Dissecting a Bacterial Collagen Domain from *Streptococcus pyogenes*

**SEQUENCE AND LENGTH-DEPENDENT VARIATIONS IN TRIPLE HELIX STABILITY AND FOLDING**

Received for publication, January 3, 2011, and in revised form, February 28, 2011. Published, JBC Papers in Press, March 28, 2011, DOI 10.1074/jbc.M110.217422

Zhuxin Yu†, Barbara Brodsky†*, and Masayori Inouye‡§

From the †Department of Biochemistry and §Center for Advanced Biotechnology and Medicine, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854 and ‡Department of Biomedical Engineering, Tufts University, Medford, Massachusetts 02155

To better investigate the relationship between sequence, stability, and folding, the *Streptococcus pyogenes* collagenous domain CL (Gly-Xaa-Yaa)79 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 helix subdomains A, B, and C 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. V-AAA was found to melt at a much lower temperature (31.0 °C) than V-ABC (V-CL), whereas V-BBB melted at almost the same temperature (~36–37 °C). When heat-denatured, the V domain enhanced refolding for all of the constructs; however, the folding rate was affected by their amino acid sequences and became reduced for longer constructs. The folding rates of all the other constructs were lower than that of the natural V-ABC 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 collagen triple helix is composed of three left-handed polyproline helices twisted into a right-handed supercoiled structure (1–4). The close packing of the three chains requires every third residue be a Gly, generating the typical (Gly-Xaa-Yaa)n repeating sequence pattern. Triple helix stability is also dependent on a high content of imino acids Pro and hydroxyproline (Hyp) and interchain backbone hydrogen bonds. 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, 6). One well characterized prokaryotic collagen is Scl2 (*Streptococcus pyogenes* collagen-like protein 2), a cell surface protein that may help 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 *Escherichia coli* (7–9). The *S. pyogenes* V-CL protein forms a stable triple helix with $T_m = \sim 36$ °C, and the collagen domain alone has the same thermal stability (7–9). 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%), whereas the N-terminal region has the least charge, and the central region...
has the most Pro (supplemental Table 1). These subdomains were expressed individually in *E. coli* fused with the N-terminal V domain to promote trimerization and folding. In addition, each subdomain was tandemly fused to construct a homodimer (AA, BB, or CC) or a homotrimer (AAA, BBB, or CCC) again appended to the N-terminal V domain. All recombinant proteins studied formed stable triple helical structures; however, their stabilities were quite different depending 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 of the constructs were obtained by inserting the PCR products into pCold II vector (15) which has an N-terminal His6 tag, using the NdeI and BamHI restriction sites. The PCR products were synthesized by Genscript, Inc. (Piscataway, NJ) and cloned into pCold II vector using NdeI and BamHI sites.

All of the constructs were confirmed by DNA sequencing and then transformed into an *E. coli* BL21 strain. Cells were shaken in 1 ml of M9-casamino acid medium with 50 μg/ml ampicillin at 37 °C for 5 h. 1 mM isopropyl-1-thio-β-d-galactopyranoside was added to induce protein expression, and the cells were then shaken at room temperature. After 16 h, 1-ml cultures were centrifuged at 4 °C, and the cell pellets thus obtained were resuspended in 100 μl of lysis buffer (20 mM NaPO4, 500 mM NaCl, pH 7.4), followed by four cycles of sonicating (30 s on and 30 s off). Subsequently, the lysates were centrifuged at 13,000 rpm at 4 °C to separate soluble and insoluble proteins. The pellet fractions were resuspended in 100 μl of 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. Because 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 of M9-casamino acid medium with 50 μg/ml ampicillin. The culture was shaken at 37 °C for 16 h and transferred into a 250-ml M9-casamino acid medium, which was then shaken at 37 °C until A590 reached 0.8. 1 mM isopropyl-1-thio-β-d-galactopyranoside was added to induce protein expression, and the cultures were shaken at room temperature. After 16 h, the cells were collected by centrifugation at 11,500 rpm at 4 °C and resuspended in 25 ml of 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 nickel-nitrilotriacetic acid-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, 100, 125, 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 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 h. CD spectra were collected from 260 nm to 195 nm at a 0.5-nm interval with an averaging time of 5 s, and each scan was repeated three times. Protein melting was examined at 220 nm with increasing temperatures, from 0 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 first derivative of the melting curve. The thermal transition was observed at both neutral pH (PBS buffer, 20 mM NaPO4, 150 mM NaCl) and at pH 2.8 (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 s. The half-time was defined as the time point when the fraction folded reached 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 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.

