that it might be due to some aggregates which dissolve by heating and take time to reform. The 32.8°C peak may represent soluble aggregates which can associate further with time, as visible aggregates were seen after incubation for several days at 4°C.

Refolding kinetics of V-C\( \Delta P \) was also monitored using MRE\(_{220nm}\). Surprisingly, the Pro-free triple-helix protein shows a faster folding rate than that seen for V-C (Fig. 4D). The faster folding in the absence of Pro residues suggests that formation of the triple-helix structure in the C domain may have been limited by cis/trans isomerization of Pro residues.

**Discussion**

The rod-like structure of the collagen triple-helix, which precludes long range interactions, in theory enables the dissecting of the full-length molecule into fragments which maintain the original triple-helix conformation. The recombinant system using a bacterial collagen protein expressed in *E. coli* allows manipulation of the sequence. In this study, the CL domain derived from *Scl2.28* protein was dissected into three fragments of equal length and studies were carried out on each fragment with and without the trimerization V domain, or in multiple tandem repeats together with the V domain. The CD analyses of these constructs indicate that all the (Gly-X-Y)\(_n\) fragments maintain a triple-helix conformation regardless of their sequences and lengths. Attachment of a V domain to the N-terminal end of the triple-helix sequences does not affect either the \( \alpha \)-helix structure of the V-domain or the triple-helix conformation of the collagen domains. The maintenance of the original conformation allows comparisons to determine the effect of amino acid sequence and triple-helix length on the stability and folding.

**Longer Triple-helix increases stability**

Increasing triple-helix length led to increased stability for A and B fragments; however, the stability reached a plateau at about 50 triplets. Since the proteins studied here contain tandem repeats of A or B fragments, the effect of the lengths of triple-helix structures on the stability can be compared without any influence from amino acid sequence. In an earlier study, it had been shown that thermal stability reached a plateau as the length of several natural and engineered *S. pyogenes* bacterial collagen constructs reached \( \sim 80 \) triplets (18). Studies on collagen-like peptides (POG)\(_n\) and (PPG)\(_n\) also indicate the presence of a maximum for stability as the length increases to more than 20 triplets (3).

**Stability depends on the amino acid sequence**

Sequence-dependent differences in the thermal stability of the isolated triple-helix domains of the same length were observed. The most stable fragment is the central fragment B, with \( T_m = 29.5°C \), followed by fragment C (25.3°C) and A (23.5°C). As the length increases, V-BB and V-BBB reach a plateau at 36.0°C, while V-AA and V-AAA plateau at a lower temperature (31.0°C). Notably, the analysis of the amino acid sequence of fragment B reveals unique features likely contributing to the stability of the fragment: it contains more Pro residues (12) than A (10) or C (7), and also has two highly stabilizing sequences, KGD and KGE. The important contribution of imino acid residues to triple-helix stability (19) was confirmed by the observed decrease in stability from 25.3°C to 15.0°C when seven Pro residues in fragment C were mutated to non-imino acid residues. In the case of V-C\( \Delta P \), the significant drop of \( T_m \) was observed in the collagen domain, which shows a thermal transition widely separated from that of the most stable V domain.

**The effect of V domain on collagen domain stability**

The V domain was added at the N-terminal end of most of the constructs since this domain has been shown to be essential for triple-helix refolding (9,11,14). The results here indicate the V domain does not affect the conformation of the triple-helix, but does increase triple-helix stability by 2-5°C for the V-A, V-B and V-C proteins. The CD thermal transition profiles of
V-A, V-B and V-C show that the melting of the triple-helix occurs first and is immediately followed by the melting of the α-helix in the V-domain. The α-helix structure in the V domain unfolds at a significantly lower temperature in V-A, V-B, and V-C constructs than when in an isolated V domain (T_m = 45.0°C). Although unfolding of both triple-helix in the collagen domain (decreasing MRE_{220nm}) and α-helix in the V domain (increasing MRE_{220nm}) can be observed in the heat denaturation experiments, the 1st derivative of the CD signal or the DSC profile show only one transition. It appears that unfolding of the triple-helix structure triggers subsequent unfolding of the V domain, at a temperature below the T_m of the V domain by itself. Although electron microscopy of V-CL by rotary shadowing shows a small globular domain at the terminus of a rod-like structure (7), suggesting that the V domain and the collagen domain are independent forming modular structures, it appears that as the triple-helix starts to unfold, the unfolded chains may have a destabilizing effect on the V domain. In V-CΔP, the transition of the collagen domain occurs at 13.9°C and the V domain at 45.0°C, demonstrating that these two domains behave completely independently without apparent interactions. It is possible that interactions may only occur when the triple-helix domain stability is closer to that of the V domain.

