

Runlin Han · Antoni Zwiefka · Clayton C. Caswell ·  
Yi Xu · Douglas R. Keene · Ewa Lukomska ·  
Zhihong Zhao · Magnus Höök · Slawomir Lukomski

## Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers

Received: 12 December 2005 / Revised: 15 February 2006 / Accepted: 18 February 2006 / Published online: 22 March 2006  
© Springer-Verlag 2006

**Abstract** Collagen triple helix, composed of the repeating Gly–Xaa–Yaa (GXY) sequence, is a structural element found in all multicellular animals and also in some prokaryotes. Long GXY polymers are highly regarded components used in food, cosmetic, biomedical, and pharmaceutical industries. In this study, we explore a new concept for the production of recombinant GXY polymers which are based on the sequence of “prokaryotic collagens”, the streptococcal collagen-like proteins Scl1 and Scl2. Analysis of 50 Scl variants identified the amino acid distribution and GXY-repeat usage that are involved in the stabilization of the triple helix in Scls. Using circular dichroism spectroscopy and electron microscopy, we show that significantly different recombinant rScl polypeptides form stable, unhydroxylated homotrimeric triple helices that can be produced both intra- and extracellularly in the *Escherichia coli*. These rScl constructs containing 20 to 129 GXY repeats had mid-point melting temperatures between 32 and 39°C. Altogether, Scl-derived collagens, which are different from the mammalian collagens, can

form stable triple helices under physiological conditions and can be used for the production of recombinant GXY polymers with a wide variety of potential applications.

### Introduction

Three polypeptide tracts, each composed of the repeating Gly–Xaa–Yaa (GXY) sequence, have the ability to form a right-handed triple helix (Ramachandran 1988). This structural element is present in collagen, a protein that is found abundantly across the animal kingdom. The stability of mammalian collagens is determined at the level of the triple helix by the requirements that a glycyl residue occupies every third position, and that proline and hydroxyproline residues are frequently present in positions X and Y, respectively (Berg and Prockop 1973).

Proteins containing GXY repeats have also been found in bacteria and phages (Smith et al. 1998; Rasmussen et al. 2003). For example, streptococcal collagen-like proteins, Scl1 and Scl2 (also known as SclA and SclB, respectively), are expressed on the cell surface of the bacterium group A *Streptococcus* (GAS) (Lukomski et al. 2000, 2001; Rasmussen et al. 2000; Ferretti et al. 2001; Rasmussen and Bjorck 2001; Whatmore 2001). GAS strains produce many Scl1 and Scl2 variants, which have the N-terminal noncollagenous variable (V) domains followed by collagen-like (CL) regions of disparate lengths that contain variable GXY repeats. Initial studies have shown that two recombinant Scl proteins (rScl) formed stable triple helices and shared similar structural organization (Xu et al. 2002). Importantly, the triple helix in Scl can be recognized by human collagen receptors, such as integrin  $\alpha_2\beta_1$ , resulting in cell signaling, indicating that Scl proteins are not only the structural but also the functional collagen mimics (Humtsoe et al. 2005).

Collagen and its denatured form, gelatin, are highly desired biomaterials that are used extensively in the medical field, as well as in cosmetic, pharmaceutical, and food industries (Asghar and Henrickson 1982; Lee et al. 2001). Industrial preparations of collagens and gelatins are

---

Antoni Zwiefka is on leave from the Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Centre of Excellence, Wrocław, Poland.

---

R. Han · A. Zwiefka · C. C. Caswell · E. Lukomska · Z. Zhao ·  
S. Lukomski (✉)  
Department of Microbiology, Immunology and Cell Biology,  
West Virginia University School of Medicine,  
Morgantown, WV 26506, USA  
e-mail: slukomski@hsc.wvu.edu  
Tel.: +1-304-2936405  
Fax: +1-304-2937823

Y. Xu · M. Höök  
Center for Extracellular Matrix Biology,  
Institute of Biosciences and Technology,  
Texas A&M University System Health Science Center,  
Houston, TX 77030, USA

D. R. Keene  
Shriners Hospital for Children,  
Portland, OR 97201, USA

typically obtained from bovine skin and other animal products and contain chemically modified heterogeneous peptides resulting from harsh extraction procedures. Moreover, they could be contaminated with infectious agents. Therefore, many eukaryotic expression systems have been employed to produce recombinant collagen- and gelatin-like polymers. These include mammalian cells (Geddis and Prockop 1993; Fertala et al. 1994), insect cells (Myllyharju et al. 1997; Tomita et al. 1997), transgenic mice (John et al. 1999; Toman et al. 1999), plants (Ruggiero et al. 2000), and yeast (Werthen et al. 1999; Toman et al. 2000).

