Preparation and characterization of monomers to tetramers of a collagen-like domain from *Streptococcus pyogenes*

Yong Y Peng, Violet Stoichevska, Linda Howell, Soren Madsen, Jerome A Werkmeister, Geoff J Dumsday & John AM Ramshaw

Preparation and characterization of monomers to tetramers of a collagen-like domain from *Streptococcus pyogenes*

Yong Y Peng, Violet Stoichevska, Linda Howell, Soren Madsen, Jerome A Werkmeister, Geoff J Dumsday, and John AM Ramshaw*

CSIRO Manufacturing Flagship; Clayton, Australia

The collagen like domain Scl2 from *Streptococcus pyogenes* has been proposed as a potential biomedical material. It is non-cytotoxic and non-immunogenic and can be prepared in good yield in fermentation. The Scl2 collagen domain is about a quarter of the length, 234 residues, of the main collagen type, mammalian type I collagen (1014 residues) that is currently used in biomedical devices. In the present study we have made constructs comprising 1 to 4 copies of the Scl2 collagen domain, plus these same constructs with a CysCys sequence at the C-terminal, analogous to that found in mammalian type III collagens. The yields of these constructs were examined from 2 L fermentation studies. The yields of both series declined with increasing size. Circular dichroism showed that the addition of further collagen domains did not lead to a change in the melting temperature compared to the monomer domain. Addition of the CysCys sequence led to a small additional stabilization of about 2-3°C for the monomer construct when the folding (V) domain was present.

Introduction

Collagens have proved effective in a variety of medical devices and have become well accepted by clinicians.[1,2] The most commonly used sources have been mammalian skin and tendon collagens, especially from bovines. However, as natural products, they can show batch to batch variability, which can affect the material properties. There have also been concerns that these animal-derived collagens may also lead to the transmission of diseases, particularly spongiform encephalopathies including "mad cow disease." Other non-mammalian tissue sources have been proposed to minimize risks of disease, including avian,[3] piscine[4] and cnidarian (jellyfish)[5] sources. The alternative, which also addresses variability and quality issues, is to examine production of recombinant collagens. Significant progress has been made in this area,[6] including use of fungal (eg: Pichia, Saccharomyces) expression. The development of these systems allows for the hydroxylation of proline residues, a characteristic of animal collagens, and helps in giving the product thermal stability.[7] However, the yields obtained may not be sufficient for widespread commercial use, although recent production of mammalian collagens in plants looks more promising as a commercial production direction.[8]

More recently, the collagen like (CL) domain Scl2 from *Streptococcus pyogenes* has been proposed as a potential biomedical material[9] that is non-cytotoxic and non-immunogenic.[10] This protein can be readily prepared in good yields, up to around 19 g/L in 2 L fermentation studies[11] and can be purified using a simple precipitation and proteolysis approach.[12] The protein can be modified, either by specific sequence changes to introduce specific cell binding and other functional domains[12-15] or by chemical modifications, for example to produce stabilized and hydrogel materials and other functional domains.[12-13]

These non-animal, bacterial collagens share the same characteristic triple-helical structure that is found in animal collagens and several other animal proteins.[16–18] However, the Scl2 collagen domain is small, about a quarter of the length, 234 residues,[19] of the abundant, mammalian interstitial collagens; eg type I, type II and type III collagens, which are just over 1000 residues each.[20] In order to make a bacterial collagen that is comparable in length to the mammalian interstitial collagen a tetramer of the S. pyogenes CL domain would be required. This new molecule, if available, would be useful for functional comparisons, especially in biomedical applications. Previously, an extended construct comprising 2 repeats of the *S. pyogenes* CL domain has been reported.[21] This suggested that there were few differences in

the properties of the CL and (CL)$_2$, such as melting temperature and association properties.[21]

In the present study we have made constructs comprising 1 to 4 copies of the *S. pyogenes* Scl2 collagen domain. The constructs with 4 CL domains are similar in length to mammalian interstitial collagens. To produce this tetramer, dimer and trimer molecules were produced in intermediate steps, and these have also been examined to give additional information on any changes that happen with increasing length. These same constructs were also made with a CysCys sequence at the C-terminal, analogous to the sequence found at the C-terminal of the triple-helical domain of mammalian type III collagens.[22] This allowed examination to see what, if any, effects the addition may have on collagen stability and whether this or other factors also affected yields.

Results

Gene constructs

The cloning strategy method (**Fig. 1**) used to make the multimers of the *S. pyogenes* Scl2-based collagen domain, CL, was successful, for both sets of constructs. Thus, 2 sets of 4 constructs were obtained, one set with and the other set without the C-terminal CysCys sequence, as shown schematically in **Figure 2**. Prior to expression studies, all constructs were confirmed by DNA sequencing.

Protein expression

All constructs were successfully expressed in the 2 L bioreactor system. It became clear, however, that the yields were greatest for the monomer constructs and that the yields decreased with the increasing size of the construct. Gel electrophoresis (**Fig. 3**) illustrated this falling yield with increased size. In part this was due to lower cell yields being obtained for constructs with larger inserts (**Fig. 4**), using comparable conditions for each expression. For

most constructs (excepting V-CL), data were collected for 2 fermentation experiments. The data shown in **Figure 4** show the ranges of values obtained. The lower cell yields do not fully account for the lower protein yields and expression must also have been less in those with the larger inserts (**Fig. 4**).

