

2007

Mechanism Of Stabilization Of A Bacterial Collagen Triple Helix In The Absence Of Hydroxyproline

Angela Mohs

Teresita Silva

Takeshi Yoshida

Ravish Amin

Slawomir Lukomski

See next page for additional authors

Follow this and additional works at: https://researchrepository.wvu.edu/faculty_publications

Digital Commons Citation

Mohs, Angela; Silva, Teresita; Yoshida, Takeshi; Amin, Ravish; Lukomski, Slawomir; Inouye, Masayori; and Brodsky, Barbara, "Mechanism Of Stabilization Of A Bacterial Collagen Triple Helix In The Absence Of Hydroxyproline" (2007). *Faculty Scholarship*. 594.

https://researchrepository.wvu.edu/faculty_publications/594

This Article is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.

Authors

Angela Mohs, Teresita Silva, Takeshi Yoshida, Ravish Amin, Slawomir Lukomski, Masayori Inouye, and Barbara Brodsky

Mechanism of Stabilization of a Bacterial Collagen Triple Helix in the Absence of Hydroxyproline*

Received for publication, May 15, 2007, and in revised form, August 9, 2007. Published, JBC Papers in Press, August 10, 2007, DOI 10.1074/jbc.M703991200

Angela Mohs[‡], Teresita Silva[‡], Takeshi Yoshida[‡], Ravish Amin[‡], Slawomir Lukomski[§], Masayori Inouye[‡], and Barbara Brodsky^{‡1}

From the [‡]Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854 and the [§]Department of Microbiology, Immunology, and Cell Biology, West Virginia University School of Medicine, Morgantown, West Virginia 26506

The *Streptococcus pyogenes* cell-surface protein Scl2 contains a globular N-terminal domain and a collagen-like domain, (Gly-Xaa-X'aa)₇₉, which forms a triple helix with a thermal stability close to that seen for mammalian collagens. 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 modeling the Scl2 highly charged repetitive (Gly-Xaa-X'aa)_n sequences. At pH 7, CD spectroscopy, dynamic light scattering, and differential scanning calorimetry of the Scl2 protein all showed a very sharp thermal transition near 36 °C, indicating a highly cooperative unfolding of both the globular and triple-helix domains. The collagen-like domain isolated by trypsin digestion showed a sharp transition at the same temperature, with an enthalpy of 12.5 kJ/mol of tripeptide. At low pH, Scl2 and its isolated collagen-like domain showed substantial destabilization from the neutral pH value, with two thermal transitions at 24 and 27 °C. A similar destabilization at low pH was seen for Scl2 charged model peptides, and the degree of destabilization was consistent with the strong pH dependence arising from the GKD tripeptide unit. The Scl2 protein contained twice as much charge as human fibril-forming collagens, and the degree of electrostatic stabilization observed for Scl2 was 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, whereas families of more specialized collagens have developed in different organisms such as the 28 distinct collagen types found in ver-

tebrates (1–3) and the ~100 cuticle collagen genes in *Caenorhabditis elegans* (4). In recent years, the range of occurrence of collagen-like sequences with Gly as every 3rd residue and a high Pro content has been extended from metazoans to >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 Hyp post-translational modification characteristic of animal collagens.

A high content of 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 3rd 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 ϕ, ψ dihedral angles found in the triple helix confers entropic stabilization to collagen (9, 10). Pro residues located in the X'aa 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 X'aa position confers a thermal stability greater than that conferred by Pro residues. Inhibition of post-translational hydroxylation leads to a decrease in the melting temperature of type I collagen by ~15 °C (12, 13), whereas recombinant human collagen homotrimers of type I $\alpha 1$ -chains that are unhydroxylated have a T_m ~ 11 °C lower than that of the hydroxylated form (14). Confirmation of Hyp stabilization of the triple helix and its position-specific nature is seen in model peptides, where T_m ~ 60 °C for (Pro-Hyp-Gly)₁₀, T_m ~ 30 °C for (Pro-Pro-Gly)₁₀, and T_m < 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 because the hydroxyl group of Hyp points outward from the triple helix and cannot form any direct intramolecular hydrogen bonds. The classic calorimetric studies of Privalov (10) showed that collagens have a very high enthalpic contribution to their stability compared with other proteins and that this calorimetric enthalpy increases with Hyp content. Because direct intramolecular hydrogen bonding is not possible,

* This work was supported by National Institutes of Health Grants R21EB007198 and GM60048 (to B. B.) and Grant AI50666 (to S. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ To whom correspondence should be addressed: Dept. of Biochemistry, Robert Wood Johnson Medical School, UMDNJ, 675 Hoes Lane, Piscataway, NJ 08854. Tel.: 732-235-4397; Fax: 732-235-4783; E-mail: brodsky@umdnj.edu.

Stabilization of Bacterial Collagen

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). Jenkins and Raines (21) suggested an alternative stereoelectronic mechanism of Hyp stabilization due to the electron-withdrawing effect of the hydroxyl group. The electron withdrawal favors an *exo*-pyrrolidine ring pucker, and recent x-ray studies on collagen-like peptides show that the *exo*/up-pucker of the imide ring in the X'aa 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 X'aa 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-helix structures (24–26). Some of the best characterized bacterial collagen-like proteins are the streptococcal collagen-like proteins Scl1 and Scl2, which are expressed on the cell surface of group A *Streptococcus* (26–31). Structurally, the extracellular portions of Scl proteins consist of an N-terminal globular domain (also known as the variable domain) attached to a rod-shaped collagen-like domain (28). Although the function of Scl 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 Scl family interacts with $\alpha_2\beta_1$ integrin to promote cell adhesion and intracellular signaling (32), indicating that bacterial collagen-like proteins show functional as well as structural similarities to human collagens. CD spectroscopy, rotary shadowing electron microscopy, and enzymatic digestion studies indicate that recombinant Scl1 and Scl2 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 Scl2, its isolated collagen domain, and a set of peptides modeling the highly charged region of this protein. The results indicate that ion pairs play a major role in stabilizing the Scl2 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 portions of the Scl2.28 protein (p163 clone) (26) was recloned into the *Escherichia coli* vector pColdIII (33), and recombinant protein was expressed in the BL21 strain. Cells were grown in M9-casamino acid medium until they reached $A_{600} = 0.8$. The cultures were shifted to a 15 °C water bath, and 1 mM isopropyl β -D-thiogalactopyranoside was added to induce protein expression. After overnight incubation, cells were harvested by centrifugation and disrupted by a French press. Cellular debris was removed by centrifugation. The expressed p163 protein was found in the supernatant as a soluble protein. The

