Self-association of *Streptococcus pyogenes* collagen-like constructs into higher order structures

Ayumi Yoshizumi,[1†] Zhuoxin Yu,[1†] Teresita Silva,[1] Geetha Thiagarajan,[1] John A. M. Ramshaw,[2] Masayori Inouye,[1] and Barbara Brodsky[1]*

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

[2]CSIRO Molecular and Health Technologies, Bayview Avenue, Clayton South, Victoria 3169, Australia

Received 24 February 2009; Revised 1 April 2009; Accepted 2 April 2009
DOI: 10.1002/pro.134
Published online 16 April 2009 proteinscience.org

Abstract: A number of bacterial collagen-like proteins with Gly as every third residue and a high Pro content have been observed to form stable triple-helical structures despite the absence of hydroxyproline (Hyp). Here, the high yield cold-shock expression system is used to obtain purified recombinant collagen-like protein (V-CL) from *Streptococcus pyogenes* containing an N-terminal globular domain V followed by the collagen triple-helix domain CL and the modified construct with two tandem collagen domains V-CL-CL. Both constructs and their isolated collagenous domains form stable triple-helices characterized by very sharp thermal transitions at 35–37°C and by high values of calorimetric enthalpy. Procedures for the formation of collagen SLS crystallites lead to parallel arrays of in register V-CL-CL molecules, as well as centrosymmetric arrays of dimers joined at their globular domains. At neutral pH and high concentrations, the bacterial constructs all show a tendency towards aggregation. The isolated collagen domains, CL and CL-CL, form units of diameter 4–5 nm which bundle together and twist to make larger fibrillar structures. Thus, although this *S. pyogenes* collagen-like protein is a cell surface protein with no indication of participation in higher order structure, the triple-helix domain has the potential of forming fibrillar structures even in the absence of hydroxyproline. The formation of fibrils suggests bacterial collagen proteins may be useful for biomaterials and tissue engineering applications.

Keywords: collagen; *Streptococcus pyogenes*; cold shock vector; self-association; fibrils; solubility

Introduction

Collagens are the major structural proteins in the extracellular matrix of multicellular animals and are defined by their characteristic triple-helix structure and their unique $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeating sequence. Evolution has led to specialized types of collagen which have important biological as well as structural functions and a family of 28 collagen types is found in humans.[1,2] Collagens are present in all vertebrates and invertebrates,[3] and more recently, collagen-like sequences with Gly at every third residue and a high Pro content have been observed in bacteria as well.[4–6] Despite the absence of post-translational modification of Pro to hydroxyproline (Hyp), several bacterial collagens have been shown to form triple-helices which are

Abbreviations: CD, Circular dichroism; CL, collagen-like domain; CL-CL, two tandem repeats of the collagen-like unit; DLS, Dynamic Light Scattering; DSC, Differential Scanning Calorimetry; Hyp, hydroxyproline; IPTG, isopropyl β-D-thiogalactopyranoside; Rh, hydrodynamic radius; SLS, Segment long spacing collagen; V, the N-terminal globular domain

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

†These two authors contributed equally to this work.

Grant sponsor: NIH; Grant numbers: R21EB007198, GM60048.

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

stable near 37°C.[5,7,8] Such proteins are amenable to production in high yield bacterial expression systems, offering the opportunity for structural and biological characterization of modified as well as original versions.

Collagen-like sequences $(Gly\text{-}Xaa\text{-}Yaa)_n$ are reported in genomes of many bacteria[4] and collagen-like proteins have been shown to be expressed in *Bacillus anthracis,*[6] *Streptococcus equi,*[9] and *Streptococcus pyogenes*.[5,10,11] For instance, *S. pyogenes* produces collagen-like proteins Scl1 and Scl2 which are simultaneously expressed on the cell surface in the logarithmic phase and help the bacteria adhere to human cells.[5,10–15] Both Scl1 and Scl2 proteins contain a signal sequence, an N-terminal variable globular domain (V), a collagen-like triple-helix domain (CL) consisting of $(Gly\text{-}Xaa\text{-}Yaa)_n$ triplet repeats and a C-terminal gram-positive cell wall attachment domain.[5] The Scl2 collagen-like domain contains a high proportion of charged residues (~30% of all residues) and the number of Gly-Xaa-Yaa triplets in the collagen-like domain varies in different serotypes, ranging from $n = 14$ to 79.[11,16] Biochemical characterization of *S. pyogenes* collagen-like proteins Scl1 and Scl2 confirmed that in the absence of any post-translational modifications these proteins form a stable triple-helix with $T_m = 36–38$°C[16] close to the stability found in the longer human collagens. The stabilization of Scl2 has been shown to be largely due to electrostatic interactions.[7] These collagen-like proteins also mimic some biological activities of collagen in terms of binding to the mammalian α11β1 and α2β1 integrin collagen receptors.[13–15] The Scl proteins may contribute to *S. pyogenes* binding and internalization in human cells through their similarity to mammalian collagens in terms of structure and binding to human tissues.[11,13–15]

Self-assembly to higher order structures is a key aspect of collagen biology in animals. The supramolecular forms of collagen provide mechanical support to the extracellular matrix and mediate cell attachment and other biological processes.[1] The most abundant collagens are found in fibrils with an axial periodicity of 67 nm arising from the staggering of adjacent molecules due to electrostatic interactions, hydrophobic interactions and hydration forces.[17,18] In addition, supramolecular structures are formed by some non-fibrillar collagens, such as the network structures of type IV collagen in basement membranes and the anti-parallel arrays of type VII collagen in anchoring fibrils attached to the basal lamina of the skin.[1] It is not known whether collagen-like proteins in bacteria self-associate to form any higher order structures *in vivo*, or whether they have the capacity to do so *in vitro*.

Here molecular and self-association studies are reported for protein constructs based on the *S. pyogenes* Scl2 collagen-like domain. The *S. pyogenes* collagen-like proteins Scl1 and Scl2 were first expressed in *E. coli* using a strep-tag II for purification,[5] and more recently Mohs *et al.* has achieved expression of Scl2 protein in a cold-shock system[7] which directs highly specific induction of the target gene and produces very high yields at low temperatures.[19] The Scl2 collagen CL domain of 79 Gly-Xaa-Yaa triplets is much shorter than the $(Gly\text{-}Xaa\text{-}Yaa)_{338}$ domain of human type I collagen, so a construct containing two tandem collagen domains CL-CL was also constructed. The V-CL and V-CL-CL constructs of the Scl2 protein were expressed in high yields sufficient for studies of solubility and self-association. Results show the proteins form stable triple-helical structures and that the triple-helical domains CL and CL-CL can self-assemble to form twisted bundles of units 4–5 nm in diameter. The ability of bacterial collagen constructs to form fibrils at neutral pH supports potential biomaterials and tissue engineering applications.