Trypsin Digestion—20 μM 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 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 h.

**RESULTS**

Subfragmentation of 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 a predicted pI value of 5.19. As shown in Fig. 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 81 residues (fragment A consisting of 27 tripeptides), the central 78 residues (fragment B, 26 tripeptides), and the C-terminal 78 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 and supplemental Table 1). Each fragment contains more acidic than basic residues so that the pH 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 GKDGQNGKDG-TPGKD. To study the length effect of the triple helices, constructs with one, two, and three tandem repeats of each frag-
ment were designed (Fig. 1A). For all constructs, 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 nickel-agarose column and analyzed by SDS-PAGE (supplemental Fig. 3).

Conformation of Collagen Constructs—CD spectra of the A, B, and C constructs, recorded from 260 to 195 nm, show typical triple helical spectra with a positive peak at 220 nm (MRE220 nm = 8000 deg cm⁻¹ dmol⁻¹; MRE198 nm = 60,000 deg cm⁻¹ dmol⁻¹) (Fig. 2A). The Rpn values (ratio of the value of positive peak to that of the negative peak) of the A, B, and C collagen fragments (0.12–0.13) are comparable with 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 α-helical CD spectrum, with Tm = 45 °C (14). The CD spectra of the A, B, and C constructs with the V domain include opposing contributions from the α-helix structure of the V domain (MRE220 nm = −20,100 deg cm⁻¹ dmol⁻¹; MRE208 nm = −21,700 deg cm⁻¹ dmol⁻¹; MRE197 nm = +24,300 deg cm⁻¹ dmol⁻¹) and the triple helix structure (MRE220 nm = +8000 deg cm⁻¹ dmol⁻¹; MRE198 nm = −60,000 deg cm⁻¹ dmol⁻¹). Note that the net spectrum depends on the relative fraction of residues in each domain; there are 74 residues in the V domain, whereas there are 81 residues in A and 78 residues in B and C. For the V-B spectrum, there is a net negative MRE220 nm = −5000 deg cm⁻¹ dmol⁻¹ and MRE198 = −27,000 deg cm⁻¹ dmol⁻¹ (Fig. 2A). As the proportion of triple helical structures relative to α-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 (Fig. 2A) (9). 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 of the constructs were determined by monitoring the intensity at the triple helix maximum, MRE220 nm, with increasing temperature at a rate of 0.1 °C/min. The thermal transition of fragment B occurs at Tm = 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 (Fig. 2B) (9). 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 MRE220 nm values due to triple helix unfolding and increasing MRE222 nm values due to the unfolding of the 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. 2, D 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 \( \alpha \)-helix in
the V domain to the triple helix in the collagen domains
decreases.

The stability of fragment A is \( T_m = 23.5 \, ^\circ \text{C} \), showing that it is
less stable than fragment B of the same length (\( T_m = 29.5 \, ^\circ \text{C} \)). The
\( T_m \) value for V-A (28.1 \(^\circ\text{C}\)) and increases further
for V-AA (\( T_m = 31.0 \, ^\circ \text{C} \)) and V-AAA (\( T_m = 31.0 \, ^\circ \text{C} \)) (supple-
mental Fig. 1, A-D). These tandem A constructs are less stable
than the V-BB and V-BBB constructs, respectively, and the \( T_m \)
values of V-AA and V-AAA never reach the V-CL \( T_m \) value.

The relative proportion of the \( \alpha \)-helix structure in the V
domain to the length of the triple helix in the collagen domains
again affects the shape of the transitions. In all cases, the
proteins with the B domains are more stable than those with the A
domains (Table 1).

The major thermal transition of fragment C is \( \sim 25.3 \, ^\circ \text{C} \),
followed by a second transition near 36.5 \(^\circ\text{C}\) (supplemental Fig.
1E); the higher transition may be due to the aggregation
because visible aggregates were observed after incubation for
a few days. The melting temperature of V-C is 28.5 \(^\circ\text{C}\), 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) \(_{n}\) triptides (Fig. 2F), and the values were shown in Table 1.