p163 protein product (recombinant Scl2 (rScl2)² protein) was precipitated by 35% ammonium sulfate. The pellet was dissolved in phosphate-buffered saline (20 mM sodium phosphate and 150 mM NaCl (pH 7.0)) and loaded onto a Sephacryl S-100 HR gel filtration column (GE Healthcare). The fractions containing rScl2 were dialyzed against 20 mM Tris-HCl (pH 8.0) containing 1 mM 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 $\epsilon_{280\text{ nm}} = 13,980\text{ M}^{-1}\text{ cm}^{-1}$ (34). For all later experiments, the protein was dialyzed into either phosphate-buffered saline or glycine buffer (20 mM glycine and 150 mM NaCl (pH 2 or 2.8)).

Trypsin Digestion of rScl2—To obtain the collagenous fragment, 1.4 mg of the rScl2 protein was digested with 150 μg of trypsin. The digested product was loaded onto a Sephacryl S-100 HR gel filtration column. The purity of the fractions was checked by mass spectrometry. Mass spectrometry of the digested product (the rScl2 collagen-like domain (rScl2-CL)) showed a heterogeneous product containing slightly different lengths of the collagen domain. A major peak seen at 22,385 Da represents the expected (Gly-Xaa-X'aa)₇₉ molecular mass of the complete collagen-like domain, whereas a second peak seen at 20,570 Da represents (Gly-Xaa-X'aa)₇₂; several minor peaks were also observed with molecular masses between those of these two major peaks.

Peptides—Peptide Ac-(GPO)₃GKDGKD(GPO)₃GY-CONH₂ (where "O" is Hyp), denoted as the GKDGKD peptide, and peptide Ac-(GPO)₃GKDGKDGQNGKDGPL(GPO)₄GY-CONH₂, denoted as the Scl2 repeat peptide, were synthesized by the Tufts University Core Facility (Boston, MA). The peptide sequence differs in one tripeptide unit (GPL versus GLP) from the Scl2 repeat sequence, 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 $\epsilon = 1400\text{ M}^{-1}\text{ cm}^{-1}$ at 275 nm on a Beckman Model DU640 spectrophotometer. Peptides were purified on a Shimadzu reversed-phase high pressure liquid chromatography system, and the identity of the peptides was confirmed by MALDI-TOF mass spectrometry.

Circular Dichroism Spectroscopy—CD spectra were recorded on an AVIV Model 62DS spectropolarimeter. Cuvettes with 0.2-mm path lengths were used to measure spectra down to 190 nm. For measurements between 210 and 260 nm, 1-mm path lengths were used. The temperature of the cells was controlled using a Peltier temperature controller. Protein solutions were equilibrated for at least 24 h at 4 °C before measurements. Wavelength scans were collected in 0.5-nm steps with a 4-s averaging time and repeated three times. For temperature-induced denaturation, the ellipticity was monitored as a function of temperature using wavelength maxima of 220 nm for the expressed proteins and 225 nm for the peptides. For thermal

² The abbreviations used are: rScl2, recombinant Scl2; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; rScl2-CL, rScl2 collagen-like domain; DSC, differential scanning calorimetry; DLS, dynamic light scattering; MRE, mean residue ellipticity; deg, degrees.

transitions of rScl2 and rScl2-CL, samples were equilibrated at each temperature for at least 5 min, and the transitions were at or close to equilibrium. The peptide melting curves were obtained under standard conditions used in our laboratory for comparison, even though equilibrium was not fully reached (35). Peptides were equilibrated for 2 min at each temperature, and the temperature was increased at an average rate of 0.1 °C/min. The data were fit assuming the trimer-to-monomer model $T \rightarrow 3M$, which showed good agreement. The T_m was determined as the temperature at which the fraction folded was equal to 0.5 in the curve fitted to the trimer-to-monomer transition. The T_m was determined with an accuracy of ± 0.5 °C.

Differential Scanning Calorimetry (DSC)—DSC experiments were recorded on a NANO DSC II Model 6100 (Calorimetry Sciences Corp.). The sample was dialyzed against either phosphate-buffered saline or glycine buffer and equilibrated at 5 °C for at least 24 h. 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 because the scans were not reversible upon cooling.

Dynamic Light Scattering (DLS)—DLS was carried out on a DynaPro Titan instrument (Wyatt Technology Corp.). 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 (R_s) via the Stokes-Einstein equation: $R_s = k_B T / (6\pi\eta D)$, where k_B is the Boltzmann constant, T is the temperature in Kelvin, η is the solvent viscosity, and D the translational diffusion coefficient. Samples were filtered through 100-nm pore size filters before measurement. For each measurement, the sample was equilibrated at the relevant temperature for 5 min. The data were fit assuming the 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 0.1 mg/ml. The excitation wavelength was 295 nm, and the emission signal was measured at 340 nm. The excitation and emission slit widths were 4 nm. The protein was heated at a heating rate of 0.16 °C/min, and the signal was measured every 0.2 min. The data were fit assuming the 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 rScl2 and Its Trypsin-resistant Collagenous Domain, rScl2-CL—A protein was expressed from a clone of the *S. pyogenes* collagen-like protein Scl2.28 variant (denoted here as Scl2) containing the N-terminal globular region (74 residues), the (Gly-Xaa-X'aa)₇₉ triple-helix domain, a short Pro-rich sequence (17 residues), and the C-terminal 8-residue *Strep*-tag II protein (28) (Fig. 1a); this protein is referred to as

rScl2. The rScl2 product was expressed in the cold-shock vector pColdIII in *E. coli* (33) and purified by ammonium sulfate precipitation followed by gel filtration chromatography and anion exchange chromatography. The purity was determined by SDS-PAGE, and the identity was confirmed by mass spectroscopy.