Collagen-like recombinant molecules were also produced in prokaryotic systems, such as *Bacillus brevis* (Kajino et al. 2000) and *Escherichia coli* (Goldberg et al. 1989; Cappello and Ferrari 1998) cultures. However, all of the above GXY polymers are based on mammalian-collagen sequences. In this study, we assess the use of the Scl-based sequences as a potential source for the production of recombinant “prokaryotic collagens” (Hook et al. 2005). The Scl sequences are analyzed for the amino acid- and GXY-repeat-usage. Several different recombinant rScl proteins are made intra- and extracellularly using an *E. coli* expression system and the triple-helical structure and thermal stability of these constructs is determined. The data presented in this study provide the necessary outline for the production of recombinant GXY polymers derived from Scl proteins.

## Materials and methods

### DNA amplification and cloning

Recombinant Scl (rScl) proteins were generated using the *Strep*-tag II expression and purification system (IBA-GmbH, Goettingen, Germany). Various *scl* alleles were polymerase chain reaction-amplified from GAS genomic DNA with primers containing synthetic extensions compatible for cloning into the *Bsa*I-digested *E. coli* vector pASK-IBA2, using high-fidelity Deep Vent Taq Polymerase (New England BioLabs, Beverly, MA, USA). The thermal conditions were optimized for each primer pair in a temperature-gradient block of a thermocycler. The *scl1*-gene based constructs originated from the following alleles: pSL144-*scl1.1* (GenBank accession number AF252861); pSL1161-*scl1.28* (AY459361); pSL176-*scl1.41* (AY452037); and pSL186-*scl1.12* (DQ309441). The *scl2*-gene constructs originated from the following alleles: pSL163-*scl2.28* (AY069936); pSL177-*scl2.4* (DQ309442); and pSL178-*scl2.77* (DQ309443). *E. coli* DH5 $\alpha$  was used in cloning experiments and all plasmids were subjected to DNA sequencing.

*E. coli* strains harboring plasmid constructs are available upon request (S. L.).

### Expression of rScl proteins in *E. coli*

The pASK-IBA2 vector is designed for periplasmic protein expression and induction is controlled by *tet*-promoter/operator system. The Scl-encoded sequence is fused at the N terminus to the OmpA signal peptide that mediates the secretion of the recombinant rScl protein to the periplasmic space. The OmpA sequence is then selectively cleaved off by an endogenous signal peptidase, thus resulting in the release of the rScl polypeptide. In addition, each rScl polypeptide is also fused at the C terminus to the short eight amino-acid-long tag, the *Strep*-tag II (WSHPQFEK), which binds to the *Strep*-Tactin–Sepharose resins with a high affinity. This short inert tag does not interfere with triple helix formation by the recombinant rScl polypeptides. *E. coli* BL21 deficient in Lon protease was used for protein expression. The identity of each rScl construct was verified by amino-terminal sequencing.

Bacterial cultures were grown in a Luria–Bertani medium (LB broth, Miller). One-liter cultures contained in four-liter flasks were propagated with a constant agitation (200 rpm) at 30°C. When OD<sub>600</sub> reached 0.5–0.6, protein expression was induced with anhydrotetracycline (0.2  $\mu$ g/ml) for 3 h. After centrifugation, cell pellets were resuspended in a high-sucrose buffer P (100 mM Tris–HCl, 1 mM EDTA pH 8.0, 500 mM sucrose) and incubated on ice for 30 min. The periplasmic fraction containing the recombinant polypeptides was then separated from the cells by centrifugation and used for affinity purification on a *Strep*-Tactin–Sepharose column. The stringency of the wash buffer (buffer W) applied to the affinity column during purification varied in sodium chloride concentration for various constructs (100 mM Tris–HCl pH 8.0, 1 mM EDTA, 150–500 mM NaCl). Purified rScl preparations were dialyzed against 25 mM HEPES, pH 8.0 and stored at –20°C. Sample purity was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) following staining with RAPIDstain (Geno Technology, St. Louis, MO, USA), and protein concentration was obtained by amino acid analysis using mass spectrometry.

The presence of rScl protein in a cell-free form in culture supernatant was tested with one rScl construct, P163 (EC), after precipitation with 55%-saturated ammonium sulphate. The pellet was dissolved in 25 mM Tris–HCl, 1 mM EDTA (pH 8.0) buffer and applied on a *Strep*-Tactin–Sepharose column.