Thermal transitions of V-A, V-B and V-C also were examined
by DSC (supplemental Fig. 2). Only one major transition is
seen for each sample, with \( T_m = 29.4, 32.4, \) and 29.4 \(^\circ\text{C}\) 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 nonequilibrium conditions (17). The DSC
profile of V-A also shows a small peak at 12.5 \(^\circ\text{C}\), 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.

Because 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). Tryptic digestion of both purified frag-
ments and cell lysates suggests that fragment A is the most
susceptible to trypsin (supplemental Fig. 3, A and E), although it
has the least number of trypsin-sensitive sites (eight sites, sup-
plemental Table 1), suggesting this may form a looser triple
helical structure. Fragments B and C are more resistant to tryp-

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal stability of the constructs containing fragments of the bacterial CL domain, as determined by CD spectroscopy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Fragment A</td>
</tr>
<tr>
<td>Fragment B</td>
</tr>
<tr>
<td>Fragment C</td>
</tr>
</tbody>
</table>

\(^a\) A small transition at about 12.5 \(^\circ\text{C}\) was also observed, which may reflect a small amount of misfolded, less stable species.

\(^b\) Thermal stabilities of V-CC and V-CCC were not determined because they were insoluble.

\(^c\) For V-CL in acidic pH, two thermal transitions were measured by DSC, 23.7 \(^\circ\text{C}\) and 27.0 \(^\circ\text{C}\) (8).

\(^d\) E. Hwang, personal communication.
to the triple helical structure starts at 0 °C. Because the kinetics shown in Fig. 3 did not fit the first, second, or third 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 \( t_{1/2}/H = 330 \) s, followed by V-A \( t_{1/2}/H = 640 \) s, and then V-C \( t_{1/2}/H = 770 \) s (Fig. 3, A and C). After 16 h, 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 of 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, 410, and 700 s, respectively) (Fig. 3, B and D). The final refolding efficiency after 16 h 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 have the same length of collagen domain (\( t_{1/2}/H = 351 \) s for V-CL versus 700 s for V-BBB). Similarly, for the V-A, V-AA, and V-AAA constructs, the folding rates and the final refolding percentage 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 V-C Construct**—Because 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 seven Pro residues) on the thermal stability was investigated. Fragment C has the least number of Pro residues compared with fragments A and B.

### Table 3

<table>
<thead>
<tr>
<th>Construct</th>
<th>Refolding of Monomer</th>
<th>CD Value</th>
<th>Refolding of Dimer</th>
<th>CD Value</th>
<th>Refolding of Trimer</th>
<th>CD Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment A</td>
<td>640 s/90.4%</td>
<td>670 s/85.0%</td>
<td>760 s/85.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragment B</td>
<td>330 s/95.6%</td>
<td>410 s/90.0%</td>
<td>700 s/89.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragment C</td>
<td>770 s/92.6%</td>
<td>N/A*</td>
<td>N/A*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Refolding of V-CC and V-CCC was not identified because they were insoluble.

**FIGURE 3.** A, folding of V-A (black squares), V-B (red circles), and V-C (green triangles), showing the effect of sequence. B, folding of B (blue reversed triangles), V-B (black squares), V-BB (red circles), V-BBB (green triangles), showing the effect of length; an expanded scale is presented in C and D, which shows only the first 5000 s.

---

**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 |
|---|---|---|
| V-Monomer | V-Dimer | V-Trimer |
| Fragment A | 640 s/90.4% | 670 s/85.0% | 760 s/85.0% |
| Fragment B | 330 s/95.6% | 410 s/90.0% | 700 s/89.2% |
| Fragment C | 770 s/92.6% | N/A* | N/A* |

---
Furthermore, all of these are at the Yaa positions in Gly-Xaa-Yaa triplets. All of these Pro residues are mutated to either Gln or Asp, creating V-C/P (Fig. 4A). 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 Asp is the most frequently used amino acid at the Yaa position if the Xaa position is occupied by Gln or Lys 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 Yaa position will lower the stability by 1.6 °C if Leu is in the Xaa position (3). On the other hand, a Pro-to-Asp substitution may decrease the stability by 3.9 °C if Lys is in the Xaa position and by 11.5 °C if Gln is in the Xaa position (3). By CD analysis, the V-C/P 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 45 °C is also observed, consistent with independent melting of the V domain in this construct.