Because the collagen triple helix is known to be resistant to trypsin, the rScl2 product was treated with trypsin at pH 7 for 1 h 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)₇₉ region and a significant amount of (Gly-Xaa-X'aa)₇₂ as detected by mass spectroscopy.

The CD spectra of the rScl2 protein at pH 7 showed a maximum near 220 nm ($MRE_{220} = +1200$ degrees (deg)·cm²·dmol⁻¹) and a minimum near 200 nm ($MRE_{200} = -45,000$ deg·cm²·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·cm²·dmol⁻¹ and $MRE_{198} = -66,000$ deg·cm²·dmol⁻¹), giving values similar to those seen for mammalian collagens (38). Subtraction of the CD spectrum of the collagenous domain from that of total rScl2 suggests that the globular domain has a negative contribution at 220 nm and a positive contribution at 197 nm, indicating the possibility of β -sheet or α -helix contributions. Secondary structure programs (e.g. NNpredict, www.cmpharm.ucsf.edu/~nomi/nnpredict.html) predicted that 46 residues or 62% of the globular domain can adopt an α -helical conformation, which would be consistent with the subtracted CD curve and with earlier structure analysis (26). The program COILS (www.ch.embnet.org/software/COILS_form.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 220 nm for rScl2 at pH 7 with increasing temperature gave a very sharp thermal transition with a T_m of 36 °C (Fig. 1b). rScl2-CL 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 unfolds 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 Trp⁶⁰ in the globular region because the only other Trp is located in the unstructured C-terminal *Strep*-tag II sequence and is not expected to show any change with temperature.

Conformational properties and thermal stability were also investigated using DLS (Fig. 1d). Both rScl2 and rScl2-CL showed a significant decrease in the hydrodynamic radius with increasing temperature. For rScl2, the Stokes radius changed from 10.4 to 4.6 nm at 36–37 °C, whereas for rScl2-CL, it decreased from 8.0 to 3.6 nm at 34–35 °C. The temperature-dependent changes in the hydrodynamic radius, together with a decrease in intensity (data not shown) upon unfolding, confirm that these molecules are undergoing a thermal transition from