Results

Production of bacterial collagen constructs: V-CL, V-CL-CL, CL, CL-CL

The high expression cold-shock vector system was used to obtain the region of the Scl2.28 collagen-like protein of *S. pyogenes* which includes an N-terminal globular domain (V) and a triple-helical domain of sequence $(Gly\text{-}Xaa\text{-}Yaa)_{79}$ (CL) [Fig. 1(A)]. The construct was designed to have an N-terminal His_6 tag for purification and a protease susceptible sequence LVPRGSP between the V and CL domains; this construct is referred to as V-CL. The sequence encoding V-CL was inserted into the cold-shock vector, pCold III. The plasmid was transformed into *E. coli* strain BL21, and V-CL was expressed at 15°C, 24°C, and 37°C; after overnight incubation, expression levels were analyzed by SDS-PAGE. As shown in Figure 1(B), the expression level was highest at 24°C, compared with 15°C and 37°C. A construct was also designed which contains the V domain followed by two repeats of the CL $(Gly\text{-}Xaa\text{-}Yaa)_n$ unit, again including an N-terminal His tag and a short linker sequence susceptible to enzymatic cleavage between the V domain and the first CL domain. This protein with two consecutive CL domains is designated V-CL-CL. The DNA sequence for V-CL-CL was inserted into the cold-shock vector, expressed and purified following the procedure described above for the V-CL protein. The expression level was again highest at 24°C [Fig. 1(C)].

A large scale purification of all constructs was carried out at room temperature. Notably, V-CL and V-CL-CL were expressed as soluble proteins without any protein in the insoluble fraction. Subsequently V-CL and V-CL-CL were purified on a Ni-sepharose column and the eluted protein (yield ~ 0.4 g/L) was detected as a single band on SDS-PAGE near the expected molecular weight position [Fig. 1(D)]. Mass spectroscopy of the V-CL product showed that it has a mass of 33389, which is comparable to the theoretical value

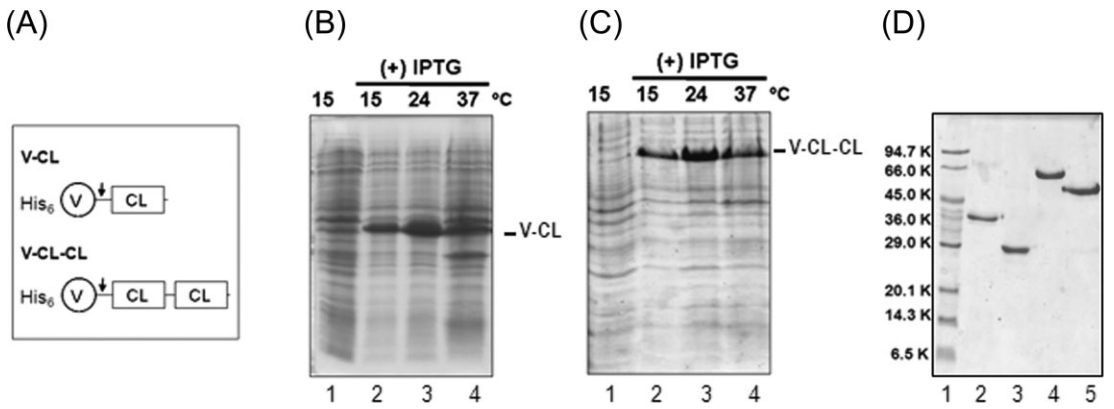

Figure 1. (A) Schematic of the design of the bacterial collagen Scl2.28 construct, showing the V-CL and V-CL-CL products; His $_6$, histidine tag (NH_2- HHHHHH-COOH); V, globular domain; ↓, enzyme cleavage site, LVPR↓GSP; CL, collagen-like domain $(Gly\text{-}Xaa\text{-}Yaa)_{78}$; The triplet sequence GKY is included at the C terminal for protein concentration determination. The linker sequence between two CL domains is GAAGVM. (B) SDS-PAGE of cells lysed after expression of V-CL. (C) SDS-PAGE of cells lysed after expression of V-CL-CL. For (B) and (C), the first lane is a control sample without 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) induction at 15°C while the lane 2, 3, and 4 are for samples with IPTG induction at 15°C, 24°C and 37°C, respectively. (D) SDS-PAGE of the purified proteins for (1) Molecular weight marker, (2) V-CL, (3) CL, (4) V-CL-CL and (5) CL-CL.

without the amino terminal methionine (33400). Purified V-CL-CL (yield ~ 0.1g/L) showed an SDS-PAGE band around 56 kDa as expected [Fig. 1(D)], and mass spectroscopy gave a mass of 55,800, which is comparable to the theoretical value, 55,701.

The CL domains were obtained from the intact proteins based on the known resistance of native collagen triple-helix to digestion by trypsin. The V-CL product was treated with trypsin at room temperature for 4 h [Fig. 2(A)], and the resulting protein was purified on a DEAE column, followed by gel filtration column chromatography. Mass spectrometry analysis of the purified CL protein was 22,850, in good agreement with the predicted molecular mass of 22,840. On SDS-PAGE, the CL band was higher than expected, which is common for rod-like triple-helix proteins [Fig. 1(D)]. The dimer collagen domain, CL-CL, was also isolated using trypsin [Fig. 2(B) and purified as described earlier [Fig. 1(D)]. Mass spectrometry of CL-CL showed 44,938, which is in good agreement with the expected value of 44,998. The ability to obtain large amounts of purified bacterial collagen protein, its collagen-like domain, and the longer construct with two collagen repeats allows a range of studies on the physical properties and self-association of these molecules.

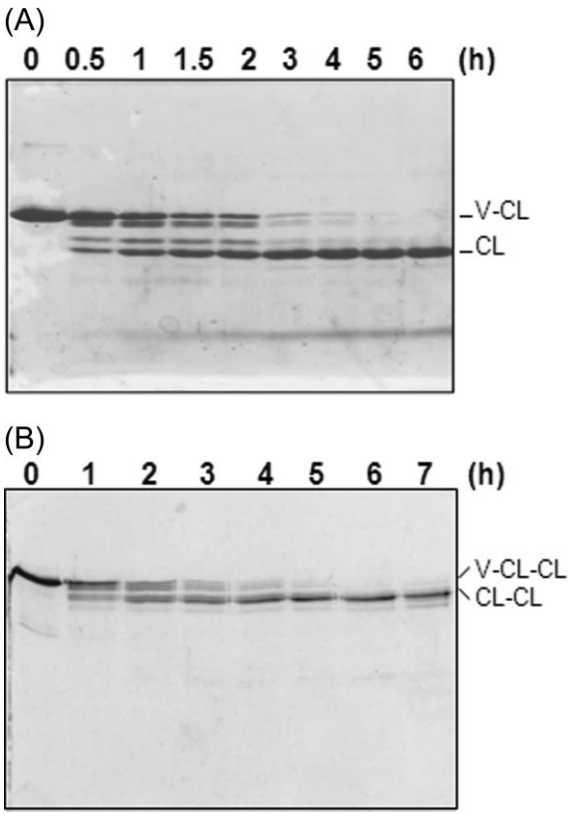

Figure 2. SDS-PAGE monitoring of the time course of the digestion of (A) V-CL and (B) V-CL-CL by trypsin at room temperature for different lengths of time in hours. At early time points, trypsin digestion of the V-CL construct produces a few intermediate products, representing the incomplete cleavage of the V domain. Trypsin digestion of the V-CL-CL construct may also produce intermediate products, but such products would not be seen due to the lower resolution of the gel at the higher molecular weight. After 4 h, complete digestion at the LVPR↓GSP occurs, as shown by mass spectroscopy.