The same P163 variant was also used to assess the feasibility of intracellular production (P163IC) of the rScl proteins. For this experiment, the *scl2.28* allele was amplified and cloned into the pASK-IBA3 vector designed for intracellular protein expression. P163IC extracted from the *E. coli* cells was purified by affinity chromatography with *Strep*-Tactin–Sepharose as described above.

## Circular dichroism spectroscopy

The triple-helical structure of rScl proteins was analyzed by circular dichroism (CD) spectroscopy as described earlier (Xu et al. 2002). Protein samples were dialyzed against 1% phosphate-buffered saline, pH 7.4. The far-UV spectra were recorded with a Jasco J720 spectropolarimeter in a thermostatically controlled cuvette with a 0.5-cm path length and data were integrated for 1 s at 0.2-nm interval with a bandwidth of 1 nm. A wavelength scan was performed for each rScl protein at each end point temperature before (4, 25°C) and after (50°C) unfolding or subsequent refolding (25°C). Thermal transition profiles were recorded by monitoring the change in the mean molar residue ellipticity at a fixed wavelength of 220 nm,  $[\theta]_{220}$ , as a function of temperature. A heating rate of 10°C/h was applied until the sample reached 50°C. The melting temperatures ( $T_m$ ) were given as the mean of  $T_m$  values from several measurements.

## Rotary shadowing and electron microscopy.

A two-domain structural organization of rScl constructs was viewed under electron microscope after rotary shadowing (Xu et al. 2002). Recombinant proteins were dialyzed against 0.1 M ammonium bicarbonate at the protein concentration of 100 µg/ml. Protein samples were mixed with glycerol to a final glycerol concentration of 70% (v/v), and 100 µl of each sample was nebulized with an airbrush onto freshly cleaved mica. The samples were dried in a vacuum and rotary shadowed with carbon/platinum using an electron beam gun tilted at an angle of 6° relative to the mica surface in a Balzers BAE 250 evaporator. The replicas were backed with carbon at 90° and placed onto copper grids. Photomicrographs were taken using a Philips 410 electron microscope.

## Results

## Analysis of the amino acid sequences of the Scl proteins

To assess sequence characteristics, total amino acid and GXY-triplet distributions were analyzed among different variants of the streptococcal collagen-like proteins Scl1 ( $n=28$ ) and Scl2 ( $n=22$ ) (Fig. 1). Seven residues (A, D, E, K, L, P, and Q) occupy position X in about 98% of all GXY repeats. Furthermore, the alanine residue in position X is more frequent in Scl1 (14.53%) than Scl2 (3.52), whereas L (0.89 vs 10.04%) and Q (0.99 vs 10.04%) are almost exclusive to Scl2 proteins. Position Y in the CL regions of both Scls was occupied by a set of nine amino acids (A, D, K, N, P, Q, R, T, and V) in 99% of all GXY repeats. While K (21.51 vs 6.78%), Q (22.24 vs 9.79%), and T (10.98 vs 5.59%) are more frequent in Scl1 variants, residues D (5.03 vs 30.63%), and N (0 vs 4.58%) are more specific for Scl2. In general, the amino acids C, H, W, and Y were never found in collagen-like regions of either Scl protein. In addition, position X was never occupied by R in Scl1 nor by I or V in Scl2, whereas position Y was never occupied by M or N in Scl1 nor by E or F in Scl2. Thus, a limited set of amino acid residues is favored in the X and Y position of the Scl proteins suggesting that these residues are important for the formation and stability of the collagen-like triple helical structures observed in these proteins.

Sixty-two distinct GXY triplets have been identified in the CL regions of all 50 Scl1 and Scl2 variants; predominant and rare GXY repeats were identified (diagram in Fig. 1). Only six GXY triplets (GEA, GPA, GKD, GEK, GPQ, and GET) occur with frequencies higher than 5% accounting for ~50% of the CL regions, with GKD being the most common (17.15%). In contrast, as many as 40 different triplets occurred with frequencies below 1%, contributing to a total of 9.4% of the CL regions. While some triplets were equally distributed among both Scl proteins, others were more frequent in either Scl1 (GEK, 13.87%) or Scl2 (GKD, 26.93%; and GPA, 10.61%) variants. Although tripeptide GKD is