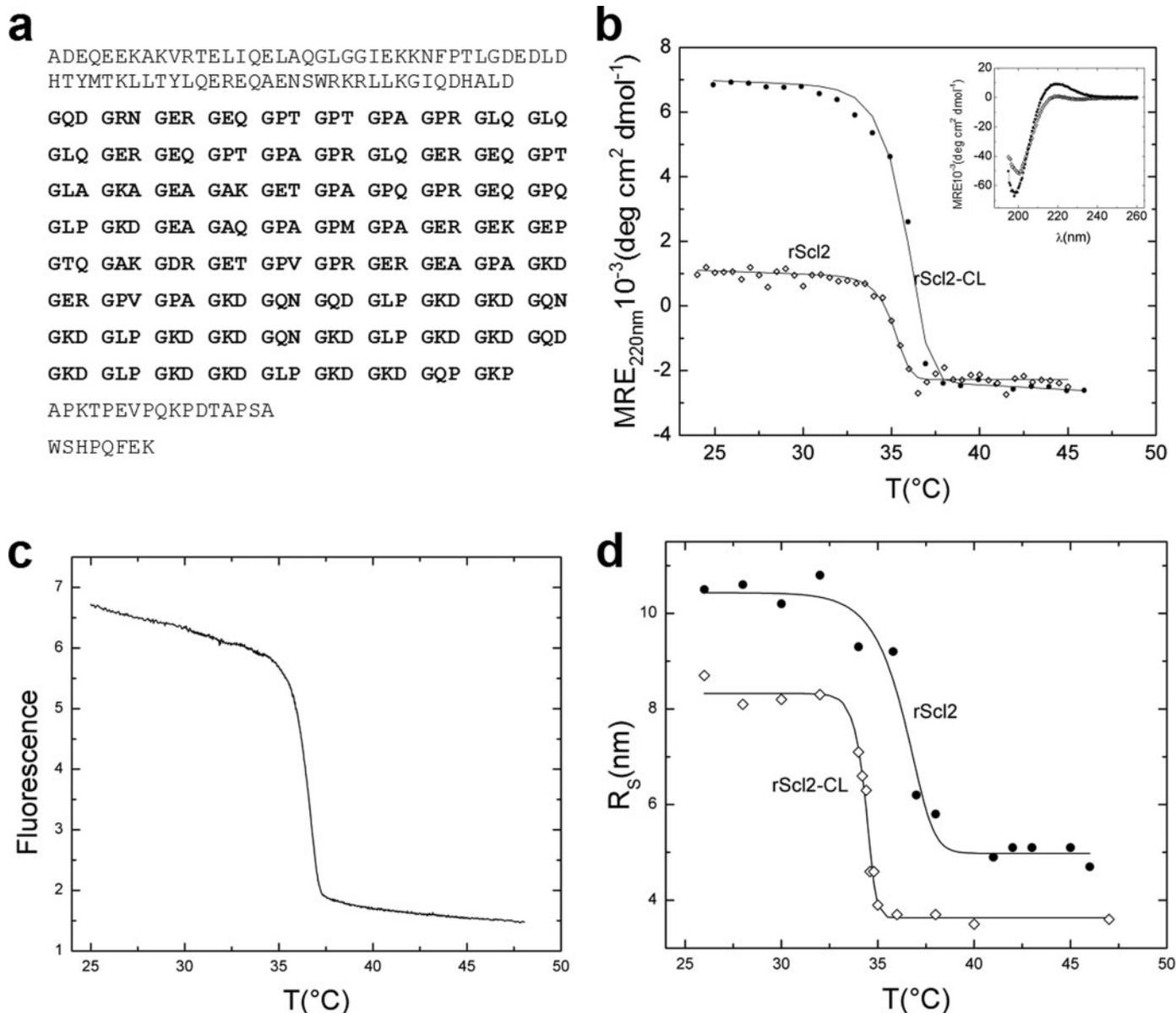


FIGURE 1. Thermal stability of the rScl2 protein and its collagenous domain rScl2-CL. *a*, amino acid sequence of the recombinant rScl2 protein, with the (Gly-Xaa-X'aa)₇₉ collagenous domain shown in *boldface*. *b*, CD thermal transitions of rScl2 (●) and rScl2-CL (◇) at pH 7, monitored at 220 nm, with their CD spectra shown as an inset. *c*, tryptophan fluorescence of rScl2 at pH 7, monitored as a function of temperature (excitation wavelength 295 nm; emission signal 340 nm). Fluorescence measurement could not be done at the rScl2-CL because there are no aromatic residues in the collagenous domain. *d*, changes in the hydrodynamic radius of rScl2 (●) and rScl2-CL (◇) at pH 7 as a function of temperature as measured by dynamic light scattering (DLS).

native trimers to unfolded monomers. The larger hydrodynamic radius for the native rScl2 protein (10.4 nm) compared with that for the rScl2-CL protein (8.0 nm) likely reflects the effect of the globular domain on the molecular shape.

DSC of the rScl2 protein at pH 7 exhibited a single very sharp transition at 37 °C, with a calorimetric enthalpy of $\Delta H_{\text{cal}} = 3499$ kJ/mol (Fig. 2*a* and Table 1). 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. 2*b*). The calorimetric enthalpy of rScl2-CL was $\Delta H_{\text{cal}} = 3400$ kJ/mol assuming an intact (Gly-Xaa-X'aa)₇₉ collagen domain of 22,385 Da. The enthalpy/tripeptide for the bacterial collagen domain was 12.5 kJ/mol of tripeptide or 4.2 kJ/residue.

Following heat denaturation, incubation of rScl2 at low temperature (neutral pH, 5 °C, 0.8 mg/ml) led to rapid refolding, resulting in the native CD signal regained within the dead time (1–2 min), whereas incubation at 20 °C led to refolding within minutes. In contrast, rScl2-CL alone showed no indication of refolding, even after 1 week at 0 °C. This is consistent with a previous 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 ~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.

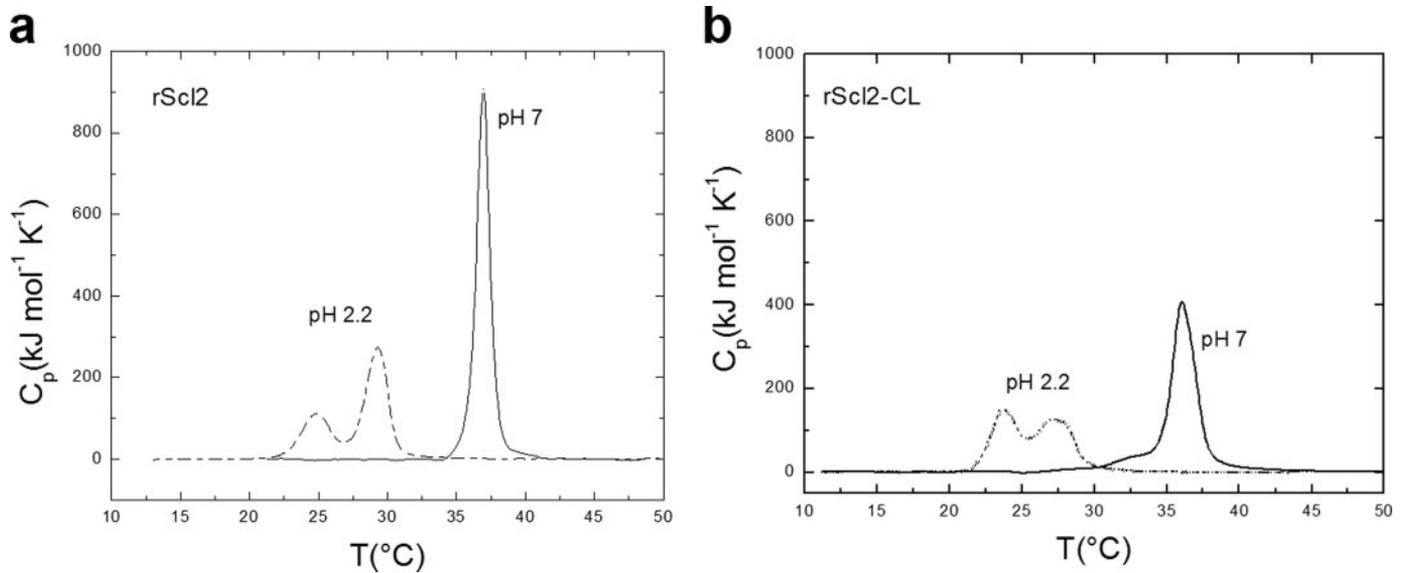


FIGURE 2. DSC of rScl2 (a) and rScl2-CL (b) at pH 7 and 2.2, showing heat capacity (C_p) as a function of temperature.