Sample analyses

Quantitative data were obtained by comparing the yield of products using gel electrophoresis[11] to a standard purified monomer (**Fig. 5**). These data showed a drop of yield with increasing length, with yields approximately halving with addition of each extra CL domain (**Fig. 5**). For most constructs (excepting V-CL), data were collected for 2 fermentation experiments. The data shown in **Figure 5** show the ranges of values obtained for each construct type.

Purification of expressed proteins

Protein samples were extracted from each preparation and the collagen components isolated by IMAC. Samples were examined by SDS-PAGE showing that treatment with pepsin readily removed the V-domain[12] (**Fig. 6**) and that this approach was sufficient to achieve good initial purification (**Fig. 6**). Some additional faint bands were still present, probably due to contamination by *E. coli* host proteins, although some could be due to aggregation or degradation of the collagen products (**Fig. 6**). For preparations that needed further purification, the eluate was concentrated and purified by gel permeation chromatography.

Circular dichroism (CD) spectroscopy

Purified samples were available for CD determinations in most, but not all cases. For the V-(CL)$_{1-4}$ series, the longest construct V-(CL)$_4$ was very insoluble and not enough could be dissolved to get a good signal for a reliable T_m determination; the
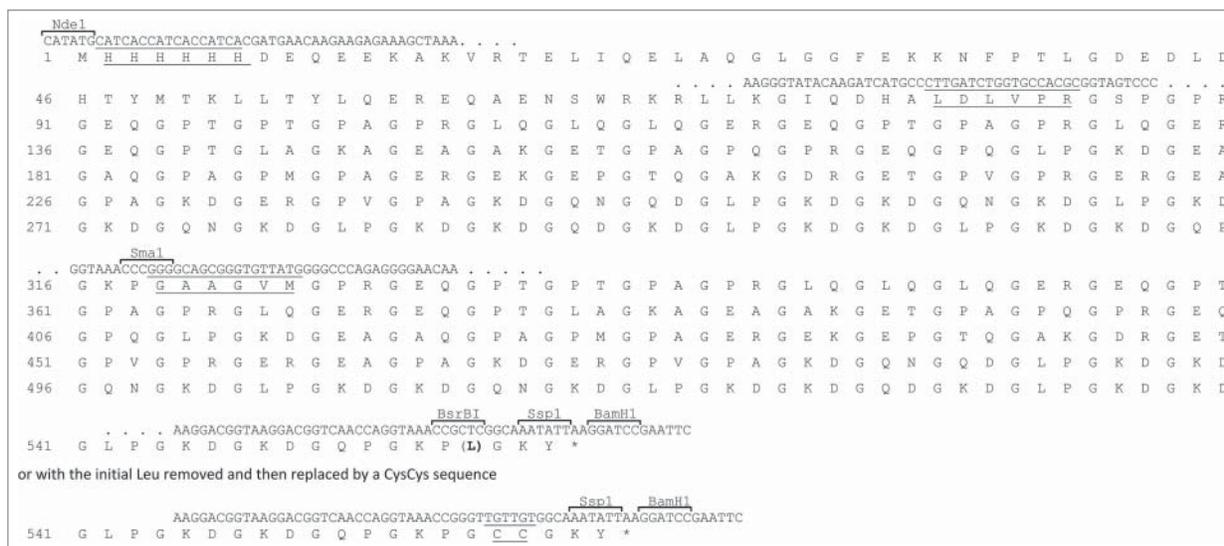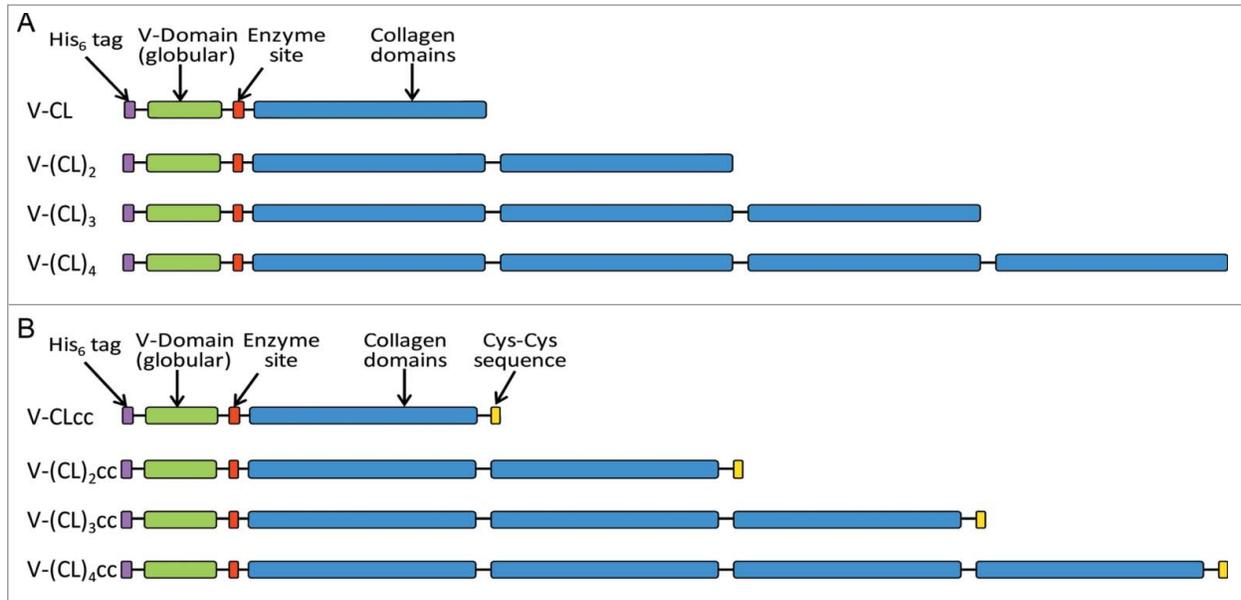

Figure 1. The protein sequence and part of the DNA sequence for the constructs used.