X\Y	A	D	E	F	I	K	M	N	P	Q	R	S	T	V	Total	Scl1	Scl2
A	0.07					1.27			1.80	4.40	0.07		0.49	0.18	8.28	14.53	3.52
D	0.21				0.11	2.36			0.39	0.21	2.71		1.13		7.12	10.23	4.77
E	5.00	0.14			0.04	7.08			0.46	2.82	3.20		5.25	0.11	24.10	33.93	16.70
I						0.21			0.04	0.70					0.95	2.19	0.00
K	0.81	17.15		0.04					2.39	0.42	0.07	0.18	0.11	0.18	21.35	10.88	29.63
L	0.18					0.04			4.89	0.85	0.04		0.04		6.04	0.89	10.04
N									0.11						0.11	0.16	0.06
P	8.77		0.39		0.04	1.97	0.28	0.04		5.18	3.24		0.88	3.80	24.59	25.89	23.79
Q		2.08						2.39	1.27	0.04	0.28				6.06	0.99	10.04
R								0.14							0.14	0.00	0.25
T						0.21			0.07	0.46					0.74	0.16	1.19
V										0.07					0.07	0.16	0.00
<b>Total</b>	15.04	19.37	0.39	0.04	0.19	13.14	0.28	2.57	11.42	15.15	9.61	0.18	7.90	4.27			
<b>Scl1</b>	15.26	5.03	0.89	0.08	0.32	21.51	0.00	0.00	9.90	22.24	8.52	0.32	10.98	4.95			
<b>Scl2</b>	15.00	30.63	0.00	0.00	0.06	6.78	0.52	4.58	12.68	9.79	10.55	0.06	5.59	3.77			

**Fig. 1** Total amino acid and GXY-triplet distributions in the CL regions of Scl1 and Scl2 proteins. Frequencies are expressed as percentile values among 28 different Scl1 and 22 Scl2 variants. The frequency of distribution of GXY triplets is shown in the diagram. GXY triplets present only in Scl1 variants are shown in *black fields*,

triplets present only in Scl2 variants are shown in *light-grey fields*, and triplets found in both Scl1 and Scl2 proteins are shown in *white fields*. Frequencies of amino acids observed in position X are shown to the *right* and those observed in position Y are shown *below* the diagram

present in the Scl1-CL regions (4.65%), it is far more predominant in the Scl2-CL regions (up to 43% in some variants). Furthermore, several GXY repeats are found only in either Scl1 (Fig. 1; black fields) or Scl2 (Fig. 1; light-grey fields) variants.

#### Extracellular and intracellular expression of rScl proteins in *E. coli*

The extracellular portion of Scl proteins, lacking the N-terminal signal sequence and C-terminal cell-wall associated region, was expressed in rScl constructs. The periplasmic compartment was chosen because Scl proteins are naturally expressed on the cell-surface by group A *Streptococcus*, and we could not exclude the possibility that triple helix formation requires the secretion of the protein. Without any optimization of the expression conditions, the yield obtained for different rScl variants varied over a wide range from 1 mg/l (P161) to 20 mg/l (P163). In general, the purity of protein preparations after single-step affinity-chromatography purification was estimated to be >90%, as judged by SDS-PAGE analysis (Fig. 2a).

The pASK-IBA2 vector employs the OmpA signal sequence for fusion transport across the cytoplasmic membrane. It has been reported that OmpF, which is another outer-membrane protein, was secreted in large quantities into the culture medium (Jeong and Lee 2002). We tested whether recombinant protein, P163EC (extracellular), was secreted into the growth medium. Following precipitation from the culture medium with 55% ammonium sulfate, a crude protein sample was subjected to the standard purification protocol on *Strep*-Tactin–Sepharose column and analyzed by SDS-PAGE (Fig. 2b). Approximately 5 mg of P163EC was recovered from 1 l of the culture medium compared to ~2 mg/l of P163 extracted from the periplasm.

Intracellular expression of a synthetic collagen-like polymer in *E. coli* has been reported, although structural characteristics of this polymer were not presented (Yin et al. 2003). The *scl2.28* allele was cloned into the pASK-IBA3 vector designed for intracellular expression. The

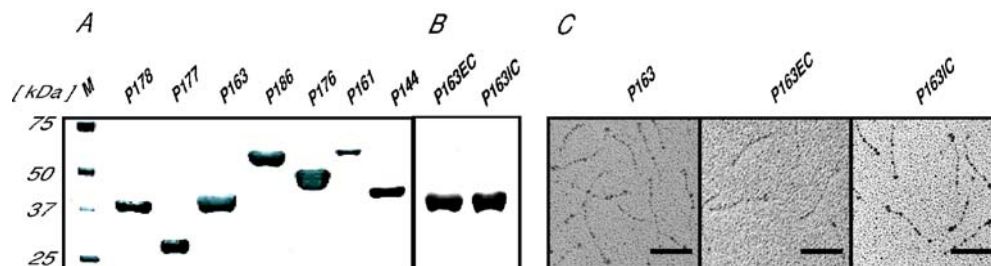
resulting P163IC (intracellular) construct was expressed and purified with *Strep*-Tactin–Sepharose at a high yield of ~20 mg of pure protein per 1 l of culture (Fig. 2b). The molecular weight of purified P163EC and P163IC preparations was the same as that of P163 produced in the periplasm.