TABLE 1

Thermal stability determined by CD spectroscopy, fluorescence, DLS, and DSC for the rScl2 protein and its collagenous domain (rScl2-CL) at pH 7 and 2.2

Protein	CD T_m		Fluorescence T_m		DLS T_m at pH 7	DSC T_m		ΔH_{cal}	
	pH 7	pH 2.2	pH 7	pH 2.2		pH 7	pH 2.2	pH 7	pH 2.2
	°C		°C		°C	°C		kJ/mol	
rScl2	35.6	24.2	36.5	26.5	36.4	37	24.2, 28	3499	2827 ^a
rScl2-CL	35.9	25.7	NA ^b	NA	34.4	36.1	23.7, 27	3400	2427 ^a

^a Sum of two peaks.

^b NA, no aromatic residues in the triple-helix domain.

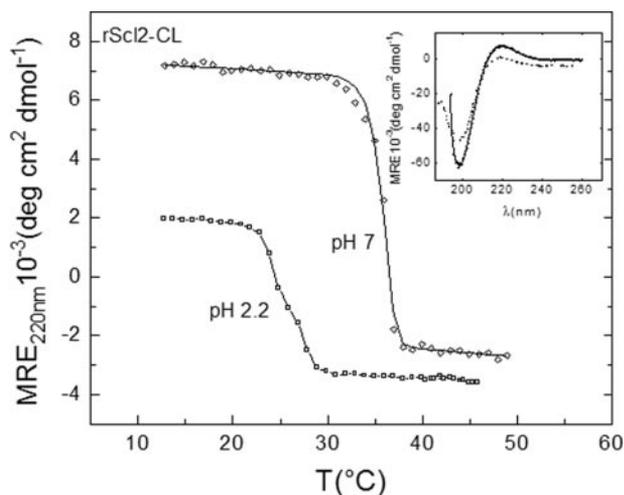


FIGURE 3. pH dependence of thermal stability monitored by CD for rScl2-CL at pH 7 and 2.2. Inset, its CD spectra; ●, pH 7; ◇, pH 2.2.

The CD spectrum of the rScl2 protein showed a significant drop in the 220 nm peak at pH 2.2 (data not shown) compared with pH 7 ($MRE_{220} = +1200 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ at pH 7 and $MRE_{220} = -1252 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ at pH 2.2). A dramatic decrease in MRE_{220} was seen for rScl2-CL, from $+7000 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ at pH 7 to $+2000$ at pH 2.2 (Fig. 3, inset). The decrease in the characteristic triple-helix 220 nm CD signal at pH 2.2 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 pH 2.2 showed a broad transition near 23–27 °C compared with the sharp 36 °C transition at pH 7 (Fig. 3). Although largely cooperative melting behavior was seen for the globular and triple-helix domains of rScl2 at a neutral pH value, in some cases, there was an indication that the globular domain unfolded slightly after unfolding of the triple helix. For instance, at pH 5, monitoring the 220 nm 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 220 nm signal likely represents denaturation of the α -helix in the globular domain subsequent to denaturation of the triple helix.

At pH 2.2, DSC scans of the rScl2 and rScl2-CL proteins showed two discrete transitions at 24 and 27 °C in contrast with the single DSC transition at 36 °C and pH 7 (Fig. 2 and 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 3400 kJ/mol for the single peak at pH 7 to 2427 kJ/mol for the sum of the two peaks at pH 2.2, suggesting that neu-

Stabilization of Bacterial Collagen

TABLE 2

Observed melting temperature values obtained by CD spectroscopy and predicted stability, together with calorimetric enthalpy, for the model peptides of the highly charged repeating portion of the collagenous region of Scl2

Peptide	Sequence	Observed T_m at pH 7	Predicted T_m at pH 7	ΔH_{cal} at pH 7	Observed T_m at acidic pH	Predicted T_m at acidic pH	ΔH_{cal} at acidic pH
		°C	°C	kJ/mol	°C	°C	kJ/mol
Scl2 repeat	Ac-(GPO) ₃ GKDGKDGQNGKDGPL(GPO) ₄ GY-CONH ₂	34.0	4.9	251	16.5 ^b	-11.4	119 ^b
GKDGKD	Ac-(GPO) ₃ GKDGKD(GPO) ₃ GY-CONH ₂	29.4	21.6	266 ^a	16.0 ^c	11	245 ^c

^a pH 8.5 and 6.

^b pH 1.4.

^c pH 2.1.

