MECHANISM OF STABILIZATION OF A BACTERIAL COLLAGEN TRIPLE-HELIX IN THE ABSENCE OF HYDROXYPROLINE

Angela Mohs1, Teresita Silva1, Takeshi Yoshida1, Ravish Amin1, Sławomir Lukomski2, Masayori Inouye1 and Barbara Brodsky1

1 Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854
2 Department of Microbiology, Immunology and Cell Biology, West Virginia University School of Medicine, Morgantown, WV 26506

Keywords: collagen, hydroxyproline, bacteria, stabilization, electrostatic

The Streptococcus pyogenes cell surface protein Scl2 contains a globular N-terminal domain and a collagen-like domain, (Gly-Xaa-X’aa)\textsubscript{n}, which forms a triple-helix with a thermal stability close to that seen for mammalian collagens. Hydroxyproline (Hyp) is a major contributor to triple-helix stability in animal collagens, but is not present in bacteria, which lack prolyl hydroxylase. To explore the basis of bacterial collagen triple-helix stability in the absence of Hyp, biophysical studies were carried out on recombinant Scl2 protein, the isolated collagen-like domain from Scl2, and a set of peptides which model the Scl2 highly charged repetitive (Gly-Xaa-X’aa)\textsubscript{n} sequences. At pH 7, CD spectroscopy, dynamic light scattering and differential scanning calorimetry of the Scl2 protein all show a very sharp thermal transition near 36°C, indicating a highly cooperative unfolding of both the globular and triple-helical domain. The collagen-like domain isolated by trypsin digestion shows a sharp transition at the same temperature, with an enthalpy of 12.5 kJ/mol tripeptide unit. At low pH, Scl2 and its isolated collagen-like domain show substantial destabilization from the neutral pH value, with two thermal transitions, at 24°C and 27°C. A similar destabilization at low pH was seen for Scl2 charged model peptides, and the degree of destabilization is consistent with the strong pH dependence arising from the GKD tripeptide unit. The Scl2 protein contains twice as much charge as human fibril forming collagens, and the degree of electrostatic stabilization observed for the Scl2 is similar to the contribution Hyp makes to the stability of mammalian collagens. The high enthalpic contribution to the stability of the Scl2 collagenous domain supports the presence of a hydration network in the absence of Hyp.

Collagens are considered to be the characteristic structural molecules of the extracellular matrix of multicellular animals. Fibril forming collagens and basement membrane collagens are ubiquitous in vertebrates and invertebrates, while families of more specialized collagens have developed in different organisms, such as the 28 distinct collagen types found in vertebrates (1-3) and the ~100 cuticle collagen genes in C. elegans (4). In recent years, the range of occurrence of collagen-like sequences with Gly as every third residue and a high Pro content has been extended from metazoans to more than 100 proteins in bacteria and bacteriophage (5). An understanding of the structure and stabilization of such bacterial collagens presents new challenges because they lack the hydroxyproline post-translational modification characteristic of animal collagens.

A high content of hydroxyproline (Hyp) is a unique stabilizing feature of animal collagens. The characteristic structural motif of all collagens is the triple helix, composed of three left-handed polyproline II-type chains (3...
residues/turn) wound around the central axis to form a right-handed superhelix (6-8). The close packing of each chain near the central axis constrains every third residue of the amino acid sequence to be Gly, generating the repeating sequence (Gly-Xaa-X’aa)n. A high content of the imino acids Pro and Hyp is found in all animal collagens, and their restricted ring conformation close to the phi, psi dihedral angles found in the triple-helix confers entropic stabilization to collagen (9,10). Pro residues located in the Y position of the (Gly-Xaa-X’aa)n sequence become post-translationally modified to Hyp by prolyl hydroxylase (11). The presence of Hyp residues in the Y position confers a thermal stability greater than Pro residues. Inhibition of post-translational hydroxylation leads to a decrease in melting temperature of type I collagen by ~15°C (12,13), while recombinant human collagen homotrimers of type I α1 chains which are unhydroxylated have a Tm value ~11°C lower than the hydroxylated form (14). Confirmation of Hyp stabilization of the triple-helix and its position specific nature is seen in model peptides, where Tm ~60°C for (Pro-Hyp-Gly)10, Tm~30°C for (Pro-Pro-Gly)10, and Tm<0°C for (Hyp-Pro-Gly)10 (15). A close correlation is observed between the thermal stability of the collagen molecule and the upper environmental temperature of the organism (16), and analysis suggests that Hyp is a major determinant of the variations in thermal stability among organisms (17).

The mechanism of Hyp stabilization of the triple-helix is controversial, since the hydroxyl group of Hyp is pointing outward from the triple-helix and cannot form any direct intramolecular hydrogen bonds. The classic calorimetric studies of Privalov showed collagens have a very high enthalpic contribution to their stability compared with other proteins, and that this calorimetric enthalpy increases with Hyp content (10). Since direct intramolecular hydrogen bonding is not possible, Privalov postulated Hyp participation in a hydration network involving available backbone carbonyl groups. Such an extensive oriented water-mediated hydrogen bonding network, supported by early NMR studies (18), is seen in the high resolution crystal structures of collagen-like peptides (19,20). The Raines laboratory suggested an alternative stereoelectronic mechanism of hydroxyproline stabilization due to the electron withdrawing effect of the hydroxyl group (21). The electron withdrawal favors an exo pyrrolidine ring pucker, and recent x-ray studies on collagen-like peptides shows that the exo/up pucker of the imide ring in the Y position is favorable for the triple-helix (22).