DSC of V-C/P (Fig. 4C) (1 °C/min) shows three peaks at 15.7, 32.8, 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 small 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 that can associate further with time, as visible aggregates were seen after incubation for several days at 4 °C.

Refolding kinetics of V-C/P was monitored using MRE220 nm. 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 that 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-Xaa-Yaa)_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 α-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 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 ~50 triplets. Because the proteins studied here contain tandem repeats of A or B (supplemental Table 1). Furthermore, all of these are at the Yaa positions in Gly-Xaa-Yaa triplets. All of these Pro residues are mutated to either Gln or Asp, creating V-C/P (Fig. 4A). 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 Asp is the most frequently used amino acid at the Yaa position if the Xaa position is occupied by Gln or Lys 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 Yaa position will lower the stability by 1.6 °C if Leu is in the Xaa position (3). On the other hand, a Pro-to-Asp substitution may decrease the stability by 3.9 °C if Lys is in the Xaa position and by 11.5 °C if Gln is in the Xaa position (3). By CD analysis, the V-C/P 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 ~45 °C is also observed, consistent with independent melting of the V domain in this construct.

DSC of V-C/P (Fig. 4C) (1 °C/min) shows three peaks at 15.7, 32.8, 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 small 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 that can associate further with time, as visible aggregates were seen after incubation for several days at 4 °C.

Refolding kinetics of V-C/P was monitored using MRE220 nm. 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.
fragments, the effect of the lengths of triple helix structures on stability can be compared without any influence from the 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 ~80 triplets (18). Studies on collagen-like peptides (Pro-Hyp-Gly)_n and (Pro-Pro-Gly)_n also indicate the presence of a maximum for stability as the length increases to >20 triplets (3).

**Stability Depends on 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, whereas V-AA and V-AAA plateau at a lower temperature (31.0 °C). 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 to 15.0 °C when seven Pro residues in fragment C were mutated to non-imino acid residues. In the case of V-CL, the significant drop of T_m was observed in the collagen domain, which shows a thermal transition widely separated from that of the more stable V domain.

**Effect of V Domain on Collagen Domain Stability—**The V domain was added at the N-terminal end of most of the constructs because this domain has been shown to be essential for triple helix refolding (6, 9, 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 followed immediately 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 the triple helix in the collagen domain (decreasing MRE_220 nm) and α-helix in the V domain (increasing MRE_220 nm) can be observed in the heat denaturation experiments, the first derivative of the CD signal or the DSC profile shows 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 unfolding chains may have a destabilizing effect on the V domain. In V-CL, 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.

**Triple Helix Subdomains of a Bacterial Collagen**

**Electrostatic Interactions Are Important for Stability at Neutral pH—**Comparison of thermal stability at acidic and neutral pH shows that electrostatic interactions make a significant contribution to stability in these highly charged triple helix structures. The stabilities at an acidic pH dropped most significantly for fragment C, which has an extremely high content of charged residues (38.3%), including many GKD and GKDGD sequences, which have been shown to have highly stabilizing effects on the triple helix structures (8). In animal collagens negatively charged residues are mainly 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 to 23 °C, whereas the V-CL protein at an acidic pH shows two thermal transitions, at 24 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 (6). The ability of the V domain to rapidly form trimers may bring the N-terminal ends of the (Gly-Xaa-Yaa), chains close enough together to entropically favor 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 junction between the globular domain and the triple helix domain. Examination of the sequences near the N terminus of A shows a Pro-rich sequence that could act as a nucleation domain. However, the overall higher Pro content and stability of the B region could make this construct more favorable 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 sequence and stability (21, 22) where the authors expressed repetitive constructs of 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 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 (11), bacterial collagen Scl2 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-ABC) protein of the same length reveals that V-BBB but not V-AAA achieves the same melting temperature as V-CL at ~36–37 °C. It appears that bacterial collagens follow the rule proposed previously 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 (6, 10, 18) have melting temperatures between 36 and 38 °C, it is reasonable to speculate that this narrow range of melting temperatures might have been evolutionarily set as 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 refolds more slowly than V-CL (V-ABC), suggests that the V-ABC arrangement in the CL domain might have been selected evolutionarily 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—We thank Eileen Hwang and Chunying Xu for helpful discussions.

REFERENCES
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

doi: 10.1074/jbc.M110.217422 originally 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 21 references, 10 of which can be accessed free at http://www.jbc.org/content/286/21/18960.full.html#ref-list-1