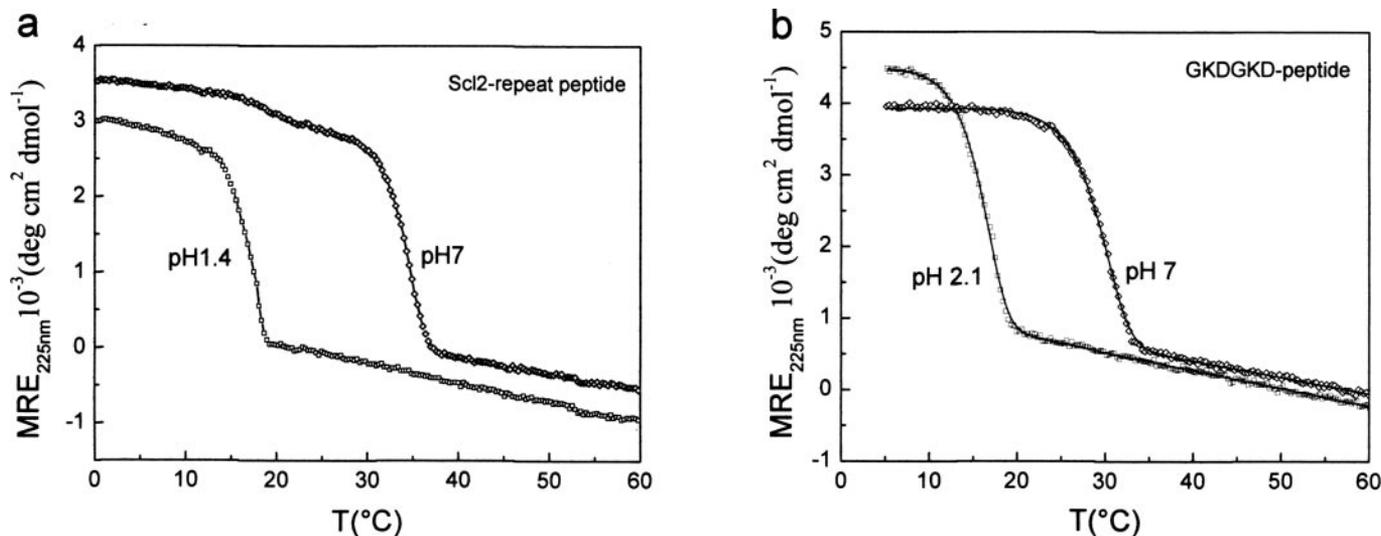


FIGURE 4. Thermal transitions monitored by CD spectroscopy of model peptides for the Scl2 C-terminal charged region. *a*, Scl2 repeat peptide at pH 7 and 1.4; *b*, GKDGKD peptide at pH 7 and 2.1.

tral pH promotes 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 pH 2.2 (Table 1). The higher of the two acidic transitions seen by DSC near 27 °C was observed, whereas the lower DSC transition near 24 °C was not. The Trp residue 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 that the more stable domain includes the N terminus of the (Gly-Xaa-X')₇₉ sequence.

Model Peptides for Regions of the 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 three full repeats of sequence GKDGKDGQNGKDGPL and several partial repeats of this sequence. One peptide included the 15-residue sequence Ac-(GPO)₃GKDGKDGQNGKDGPL(GPO)₄GY-CONH₂ (designated the Scl2 repeat peptide), whereas the other contained GKDGKD in a host-guest design, Ac-(GPO)₃GKDGKD(GPO)₃GY-CONH₂ (designated the GKDGKD peptide) (Table 2).

Using the T_m values obtained from guest tripeptide units in the host peptide (Gly-Pro-Hyp)₈, it was predicted that the Scl2 repeat peptide would have a stability of 4.9 °C (/jupiter.umd.edu/collagen_calculator/) (36). Surprisingly, the Scl2 repeat peptide formed a stable triple helix at neutral pH with $MRE_{225} = 3500$ deg·cm²·dmol⁻¹ and $T_m = 34.0$ °C (Fig. 4), demonstrating that the

triple-helix structure of the Scl2 repeat peptide is far more stable than expected. DSC indicated this peptide to have a calorimetric enthalpy of 251 kJ/mol at pH 7. At pH 1.4, the T_m decreased to 16.5 °C, and the calorimetric enthalpy decreased to 119 kJ/mol.

The GKDGKD peptide in the host-guest peptide design formed a triple helix at neutral pH with $MRE_{225} = 4000$ deg·cm²·dmol⁻¹. The observed T_m for the GKDGKD peptide was 29.4 °C, which is again greater than that predicted (predicted $T_m = 21.6$ °C) (36), with $\Delta H_{cal} = 266$ kJ/mol (Table 2). Measurements of thermal stability by CD at different pH values indicated that the T_m decreased below pH 4, with $T_m = 16.0$ °C at pH 2.1 and 15.1 °C at pH 1.4. The calorimetric enthalpy was not substantially lower at acidic pH, with $\Delta H_{cal} = 245$ kJ/mol at pH 2.1. The DSC scans showed higher T_m 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 *S. 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 T_m of BclA may relate to its high ~25% Pro

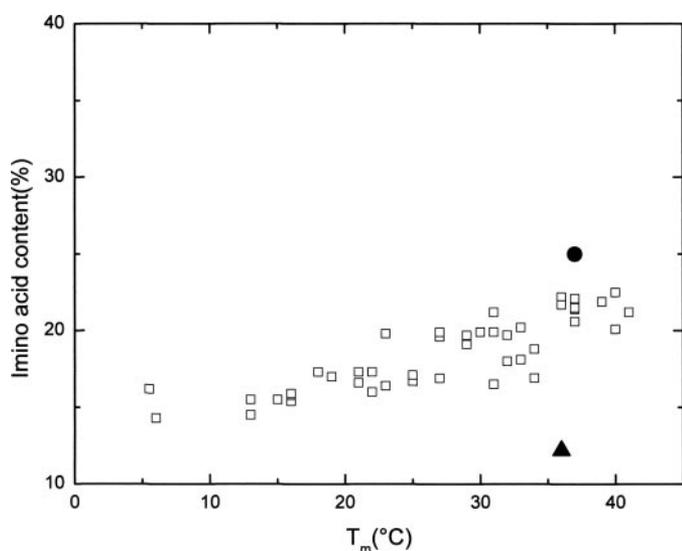


FIGURE 5. Plot of T_m values against imino acid content for a wide variety of animal collagens (\square), showing the bacterial collagen-like protein Scl2 (\blacktriangle) and the *B. anthracis* collagen-like protein BclA (\bullet). The values for the animal collagens were taken from Ref. 10.

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 $\sim 12\%$ 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 that vary in length from 40 to 120 tripeptide units have T_m values in the range of $34\text{--}37^\circ\text{C}$, indicating that there is no longer a length dependence.