Prolyl hydroxylase is considered to be present only in multicellular organisms, although a recent report indicates its presence in the yeast Hansenula polymorpha (23). Bacteria lack this enzyme and cannot post-translationally modify Pro in the Y position of the (Gly-Xaa-X’aa)n sequence to form Hyp. Despite the absence of Hyp, several expressed bacterial proteins with collagen-like sequences have been shown to form stable triple-helical structures (24-26). Some of the best characterized bacterial collagen-like proteins are the streptococcal collagen-like proteins, ScI1 and ScI2, which are expressed on the cell surface of group A Streptococcus (26-31). Structurally, the extracellular portions of ScI proteins consist of an N-terminal globular domain (also known as the variable domain) attached to a rod-shaped collagen-like (CL) domain (28). Although the function of ScI proteins is not known, they have been reported to be involved in the adherence of Streptococcus pyogenes to human cells and tissues (27,29). One member of the ScI family interacts with α2β1 integrin to promote cell adhesion and intracellular signaling (32), indicating bacterial collagen-like proteins show functional as well as structural similarities to human collagens. Circular dichroism spectroscopy, rotary shadowing electron microscopy and enzymatic digestion studies indicate that recombinant ScI1 and ScI2 proteins adopt a stable collagen triple-helix structure (26).

To explore the basis of bacterial collagen triple-helix stability in the absence of Hyp, biophysical studies were carried out on recombinant ScI2, its isolated collagen domain, and a set of peptides which model the highly charged region of this protein. These studies indicate that ion pairs play a major role in stabilizing the ScI2 triple-helix and that enthalpic stabilization, likely to involve interactions of polar groups with an ordered hydration network, also makes an important contribution.

**EXPERIMENTAL PROCEDURES**
Protein expression and purification - The fragment of the scl2.28 allele (Q8RLX7) encoding the combined globular and collagen-like (CL) portion of Scl2.28 protein (p163 clone) (26) was recloned into the E. coli vector pColdIII (33), and recombinant protein was expressed in BL21 strain. Cells were grown in M9-Casaminoacid medium until they reached an OD₆₀₀ of 0.8. The cultures were shifted to a 15°C water bath and 1mM IPTG was added to induce protein expression. After overnight incubation, cells were harvested by centrifugation and were disrupted by a French press. Cell debris was removed by centrifugation. The expressed p163 protein was found in the supernatant, as a soluble protein. The p163 protein product, denoted as the rScl2 protein, was precipitated by 35% ammonium sulfate. The pellet was dissolved in PBS buffer (pH 7.0) and loaded onto a Sephacryl S-100HR gel filtration column (Pharmacia). The fractions containing rScl2 were dialyzed against 20mM Tris-HCl buffer (pH8.0) containing 1mM EDTA and 5% glycerol and further purified using a DEAE Sephadex anion exchange column. Protein purity was checked by SDS-PAGE and MALDI-TOF mass spectrometry, and protein concentration was determined using an extinction coefficient of ε(280nm) = 13980 (M⁻¹ cm⁻¹) (34). For all later experiments, the protein was dialyzed into either PBS buffer (20mM sodium phosphate, 150mM NaCl, pH7) or glycine buffer (20mM glycine, 150mM NaCl, pH2 or pH2.8).

Trypsin digestion of rScl2 - To obtain the collagenous fragment, 1.4mg of the rScl2 protein was digested with 150μg of trypsin. The digested product was loaded onto a Sephacryl S-100HR gel filtration column (Pharmacia). The purity of the fractions was checked by mass spectrometry. Mass spectrometry of the digested product, designated as rScl2-CL, showed a heterogeneous product containing slightly different lengths of the collagen domain. A major peak seen at 22385Da represents the expected (Gly-Xaa-X’aa)₇₉ molecular weight of the complete collagen like domain, while a second peak seen at 20570Da represents (Gly-Xaa-X’aa)₇₂; several minor peaks were also observed with molecular weights between these two major peaks.

Peptides - The peptide Ac-GPOGPOGPOGKDGKDGQNGKDGPLGPOGPOGPOGY-NH₂, denoted as the GKDGDKD peptide, and peptide Ac-GPOGPOGPOGKDGKDGQNGKDGPLGPOGPOGPOGY-NH₂, denoted as the Scl2-repeat peptide, were synthesized by Tufts University Core Facility (Boston, MA). The peptide sequence differs in one tripeptide unit GPL versus GLP from the Scl2-repeat, but the charges are unchanged and the stability is expected to differ by 2°C. The peptides include a Tyr at the C-terminus for concentration determination, using the molar extinction coefficient ε=1400M⁻¹cm⁻¹ at 275nm on a Beckmann model DU640 spectrophotometer. Peptides were purified on a Shimadzu reversed-phase HPLC system and the identity of the peptides confirmed by MALDI-TOF mass spectrometry.