Figure 2. Schematic of expression constructs for multiple repeats of *S. pyogenes* Scl2 collagen domain. (**A**) Without crosslinker, and (**B**) with a C-terminal Cys-Cys crosslinking sequence.

same was the case after the V- domain was removed. For all samples with the C-terminal Cys-Cys sequence (**Fig. 2**) lack of solubility was more of an issue, with possible aggregation associated with disulphide interchange between individual molecules. For the V-(CL)$_{1-4}$cc series, the longest construct V-(CL)$_4$ cc was too

insoluble for a reliable T_m determination; after the V- domain was removed the (CL)$_4$cc sample was too insoluble to allow any T_m determination.

The CD data that could be collected for samples with adequate solubility (**Table 1**) showed V-CL with a T_m of 37.5°C. This value is about 0.7°C higher than previously reported (36.8°C).[21] A T_m was determined for V-(CL)$_2$ of 37.6°C (**Table 1**), which was also slightly higher than a value (37.1°C)
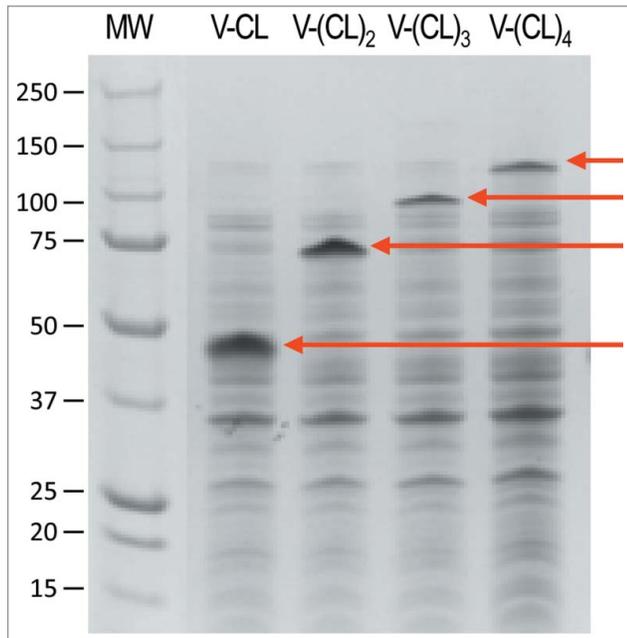
Figure 3. SDS-PAGE evaluation of extracts of samples from equal amounts of cells after completion of fermentation for the constructs V-CL, V-(CL)$_2$, V-(CL)$_3$, V-(CL)$_4$. MW indicates molecular weight markers, given as kDa.
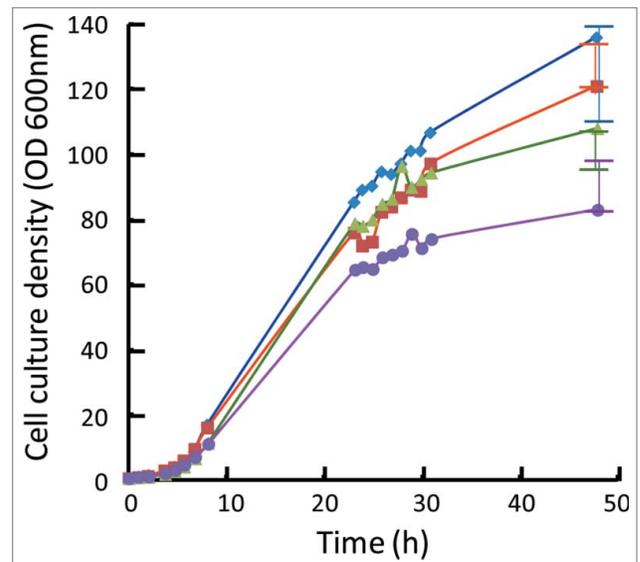
Figure 4. Cell growth for the constructs V-CL [♦], V-(CL)$_2$ [■], V-(CL)$_3$ [▲] and V-(CL)$_4$ [●], from the initial series of fermentations. The range observed for subsequent experiments is indicated, where n = 2, except for V-CL where n = 4.
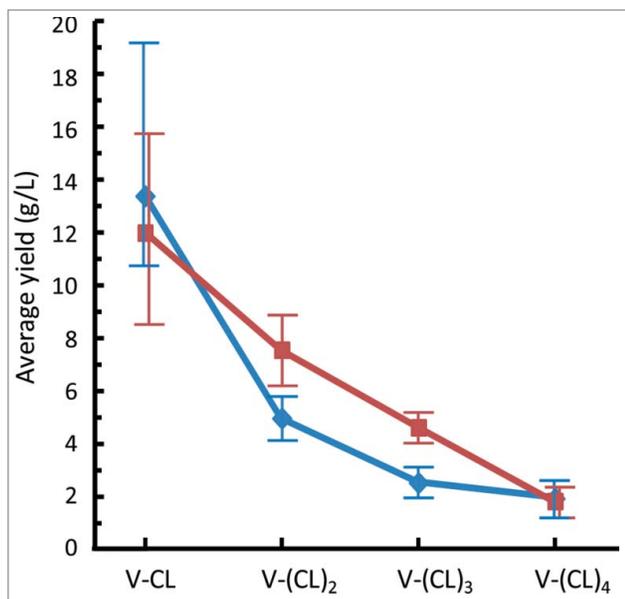
Figure 5. Average estimated protein expression yields for the constructs V-CL, V-(CL)$_2$, V-(CL)$_3$, and V-(CL)$_4$ [♦] and those modified to have a Cys-Cys sequence at the C-terminal, V-CLcc, V-(CL)$_2$cc, V-(CL)$_3$cc, and V-(CL)$_4$cc [■], shown with the ranges of observed values. Averages of values are given, where n = 2 for all constructs except for V-CL where n = 6. For n = 6, the mean and standard deviation was 13.3 ± 2.9.
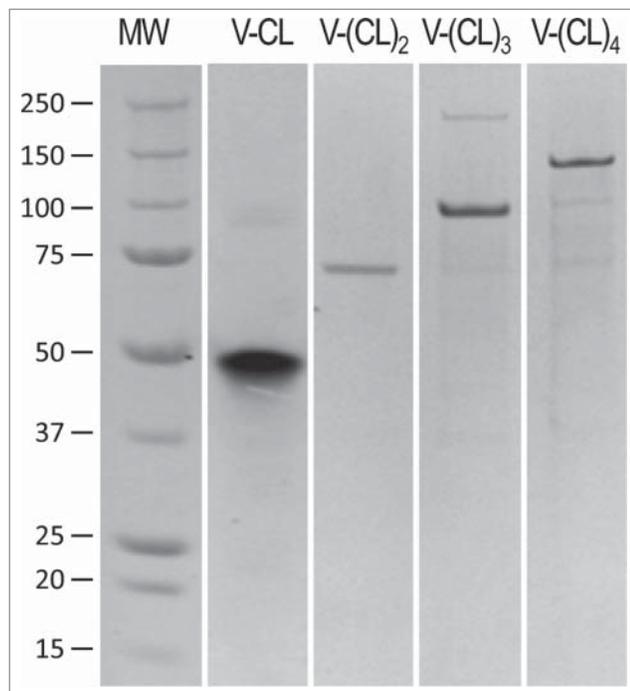
Figure 6. A collage of SDS-PAGE evaluation of samples after chromatography on Ni charged HyperCel-Sepharose metal ion affinity resin. Constructs V-CL, V-(CL)$_2$, V-(CL)$_3$, V-(CL)$_4$. MW indicates molecular weight markers, given as kDa.