The thermal transitions of Scl2 and its collagen domain are as sharp as those of 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 $\sim 20\%$ imino acids in animal fibril-forming collagens (Table 3). In the collagenous domain of Scl2, the Pro residues are found preferentially in the Xaa positions, and this seems to be true for all prokaryotic collagen-like proteins, with the frequency of Pro residues in the Xaa position exceeding 30% , whereas the frequency of Pro in the X'aa position is $<5\%$.³ The percent of hydrophobic residues is relatively low in both Scl2 (7.5%) and human fibril-forming collagens ($\sim 6\%$), 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 $\sim 15\%$ for fibrillar collagens (Table 3). Because all residues in the Xaa and X'aa 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-helix 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) $_8$, the propensities of all 20 amino acids in the Xaa position were evaluated as a guest Gly-X-Hyp triplet and in the X'aa position as a guest Gly-Pro-Y triplet, whereas 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.* GPKGEO. Examination of the (Gly-Xaa-X'aa) $_{79}$ sequence of the Scl2 collagen domain in light of these studies indicates a stable domain for tripeptide units 5–53, which contain 15% Pro content, hydrophobic stabilization of the form GLQGLQGLQ, and favorable charge interactions (*e.g.* KGD and KGE). The C-terminal end of the collagenous sequence of Scl2 (tripeptide units 54–79) is repetitive and highly charged, containing three repeats of GKDGKDGQNGKDGLP (40% charged residues) and several partial repeats. Other strains of *Streptococcus* have triple-helix domains ranging from 10 to 220 tripeptide units and many contain repeating sequences related to the repeat in Scl2 (25, 43). An algorithm based on host-guest peptides (jupiter.umdj.edu/collagen_calculator/) (36) predicts this highly charged repeating region to be very unstable and unable to form a triple helix. However, the peptide data reported here indicate that 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 that 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 GKDGKDGQNGKDGLP 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 1 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 ($\Delta T_m \sim 10\text{--}13^\circ\text{C}$) 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 at acid pH than at neutral pH, whereas the Scl2 repeat peptide and the GKDGKD peptide both show significant destabilization at pH 2.2 relative to pH 7. 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 from 35.8°C at neutral pH to 30.5°C

³ M. Pawlowski and J. M. Bujnicki, personal communication.

TABLE 3

Comparison of the amino acid features of rScl2-CL with human type II fibrillar collagen (P02458)

Collagen	Imino acid	Hydrophobic	Charged	Charged pairs
rScl2-CL (Gly-Xaa-X'aa) ₇₉	12% Pro: 20 Pro in Xaa position, 9 Pro in X'aa position	%	%	
Tripeptide units 5–53	15% Pro: 20 Pro in Xaa position, 2 Pro in X'aa position	7.5	30	16 GKD, 6 GER, 1 GEK
Tripeptide units 54–79	9% Pro: 0 Pro in Xaa position, 7 Pro in X'aa position	4.7	23	5 GER, 2 GKD, 1 GEK/GDR
Type II (Gly-Xaa-X'aa) ₃₃₈	21% imino acids: 103 Pro in Xaa position, 112 Hyp in X'aa position	6.4	40	14 GKD
		5.9	16	11 GER, 4 GKD, 4 GDR, 4 GEK

TABLE 4

Comparison of calorimetric enthalpy/tripeptide for collagens and model peptides

	ΔH_{cal} kJ/mol
Collagen	
Type I	20.1 ^a
Type II	22.5 ^b
Scl2-CL	12.5
Model peptide	
(POG) ₁₀	13.9 ^c
GKDGKD	10.2
Scl2 repeat (PPG) ₁₀	6.6 6

^a Ref. 10.^b Ref. 45.^c Ref. 35.

at pH 2.2 (37). This 5.3 °C decrease in the ΔT_m from pH 7 to 2.2 for GKD is consistent with the observed 17.5 °C decrease from neutral to pH 2.2 in the T_m of the Scl2 repeat peptide, which contains three GKD units ($3 \times -5.3 \text{ °C} = -15.9 \text{ °C}$), and the observed 13.4 °C decrease from neutral to pH 2.1 in the T_m of the GKDGKD peptide, which contains two GKD units ($2 \times -5.3 \text{ °C} = -10.6 \text{ °C}$). The double peaks seen for thermal transitions of both rScl2 and rScl2-CL upon DSC suggest that, under low pH conditions, there are two less stable triple-helix domains. Their lower stability (24 and 27 °C) is likely to reflect their shorter length as well as their high GKD content. The C-terminal repeating unit GKDGKDGQNGKDGLP is 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 that hydrogen bonding is a major contributing factor to stability. Privalov (10) observed that the enthalpy increased with increasing Hyp content and proposed that this is a consequence of an ordered water-mediated hydrogen bonding network. 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 of tripeptide (Table 4), which is less than the 18–22 kJ/mol of tripeptide seen for animal collagens near neutral pH (10, 45), but in the same range as seen for (Pro-Hyp-Gly)₁₀ (35). 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 the X'aa position as a major mechanism of modulating triple-helix stabilization, whereas prokaryotes have evolved different molecular strategies of triple-helix stabilization that differ not only from their eukaryotic counterparts, but also between different

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 T_m of (Gly-Pro-Pro)₁₄ is >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. 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 an 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. Because 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 that do self-assemble to higher order structures. The availability of triple-helix proteins with the 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

1. Exposito, J. Y., Cluzel, C., Garrone, R., and Lethias, C. (2002) *Anat. Rec.* **268**, 302–316
2. Kielty, C. M., and Grant, M. E. (2002) in *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Royce, P. M., and Steinmann, B., eds) 2nd Ed., pp. 159–222, Wiley-Liss, Inc., New York
3. Veit, G., Kobbe, B., Keene, D. R., Paulsson, M., Koch, M., and Wagener, R. (2006) *J. Biol. Chem.* **281**, 3494–3504
4. Kramer, J. M. (1994) *FASEB J.* **8**, 329–336
5. Rasmussen, M., Jacobsson, M., and Bjorck, L. (2003) *J. Biol. Chem.* **278**, 32313–32316
6. Brodsky, B., and Persikov, A. V. (2005) *Adv. Protein Chem.* **70**, 301–339
7. Ramachandran, G. N., and Kartha, G. (1955) *Nature* **176**, 593–595
8. Rich, A., and Crick, F. H. (1961) *J. Mol. Biol.* **3**, 483–506
9. Josse, J., and Harrington, W. F. (1964) *J. Mol. Biol.* **93**, 269–287
10. Privalov, P. L. (1982) *Adv. Protein Chem.* **35**, 1–104
11. Myllyharju, J., and Kivirikko, K. I. (2004) *Trends Genet.* **20**, 33–43
12. Berg, R. A., and Prockop, D. J. (1973) *Biochem. Biophys. Res. Commun.* **52**, 115–120
13. Jimenez, S., Harsch, M., and Rosenbloom, J. (1973) *Biochem. Biophys. Res.*