Circular dichroism spectroscopy - Circular dichroism (CD) spectra were recorded on an AVIV model 62DS spectropolarimeter. Cuvettes of 0.2mm path length were used to measure spectra down to 190nm. For measurements between 210nm and 260nm, 1mm path lengths were used. The temperature of the cells was controlled using a Peltier temperature controller. Protein solutions were equilibrated for at least 24 hrs at 4°C before measurements. Wavelength scans were collected in 0.5nm steps with 4s averaging time and repeated 3 times. For temperature-induced denaturation, the ellipticity was monitored as a function of temperature, using the wavelength maximum of 220nm for the expressed proteins and 225nm for the peptides. For thermal transitions of rScl2 and rScl2-CL, samples were equilibrated at each temperature for at least 5min and the transitions are at or close to equilibrium. The peptide melting curves are obtained under standard conditions used in our laboratory for comparison, even though equilibrium is not fully reached (35). Peptides were equilibrated for 2 minutes at each temperature and the temperature was increased at an average rate of 0.1°C/min. The data were fit assuming a trimer to monomer model T→3M, which showed good agreement. The Tₘ value was determined as the temperature where the fraction folded is equal to 0.5 in the curve fitted to the trimer to monomer transition. The Tₘ value is determined with an accuracy of ±0.5°C.

Differential Scanning Calorimetry - Differential scanning calorimetry (DSC) experiments were recorded on a NANO-DSC II model 6100 (Calorimetry Sciences Corporation). The sample
was dialyzed against either PBS buffer (20mM sodium phosphate, 150mM NaCl, pH7) or glycine buffer (20mM glycine buffer, 150mM NaCl, pH2 or pH2.8) and equilibrated at 5°C for at least 24 hours. Sample solutions were loaded at 5°C into the cell and heated at a rate of 1°C/min. The enthalpy was calculated from the first scan, since the scans were not reversible upon cooling.

**Dynamic light scattering and static light scattering** - Dynamic light scattering (DLS) and static light scattering were simultaneously carried out on a DynaPro Titan instrument (Wyatt). The instrument calculates the intensity-intensity autocorrelation function \( g(\tau) \) from the scattered light at an angle of 90°. The translational diffusion coefficient was obtained from \( g(\tau) \) and converted into the Stokes’ radius via the Stokes-Einstein equation

\[
R_s = \frac{k_B T}{6\pi \eta D},
\]

where \( k_B \) is the Boltzmann constant, \( T \) the temperature in Kelvin, \( \eta \) the solvent viscosity and \( D \) the translational diffusion coefficient. Samples were filtered through 100nm pore size filters before measurement. For each measurement, the sample was equilibrated at the relevant temperature for 5min. The data were fit assuming a trimer to monomer model \( T \rightarrow 3M \). The \( T_m \) was determined with an accuracy of ±1°C.

**Fluorescence** - Fluorescence measurements were done on an Aminco Bowman Series 2 luminescence spectrometer, with a protein concentration of 0.1mg/ml. The excitation wavelength was 295nm and the emission signal was measured at 340nm. The excitation and emission slit width were 4nm. The protein was heated with a heating rate of 0.16°C/min and the signal measured every 0.2min. The data were fit assuming a trimer to monomer model \( T \rightarrow 3M \). The \( T_m \) was determined with an accuracy of ±0.2°C.

**Collagen stability calculations** - Predictions of \( T_m \) values for peptides and of the relative stability along a collagen molecule were based on the algorithm derived from host-guest peptides (36). The pH dependent changes in the stability contributions of charged tripeptide sequences, such as GKD, were based on previously published data (37).

**RESULTS**

Stability of recombinant rScl2 and its trypsin resistant collagenous domain rScl2-CL - A protein was expressed from a clone of the *Streptococcus pyogenes* collagen-like protein Scl2.28 variant (denoted here as Scl2) containing the N-terminal globular region (74 residues), the (Gly-Xaa-X'aa)\(_9\) triple-helix domain, a short Pro-rich sequence (17 residues), and the C-terminal 8-residue Strep-tag II (28) (Fig. 1a); this protein is referred to as rScl2. The rScl2 product was expressed in a cold shock vector pcoldIII in *E. coli* (33) and purified by ammonium sulfate precipitation and anion exchange chromatography. The purity was determined by SDS-PAGE and the identity was confirmed by mass spectroscopy.

Since the collagen triple-helix is known to be resistant to trypsin, the rScl2 product was treated with trypsin at pH7 for one hour at room temperature to isolate the collagenous domain of this protein. The trypsin digested product, designated rScl2-CL, contained the intact (Gly-Xaa-X’aa)\(_9\) region, and a significant amount of (Gly-Xaa-X’aa)\(_{72}\), as detected by mass spectroscopy.

The circular dichroism (CD) spectra of the rScl2 protein at pH7 showed a maximum near 220nm (MRE\(_{220}\) = +1200deg.cm\(^2\).dmol\(^{-1}\)) and a minimum near 200nm (MRE\(_{200}\) = -45000 deg.cm\(^2\).dmol\(^{-1}\)) (Fig. 1b). The positions of the maximum and minimum are typical of the collagen triple-helix, but with lower magnitudes. The isolated collagen domain, rScl2-CL had similar CD features, but with much larger magnitudes (MRE\(_{220}\) = +7400 deg.cm\(^2\).dmol\(^{-1}\); MRE\(_{198}\) = -66,000 deg.cm\(^2\).dmol\(^{-1}\)), giving values similar to those seen for mammalian collagens (38). Subtraction of the CD spectrum of the collagenous domain from total rScl2 CD spectrum suggests the globular domain has a negative contribution at 220nm and a positive contribution at 197nm, indicating the possibility of \( \beta \)-sheet or \( \alpha \)-helix contributions. Secondary structure programs (e.g. www.cm pharm.ucsf.edu/~nomi/nnpredict.html, NNPREDICT) predicted 46 residues or 62% of the globular domain can adopt an \( \alpha \)-helical conformation, which would be consistent with the subtracted CD curve and with earlier structure analysis (26). The program COILS (http://www.ch.embnet.org/software/COILS_for.html) predicts a high coiled-coil propensity for residues 1-15 and 43-60, consistent with the
observation that coiled-coil domains are often found adjacent to triple-helices and may fill a nucleating role (39).