Electrostatic interactions are important for stability at neutral pH-
Comparison of the thermal stability at acid and neutral pH shows that electrostatic interactions make a significant contribution to stability in these highly charged triple-helix structures. The stabilities at acid pH dropped most significantly for fragment C which has an extremely high content of charged residues (38.3 %) including many GKD and GKDGKD sequences, which have been shown to have highly stabilizing effects on the triple-helix structures (8). In animal collagens negatively charged residues are predominately at Xaa positions and positively charged residues at Yaa positions, however it is clear from these studies that positively charged residues in the Xaa position adjacent to negatively charged residues in the Yaa position also can have a highly stabilizing effect on collagen. All V-A, V-B and V-C constructs show only one transition at acid pH from 14°C to 23°C, while the V-CL protein at acid pH shows two thermal transitions, at 24°C and 27°C, suggesting there may be some different domain structures.

Triple-helix folding is affected by sequence and length-
The A, B and C domains alone are unable to refold after heat denaturation; however when the V domain was added N-terminal to each fragment, they gained the ability to refold and attain the expected triple-helix CD spectra. The requirement for the trimerization V domain to facilitate triple-helix formation is reminiscent of the folding of human collagens which frequently requires a C-terminal trimerization domain (20). Although the V domain is located originally adjacent to the N-terminal A fragment in the V-CL molecule, this V domain is also capable of promoting folding of these much shorter triple-helical segments and can do this when adjacent to B and C fragments as well as A. This is consistent with the previous report that the V domain is capable of promoting folding of the heterologous Clostridium perfringens triple-helix domain (11). The ability of the V domain to rapidly form trimers may bring the N-terminal ends of the (Gly-Xaa-Yaa)_n chains close enough together to entropically favour the triple-helix formation. It is likely that this initial triple-helix formation at the N-terminal end propagates to the C-terminal end.

Interestingly, the refolding rate observed for the V-B construct is faster than seen for the V-A protein containing the original globular-triple-helix junction. Examination of the sequences near the N-terminus of A shows a Pro-rich sequence which could act as a nucleation domain (E. Hwang, pers. comm.). However, the overall higher Pro content and stability of the B region could make this construct more favourable for rapid folding. Surprisingly, the refolding of V-CΔP was even faster than V-C, possibly because the elimination of Pro from the sequence resulted in the reduction of the time required for Pro cis-trans isomerization. Refolding studies also show that the folding rate decreases with increasing lengths of the collagen domain. This length effect could be due to the increased number of cis-trans isomerization events required for triple-helix formation as the chain gets longer or to an increased potential for misfolding.

Comparison with animal collagen-
Manipulation of animal collagen sequences has been reported to investigate the relation between
Triple-helix subdomains of a bacterial collagen sequence and stability (21,22), where the authors expressed repetitive constructs of D period segments of type II collagen, using a mammalian cell expression system. In the present work, we have done similar studies using a bacterial collagen domain, where the expression of bacterial collagens with different sequences and lengths is much easier than for mammalian collagens in mammalian cells. Unlike animal collagens with the more stable triple-helix domains at both ends where they may serve as clamps for folding and stability (23), bacterial collagen ScII has its most stable triple-helix region in the middle, with less stable regions at both N- and C-terminal ends.