- Commun.* **52**, 106–114
14. Perret, S., Merle, C., Bernocco, S., Berland, P., Garrone, R., Hulmes, D. J., Theisen, M., and Ruggiero, F. (2001) *J. Biol. Chem.* **276**, 43693–43698
 15. Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1982) *Arch. Biochem. Biophys.* **219**, 198–203
 16. Rigby, B. J. (1968) *Nature* **219**, 166–167
 17. Burjanadze, T. V. (1982) *Biopolymers* **21**, 1489–1501
 18. Mighelsen, C., and Berendsen, H. J. C. (1973) *J. Chem. Phys.* **59**, 296–305
 19. Bella, J., Brodsky, B., and Berman, H. M. (1995) *Structure (Lond.)* **3**, 893–906
 20. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science* **266**, 75–81
 21. Jenkins, C. L., and Raines, R. T. (2002) *Nat. Prod. Rep.* **19**, 49–59
 22. Vitagliano, L., Berisio, R., Mazzarella, L., and Zagari, A. (2001) *Biopolymers* **58**, 459–464
 23. de Bruin, E. C., Werten, M. W., Laane, C., and de Wolf, F. A. (2002) *FEMS Yeast Res.* **1**, 291–298
 24. Boydston, J. A., Chen, P., Steichen, C. T., and Turnbough, C. L., Jr. (2005) *J. Bacteriol.* **187**, 5310–5317
 25. Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, M., and Lukowski, S. (2006) *Appl. Microbiol. Biotechnol.* **72**, 109–115
 26. Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., and Lukowski, S. (2002) *J. Biol. Chem.* **277**, 27312–27318
 27. Lukowski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G., and Musser, J. M. (2000) *Infect. Immun.* **68**, 6542–6553
 28. Lukowski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, E. A., and Musser, J. M. (2001) *Infect. Immun.* **69**, 1729–1738
 29. Rasmussen, M., and Bjorck, L. (2001) *Mol. Microbiol.* **40**, 1427–1438
 30. Rasmussen, M., Eden, A., and Bjorck, L. (2000) *Infect. Immun.* **68**, 6370–6377
 31. Whatmore, A. M. (2001) *Microbiology (Read.)* **147**, 419–429
 32. Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukowski, S., and Wary, K. K. (2005) *J. Biol. Chem.* **280**, 13848–13857
 33. Qing, G., Ma, L. C., Khorchid, A., Swapna, G. V., Mal, T. K., Takayama, M. M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G. T., Ikura, M., and Inouye, M. (2004) *Nat. Biotechnol.* **22**, 877–882
 34. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326
 35. Persikov, A. V., Xu, Y., and Brodsky, B. (2004) *Protein Sci.* **13**, 893–902
 36. Persikov, A. V., Ramshaw, J. A., and Brodsky, B. (2005) *J. Biol. Chem.* **280**, 19343–19349
 37. Chan, V. C., Ramshaw, J. A., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997) *J. Biol. Chem.* **272**, 31441–31446
 38. Brodsky-Doyle, B., Leonard, K. R., and Reid, K. B. (1976) *Biochem. J.* **159**, 279–286
 39. McAlinden, A., Smith, T. A., Sandell, L. J., Ficheux, D., Parry, D. A., and Hulmes, D. J. (2003) *J. Biol. Chem.* **278**, 42200–42207
 40. Bann, J. G., Peyton, D. H., and Bachinger, H. P. (2000) *FEBS Lett.* **473**, 237–240
 41. Daubenspeck, J. M., Zeng, H., Chen, P., Dong, S., Steichen, C. T., Krishna, N. R., Pritchard, D. G., and Turnbough, C. L., Jr. (2004) *J. Biol. Chem.* **279**, 30945–30953
 42. Suto, K., and Noda, H. (1974) *Biopolymers* **13**, 2477–2488
 43. Beres, S. B., Sylva, G. L., Sturdevant, D. E., Granville, C. N., Liu, M., Ricklefs, S. M., Whitney, A. R., Parkins, L. D., Hoe, N. P., Adams, G. J., Low, D. E., DeLeo, F. R., McGeer, A., and Musser, J. M. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11833–11838
 44. Hyashi, T., Curran-Patel, S., and Prockop, D. J. (1979) *Biochemistry* **18**, 4182–4187
 45. Morris, N. P., Watt, S. L., Davis, J. M., and Bachinger, H. P. (1990) *J. Biol. Chem.* **265**, 10081–10087
 46. Ackerman, M. S., Bhate, M., Shenoy, N., Beck, K., Ramshaw, J. A., and Brodsky, B. (1999) *J. Biol. Chem.* **274**, 7668–7673
 47. Bulleid, N. J., Dalley, J. A., and Lees, J. F. (1997) *EMBO J.* **16**, 6694–6701
 48. Xu, Y., Hyde, T., Wang, X., Bhate, M., Brodsky, B., and Baum, J. (2003) *Biochemistry* **42**, 8696–8703
 49. Kar, K., Amin, P., Bryan, M. A., Persikov, A. V., Mohs, A., Wang, Y.-H., and Brodsky, B. (2006) *J. Biol. Chem.* **281**, 33283–33290