Structural characterization of V-CL, CL, V-CL-CL and CL-CL molecules

To investigate conformational features, circular dichroism spectroscopy (CD) was carried out on the intact V-CL and V-CL-CL proteins and their isolated collagenous domains. The CD spectra of both V-CL and V-CL-CL proteins in PBS buffer at pH 7 show typical collagen features, with a maximum at 220 nm and a

Table I. *Physical Properties of Bacterial Collagen in PBS (pH 7) and Bovine Skin Collagen (pH 3)*

	DLS	CD				DSC		
	Rh[a] (nm)	T_m (°C)	MRE_{198} (deg cm^2 dmol^{-1})	MRE_{220} (deg cm^2 dmol^{-1})	Rpn[b]	T_m (°C)	ΔH_{cal} (kJ/mol)	ΔH_{cal} / tripeptide (kJ/mol)
V-CL	10.2 ± 1.1	36.8	−38,400	1500	0.04	37.9	2730	ND
CL	8.0 ± 1.6	35.2	−62,500	8000	0.13	37.0	2820	11.8
V-CL-CL	49.6 ± 22.4	37.1	−54,500	5400	0.10	37.5	4290	ND
CL-CL	17.0 ± 1.1	36.5	−53,500	6100	0.11	37.9	5300	11.2
BSC[c]	25.0[d]	37.5	−50,000[e]	5261	0.11	37.0	5187	15.4

[a] The hydrodynamic radius obtained by dynamic light scattering.
[b] Rpn is the ratio of the positive peak at 220nm to the negative peak at 198nm.
[c] Pepsin treated-bovine skin collagen type I in 0.5 M acetic acid, since it precipitates in PBS. (Karunakar Kar, unpublished data).
[d] Ref. 20.
[e] Ref. 21.

minimum at 198 nm (Table I). When the purified CL and CL-CL domains are examined, the CD peaks are at similar locations but with much higher intensities. Accurate concentration values could not be determined for the isolated collagen domains which have no aromatic residues, but concentration estimates by weight lead to an estimated MRE_{220} ~8000 deg cm^2 dmol^{-1} (Table I), a value even higher than seen for animal collagens. The parameter Rpn, which is the ratio of the intensity of the positive peak near 220 nm over the intensity of the negative peak near 198 nm, has been shown to be a useful measure of the collagen triple-helix conformation.[22] The Rpn values for the isolated collagen domains (Rpn = 0.13 for CL and Rpn = 0.11 for CL-CL) are similar to the Rpn value observed for animal collagens, indicating the collagenous domains form a fully folded triple-helix (Table I).

The thermal stabilities of V-CL, V-CL-CL and the isolated CL and CL-CL domains were examined by monitoring the change in the CD peak at 220 nm with increasing temperature (Fig. 3). Very sharp thermal transitions are observed for all samples, with similar values of T_m = 36.8°C for V-CL, T_m = 37.1°C for V-CL-CL, T_m = 35.2°C for CL and T_m = 36.5°C for CL-CL. The increased length of the collagenous domain appears to cause only a slight increase in thermal stability. The presence of a single sharp transition for V-CL and V-CL-CL just slightly higher than the T_m values of CL and CL-CL indicates the V domain confers little stabilization and that it unfolds simultaneously with the collagenous domain under these conditions. Very similar thermal transitions at 37.0–37.9°C are seen by differential scanning calorimetry (DSC), with a calorimetric enthalpy of ΔH_{cal} = 2730 kJ/mol for V-CL, ΔH_{cal} = 4290 kJ/mol for V-CL-CL, ΔH_{cal} = 2820 kJ/mol for CL, and ΔH_{cal} = 5300 kJ/mol for CL-CL (Fig. 4; Table I). As expected, the enthalpy for the CL-CL domain is almost twice as great as that for the single CL domain. Calculation of the calorimetric enthalpy per tripeptide unit for the CL and CL-CL proteins gives 11–12 kJ/mol tripeptide (Table I), values somewhat less than seen for collagens[23] (e.g. 15.4 kJ/mol tripeptide for calf skin collagen). Based on previous analyses and the water mediated hydrogen bonding of charged side chains seen in crystal structures, it is likely that the high enthalpy values are due to hydrogen bonding involving an extensive hydration network.[23–25]

The ability of the proteins to refold after heat denaturation was also examined, by monitoring the CD signal at 220 nm as the sample is cooled from 70°C to 0°C at the same rate as the heating rate (Fig. 3, see reverse arrows). A substantial amount of their original CD signal is regained by V-CL (60%) and V-CL-CL (40%) during this process, indicating these

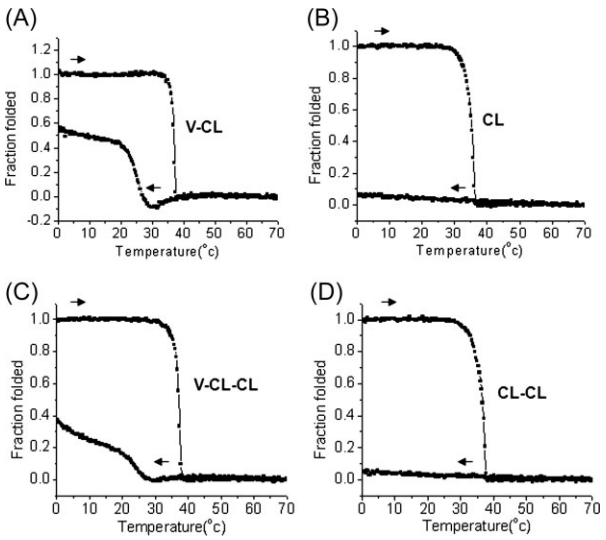

Figure 3. Thermal transitions of the Scl2 constructs determined by monitoring CD signal at 220 nm for (A) V-CL, (B) CL , (C) V-CL-CL and (C) CL-CL. In each figure, the arrow indicates the direction of temperature change with → for the unfolding curve with increasing temperature and ← for the refolding curve with decreasing temperature. The concentrations of all samples were 1 mg/mL and heating rates were ~0.1°C/min in both directions.