Comparison of the repetitive V-AAA and V-BBB proteins with the original more diverse V-CL (V-A-B-C) protein of the same length reveals that V-BBB, but not V-AAA achieves the same melting temperature as V-CL at around 36-37°C. It appears that bacterial collagens follow the rule previously proposed from the study on animal collagens (22) that the thermostability of collagen triple helix depends on the most thermostable domain. Considering that almost all other bacterial collagens studied so far (10,11,18) have melting temperatures between 36°C and 38°C, it is reasonable to speculate that this narrow range of melting temperatures might have been evolutionarily set since bacterial collagens may be important for the interaction of these bacteria with animal hosts. Intriguingly, the fact that V-BBB refolds faster than V-AAA, but still slower than V-CL (V-A-B-C), suggests that V-ABC arrangement in the CL domain might have been evolutionarily selected for the most efficient folding. Further studies of the intrinsic properties of the individual fragments and their arrangement will provide important insights into the biological roles of bacterial collagens, and genetic engineering of bacterial collagens for biomaterials.

Acknowledgments
The authors acknowledge Eileen Hwang and Chunying Xu for helpful discussions. This work was supported by a grant, GM60048 from the National Institutes of National Health (to BB).

REFERENCES
6. Xu, C., Yu, Z., Inouye, M., Brodsky, B., and Mirochnitchenko, O. Biomacromolecules 11, 348-356
Figure legends

Figure 1. (A) Schematic diagram of the constructs including fragments of the original S. pyogenes Scl2 CL domain, showing the rod-like triple-helix domain, the globular V domain and a small diamond to represent the TEE sequence (Translation Enhancing Element, from pCold vector (15)) and His<sub>6</sub> tag at the N-terminus. Between V and collagen domains there is a linker with a sequence of LVPRGSP (9); (B) Amino acid sequences of the V domain and the individual A, B and C fragments, with Gly residues underlined to highlight the repeating (Gly-Xaa-Yaa)<sub>n</sub> nature in the triple-helix segments.

Figure 2. (A) Circular dichroism spectra of the V domain (dash line), the B fragment (solid line), V-B (square), V-BB (circle) and V-BBB (triangle); (B) CD melting transition of the B fragment (solid line) and the transition of the isolated V domain (dash line); (C) CD melting transition of V-B; (D) CD melting transition of V-BB; (E) CD melting transition of V-BBB (black square), compared with V-CL (circle); (F) A plot of the melting temperature T<sub>m</sub> of the VA, VAA, VAAA (black square), V-B, V-BB, V-BBB (circle), and V-C (triangle) as a function of the number of (Gly-Xaa-Yaa)<sub>n</sub> units, with the value for the full length V-CL (reversed triangle).

Figure 3. (A) Folding of V-A (black square), V-B (red circle) and V-C (green triangle), showing the effect of sequence; (B) Folding of B (blue reversed triangle), V-B (black square), V-BB (red circle), V-BBB (green triangle), showing the effect of length; an expanded scale is presented in panels (C) and (D), which show only the first 5,000 seconds.

Figure 4. (A) Amino acid sequence of the protein constructed from the C fragment, with all Pro residues mutated to other residues, designated the V-C<sub>ΔP</sub> protein. The TEE sequence, His<sub>6</sub> tag, V domain and the linker sequence are shown in italics. GSP from the linker sequence was removed, and the amino acid residues which replaced Pro were bold; (B) The thermal transition of the V-C<sub>ΔP</sub> (circle) compared with V-C (black square); (C) DSC of V-C<sub>ΔP</sub> (solid line) compared with V-C (dotted line); (D) Folding of V-C<sub>ΔP</sub> (circle) compared with V-C (black square).

Tables

<table>
<thead>
<tr>
<th>Fragment X</th>
<th>X</th>
<th>V-X</th>
<th>V-X-X</th>
<th>V-X-X-X</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23.5 °C</td>
<td>28.1 °C</td>
<td>31.0 °C</td>
<td>31.0 °C</td>
</tr>
<tr>
<td>B</td>
<td>29.5 °C</td>
<td>31.5 °C</td>
<td>36.0 °C</td>
<td>37.6 °C</td>
</tr>
<tr>
<td>C</td>
<td>25.3 °C</td>
<td>28.5 °C</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>: Thermal stabilities of V-CC and V-CCC were not determined since they were insoluble.
Table 2. Comparison of the thermal stabilities at neutral and acid pH, as determined by CD spectroscopy.