Monitoring the change in CD ellipticity at 220nm for rScl2 at pH7 with increasing temperature gave a very sharp thermal transition with a T_m value of 36°C (Fig. 1b). The rScl2-CL domain showed a similar sharp transition at 36°C. The absence of a second independent thermal transition in rScl2 suggests that the non-collagenous globular domain is unfolding at the same time as the triple-helix in a highly cooperative transition under these conditions. This is consistent with the Trp fluorescence monitored as a function of temperature at neutral pH, which showed a single transition at 36°C (Fig. 1c). It is likely the fluorescence arises from the Trp60 in the globular region since the only other Trp is located in the unstructured C-terminal Strep-tag sequence and is not expected to show any change with temperature.

Conformational properties and thermal stability were also investigated using dynamic light scattering (Fig. 1d). Both rScl2 and rScl2-CL showed a significant decrease in hydrodynamic radius with increasing temperature. For rScl2, the Stokes radius changed from R_s= 10.4nm to 4.6nm at 36-37°C, while the Stokes radius decreased from R_s= 8.0nm to 3.6nm at 34-35°C for the isolated collagenous domain, rScl2-CL. The temperature dependent changes in hydrodynamic radius together with a decrease in intensity (data not shown) upon unfolding confirm these molecules are undergoing a thermal transition from native trimers to unfolded monomers. The larger hydrodynamic radius for the native rScl2 protein (10.4nm) compared with rScl2-CL protein (8.0nm) is likely to reflect the effect of the globular domain on the molecular shape.

Differential scanning calorimetry (DSC) of the rScl2 protein at pH7 exhibited a single very sharp transition at 37°C, with a calorimetric enthalpy ΔH_cal = 3499kJ/mol (Table 1, Fig 2a). The collagenous domain rScl2-CL showed a somewhat broader transition at the same temperature, with a small shoulder that is likely due to some heterogeneity in the digestion products (Fig. 2b). The calorimetric enthalpy of the collagenous domain rScl2-CL is ΔH_cal = 3400 kJ/mol assuming an intact (Gly-Xaa-X'aa)_79 collagen domain of MW = 22385. The enthalpy per tripeptide for the bacterial collagen domain is 12.5 kJ/mol tripeptide, or 4.2kJ/residue.

Following heat denaturation, incubation of rScl2 at low temperature (neutral pH, 5°C, c=0.8mg/ml) led to rapid refolding, resulting in the native CD signal regained within the deadtime (1-2 minutes), while incubation at 20°C led to refolding within minutes. In contrast, the collagenous domain rScl2-CL alone showed no indication of refolding, even after a week at 0°C. This is consistent with the earlier report of the necessity of the N-terminal globular domain for formation of triple-helices (26).

Effect of pH on stability - The collagenous region of Scl2 contains a high proportion of charged residues, with approximately 30% of the (Gly-Xaa-X'aa)_79 sequence occupied by Asp, Glu, Lys and Arg. The effects of pH on the stability and conformation of rScl2 and its isolated collagenous domain were characterized.

The CD spectrum of the rScl2 protein showed a significant drop in the 220nm peak at pH2.2 (data not shown) compared with pH7 (MREW220 = +1200deg.cm^2.dmol^-1 at pH7; MREW220 = -1252 deg.cm^2.dmol^-1 at pH2.2). A dramatic decrease in MREW220 was seen for the collagenous domain rScl2-CL, decreasing from +7000 at pH7 to +2000deg.cm^2.dmol^-1 at pH2.2 (Fig. 3, inset). The decrease in the characteristic triple-helix 220nm CD signal at pH2.2 indicates that the protonation and loss of negative charge on Glu and Asp residues lead to a significant loss of triple-helix content.

Low pH also led to a significant decrease in triple-helix stability. The CD thermal transition at pH2.2 showed a broad transition near 23-27°C, compared with the sharp 36°C transition at pH7 (Fig. 3). Although largely cooperative melting behavior was seen for the globular and triple-helix domain of rScl2 at neutral pH value, in some cases there was an indication that the globular domain unfolds slightly after unfolding of the triple-helix. For instance, at pH5, monitoring the 220nm ellipticity as a function of temperature showed a sharp decrease at 36°C followed by a small increasing signal at higher temperatures (data not shown). The increasing 220nm signal is likely to represent denaturation of α-helix in the globular domain subsequent to the denaturation of the triple-helix.

At pH2.2, DSC scans of rScl2 and rScl2-CL proteins showed two discrete transitions at 24°C and 27°C, in contrast to the
single DSC transition at 36°C for pH7 (Fig. 2; Table 1). Again, the decrease in thermal stability at low pH suggests that ion pairs make an important contribution to triple-helix stability. The appearance of two discrete peaks for the collagenous domain suggests two independent folding/unfolding domains at low pH. There was some indication of a biphasic transition in the CD denaturation curve, although it was fit to a single transition (Fig. 3). There was a significant loss of calorimetric enthalpy in going from neutral to acidic pH (Table 1), from 3146kJ/mol for the single peak at pH7 to 2427kJ/mol for the sum of the two peaks seen at pH2.2, suggesting that neutral pH is promoting hydrogen bonding, as well as electrostatic interactions.