Table 1. Thermal stabilities of various non-animal collagen multimer constructs

Construct	T$_m$	Construct	T$_m$
V-CL	37.5	CL	37.6
V-(CL)$_2$	37.6	(CL)$_2$	37.7
V-(CL)$_3$	37.5	(CL)$_3$	37.0
V-(CL)$_4$	(36)	(CL)$_4$	(37)
V-CLcc	40.4	CLcc	37.5
V-(CL)$_2$cc	38.1	(CL)$_2$cc	37.5
V-(CL)$_3$cc	38.0	(CL)$_3$cc	37.5
V-(CL)$_4$cc	(38.0)	(CL)$_4$cc	n.s.

Values in parentheses were considered less certain due to poor solubility (see text).
n.s. Not sufficiently soluble to obtain data.

previously reported.[21] An essentially similar value, of 37.5°C, was also obtained for V-(CL)$_3$, while for V-(CL)$_4$ a slightly lower value was obtained, but this lower value may have arisen because of the low intensity of the spectrum, caused by poor sample solubility. When the V-domain was removed from these samples by proteolysis, the T$_m$ values for CL and (CL)$_2$ remained essentially unchanged (**Table 1**). The solubility of the proteins lacking the V-domain seemed less than for those that included the V-domain, but this was not quantified. Thus, there were more difficulties getting good CD spectra for (CL)$_3$ and particularly for (CL)$_4$. The T$_m$ for (CL)$_3$ was 0.5°C lower than for V-(CL)$_3$, while the estimated T$_m$ for (CL)$_4$ was similar to that for (CL)$_3$ and was slightly higher than for the V-(CL)$_4$ (**Table 1**).

Using unreduced samples, where oxidation was achieved by exposing to air for 72 h, a T$_m$ value for V-CLcc of 40.4°C was obtained (**Table 1**), some 2.9°C higher than when the CysCys was not present. However, the T$_m$ obtained after removal of the V-domain, removing the additional stabilizing at the N-terminal, gave a product CLcc that had essentially the same stability, 37.5°C, as the construct lacking the CysCys sequence. The stabilizing effect on larger constructs containing both the CysCys sequence and the V-domain was less evident, with the T$_m$ values for V-(CL)$_2$cc and V-(CL)$_3$cc being only marginally more stable by 0.5°C than the corresponding sequences lacking the CysCys sequences (**Table 1**). As with V-CLcc, removal of the V-domain from other constructs gave products where the T$_m$ was essentially no different to the equivalent constructs lacking the CysCys sequences (**Table 1**).