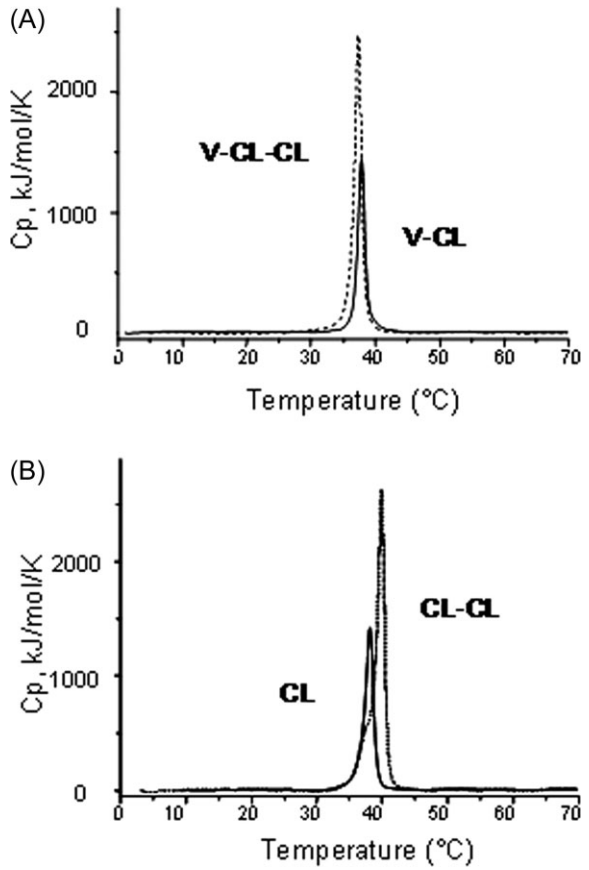

Figure 4. Differential scanning calorimetry (DSC) of (A) V-CL and V-CL-CL and (B) CL and CL-CL at pH7.0 in PBS, showing heat capacity (C*p*) as a function of temperature. The concentrations of all samples were 1 mg/mL with a heating rate of 1°C /min. A small shoulder is seen in the DSC of CL-CL that is likely due to some heterogeneity in the digestion products.

molecules are capable of refolding. Longer incubation at 0°C following the slow refolding did not increase the signal, but a temperature jump from 70°C to 0°C over a 1–2 min period led to almost full recovery of the original triple-helix signal. The initial drop in the 220 nm signal for V-CL and V-CL-CL upon cooling (see Fig. 3) likely represents refolding of the V globular domain prior to triple-helix folding. Sequence analysis suggests this V domain contains coiled coil α-helix structure,[5,7] and formation of α-helix would result in a drop in intensity near 220 nm. No signal is regained upon cooling of the CL and CL-CL proteins. These results indicate that the globular V domain is essential for refolding of these isolated collagen domains.

Self association of bacterial collagen products

Since the solubility of animal collagens is related to their self-association to form fibrils, solubility studies were carried out to compare bacterial collagen-like proteins with pepsin extracted bovine skin collagen at different pH and temperature values (Table II). The V-CL, CL, V-CL-CL, and CL-CL proteins are completely soluble in 0.1*M* acetic acid at 4°C, similar to the high solubility in acetic acid observed for bovine skin collagen and other animal collagens. In PBS in the cold (1 mg/mL), both V-CL and V-CL-CL are partially soluble, in contrast to the very high solubility of the isolated collagenous CL and CL-CL domains, suggesting that some association may be mediated by the globular V domains under these conditions. But the isolated CL and CL-CL domains become less soluble as the concentration is increased to 2 mg/mL and as the temperature is increased to 24°C. Bovine collagen is highly soluble in PBS (pH 7.0) at 4°C and 24°C, but shows a sharp decrease in solubility at higher temperatures due to fibril formation. Bovine collagen is insoluble near its isoelectric point (pI ~9) in Glycine buffer (pH 8.6) at 4°C and 24°C, while the bacterial collagens (pI ~ 5.5–6.1) are highly soluble at this pH value.

Segment long spacing crystallites (SLS), a lateral aggregate of parallel collagen molecules in register, have been used to visualize the molecular length and charge distribution of animal collagens.[26–28] Dialysis of bovine skin collagen against 0.2% ATP at pH 3, the procedure typically used to form SLS,[26–28] yields crystallites 280 nm in length with a clear set of many discrete positively stained bands which correspond to the positions of charged residues [Fig. 5(A)]. This protocol was applied to bacterial collagen protein products. The isolated collagen domain CL formed a network like

Table II. *Solubility of Bacterial Collagen Proteins and Bovine Skin Collagen in Different Buffers and pH Values (1 mg/mL)*

	0.1*M* acetic acid (pH 2.9)[a]	50 m*M* sodium acetate (pH 5.0)		PBS (pH 7.0)			50 m*M* glycine NaOH (pH 8.6)	
Protein	4°C (%)	4°C (%)	24°C (%)	4°C (%)	24°C (%)	30°C (%)	4°C (%)	24°C (%)
V-CL	98.1	76.5	70.6	66.0	69.6	ND	100	100
CL	100	83.6	84.9	100	80.0	ND	100	87.5
V-CL-CL	100	79.4	84.4	80.0	92.7	ND	75.0	82.3
CL-CL	100	ND	ND	100	63.6	65.6	ND	ND
BSC[b]	100	65.9	67.5	98.1	93.7	54.0	0	0

[a] The denatured temperature of V-CL and CL in acid pH is close to 24°C, so solubility studies were not carried out in acid pH at 24°C.[7]
[b] Pepsin treated bovine skin collagen type I.

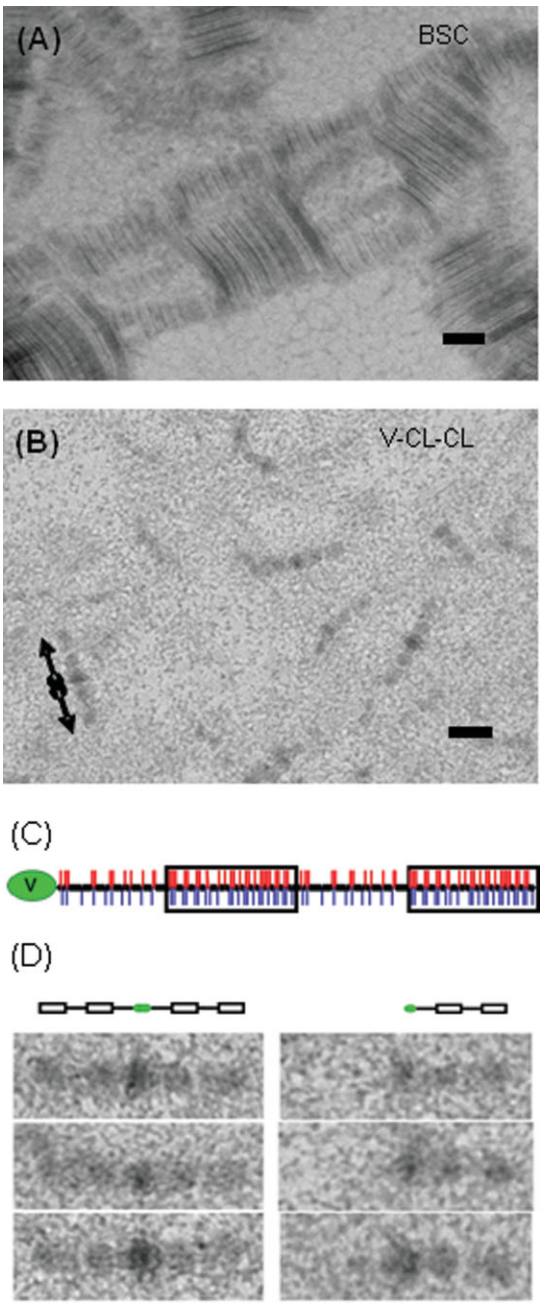

Figure 5. Electron micrographs of samples prepared for Segment Long Spacing collagen (0.2% ATP, at pH 3.0). (A) pepsin-treated bovine skin type I collagen and (B) V-CL-CL, with samples positively stained with uranyl acetate and phosphotungstic acid. The bar indicates 100 nm. (C) Schematic illustration of the location of positive charges (top bars) and negative charges (bottom bars) along the V-CL-CL chain, showing the location of a very high charge concentration (boxed) within each CL unit and the globular V domain. (D). Panel of SLS patterns of parallel, in register array of dimeric head-to-head V-CL-CL molecules (left) and monomer V-CL-CL molecules (right), with a schematic of the location of the high charged density regions shown above. On the left are head-to-head V domains giving a centrosymmetric pattern with the two CL domains on either side; on the right, the SLS patterns corresponding to a single V-CL-CL domain are shown.