The thermal transition of rScl2 monitored by Trp fluorescence was shifted from 36.0°C at neutral pH to 26.5°C at pH2.2 (Table 1). The higher of the two acidic transitions seen by DSC near 27°C was observed, while the lower DSC transition near 24°C was not. The Trp appears to unfold with the more stable of the two acidic triple-helix domains. The location of Trp60 at the C-terminus of the globular sequence close to the N-terminal end of the triple-helix sequence, suggests the more stable domain includes the N-terminus of the (Gly-Xaa-X'aa)79 sequence.

Peptide models for regions of Scl2 collagen domain - Peptides were designed to clarify the role of the highly charged repeating sequences found near the C-terminal end of the collagenous domain of Scl2. There are 3 full repeats of the sequence GKDGDQGKGDGPL and several partial repeats of this sequence. One peptide includes the 15 residue sequence Ac(GPO)3GKDGKD(GPO)3GYCONH2 (designated GKDGKD peptide), while the others contain GKDGDK in a host-guest design Ac(GPO)3GKDGDK(GPO)3GYCONH2 (designated GKDGDK peptide) (Table 2).

Using the Tm values obtained from guest tripeptide units in the host peptide (Gly-Pro-Hyp)n, it was predicted that the Scl2-repeat peptide would have a stability of 4.9°C (36); http://jupiter.umdnj.edu/collagen_calculator/. Surprisingly, the Scl2-repeat peptide formed a stable triple-helix at neutral pH with an MRE225 = 3500 deg cm² dmol⁻¹ and Tm = 34.0°C (Fig. 4), demonstrating that the triple-helix structure of the Scl2-repeat is far more stable than expected. Differential scanning calorimetry indicates this peptide has a calorimetric enthalpy of 251kJ/mol at pH7. At pH1.4, the Tm decreased to 16.5°C, and the calorimetric enthalpy decreased to 118 kJ/mol.

The GKDGDK peptide in the host-guest peptide design forms a triple-helix at neutral pH with MRE225 = 5500 deg.cm².dmool⁻¹. The observed Tm for the GKDGDK peptide was 30.2°C, which is again greater than that predicted (predicted Tm = 21.6°C) (36), and its calorimetric enthalpy is 266 kJ/mol (Table 2). Measurements of thermal stability by CD at different pH values indicated that the Tm value decreased below pH4, with Tm = 18.7°C at pH2.8 and 15.1°C at pH1.4. The calorimetric enthalpy was not substantially lower at acidic pH, with ΔHcal = 245 kJ/mol at pH2.1. The DSC scans showed higher Tm values than CD transitions, which is due to the higher scan rate and non-equilibrium conditions for these peptides (36).

DISCUSSION

The thermal stability of animal collagens plays a critical role in their biological function, in terms of biosynthesis, their role in the extracellular matrix, and degradation. In animals, the stability of collagen has been correlated with overall imino acid content and in particular with Hyp content (9,17). Bacterial collagen-like proteins contain no Hyp, yet the collagenous domains of Streptococcus pyogenes Scl proteins and the Bacillus anthracis BclA collagen-like proteins form triple-helix structures with a stability near 37°C, close to that seen for mammalian collagens (24-26). The high Tm value of BclA may relate to its high ~ 25% Pro content, and the O-glycosylation of the abundant Thr residues by novel oligosaccharides (24,40,41). However, the Scl2 collagen-like domain has a relative low ~12% Pro content. Placing the Scl2 collagenous domain on a plot of thermal stability versus imino acid content for animal collagens, indicates that its Pro content is not predicted to lead to a stable collagen triple-helix (Fig. 5). This plot shows that animal collagens generally use a similar mechanism of stabilization, in which imino acid content and Hyp content play a key role, whereas the collagenous domain of Scl2 must have differences in the major interactions leading to triple-helix stability. Stability of the triple-helix is also dependent on length when the number of (Gly-Xaa-X'aa)n units is small, e.g. n=10, 20 or 30 (42).
However, Han et al (25) have shown that collagen-like regions of Scl1 and Scl2 proteins from different strains which vary in length from 40-120 tripeptides units have $T_m$ values in the range of 34-37°C, indicating there is no longer a length dependence.

The thermal transitions of Scl2 and its collagen domain are as sharp as seen for bovine collagen, so high cooperativity and stability are realized in the absence of Hyp. The 12% Pro content of the collagen-like domain of Scl2 compares with ~20% imino acids in fibril forming animal collagens (Table 3). In the collagenous domain of Scl2, the Pro residues are found preferentially in the X positions, and this seems to be true for all prokaryotic collagen-like proteins, with the frequency of Pro residues in the X position exceeding 30%, while the frequency of Pro in the Y position is less than 5% (Pawlowski and Bujnicki, pers. comm.). The percent of hydrophobic residues is relatively low in both Scl2 (7.5%) and human fibril forming collagens (~6%), consistent with the lack of hydrophobic core in the triple-helix. The most striking difference lies in the charged residue content, with the Scl2 collagen domain having 30% charged residues compared with ~15% for fibrillar collagens (Table 3). Because all residues in the X and Y positions of a collagen triple-helix are largely exposed to solvent, this high percentage of charged amino acids can be accommodated in the triple-helix structure.