Discussion

The present study has demonstrated a range of polymer constructs, up to 4 domains, for the CL domain from *S. pyogenes* Scl2, some also including a C-terminal CysCys sequence. The polymers with 4 CL units were of comparable size to interstitial collagens,[20] for example type I and II collagens and type III collagen for the V-CL$_4$cc construct. The inclusion of the C-terminal Cys-Cys sequence was selected as it is found at the C-terminal of mammalian type III collagen sequences,[22] where it can form an intra-molecular cross-linked knot that keeps the chains linked

together in register and allows type III collagen refolding if denaturation occurs.[23] It may also contribute to an increased stability for a construct, especially for shorter constructs.[24]

The protein yields of these various constructs dropped with increasing size, approximately halving with each CL addition. A similar effect has also been reported for globular proteins,[25] and covered a much larger molecular weight range of up to 6000 amino acids in the largest construct. In the present study, the expression was not optimized for potentially better yields from larger constructs, but it seems unlikely that yields such as those obtained for the monomeric form would be achieved. Thus, for recombinant expression of non-animal, bacterial protein for biomedical applications, expression of the monomer is best, and if higher molecular weight material is needed, for example to assist in fabrication, then post-purification chemical crosslinking can be used.

The thermal stability, T_m, was determined for most constructs. In some cases, for example V-CL, CL, V-(CL)$_2$ and (CL)$_2$, the T_m values were slightly higher than previously reported. This may arise from use of a different method to determine the value. In the present study a first derivative was used,[26] whereas, in other studies, the point on the melting curve at 50% melting was determined. Similar, slight differences between these methods have previously been reported, with the first derivative approach giving values up to 1°C higher.[26] The use of a first derivative to determine a mid-point of the melting curve was chosen for all determinations. In some samples, particularly those with more copies of the CL domain in the structure and those with C-terminal CysCys sequences, sample insolubility became a significant issue and led to CD spectra of low intensity. Use of a first derivative approach assisted in their interpretation.

The stability data from the present study have confirmed the previous observation that the V-domain did not enhance the triple helical stability[21] and have indicated that the T_m values were not dependent on the construct length. Previous studies on collagen-like structures have shown that a minimum length of (Gly-Xaa-Yaa)$_n$ triplets is necessary to form a stable triple helix. The smallest sequence length is for peptides with a high imino content and hydroxyproline in the Yaa position, and is around 6 triplets (ie: n = 6).[27] For peptide with this composition, the stability increases with increasing length. These data start to level of as n increases, for example when n > 12–15, with the stability fitting to an exponential curve.[27] For the CL polymers examined in the present study, increasing the length above monomer did not lead to a change in the triple helix stability (**Table 1**). Changes in T_m have been observed previously with shorter constructs. The triple helix length of collagen-like proteins from different *S. pyogenes* strains vary in length,[9] with samples containing ∼60 or fewer repeats (but still with the V-domain) showing reduced stability compared to the 78 repeat CL domain used in this study. Further variants, shorter than the present monomer CL, have also been prepared as recombinant constructs[28] and also show reduced stabilities. Longer structures from S. pyogenes, for example with 129 repeats show a similar stability to the CL and (CL)$_2$ structures of the present study.[9] Thus, for all longer constructs above ∼70 repeats, the stability is similar to that of mammalian collagens, but in the present case these recombinant proteins lack any hydroxyproline.

Lack of sample solubility was an issue for constructs containing the C-terminal Cys-Cys sequences, especially those with the V-domain removed by proteolysis. This was possibly due to disulfide exchange allowing insoluble aggregates of the constructs to form. Solubility could be increased slightly through addition of a reducing agent, but samples with this treatment were precluded from determination of T_m values as the constructs no longer had the intramolecular 'barrel' of disulphide bond found in type III collagen.[22] In the V-CLcc construct, stabilization of the triple helix at both the C-terminal end, by the disulfide 'barrel' and at the N-terminal through the V-domain, increases the overall stability. However, the T_m obtained after removal of the V-domain, removing the additional stabilizing at the N-terminal gave a product CLcc that had essentially the same stability, 37.5°C, as the construct lacking the CysCys sequence. The stabilizing effect on larger constructs containing both the CysCys sequence and the V-domain was less evident, with the T_m values for V-(CL)$_2$cc and V-(CL)$_3$cc being only marginally more stable by 0.5°C than the corresponding sequences lacking the CysCys sequences (**Table 1**). As with V-CLcc, removal of the V-domain from other constructs gave products where the T_m was essentially no different that for the equivalent constructs lacking the CysCys sequences (**Table 1**).

The CysCys sequence that has been added to the present constructs was first observed in type III collagen where through inter chain, intramolecular disulphide bonds it links all 3 collagen chains.[22] This sequence motif is also found in the N-terminal propeptide of type III collagen, along with 5 intrachain disulphide bonds, contributing to a significant increase in stability, of 19°C, when compared with fully reduced and carboxymethylated protein.[29] This sequence motif has been added to synthetic peptide constructs,[24,30] leading to notable increases in stability. In one example, where the recombinant construct (GlyProPro)$_{10}$Cys$_2$ was studied,[30] the inclusion of the CysCys sequence led to a T_m of 82°C, which is ≈50°C more than (GlyProPro)$_{10}$.[31] In the present study, where much larger constructs were studied, the enhancement of the T_m was generally low or not measurably different, except in the shortest construct, V-CLcc, when the V-domain was also present. Nevertheless, the increase in stability provided by the presence of the CysCys sequence in the V-CLcc construct may be useful in expression, production and for purification when the V-domain is still present,[11,12] whereas no extra stability is present after removal of the V-domain.

With type III collagen the presence of the disulfide 'barrel' allows refolding of the thermally denatured molecule, while refolding of other mammalian, interstitial collagen proceeds with difficulty.[23] Denatured bacterial CL sequences do not refold in the absence of a V-domain,[21] so addition of the disulfide 'barrel' to bacterial collagens may assist if refolding studies are intended on molecules where hydroxyproline is not present. In addition, the presence of the additional CysCys residues on the CL domain, with the V-domain removed, could still be an advantage as it could allow site specific modification of the collagen sequence with entities, such as sugars, non-standard amino acids

or cyclic peptides, which cannot be readily achieved by recombinant methods.