structure of thin filaments with no discrete entities, while preparations of V-CL showed numerous narrow aggregates with a very heavily stained central band (Supporting Information Fig. 1). The SLS preparations of V-CL-CL show discrete aggregates with lengths clustered around 149 ± 18 and 277 ± 36 nm [Fig. 5(B)]. Given the expected dimensions of the CL-CL domain (140 nm) and the V domain (reported as 3.9 nm[5]), it is likely that the 150 nm crystallites are formed by an array of parallel single molecules of V-CL-CL while the centrosymmetric 280 nm crystallites could be formed by two antiparallel molecules of V-CL-CL joined at their globular domains (central dark band). Alignment of the V-CL-CL aggregates with the distribution of charged residues in the collagen-like CL domain is consistent with this interpretation and shows that the broad bands of the aggregates line up with an extremely high density of basic and acidic amino acids at the C-terminus of each CL repeat [Fig. 5(C,D)]. It is interesting to find that proteins with the collagen triple-helix conformation which lack Hyp and have such a high charged concentration and low isoelectric point have the capacity to form SLS assemblies.

Self-association of the expressed bacterial collagen proteins at neutral pH was studied by dynamic light scattering (DLS) and electron microscopy. DLS studies were carried out on all samples to monitor the size distribution profile and aggregation in PBS, pH 7, at low temperature (Table I). At 4°C, V-CL and the CL domain have hydrodynamic radii (R_h) of 10.2 ± 1.1 nm and 8.0 ± 1.6 nm, respectively, in good agreement with previous studies on a closely related construct.[7] The hydrodynamic radius of CL-CL is R_h = 17 ± 1.1 nm, consistent with a molecule that is twice as long as the CL trimer molecule. In contrast, DLS studies of V-CL-CL gave R_h = 49.6 ± 22.4 nm, which is higher than expected for a single trimeric V-CL-CL molecule and suggests the presence of some aggregated form. Since aggregation is not seen for CL-CL constructs, it is likely that the self-association is promoted by intermolecular interactions between the globular V domains.

At concentrations of 2 mg/mL, bacterial collagen products show some precipitation in PBS at 4°C and these turbid solutions were examined by transmission electron microscopy (Fig. 6). V-CL precipitates appear as poorly ordered densely stained clusters while V-CL-CL appear to have some more regular fibrillar nature embedded in disordered aggregates. In contrast, the CL and CL-CL precipitates show the presence of fibrillar structures. There appear to be discrete aggregated units which are in register arrays of CL-CL molecules, since their length corresponds to the molecular length (~140 nm) and their diameter (4–5 nm) indicates more than one molecule. These units are bundled to form wider and longer fibrillar structures which appear twisted and staggered. Some subunit structure

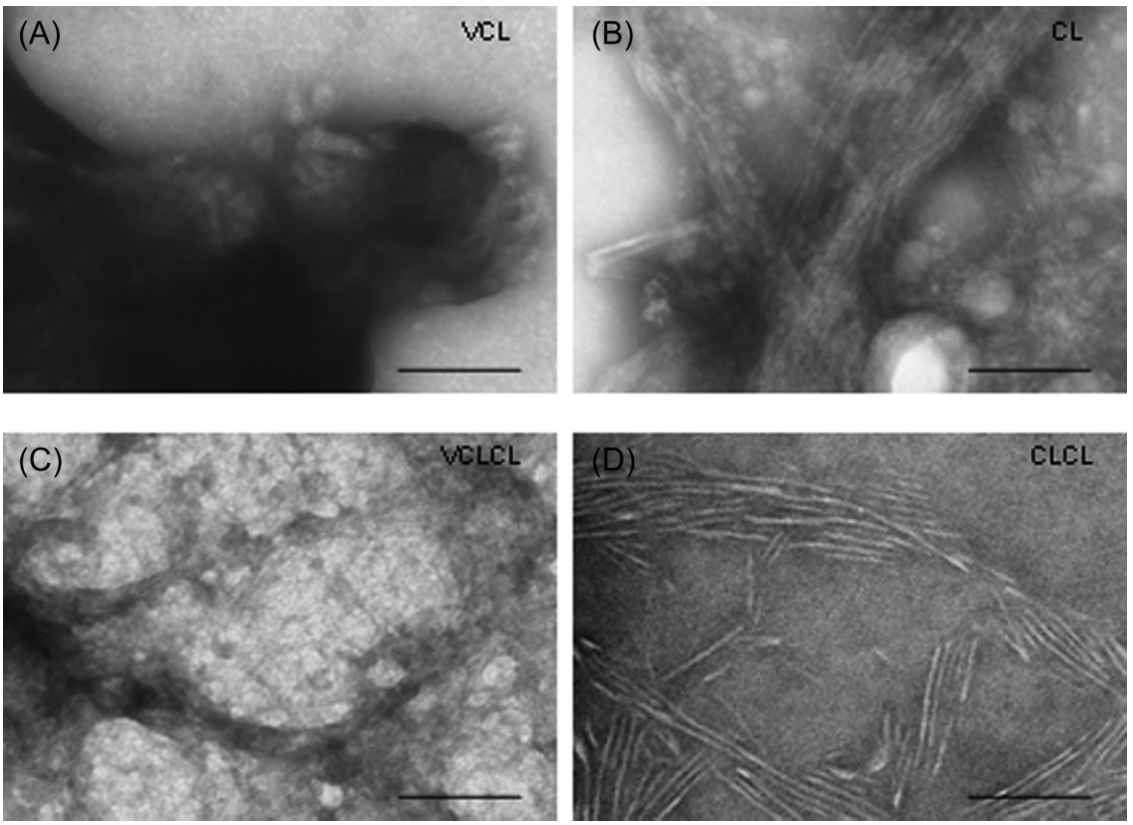

Figure 6. Electron micrographs of precipitates of (A) V-CL, (B) CL, (C) V-CL-CL, and (D) CL-CL at neutral pH. The precipitates were obtained from 2 mg/mL protein solutions after dialyzing against PBS buffer at 4°C and centrifuged at 14, 000 rpm for 10 min. All samples were negatively stained. The bar indicates 100 nm.

and supercoiling of units is also seen in the CL aggregates.