Evaluation of the $T_m$ values of triple-helical host-guest peptides has provided insight into the contributions of varying Gly-Xaa-X'aa amino acid sequences to stability (36). Within the host peptide (Pro-Hyp-Gly)$_n$, the propensities of all 20 amino acids in the X position were evaluated as a guest Gly-X-Hyp triplet and in the Y position as a guest Gly-Pro-Y triplet, while more complex interactions were studied as guest Gly-Xaa-X'aa tripeptides with complementary charges (e.g. GEK, GKD, etc) or in guest hexapeptide sequences, e.g. GPKGE0. Examination of the (Gly-Xaa-X'aa)$_{79}$ sequence of the Scl2 collagen domain in light of these studies indicates a stable domain for tripeptides 5-53, which contains 15% Pro, hydrophobic stabilization of the form GLQGLQGLQ and favorable charge interactions e.g. KGD, KGE. The C-terminal end of the collagenous sequence of Scl2 (tripeptide units 54-79) is repetitive and highly charged, containing 3 repeats of GKDGKDGQNGKDGLP (40% charged residues) and several partial repeats. Other strains of Streptococcus have triple-helix domains ranging from 10-220 tripeptides units, and many contain repeating sequences related to the repeat in Scl2 (25,43). An algorithm based on host-guest peptides ((36), http://jupiter.umdnj.edu/collagen_calculator/) predicts this highly charged repeating region to be very unstable and unable to form a triple-helix. However, the peptide data reported here indicate this repeating sequence and the GKDGKD sequence are much more stable than expected (Table 2). Although the collagen stability calculator is reasonably successful at predicting the stability of peptides which include frequent Gly-Pro-Hyp triplets and sequences from human collagens (36), it is far off in its predictions of these highly charged bacterial collagen sequences. It appears that a tripeptide sequence such as GQN or GKD in the context of the imino acid poor sequence GKDGKDQNGKDGLP does not lead to the large destabilization seen in the context of surrounding Gly-Pro-Hyp triplets. These highly charged and polar repeating sequences contribute to the stability of the triple-helix even though they lack Hyp and have only one Pro in each 15 residue repeat. These studies point out the limitations of the collagen stability calculator, because of its simplified basis sets, but the data from these studies are being incorporated into future calculations to improve its prediction accuracy.

The destabilizing effect of low pH on rScl2-CL ($\Delta T_m ~ 10-13°C$) indicates an important contribution of electrostatic interactions to stability, and contrasts with the relatively small effect of acidic pH on animal collagens (44). It is interesting to note that the degree of electrostatic stabilization observed in rScl2-CL is similar to the contribution that Hyp makes to the stability of mammalian collagens (12-14). Peptide studies indicate the molecular basis for this strong pH dependence in the rScl2 triple-helix. The peptide (Pro-Hyp-Gly)$_{10}$ has a 3°C higher $T_m$ value at acid pH than at neutral pH, while the Scl2-repeat model peptide and the GKDGKD peptide both showed significant destabilization at pH2.2 relative to pH7. The pH dependence of both peptides can be explained by the presence of GKD triplets. The tripeptide sequence GKD in a host-guest system shows a decrease in its $T_m$ value from 35.8°C at neutral pH to 30.5°C at pH2.2 (37). This $\Delta T_m$
(pH7→2.2)= -5.3°C for GKD is consistent with the observed 17.5°C decrease from neutral to pH2.2 in the Tm value of Scl2-repeat peptide which contains 3 GKD units [3 x -5.3°C= -15.9°C] and the observed 10.7°C decrease from neutral to pH2.2 in the Tm value of the GKDGKD peptide which contains 2 GKD units [2 x -5.3°C= -10.6°C]. The double peaks seen for thermal transitions of both rScl2 and rScl2-CL on DSC suggest that under low pH conditions, there are two less stable triple-helical domains. Their lower stability (24°C, 27°C) is likely to reflect their shorter length as well as their high GKD content. The C-terminal repeating units GKDGDQNGKDGLP are likely to be more destabilized at low pH than the N-terminal part of the collagenous domain.

Animal collagens have a very high enthalpy of denaturation, indicating hydrogen bonding is a major contributing factor to stability. Privalov observed that the enthalpy increased with increasing Hyp content, and proposed this is a consequence of an ordered water mediated hydrogen bonding network (10). Such a hydration network has been observed in the crystal structure of various collagen model peptides, and Hyp residues appear to be linchpins of this network (6,19). The calorimetric enthalpy of the collagenous domain of Scl2 is ~ 12.5 kJ/mol tripeptide, which is less than the 18-22kJ/mol tripeptide seen for animal collagens near neutral pH (10,45), but in the same range as seen for (Pro-Hyp-Gly)10 (35) (Table 4). This relatively high calorimetric enthalpy for Scl2-CL suggests that a hydration network is likely to be present in the bacterial collagen domain even in the absence of Hyp, and may involve the numerous polar and charged residues in this domain.