For materials for biomedical applications large scale production could be necessary. The present study suggests that only the constructs with one or 2 CL domains would likely be suitable for large scale production. The addition of the extra CL domains did not change the T_m for the larger constructs, but did reduce their solubility and ease of purification. The addition of the CysCys sequences also did not lead to a notable change in triple helical stability, although for the monomer construct, V-CLcc, a small increase was observed that may be of benefit in large scale production and gives a location for site specific chemical modification.

Materials and Methods

Gene constructs

The DNA sequence for the fragment of the *scl2.28* allele (Q8RLX7) of *S. pyogenes* encoding the combined globular and collagen-like portions of the Scl2.28 protein, but lacking the C-terminal attachment domain was as previously described.[10] A sequence for a His$_6$-tag was added at the N-terminal of the sequence. Also a thrombin/trypsin cleavage/spacer sequence, LVPRGSP, was added between the N-terminal globular domain (V) and the following (Gly-Xaa-Yaa)$_n$ collagen-like domain (CL). This sequence had unique 5′ *Nde*I and a 3′-*Bam*H1 at the ends of the full coding sequence (**Fig. 1**) as well as unique 5′ *Sma*I and 3′ *Bsr*BI sites at either end of the CL domain (**Fig. 1**). The DNA for this design was synthesized commercially with codon optimisation for expression in *E. coli* (GeneArt) (**Fig. 1**).

A second construct, comprising the above sequence, and including the same unique restriction sites, with a second CL domain linked via a GAAGVM linker sequence which was followed by the single 3′ *Bsr*BI site, was also synthesized commercially with codon optimization for expression in *E. coli* (GeneArt) (**Fig. 1**).

The DNA sequences for trimeric and tetrameric constructs of the CL domain of *S. pyogenes* with each of the adjacent CL⁻domains linked through GAAGVM linker sequences were made. These were engineered using the dimeric construct described above. This construct contained unique 5′ *Sma*I and 3′ *Bsr*BI sites which were used to excise the GAAGVM-CL C-terminal sequence from the dimeric construct. This purified insert was then cloned back, first into the dimeric sequence and subsequently into the trimeric sequence using the *Sma*I restriction site as a blunt end fragment. The bases CTC (Leu) which were part of the *Bsr*BI site were removed using site directed mutagenesis so that the C-terminal Gly-Xaa-Yaa protein sequence of the CL domain was not altered at the 3′ end of the sequence, and did not contain any interruption of the (Gly-Xaa-Yaa)$_n$ repeating sequence. This was done using oligos, such that the initial sequence,

5′ CCTGGTAAACCGCTCGGCAAATATTAAGGA
P G K P L G K Y *
had the Leu residue removed by using oligos,

5′ CCTGGTAAACCGGGCAAATATTAAGGA
P G K P G K Y *
and
3′ TCCTTAATATTTGCCCGGTTTACCAGG

Since the trimeric and subsequently the tetrameric inserts were being cloned in as a blunt end fragment, colonies were chosen, grown up in 1 × YT media (8 g tryptone, 5 g yeast extract and 2.5 g NaCl per liter) with ampicillin (100 μg/ml), midi preps were carried out to select clones that included the additional CL sequence in the correct orientation. The clones were then confirmed by sequencing. All 4 constructs were then cloned into the pColdIII vector using terminal restriction sites 5′ *Nde*I and 3′ *Bam*HI.

In addition, site directed mutagenesis was used on all 4 gene constructs to add a C-terminal CysCys sequence. These modifications all used the same oligos,

5′ CAGCCTGGTAAACCGGGTTGTTGTGGCAAATAT-TAAGGATCCGAA
Q P G K P G C C G K Y *
and
3′ TTCGGATCCTTAATATTTGCCACAACAACCCGG-TTTACCAGGCTG

After mutagenesis, the 4 new clones were sequenced to ensure that the GCC residues had been added to the 3′ end of the CL gene. Upon sequence confirmation, the terminal restriction sites 5′ *Nde*I and 3′ *Bam*HI were used to isolate the sequences which were cloned into the pCOLDIII vector.

Protein expression

DNA sequences were sub-cloned into the pColdIII vector (Takara Bio #3363) using the 5′ *Nde*I and 3′ *Bam*HI restriction sites that had been introduced at the ends of each sequence. The PCR colony screening technique was then used to detect positive clones. Selected clones for each of the 8 constructs were grown up in 100 ml culture volumes and Qiagen midi preps then used to expand the vector quantities.

For expression, selected positive clones were transformed into the *E. coli* host BL21-DE3. Single colonies of transformed cells taken from a fresh transformation plate were first grown in shaker flasks as primary seed cultures in 10 ml 2 × YT Media (16 g tryptone, 10g yeast extract and 5g NaCl per liter) containing 10 g/L glucose and 200 μg/mL ampicillin at 37°C shaking at 200 rpm for 24 h until the optical density at A600nm of the culture reached between 3–6. Samples of primary seed culture (0.5 mL) were used to inoculate 500 mL of 2xYT (in a 2 L Erlenmeyer flask) containing 10 g/L glucose and 200 μg/mL ampicillin. This secondary seed culture was incubated at 37°C shaking at 200 rpm for 16 h. These preparations were used as the inoculums for 2 L fermentation.