Discussion

Increasing concerns over the risk of contamination by infectious agents as well as challenges in product standardization have led to the active development of expression systems for recombinant collagens. Recombinant human collagens have been produced in animal cells, mammary glands of transgenic mice, silkworms, insect cells, and in tobacco plants.[29,30] Commercial production of full length hydroxylated collagens has been achieved using the yeast *Pichia pastoris,* where there is co-expression of the prolyl 4-hydroxylase genes with the human collagen gene.[31] An alternative strategy is to express a bacterial collagen-like protein which is not dependent on the presence of hydroxyproline for stability. The bacterial collagen-like protein construct V-CL and its modified V-CL-CL construct based on the *Streptococcus pyogenes* Scl2 protein have been successfully expressed in *E. coli* and shown to form triple-helices of stability $T_m \sim 35–37$°C. A His tag at the N-terminus allows batch affinity purification on a Nickel column. The Arg-containing sequence inserted between the globular domain and the triple-helix is highly susceptible to trypsin cleavage, leading to a homogeous triple-helix collagenous domain and elimination of the heterogeneity reported previously.[7] The ability to obtain large yields of bacterial collagen protein is due to its expression at temperatures well below 37°C in the cold shock system, since the unfolded monomer is susceptible to non-specific enzyme digestion while the folded trimer is highly enzyme resistant. Expression levels of ~800 mg/L were obtained for V-CL, compared to 0.6–20 mg/L yields in mammalian cells, 40 mg/L for insect cells, and 1 g/L yields in the Pichia system.[29] The bacterial system, with no requirement for prolyl hydroxylase, may prove easier to manipulate and to scale up to commercial production than other currently used recombinant collagen systems.

The triple-helix stability of these S. *pyogenes* constructs is achieved without the presence of hydroxyproline. Previous reports of decreased stability at acid pH and studies on model peptides indicate that a major compensation for the lack of hydroxyproline is stabilizing electrostatic interactions involving the many repeating highly charged $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequences.[7] The high calorimetric enthalpy observed for both CL and CL-CL collagenous domains indicates a high degree of hydrogen bonding which is likely to be mediated by hydration,[23,24] suggesting that the characteristic high enthalpy values of collagen and its extensive hydration network may be present even in the absence

of Hyp residues. Our studies show that the creation of a longer triple-helix domain (CL-CL) has little effect on thermal stability. Triple-helix stability of peptides is known to depend on peptide length until a minimum size is reached,[32] and the lengths of the Scl2 constructs and variants reported here and by Han *et al.*[16] appear to be longer than that minimum size.

The SLS parallel array of collagen molecules has been a useful tool in analyzing the charge distribution in vertebrate collagens, in clarifying the molecular packing in fibrils, and in characterizing the length and charges in invertebrate collagens.[26–28] Application of this preparation procedure to bacterial collagen products suggests that V-CL-CL proteins form SLS crystallites of the length expected for monomer and for dimers connected through their globular domains, similar to dimeric SLS forms that have been reported for animal collagens.[33] The few broad bands seen in these patterns are consistent with the very high density of charges in the collagen domain.

Studies reported here show Hyp-free bacterial collagen-like products can self-associate to form higher order structures at neutral pH. Molecules with the globular domain V-CL and V-CL-CL have less ordered aggregates while the isolated triple-helix domains, CL and CL-CL, both form supramolecular structures with a clearly fibrillar nature, appearing as bundles of some smaller basic unit. The effect of tandem repeats of the collagenous domain in CL-CL was not clear, although the fibrils had a better defined appearance. Increased length of collagen model peptides does lead to an increased propensity for aggregation,[34] and tandem repeats of a type II collagen segment were shown in the Fertala laboratory to exhibit some differences in terms of stability, the capacity to form fibrils with the axial D period, and their ability to promote chondrocyte growth and migration.[35,36]

The forces involved in directing collagen fibrillogenesis include hydration forces,[18] electrostatic interactions, hydrophobic interactions[17] and Hyp-mediated interactions.[25,37] Recent studies on collagen model peptides suggest non-specific lateral assembly is mediated by hydration and Hyp-Hyp interactions,[34] while more specific axial interactions and defined morphology are determined by charged and hydrophobic residues.[38,39] It is noteworthy that bacterial collagen forms fibrillar structures in the absence of Hyp. This ability for fibril formation without Hyp was also reported for unhydroxylated human collagen expressed in transgenic tobacco plants under low ionic strength conditions.[40] Electrostatic interactions are likely to play an important role in triple-helix self-association for these bacterial collagen modules and the hydration network suggested by its high enthalpy values may also be important for self-assembly into fibrillar structures. The process of self-association by CL-CL domains shows some temperature dependence, with more aggregation at 24°C compared with 0°C, but the dramatic acceleration of animal collagen fibrillogenesis near the collagen molecular T_m value is not seen, possibly because it is mediated by interactions involving Hyp.

The *S. pyogenes* Scl2 is a cell surface protein with no indication of participation in higher order structures, yet the Hyp-deficient collagen-like domain is itself capable of self-assembly to form fibrils. The results suggest some intrinsic propensity of collagen type molecules to self-associate. The fibrils formed are small in diameter compared with those found in human tissues and do not show the characteristic regular axial banding pattern seen in animal collagen fibrils. It is possible that periodic, larger collagen fibrils more suitable for biomaterial and tissue engineering applications can be achieved through manipulation of the bacterial collagen sequence, by selectively distributing the charged residues, further increasing the length of the collagen domain, or adding other matrix components. The formation of fibrils by the bacterial collagen-like proteins as reported here is an encouraging sign that these proteins may be useful for biomaterials and tissue engineering applications.

Materials and Methods

Construction of pColdIII-V-CL and pColdIII- V-CL-CL

pColdIII- V-CL was constructed using pColdIII-163 encoding the p163 polypeptide based on Scl 2.28.[5,7] The His_6 tag sequence was introduced at the N-terminal end of the P163 polypeptide sequence described in Xu *et al.*[5] A LVPRGSP sequence was inserted between the N-terminal globular domain (V) and collagen-like domain (CL) sequences by PCR, to act as a potential cleavage site for thrombin or trypsin. The cleavage by thrombin was much slower and required much more protease than found for trypsin, so all preparations of collagen domains were done using trypsin. pColdIII-V-CL-CL was constructed based on an N-terminal V domain followed by two $(Gly\text{-}X\text{-}Y)_{79}$ units with two additional triplets GAAGVM between two CL domains. All constructed plasmids were confirmed by DNA sequencing.

Expression analysis of V-CL and V-CL-CL

pColdIII-V-CL and V-CL-CL were expressed in *E. coli* BL21 strain. Cells were grown in 1.5 mL M9-Casamino Acid with ampicillin (50 μg/mL) medium at 37°C until A_{600} reached 0.8. The culture was split into three equal portions (0.5 mL each), which were then incubated at 15°C, 24°C and 37°C, respectively and 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce protein expression. After overnight incubation, cells were harvested by centrifugation. Subsequently, the cell pellets were analyzed by SDS-PAGE.