Eukaryotic collagens utilize Pro hydroxylation in position Y as a major mechanism of modulating triple helix stabilization, while prokaryotes have evolved different molecular strategies of triple helix stabilization which differ not only from their eukaryotic counterparts but also between different members of prokaryotic collagen-like proteins (5,24,25). In theory, bacteria could use Gly-Pro-Pro tripeptide units, which are very stabilizing sequences [e.g. the Tm value of (Gly-Pro-Pro)14>40°C, (42)], but Gly-Pro-Pro tripeptides are rarely found in bacterial collagen-like proteins, and it is possible that such sequences present toxicity problems in bacteria (pers. obs.). The studies reported here highlight the flexibility of the triple-helix motif which can attain 37°C stability by interactions not involving Hyp. Imino acid free tripeptide sequences with polar residues, such as Gly-Gln-Asn, are much less destabilizing in the Scl2 sequence than in a Pro/Hyp rich environment. Once the length is longer than 40 tripeptide sequences, the Scl2 protein uses a variety of electrostatic interactions, interchain hydrogen bonds, and a hydration mediated hydrogen bonding network as alternative to the Hyp stabilization in animal collagens.

These studies have focused on the role of Hyp in stabilizing the collagen triple-helix, but Hyp also plays important roles in nucleation and folding of the collagen triple-helix domain (46-48) and in the self-association of triple-helices (49). The triple-helix domain of rScl2 does not refold by itself, but requires the N-terminal globular domain, which contains a coiled-coil motif, for trimerization and folding. It is not yet clear whether alternative interactions are involved in the folding process in the absence of Hyp. Since Scl2 is a membrane bound cell surface protein, self-association to a supramolecular structure is not likely to be necessary for its biological role. It is possible that Hyp is a requirement for all collagen-like proteins which do self-assemble to higher order structures. The availability of triple-helix proteins with same stability as mammalian collagens but without Hyp provides opportunities for the use of expressed bacterial collagenous domains to create well-defined and novel biomaterials.

REFERENCES

Acknowledgements - This work was supported by NIH grants R21EB007198 (B.B.), GM60048 (B.B.) and AI50666 (S.L.).

1 The abbreviations used are: CD, circular dichroism; DSC, differential scanning calorimetry; DLS, dynamic light scattering. The three letter code for amino acids is used, with Hyp designating hydroxyproline; single amino acid notation is used to indicate the sequence of peptides, with O used for hydroxyproline.

FIGURE LEGENDS

Figure 1. Thermal stability of the rScl2 protein and its collagenous domain rScl2-CL. (a) Amino acid sequence of the recombinant rScl2 protein, with the (Gly-Xaa-X’aa)79 collagenous domain shown in bold. (b) CD thermal transitions of rScl2 (filled circles) and rScl2-CL (open diamonds) at pH7, monitored at 220nm, with their CD spectra shown as an inset. (c) Tryptophan fluorescence of rScl2 at pH7, monitored as a function of temperature (excitation wavelength 295nm; emission signal 340nm). Fluorescence measurement could not be done at the rScl-CL because there are no aromatic residues in the collagenous domain. (d) Changes in the hydrodynamic radius of rScl2 (filled circles) and rScl2-CL (open diamonds) at pH7 as a function of temperature as measured by dynamic light scattering (DLS).

Figure 2. Differential scanning calorimetry of (a) rScl2 at pH7 and pH2.2; and (b) rScl2-CL at pH7 and pH2.2.

Figure 3. The pH dependence of thermal stability monitored by CD for rScl2-CL at pH7 and pH2.2, with an insert of their CD spectra.

Figure 4. Thermal transitions monitored by CD spectroscopy of model peptides for the Scl2 C-terminal charged region. (a) Scl2-repeat peptide at pH7 and pH2.8; (b) the GKGDKD peptide at pH7 and pH2.2.

Figure 5. Plot of Tm values against imino acid content for a wide variety of animal collagens, showing the bacterial collagen-like proteins Scl2 (filled triangle) and Bacillus anthracis collagen-like protein BclA (filled circle). The values for the animal collagens were taken from (10).
Table 1. Thermal stability determined by CD spectroscopy, fluorescence, dynamic light scattering (DLS) and differential calorimetry (DSC) for the rScl2 protein and its collagenous domain rScl2-CL, at pH7 and pH2.2.

<table>
<thead>
<tr>
<th>Protein</th>
<th>CD $T_m$ (°C)</th>
<th>Fluorescence $T_m$ (°C)</th>
<th>DLS $T_m$ (°C)</th>
<th>DSC $T_m$ (°C)</th>
<th>$\Delta H$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rScl2</td>
<td>pH 7 24.2</td>
<td>pH 7 36.5</td>
<td>pH 7 36.4</td>
<td>pH 7 3499</td>
<td>pH 7 3499</td>
</tr>
<tr>
<td>rScl2-CL</td>
<td>pH 2.2 25.7</td>
<td>pH 2.2 26.5</td>
<td>pH 2.2 37</td>
<td>pH 2.2 28</td>
<td>pH 2.2 28</td>
</tr>
</tbody>
</table>