For production in 2L stirred tank bioreactors vessels were connected to Biostat B (Sartorius Stedim) control systems, as previously described.[12] Briefly, the initial volume of medium in the fermentor was 1.6 L and seed culture was added to attain an

A600nm of 0.25. A high cell density fed-batch process was used, with glucose used as the carbon source. The feed solution was 400 mL of 660 g/L glucose solution to which 40 mL of 1 M $MgSO_4.7H_2O$ was added. The feed flow rate was 15 mL/h, initiated 8.5 h after inoculation.

For induction of expression, the culture was cooled (over a 20 minute period) to 25°C, 24 h after inoculation. This temperature change activated the cold shock component of the vector and protein expression was induced by addition of 1mM IPTG. After 10 h incubation at 25°C, the temperature was further decreased to 15°C and followed by a further 14 h incubation. At completion, the cells, which contain the expressed products, were collected by centrifugation.[12]

Sample analyses

Samples were taken from the cultures throughout each process, as previously described.[12] Cell mass was determined by measuring the optical density of diluted samples at 600 nm. Cell pellets were lysed by addition of 40 μL of Bugbuster[TM] (Merck #71456) and incubated for 1 h. After the incubation, samples reduced by 5% (v/v) 2-mercaptoethanol were examined by SDS PAGE[32] using NuPAGE (Invitrogen NP0322) 4–12% Bis-Tris gels with MES running gel buffer, at 180V for 60 min, followed by staining with Coomassie Blue R-250. After destaining, the product bands were quantified by densitometric analysis using Multi Gauge V3.0 FujiFilm software and compared to known standard loadings of purified V-CL protein.

Purification of expressed proteins

Recombinant protein products were extracted from wet cell paste (20 ml buffer per gram) by sonication in 20 mM sodium phosphate buffer, pH8.0, using a Misonix S4000 instrument, with an Enhance Booster #1 probe, at 30 A (instrument scale) for 6 × 5 min on ice.

After centrifugation (12,000 × g for 30 min), clarified supernatant was taken to 20 mM sodium phosphate, 150 mM NaCl and 50 mM imidazole buffer, pH8.0, and absorbed onto a Ni charged HyperCel-Sepharose metal ion affinity resin (Pall Life Sciences 20093-028). Elution was by the same buffer, but containing 500 mM imidazole. Eluted fractions containing recombinant protein were pooled, concentrated and exchanged into 20 mM sodium phosphate buffer, pH8.0, using a 10kDa cross-flow filtration membrane (Pall Life Sciences #05010T12). This fraction was further purified in the same buffer on a Sephacryl S200 26/60 column (GE Healthcare #17-1195-01). When needed, samples were further dialysed against 20 mM acetic acid prior to proteolysis to remove the V-domain or for freeze drying. Samples from construct containing C-terminal CysCys sequences were allowed to stand in air for >72 to enhance disulfide bond formation. For removal of the V-domain, samples in 20 mM acetic acid, adjusted to pH 2.0 with HCl, were digested with 0.01% pepsin at 4°C for 16 h. Samples were then re-purified in 20 mM sodium phosphate buffer, pH8.0, on a Sephacryl S200 26/60 column. All steps were performed at temperatures less than the melting temperature of the triple-helix, typically at 4°C.

Sample purity was assessed by SDS PAGE[32] using NuPAGE (Invitrogen) gels as described above.

Circular dichroism spectroscopy

CD spectra were collected for purified protein samples, approximately 0.3 mg/ml (when attainable) in 20 mM sodium phosphate, pH 7.6, using 1 mm path length cells in a JASCO J-815 instrument. Melting experiments were performed from 10°C to 60°C with a temperature increase of 0.1°C per minute and measured at 220 nm. Melting temperatures (T_m) were determined from the first derivative of the melting curve.[26] Data were averaged from at least 3 separate determinations.

Accession numbers for cDNA constructs

The cDNA sequences used in this study have been lodged with the European Nucleotide Archive and have the accession numbers as indicated. The accession numbers for the original sequence on which these new constructs were based is AY069936.1

1. LM999958—a synthetic construct based on the *S. pyogenes* bacterial collagen Scl2 fragment.
2. LM999959—a synthetic construct based on 2 repeats of the CL domain from S. pyogenes bacterial collagen Scl2 fragment.
3. LM999960—a synthetic construct based on 3 repeats of the CL domain the *S. pyogenes* bacterial collagen Scl2 fragment.
4. LM999961—a synthetic construct based on 4 repeats of the CL domain the *S. pyogenes* bacterial collagen Scl2 fragment.
5. LM999962—a synthetic construct based on the *S. pyogenes* bacterial collagen Scl2 fragment, with an added C-terminal CysCys sequence.
6. LM999963—a synthetic construct based on 2 repeats of the CL domain the *S. pyogenes* bacterial collagen Scl2 fragment, with an added C-terminal CysCys sequence.
7. LM999964—a synthetic construct based on 3 repeats of the CL domain the *S. pyogenes* bacterial collagen Scl2 fragment, with an added C-terminal CysCys sequence.
8. LM999965—a synthetic construct based on 4 repeats of the CL domain the *S. pyogenes* bacterial collagen Scl2 fragment, with an added C-terminal CysCys sequence.

References

1. Ramshaw JAM, Werkmeister JA, Glattauer V. Collagen-based biomaterials. Biotechnol Genet Eng Rev 1996; 13:335-82; PMID:8948117; http://dx.doi.org/10.1080/02648725.1996.10647934

2. Ramshaw JAM, Peng YY, Glattauer V, Werkmeister JA. Collagens as biomaterials. J Mat Res Mat Med 2009; 20:S3-8; http://dx.doi.org/10.1007/s10856-008-3415-4

3. Peng YY, Werkmeister JA, Glattauer V, Ramshaw JAM. Evaluation of the immunogenicity of avian collagen for biomedical applications. J. Biomed. Materials Res 2010; 93A:1235-44.