#### Structural organization of rScl proteins

Electron microscopy of rotary shadowed samples has been used for the investigation of structural organization of proteins with collagenous domains (Engel and Furthmayr 1987). It was previously demonstrated that rScl proteins may form lollipop-like structures with the lollipop stalks being the CL regions and the globular heads, the V regions (Xu et al. 2002). Examination of the rotary shadowed samples of all rScl proteins shown in Fig. 2a revealed a characteristic lollipop-like structural organization (data not shown). The length of the stalks was consistent with the calculated length of the CL regions in a collagen-like triple helical conformation. Likewise, all three P163 preparations obtained from the periplasm, precipitated from the culture medium, and expressed intracellularly shared similar structural organization as viewed by electron microscopy following rotary shadowing (Fig. 2c).

#### Triple helix formation by rScl proteins

The triple helix formation by each recombinant rScl construct was studied by circular dichroism (CD) spectroscopy (Fig. 3). The Scl variants selected for cloning and expression differed significantly in their CL-region length (34–79 GXY repeats) and primary sequence. Triple helix formation was assessed by the presence of the characteristic molar ellipticity maximum at 220 nm,  $[\theta]_{220}$  (Brodsky-Doyle et al. 1976). The CD spectrum of each rScl sample recorded at 25°C resembled that of a collagen triple helix, with an increased ellipticity at ~220 nm (Fig. 3a). The increase was not seen when the CD spectra were recorded at 50°C, indicating that the triple helix structure had unfolded. When protein samples were cooled



**Fig. 2** Recombinant rScl proteins. Various Scl1- and Scl2-derived recombinant proteins were expressed in *E. coli*, purified by affinity chromatography with *Strep*-Tactin–Sepharose, and analyzed in 10% SDS-PAGE stained with RAPIDstain. *M* molecular size marker. **a** rScl constructs produced in *E. coli* periplasm. **b** Protein fraction secreted into the culture medium precipitated with 55% ammonium sulfate

(P163EC) or P163IC construct expressed intracellularly. Crude preparations of P163EC and P163IC were purified by affinity chromatography with *Strep*-Tactin–Sepharose. **c** Structural organization of P163, P163EC, and P163IC as viewed by electron microscopy after rotary shadowing. All three preparations share similar two-domain lollipop-like organization. Bars 50 nm

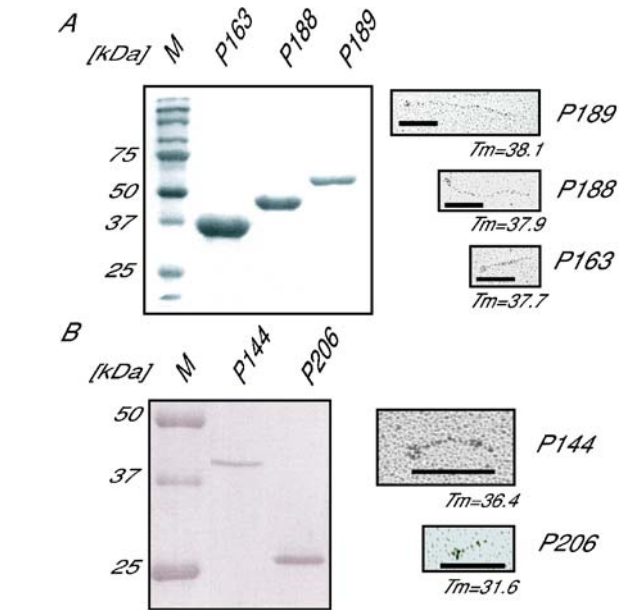
to 4°C from 50°C, the ellipticity at 220 nm gradually increased, and the CD spectra again showed the characteristics of a collagen-like triple helix (for clarity, unfolding and refolding curves are not shown on the plot). The differences between  $[\theta]_{220}$  values observed for various rScls reflect the contribution of the secondary structures in their noncollagenous regions.