$^a$ sum of two peaks
$^b$ no aromatic residues in triple-helix domain
Table 2. Observed melting temperature values obtained by CD spectroscopy and predicted stability, together with calorimetric enthalpy, for the peptide models of the repeating, highly charged portion of the collagenous region of Scl2.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Obs. $T_m$ pH 7</th>
<th>Pred. $T_m$ pH 7</th>
<th>$\Delta H_{cal}$ pH 7</th>
<th>Obs. $T_m$ pH 1.4</th>
<th>Pred. $T_m$ pH 1.4</th>
<th>$\Delta H_{cal}$ pH 1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scl2-repeat</td>
<td>34.0°C</td>
<td>4.9°C</td>
<td>251 kJ/mol</td>
<td>16.5°C</td>
<td>-11.4°C</td>
<td>119 kJ/mol</td>
</tr>
<tr>
<td>Ac(GPO)$_3$-GKDGDQNGKDGPL(GPO)$_4$-GY-CONH$_2$</td>
<td>29.4°C</td>
<td>21.6°C</td>
<td>266 kJ/mol$^a$</td>
<td>18.7°C</td>
<td>11°C</td>
<td>245 kJ/mol$^b$</td>
</tr>
<tr>
<td>GKDGDQ</td>
<td>pH 8.5 and pH 6</td>
<td>pH 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ac(GPO)$_3$-GKDGDQ-(GPO)$_3$-GY-CONH$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ pH 8.5 and pH 6

$^b$ pH 2.1
Table 3. Comparison of amino acid features of the Scl2 collagen-like domain with human type II fibrillar collagen (P02458).

<table>
<thead>
<tr>
<th>Collagen</th>
<th>Imino acid</th>
<th>% Hydrophobic</th>
<th>% Charged</th>
<th>Charged pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scl2-collagen domain</td>
<td>12% Pro</td>
<td>7.5</td>
<td>30</td>
<td>16 GKD</td>
</tr>
<tr>
<td>(Gly-Xaa-X’aa)$_{79}$</td>
<td>20 Pro (X position)</td>
<td></td>
<td></td>
<td>6 GER</td>
</tr>
<tr>
<td></td>
<td>9 Pro (Y position)</td>
<td></td>
<td></td>
<td>1 GEK</td>
</tr>
<tr>
<td>tripeptides 5-53</td>
<td>15% Pro</td>
<td>4.7</td>
<td>23</td>
<td>5 GER</td>
</tr>
<tr>
<td></td>
<td>20 Pro (X position)</td>
<td></td>
<td></td>
<td>2 GKD</td>
</tr>
<tr>
<td></td>
<td>2 Pro (Y position)</td>
<td></td>
<td></td>
<td>1 GEK, GDR</td>
</tr>
<tr>
<td>tripeptides 54-79</td>
<td>9% Pro</td>
<td>6.4</td>
<td>40</td>
<td>14 GKD</td>
</tr>
<tr>
<td></td>
<td>0 Pro (X position)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 Pro (Y position)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II collagen</td>
<td>21% imino acids</td>
<td>5.9</td>
<td>16</td>
<td>11 GER</td>
</tr>
<tr>
<td>(Gly-Xaa-X’aa)$_{338}$</td>
<td>103 Pro (X position)</td>
<td></td>
<td></td>
<td>4 GKD</td>
</tr>
<tr>
<td></td>
<td>112 Hyp (Y position)</td>
<td></td>
<td></td>
<td>4 GDR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4 GEK</td>
</tr>
</tbody>
</table>
Table 4. Comparison of calorimetric enthalpy per tripeptide for collagens and model peptides

<table>
<thead>
<tr>
<th></th>
<th>$\Delta H_{\text{cal}}$(kJ/mol)/tripeptide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collagens</strong></td>
<td></td>
</tr>
<tr>
<td>Type I collagen</td>
<td>$20.1^a$</td>
</tr>
<tr>
<td>Type II collagen</td>
<td>$22.5^b$</td>
</tr>
<tr>
<td>Scl2-CL</td>
<td>$12.5$</td>
</tr>
<tr>
<td><strong>Model peptides</strong></td>
<td></td>
</tr>
<tr>
<td>(POG)$_{10}$</td>
<td>$13.9^c$</td>
</tr>
<tr>
<td>GKDGDGD</td>
<td>$10.2$</td>
</tr>
<tr>
<td>Scl2</td>
<td>$6.6$</td>
</tr>
<tr>
<td>(PPG)$_{10}$</td>
<td>$6$</td>
</tr>
</tbody>
</table>

$^a$ ref (10)

$^b$ ref (45)

$^c$ ref (35)
Figure 1

(a)
ADEQEEAKVRELEIQELAQGLGGIIEKKNFPTLGDGDLDHSTYMTKLTLT
YLQERFQAENSSWRKRLKGIQDHALD

GQD GRN GER GEQ GPT GPA GPR GLQ GLQ
GLQ GER GEQ GPT GPA GPR GLQ GER GEQ GPT
GLA GKA GEA GAK GET GPA GPQ GPR GEQ GPQ
GLP GKD GEA GAQ GPA GPM GPA GER GEK GEP
GTQ GAK GKR GET GPV GPR GER GEA GPA GKD
GER GPV GPA GKD GQN GQD GLP GKD GKD GQN
GKD GLP GKD GKD GQN GKD GLP GKD GKD GQD
GKD GLP GKD GKD GLP GKD GKD GQP GKP
APKTPERVQKPDTAPSA
WSHPQFEK

(b)
Figure 2.

(a)

(b)
Figure 3.

Figure 4.

(a)

(b)
Figure 5.
Mechanism of stabilization of a bacterial collagen triple-helix in the absence of hydroxyproline
Angela Mohs, Teresita Silva, Takeshi Yoshida, Ravish Amin, Slawomir Lukomski, Masayori Inouye and Barbara Brodsky

J. Biol. Chem. published online August 10, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M703991200

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

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2007/08/10/jbc.M703991200.citation.full.html#ref-list-1