Large-scale preparation of V-CL and V-CL-CL

A single colony of *E. coli* BL21 cells harboring pCold III- V-CL and pCold III- V-CL-CL was innoculated into 10 mL of M9-Casamino Acid medium with ampicillin (50 μg/mL) at 37°C for 12 h. The cultures were then transferred into 1 L of M9-Casamino Acid medium with ampicillin (50 μg/mL) and incubated at 37°C until A_{600} reached 1.2. Cells were harvested by centrifugation and resuspended in 1 L by 2× concentrated L broth with ampicillin (50 μg/mL). Each culture was shifted to room temperature and 1 m*M* IPTG was added to induce protein expression. After overnight expression, cells were harvested by centrifugation and disrupted by a French press. Cellular debris was removed by centrifugation at 4°C. Both V-CL and V-CL-CL were found in the supernatant fraction as soluble protein. The supernatant fraction was loaded onto a Ni-sepharose resin column (75 mL bed volume; GE Healthcare) and equilibrated with the binding buffer [20 m*M* phosphate buffer (pH7.4) containing 500 m*M* NaCL and 20 m*M* imidazole] at room temperature.

After washing with the binding buffer, each protein was eluted using the eluting buffers (the binding buffer plus 58, 96, 115, and 400 m*M* imidazole) in a stepwise manner. Both V-CL and V-CL-CL eluted at 96–115 m*M* imidazole. Their purity was checked by SDS-PAGE and the concentration was determined using an extinction coefficient of $\varepsilon_{275} = 9635\ M^{-1}\ cm^{-1}$.[41]

Purification of CL and CL-CL domain

Ten microgram of purified V-CL was dialysed against 50 m*M* glycine Buffer (pH 8.6), and then digested with 100 μg trypsin at room temperature. The digested products were loaded onto a DEAE Sephadex anion exchange column at room temperature. The fractions containing CL domain as determined by SDS-PAGE were further purified using a Superdex^{TM} 200 gel filtration column (GE Healthcare). Protein purity was checked by SDS-PAGE and MALDI-TOF mass spectrometry. The tandem collagen domain CL-CL was prepared using the same method as described for the CL domain. To confirm the triple helix nature of CL and CL-CL, trypsin digestion was carried out for 1 h at 25°C and 37°C. Protein concentrations of CL and CL-CL were determined using circular dichroism spectra, assuming a mean residue ellipticity of $MRE_{220} = 6000$ deg cm^2 dmol^{-1}.[21]

Circular dichroism (CD) spectroscopy

CD spectra were obtained on an AVIV Model 62DS spectropolarimeter with a Peltier temperature controller, using cuvettes with 1 mm path length. Protein solutions were equilibrated for at least 24 h at 4°C before measurements. Wavelength scans were collected from 190 to 260 nm in 0.5 nm steps with a 5 s averaging time and repeated three times. CD thermal transitions were monitored at the 220 nm maximum as a function of temperature. The samples were equilibrated for at least 2 min at each temperature, and the temperature was increased at an average rate of 0.1°C/min. 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. For refolding studies, the ellipticity at 220 nm was monitored as the temperature was decreased from 70°C to 0°C with an average decreasing rate of 0.1°C/min. The percentage of refolding was defined as the ratio of the CD signal regained at 0°C after refolding compared with the original CD signal at 0°C before melting.

Differential scanning calorimetry (DSC)

DSC experiments were recorded on a NANO DSC II Model 6100 (Calorimetry Sciences Corp). Each sample was dialyzed against phosphate-buffered saline (pH 7.0). Sample solutions were loaded at 0°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. The thermal transition temperatures measured by DSC at a heating rate of 1°C/min are always slightly higher (1–2°C) than the transition temperature observed by CD measured at an average heating rate of 0.1°C/min.[42] This suggests that the thermal transitions are close to, but not completely at equilibrium.

Dynamic light scattering (DLS)

DLS measurements were performed using a DynaPro Titan instrument (Wyatt Technology Corp., Santa Barbara, CA) equipped with a temperature controller using a 12-μL quartz cuvette. All samples were dialyzed against phosphate-buffered saline (pH 7.0), centrifuged and then filtered through 0.1 μm Whatman Anotop filters before measurements. To obtain the hydrodynamic radius (R_h), the intensity autocorrelation functions were analyzed by Dynamic software (Wyatt Technology Corp).

Solubility studies

The purified V-CL, CL, V-CL-CL preparations, and bovine skin collagen were dialyzed at 4°C for 24 h against four different buffers: 0.1*M* acetic acid (pH 2.9) at 4°C; 50 m*M* sodium acetate buffer (pH 5.0); phosphate-buffered saline (pH 7.0) and 50 m*M* glycine buffer (pH 8.6). These samples were also dialyzed at 24°C against all buffers except for acetic acid, since the V-CL and CL proteins were shown to denature at this temperature.[7] The CL-CL protein was dialyzed against PBS at 4°C, 24°C, and 30°C for 24 h, and also against 0.1*M* acetic acid at 4°C. After dialysis, each sample was centrifuged at 14,000 rpm for 10 min to remove insoluble materials. Solubility was determined by measuring the concentration in the supernatant, using the Tyr extinction coefficient of $\varepsilon_{275} = 9635M^{-1}$ cm^{-1} for V-CL and V-CL-CL and using the CD

spectrum for CL, CL-CL and bovine skin collagen assuming a mean residue ellipticity of MRE_{220} = 6000 deg cm² $dmol^{-1}$.[21] Lyophilyzed pepsin-treated bovine skin collagen was dissolved in 0.1 M acetic acid and stirred for 12 h at 4°C. After spinning down at 14,000 rpm for 10 min to remove the cross-linked collagen, the sample was used for solubility studies.

Segment long spacing (SLS) crystallites

To form SLS aggregates, 0.1–0.2 mg/mL of V-CL , CL and V-CL-CL were dialyzed at 4°C against 0.5*M* acetic acid , and then against 0.2% ATP in 0.01 M acetic acid for 2 days. Sample were placed on carbon-coated grids, positively stained with 0.4% phosphotungstic acid (pH 3.5), and examined with a Philips CM12 electron microscope.[26,27]

Electron microscopy

Precipitates of V-CL, CL , V-CL-CL and CL-CL were analyzed under an electron microscope after negative staining. The precipitates were obtained after dialyzing against PBS buffer (20 m*M* $NaPO_4$,150 m*M* NaCl, pH 7.0) and centrifuged at 14,000 rpm for 10 min with a Sorval RC5B centrifuge. After resuspending the precipitates with the buffer, 5 μL samples were adsorbed onto carbon-coated grids for 30 sec, and stained for 5 sec on 1 drop of 0.5% uranyl acetate or 1% PTA (pH 7.0). The grids were dried at room temperature. Specimens were analyzed under a Philips CM12 electron microscope.

Electronic Supporting Information Figure 1(A,B) shows the electron micrographs obtained for CL and V-CL after treatment for SLS preparation (supp_Yoshizumi et al.).

Acknowledgment

Dr. Takeshi Yoshida and Dr. Angela Mohs played an important role in the early design and purification of the bacterial constructs and the authors thank them for their valuable help. Authors also acknowledge Dr. Oleg Mirochnitchenko for helpful discussions and thank Dr. Karunakar Kar for the bovine skin collagen parameters. Dr. Slawomir Lukomski kindly supplied the original p163 clone.