### Thermal stability of the rScl proteins

Thermal unfolding of the collagen triple helix in the rScl proteins was measured by monitoring the CD at 220 nm as a function of increasing temperature (Fig. 3b). Most rScl variants showed a triple helix to random coil transition within a narrow temperature range. The thermal unfolding was reversible because CD spectra at 220 nm increased upon cooling the samples back to 4°C (refolding curves are not shown on the plot). The midpoint melting temperatures,  $T_m$ , for all rScl variants ranged between 34.1°C measured for P161 and 39.1°C for P178 (Fig. 3c). In general, higher  $T_m$  values were observed for rScl proteins with longer CL regions, although noncollagenous domains seemed to contribute to the stability of these proteins.

### Structure and stability of rScls with long and short CL regions

We generated rScl constructs with extended CL regions. A DNA fragment encoding 25 GXY repeats from the interior of the P163-CL region was amplified with primers flanked by the *BsaI*-cleavage sites and cloned at the single *BsaI* site of the P163-CL region. A single 25× GXY insertion resulting in P188 with a total of 104 triplets, and a double insertion resulting in a 129-GXY-long CL region in P189 were identified (Fig. 4a). As expected, both variants formed lollipops with elongated stalks, and increased thermal stability of the P188 and P189 was recorded as  $T_m=37.9^\circ\text{C}$  and  $T_m=38.1^\circ\text{C}$ , respectively, compared with  $T_m=37.7^\circ\text{C}$  of the original P163 construct (Fig. 3c).

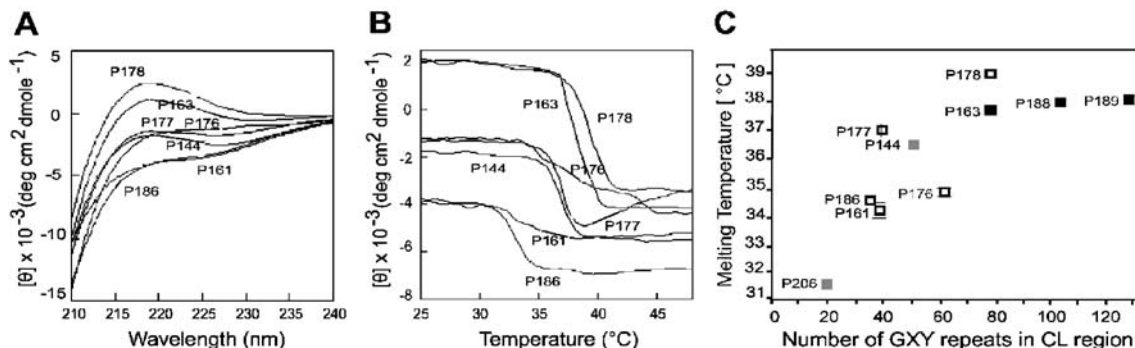


**Fig. 4** Expression and structure of extended recombinant rScl variants. **a** P163-CL region (79 GXY repeats) extended by inserting 25 (P188, 104 GXY repeats total) and 50 (P189, 129 GXY repeats) additional GXY repeats. **b** In-frame deletion of 30 GXYs in P144-CL region (P206, 20 GXY repeats). Recombinant rScl constructs were expressed in the periplasm, purified with *Strep*-Tactin-Sephacrose, and analyzed in 10% SDS-PAGE. *M* Molecular size markers. Structural organization viewed by electron microscopy after rotary shadowing. *Bars* 50 nm

Similarly, an internal in-frame deletion of 30 GXYs in the P144-CL region decreased the  $T_m=36.4^\circ\text{C}$  recorded for P144 to  $T_m=31.6^\circ\text{C}$  of the resulting P206 (Fig. 3c), although structural organization remained the same (Fig. 4b).

## Discussion

Eukaryotic and prokaryotic collagens evolved different molecular mechanisms to stabilize the triple helix. While collagen triple helix is stabilized in eukaryotes by the



**Fig. 3** Circular dichroism spectra and thermal stability. **a** Wavelength scans of recombinant rScl proteins at 25°C. **b** Thermal unfolding profiles recorded for the same constructs at 220 nm as a function of increasing temperature with a slope of 10 C/h. **c** Correlation between

thermal stability and number of GXY repeats in the CL region of rScl polypeptides. *Black squares* P163 and its extended derivatives P188 and P189. *Grey squares* P144 and its in-frame deleted derivative P206. The midpoint melting temperatures  $T_m \pm \text{SD}$  are shown

presence of hydroxyprolines (Hyp) in the Y position, the mechanisms stabilizing “prokaryotic collagens”, which naturally lack Hyp residues, are poorly understood. Recent sequence analysis proposed that prokaryotic GXY tracts are mainly stabilized by threonine and glutamine residues often observed in position Y. In some bacterial collagen-like proteins, threonines occur with frequencies above 50% of all Y positions, as compared with only 2.6% found in human collagens (Rasmussen et al. 2003). Threonine in this position may substitute for hydroxyproline in the formation of hydrogen bonds (Kramer et al. 1999). A similar mechanism may be true for glutamine because all three amino acids, e.g., hydroxyproline, threonine, and glutamine are hydrogen donors. However, threonines occur in the Y position of ~8% of the Scl-GXY repeats (Fig. 1), and are twice as frequent among Scl1 compared to Scl2 variants. In aggregate, the most predominant amino acids found in position X of Scls are K (21.35%), E (24.1%), and P (24.59%) whereas position Y is mainly occupied by A (15.04%), Q (15.15%), and D (19.37%).