<table>
<thead>
<tr>
<th>Fragment</th>
<th>$T_m$ at pH 7.0</th>
<th>$T_m$ at pH 2.8</th>
<th>charged residues % in the collagenous domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23.5 °C</td>
<td>22.9 °C</td>
<td>21.8%</td>
</tr>
<tr>
<td>V-A</td>
<td>28.1 °C</td>
<td>23.2 °C</td>
<td>21.8%</td>
</tr>
<tr>
<td>B</td>
<td>29.5 °C</td>
<td>18.9 °C</td>
<td>28.2%</td>
</tr>
<tr>
<td>V-B</td>
<td>31.5 °C</td>
<td>22.4 °C</td>
<td>28.2%</td>
</tr>
<tr>
<td>V-CL</td>
<td>35.6 °C</td>
<td>24.2 °C</td>
<td>29.5%</td>
</tr>
<tr>
<td>C</td>
<td>25.3 °C</td>
<td>12.5 °C</td>
<td>38.3%</td>
</tr>
<tr>
<td>V-C</td>
<td>28.5 °C</td>
<td>14.4 °C</td>
<td>38.3%</td>
</tr>
</tbody>
</table>

*: For V-CL in acid pH, two thermal transitions were measured by DSC, 23.7°C and 27.0°C (8).

Table 3. Refolding of the bacterial collagen fragments showing the half times of folding and the final percentages of structures recovered as determined by CD spectroscopy.

<table>
<thead>
<tr>
<th>Fragment X</th>
<th>V-X</th>
<th>V-X-X</th>
<th>V-X-X-X</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>640s / 90.4%</td>
<td>670s / 85.0%</td>
<td>760s / 85.0%</td>
</tr>
<tr>
<td>B</td>
<td>330s / 95.6%</td>
<td>410s / 90.0%</td>
<td>700s / 89.2%</td>
</tr>
<tr>
<td>C</td>
<td>770s / 92.6%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*: Refolding of V-CC and V-CCC was not identified since they were insoluble.
Fig. 1
A.

V: ADEQEKKAKVRLTELIQELAQGLGIEKKNFPTLGEDLDHITYMTKLYLLQEREQAESENSWKRLLKGQDHALD

Fragment A (triplets 1-27):
GODGRNEQEGQPTGPAGPRGLOQLOQGERFGQPTGPAGPRGLOQERFGQPTGLAGKAGEAGAKGETGAPGQP

Fragment B (triplets 28-53):
GPRGEQGPGQGPGLGKDGEAGAQGPAGMPGAPGERGEKGPGETQAKGDRGETGVPGPRGGERGEAGPAGKDGERGPVGA

Fragment C (triplets 54-79):
GKDQNGQDGLPKKDGKQDGQNGKDLPGKDQKDQDGKDQDGKDQDGKDQDGKDQDGKDQDGKDQDGKDQDGKDPGKP
Fig. 3

A.

B.

C.

D.
Fig. 4

A.  

V-CAp:
MNHKVHHHHMADEQEEKAKVRTELIQELAQGLGGIEKKNFPTLGEDELDHTYMTKLLLYLQEREQAENS
WRKRLLKGIQDHALDLVPR
GKDQNGQDGQLQKDGDGQNGKDGLQKGDKDGQNGKGDLQKGKDGDGDQLQGDGKDQDGKD
LQGKDGDQDGKD

B.  

C.  

D.
Dissecting a bacterial collagen domain from *Streptococcus pyogenes*: Sequence and length dependent variations in triple-helix stability and folding

Zhuoxin Yu, Barbara Brodsky and Masayori Inouye

*J. Biol. Chem. published online March 28, 2011*

Access the most updated version of this article at doi: 10.1074/jbc.M110.217422

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/M110.217422.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2011/03/28/jbc.M110.217422.full.html#ref-list-1