References

1. Kielty CM, Grant ME, The Collagen family: structure, assembly and organization in the extracellular matrix. In: Royce PM, Steinmann B, Eds (2002) Connective tissue and its heritable disorders: molecular, genetic, and medical aspects, 2nd Ed., New York: Wiley-Liss, pp 159–222.
2. Veit G, Kobbe B, Keene DR, Paulsson M, Koch M, Wagener R (2006) Collagen XXVIII, a novel von Willebrand Factor A domain-containing protein with many imperfections in the collagenous domain. J Biol Chem 281: 3494–3504.
3. Exposito JY, Cluzel C, Garrone R, Lethias C (2002) Evolution of collagens. Anat Rec 268:302–316.
4. Rasmussen M, Jacobsson MB, Jorck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 278:32313–32316.
5. Xu Y, Keene DR, Bujnicki JM, Höök M, Lukomski S (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem 277:27312–27318.
6. Sylvestre P, Couture-Tosi E, Mock M (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. Mol Microbiol 45:169–178.
7. Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, Brodsky B (2007) Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. J Biol Chem 282:29757–29765.
8. Boydston J, Chen P, Steichen C, Turnbough C (2005) Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. J Bacteriol 187:5310–5317.
9. Karlström Å, Jacivsson K, Flock M, Folk JI, Guss B (2004) Identification of a novel collagen-like protein Scl C, in *Streptococcus equi* using signal sequence phage display Vet Microbiol 104:179–188.
10. Rasmussen M, Edén A, Björck L (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. Infect Immun 68:6370–6377.
11. Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Ireland RM, Reid SD, Adams GG, Musser JM (2000) Identification and characterization of the scl gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. Infect Immun 68:6542–6553.
12. Whatmore AM (2001) *Streptococcus pyogenes* sclB encodes a putative hypervariable surface protein with a collagen-like repetitive structure. Microbiology 147:419–429.
13. Caswell CC, Lukomska E, Seo NS, Höök M, Lukomski S (2007) Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the α2β(1) integrin enhances pathogen survival and re-emergence. Mol Microbiol 64:1319–1331.
14. Caswell CC, Barczyk M, Keene DR, Lukomska E, Gullberg DE, Lukomski S (2008) Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins α2β1 and α11β1. J Biol Chem 283:36168–36175.
15. Humtsoe JO, Kim JK, Xu Y, Keene DR, Höök M, Lukomski S, Wary KK (2005) A streptococcal collagen-like protein interacts with the α2β1 integrin and induces intracellular signaling. J Biol Chem 280:13848–13857.
16. Han R, Zwiefka A, Caswell CC, Xu Y, Keene DR, Lukomska E, Zhao Z, Höök M, Lukomski S (2006) Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. Appl Microbiol Biotechnol 72:109–115.
17. Hulmes DJ, Miller A, Parry DA, Piez KA, Woodhead-Galloway J (1973) Analysis of the primary structure of collagen for the origins of molecular packing. J Mol Biol 79:137–148.
18. Leikin S, Rau DC, Parsegian VA (1994) Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices. Proc Natl Acad Sci USA 91:276–280.
19. Qing G, Ma LC, Khorchid A, Swapna GV, Mal TK, Takayama MM, Xia B, Phadtare S, Ke H, Acton T, Montelione GT, Ikura M, Inouye M (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. Nat Biotechnol 22:877–882.
20. Fletcher GC (1976) Dynamic light scattering from collagen solutions. I. Translational diffusion coefficient and aggregation effects Biopolymers 15:2201–2217.
21. Brodsky-Doyle B, Leonard KR, Reid KB (1976) Circular-dichroism and electron-microscopy studies of human

subcomponent C1q before and after limited proteolysis by pepsin. Biochem J 159:279–286.
22. Feng Y, Melacini G, Taulane JP, Goodman M (1996) Acetyl-Terminated and template-assembled collagen-based polypeptides composed of Gly-Pro-Hyp sequences. 2. Synthesis and conformational analysis by circular dichroism, ultraviolet absorbance, and optical rotation. J Am Chem Soc 1996;118:10351–10358.
23. Privalov P (1982) Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem 35:1–104.
24. Bella J, Eaton M, Brodsky B, Berman HM (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 266:75–81.
25. Kramer RZ, Venugopal MG, Bella J, Mayville P, Brodsky B, Berman HM (2000) Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. J Mol. Biol 301:1191–1205.
26. Murray LW, Tanzer ML (1983) Characterization of a large fragment from annelid cuticle collagen and its relationship to the intact molecule. Coll Relat Res 3:445–458.
27. Kühn K (1982) Segment-long-spacing crystallites, a powerful tool in collagen research. Coll Relat Res 2:61–80.
28. Brodsky B, Eikenberry EF (1982) Characterization of fibrous forms of collagen. Methods Enzymol 82:127–174.
29. Olsen DR, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perälä M, Hämäläinen ER, Jarvinen M, Polarek J (2003) Recombinant collagen and gelatin for drug delivery. Adv Drug Deliv Rev 55:1547–1567.
30. Ruggiero F, Koch M (2008) Making recombinant extracellular matrix proteins. Methods 45:75–85.
31. Nokelainen M, Tu H, Vuorela A, Notbohm H, Kivirikko KI, Myllyharju J (2001) High-level production of human type I collagen in the yeast Pichia pastoris. Yeast 18:792–806.
32. Persikov A, Ramshaw J, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. J Biol Chem 280:19343–19349.
33. Mallinger R, Schmut O (1988) Reaggregation behavior of different types of collagen in vitro: variations in the occurrence and structure of dimeric segment long-spacing collagen. J Ultrastruct Mol Struct Res 98:11–18.
34. Kar K, Amin P, Bryan MA, Persikov AV, Mohs A, Wang YH, Brodsky B (2006) Self-association of collagen triple helic peptides into higher order structures. J Biol Chem 2006;281:33283–33290.
35. Steplewski A, Hintze V, Fertala A (2007) Molecular basis of organization of collagen fibrils. J Struct Biol 157: 297–307.
36. Ito H, Steplewski A, Alabyeva T, Fertala A (2006) Testing the utility of rationally engineered recombinant collagen-like proteins for applications in tissue engineering. J Biomed Mater Res A 76:551–560.
37. Vitagliano L, Berisio R, Mazzarella L, Zagari A (2001) Structural bases of collagen stabilization induced by proline hydroxylation. Biopolymers 58:459–464.
38. Kar K, Wang YH, Brodsky B (2008) Sequence dependence of kinetics and morphology of collagen model peptide self-assembly into higher order structures. Protein Sci 17:1086–1095.
39. Rele S, Song Y, Apkarian RP, Qu Z, Conticello VP, Chaikof EL (2007) D-periodic collagen-mimetic microfibers. J Am Chem Soc 129:14780–14787.
40. Perret S, Merle C, Bernocco S, Berland P, Garrone R, Hulmes DJS, Theisen M, Ruggiero F (2001) Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. J Biol Chem 276: 43693–43698.
41. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319–326.
42. Persikov AV, Xu Y, Brodsky B (2004) Equilibrium termal transition of collagen model peptides. Protein Sci 13: 893–902.