The experiments with model synthetic “host” peptides containing the so-called “guest” residues defined triplets stabilizing the triple helix despite a lack of hydroxyproline residue (Ramshaw et al. 1998; Persikov et al. 2000). We found that Scl proteins incorporated several of such triplets including GPA (8.77%), GPQ (5.18%), or GPR in Scl2 (4.58%). In addition, the stabilizing context of strings of residues with opposite charges, such as KGE/D (Persikov et al. 2005), is often found in Scl proteins. Based on this analysis, the thermal stability of the Scl-derived triple helices can be manipulated allowing for the construction of custom-designed GXY polymers.

Unhydroxylated recombinant gelatin-like and collagen-like polymers, derived from mammalian collagens, have been designed and produced in yeasts, transgenic plants, and bacteria (Goldberg et al. 1989; Cappello and Ferrari 1998; Werten et al. 1999; Kajino et al. 2000; Ruggiero et al. 2000). These long unhydroxylated GXY polymers derived from type I collagen had significantly lower melting temperatures compared with the native molecule (30 vs 41°C) (Perret et al. 2001). It is interesting to note that the rScl constructs with significantly shorter GXY tracts had mid-point melting temperatures of about 35–39°C, which is approximately 5°C higher than long-unhydroxylated molecules derived from human- or bovine-collagen sequences, and recombinant P206 containing only 20 GXY repeats formed a stable triple helix ( $T_m \approx 32^\circ\text{C}$ ). Furthermore, a significant increase in the CL-region length (79–129 GXYs) did not significantly increase the thermal stability of their triple helices (37.7–38.1°C). Thus, thermal stability of the GXY polymers derived from natural Scl variants will largely oscillate around body temperature and can be increased by the enrichment with triple-helix stabilizing triplets such as GPR, GPQ, and GPA.

In this study, we showed that recombinant rScl peptides form triple helices when expressed intracellularly and extracellularly, both in the periplasm and secreted into the medium. The overall yield, without optimization, varied greatly for different variants between 1–20 mg/l of

culture, although rScl could be recovered in significant quantities from culture medium, thereby potentially increasing the production yield. In addition, recloning of the P163 sequence into the pCold III vector resulted in increased expression above 30 mg/l (B. Brodsky, personal communication). Recent expression studies of collagen-like polymers in *E. coli* indicated that media, growth conditions, and construct design may greatly enhance the production yield (Sang 1996; Hannig and Makrides 1998; Yin et al. 2003). Thus, it is very likely that the production yield of Scl-based collagens can be significantly increased.

**Acknowledgements** This work was supported by Public Health Service grant AI50666 from the National Institutes of Health (to S. L.). We thank S. F. Tufa for technical assistance.

## References

- Asgar A, Henrickson RL (1982) Chemical, biochemical, functional, and nutritional characteristics of collagen in food systems. *Adv Food Res* 28:231–372
- Berg RA, Prockop DJ (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline stabilizing the triple-helix of collagen. *Biochem Biophys Res Comm* 52:115–120
- 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
- Cappello J, Ferrari FA (1998) High molecular weight collagen-like protein polymers. US Patent 5773249
- Engel J, Furthmayr H (1987) Electron microscopy and other physical methods for the characterization of extracellular matrix components: laminin, fibronectin, collagen IV, collagen VI, and proteoglycans. *Methods Enzymol* 145:3–78
- Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S, Lai HS, Lin SP, Qian Y, Jia HG, Najjar FZ, Ren Q, Zhu H, Song L, White J, Yuan X, Clifton SW, Roe BA, McLaughlin R (2001) Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci USA* 98:4658–4663
- Fertala A, Sieron AL, Ganguly A, Li SW, Ala-Kokko L, Anumula KR, Prockop DJ (1994) Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). *Biochem J* 298:31–37
- Geddis AE, Prockop DJ (1993) Expression of human COL1A1 gene in stably transfected HT1080 cells: the production of a thermostable homotrimer of type I collagen in a recombinant system. *Matrix* 13:399–405
- Goldberg I, Salerno AJ, Patterson T, Williams II (1989) Cloning and expression of a collagen-analog-encoding synthetic gene in *Escherichia coli*. *Gene* 80:305–314
- Hannig G, Makrides SC (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol* 16:54–60
- Hook M, Lukowski S, Xu Y (2005) Prokaryotic collagen-like proteins and uses thereof. US Patent 6953839
- Humtsoe JO, Kim JK, Xu Y, Keene DR, Höök M, Lukowski S, Wary KK (2005) A streptococcal collagen-like protein interacts with the  $\alpha 2\beta 1$  integrin and induces intracellular signaling. *J Biol Chem* 280:13848–13857
- Jeong KJ, Lee SY (2002) Excretion of human  $\beta$ -endorphin into culture medium by using outer membrane protein F as a fusion partner in recombinant *Escherichia coli*. *Appl Environ Microbiol* 68:4979–4985

- John DC, Watson R, Kind AJ, Scott AR, Kadler KE, Bulleid NJ (1999) Expression of an engineered form of recombinant procollagen in mouse milk. *Nat Biotechnol* 17:385–389
- Kajino T, Takahashi H, Hirai M, Yamada Y (2000) Efficient production of artificially designed gelatins with a *Bacillus brevis* system. *Appl Environ Microbiol* 66:304–309
- Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat Struct Biol* 6:454–457
- Lee CH, Singla A, Lee Y (2001) Biomedical applications of collagen. *Int J Pharm* 221:1–22
- 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
- Lukomski S, Nakashima K, Abdi I, Cipriano VJ, Shelvin BJ, Graviss EA, Musser JM (2001) Identification and characterization of a second extracellular collagen-like protein made by group A *Streptococcus*: control of production at the level of translation. *Infect Immun* 69:1729–1738
- Myllyharju J, Lamberg A, Notbohm H, Fietzek PP, Pihlajaniemi T, Kivirikko KI (1997) Expression of wild-type and modified proalpha chains of human type I procollagen in insect cells leads to the formation of stable [alpha 1(I)]2alpha 2(I) collagen heterotrimers and [alpha 1(I)]3 homotrimers but not [alpha 2(I)]3 homotrimers. *J Biol Chem* 272:21824–21830
- 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
- Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2000) Amino acid propensities for the collagen triple-helix. *Biochemistry* 39:14960–14967
- Persikov AV, Ramshaw JAM, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. *J Biol Chem* 280:19343–19349
- Ramachandran GN (1988) Stereochemistry of collagen. *Int J Pept Protein Res* 31:1–16
- Ramshaw JAM, Shah NK, Brodsky B (1998) Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol* 122:86–91
- Rasmussen M, Bjorck L (2001) Unique regulation of SclB—a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* 40:1427–1438
- Rasmussen M, Eden A, Bjorck L (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun* 68:6370–6377
- Rasmussen M, Jacobsson M, Bjorck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* 278:32313–32316
- Ruggiero F, Exposito J-Y, Bournat P, Gruber V, Perret S, Comte J, Olgarnier B, Garrone R, Theisen M (2000) Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett* 469:132–136
- Sang YL (1996) High cell-density culture of *Escherichia coli*. *Trends Biotechnol* 14:98–105
- Smith MCM, Burns N, Sayer JR, Sorrel JA, Casjens SR, Hendrix RW (1998) Bacteriophage collagen. *Science* 279:1834
- Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, Berg RA, Hitzeman RA (2000) Production of recombinant human type I procollagen trimers using a four-gene expression system in the yeast *Saccharomyces cerevisiae*. *J Biol Chem* 275:23303–23309
- Toman PD, Pieper F, Sakai N, Karatzas C, Platenburg E, de Wit I, Samuel C, Dekker A, Daniels GA, Berg RA, Platenburg GJ (1999) Production of recombinant human type I procollagen homotrimer in the mammary gland of transgenic mice. *Transgenic Res* 8:415–427
- Tomita M, Kitajima T, Yoshizato K (1997) Formation of recombinant human procollagen I heterotrimers in a baculovirus expression system. *J Biochem* 121:1061–1069
- Werten MWT, van den Bosch T, Wind RD, Mooibroek H, de Wolf FA (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast* 15:1087–1096
- Whatmore AM (2001) *Streptococcus pyogenes* sclB encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* 147:419–429
- 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
- Yin J, Lin J, Li W, Wang DIC (2003) Evaluation of different promoters and host strains for the high-level expression of collagen-like polymer in *Escherichia coli*. *J Biotechnol* 100:181–191