4. Hoyer B, Bernhardt A, Heinemann S, Stachel I, Meyer M, Gelinsky M. Biomimetically mineralized salmon collagen scaffolds for application in bone tissue engineering. Biomacromolecules 2012; 13:1059-66; PMID:22264350; http://dx.doi.org/10.1021/bm201776r

5. Addad S, Exposito JY, Faye C, Ricard-Blum S, Lethias C. Isolation, characterization and biological evaluation of jellyfish collagen for use in biomedical applications. Mar Drugs 2011; 9:967-83; PMID:21747742; http://dx.doi.org/10.3390/md9060967

6. Werkmeister JA, Ramshaw JAM. Recombinant protein scaffolds for tissue engineering. Biomed Mater 2012; 7:012002; PMID:22262725; http://dx.doi.org/10.1088/1748-6041/7/1/012002

7. Rosenbloom J, Harsch M, Jimenez S. Hydroxyproline content determines the denaturation temperature of chick tendon collagen. Arch Biochem Biophys 1973; 158:478-84; PMID:4592982; http://dx.doi.org/10.1016/0003-9861(73)90539-0

8. Shoseyov O, Posen Y, Grynspan F. Human recombinant type I collagen produced in plants. Tissue Eng Part A 2013; 19:1527-33; PMID:23252967; http://dx.doi.org/10.1089/ten.tea.2012.0347

9. Han R, Zwiefka A, Caswell CC, Xu Y, Keene DR, Lukomska E, Zhao Z, Höök M, Lukomski S. Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. Appl Microbiol Biotechnol 2006; 72:109-15; PMID:16552563; http://dx.doi.org/10.1007/s00253-006-0387-5

10. Peng YY, Yoshizumi A, Danon SJ, Glattauer V, Prokopenko O, Mirochnitchenko O, Yu Z, Inouye M, Werkmeister JA, Brodsky B, et al. A Streptococcus pyogenes derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. Biomaterials 2010; 31:2755-61; PMID:20056274; http://dx.doi.org/10.1016/j.biomaterials.2009.12.040

11. Peng YY, Howell L, Stoichevska V, Werkmeister JA, Dumsday GJ, Ramshaw JAM. Pilot production of a collagen-like protein from Streptococcus pyogenes for biomedical applications. Microbial Cell Fact 2012; 11:146; http://dx.doi.org/10.1186/1475-2859-11-146

12. Peng YY, Stoichevska V, Madsen S, Howell L, Dumsday GJ, Werkmeister JA, Ramshaw JAM. A simple cost effective methodology for large scale purification of recombinant non-animal collagens. Appl Micro Biotech 2014; 98:1807-15; http://dx.doi.org/10.1007/s00253-013-5475-8

13. Cosgriff-Hernandez E, Hahn MS, Russell B, Wilems T, Munoz-Pinto D, Browning MB, Rivera J, Höök M. Bioactive hydrogels based on designer collagens. Acta Biomater 2010; 6:3969-77; PMID:20466083; http://dx.doi.org/10.1016/j.actbio.2010.05.002

14. An B, DesRochers TM, Qin G, Xia X, Thiagarajan G, Brodsky B, Kaplan DL. The influence of specific binding of collagen-silk chimeras to silk biomaterials on hMSC behavior. Biomaterials 2013; 34:402-12; PMID:23088839; http://dx.doi.org/10.1016/j.biomaterials.2012.09.085

15. An B, Abbonante V, Yigit S, Balduini A, Kaplan DL, Brodsky B. Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. J Biol Chem 2014; 289:4941-51; PMID:24375478; http://dx.doi.org/10.1074/jbc.M113.530808

16. Brodsky B, Ramshaw JAM. The collagen triple-helix structure. Matrix Biol 1997; 15:545-54; PMID:9138287; http://dx.doi.org/10.1016/S0945-053X(97)90030-5

17. Yu Z, An B, Ramshaw JA, Brodsky B. Bacterial collagen-like proteins that form triple-helical structures. J Struct Biol 2014;186:451-61; PMID:24434612; http://dx.doi.org/10.1016/j.jsb.2014.01.003

18. Ramshaw JA, Werkmeister JA, Dumsday GJ. Bioengineered collagens: Emerging directions for biomedical materials. Bioengineered 2014; 5:227-426; [Epub ahead of print]; PMID:24717980; http://dx.doi.org/10.4161/bioe.28791

19. Xu Y, Keene DR, Bujnicki JM, Höök M, Lukomski S. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem 2002; 277:27312-18; PMID:11976327; http://dx.doi.org/10.1074/jbc.M201163200

20. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol 2011; 3:a004978; PMID:21421911; http://dx.doi.org/10.1101/cshperspect.a004978

21. Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JAM, Inouye M, Brodsky B. Self-association of Streptococcus pyogenes collagen-like constructs into higher order structures. Protein Sci 2009; 18:1241-51; PMID:19472339; http://dx.doi.org/10.1002/pro.134

22. Glanville RW, Allmann H, Fietzek PP. Cysteine residues mark the C-terminal triple-helicalnon-triple-helical junction in type III collagen. Hoppe Seylers Z Physiol Chem 1976; 357:1663-65; PMID:1033911

23. ChandraRajan J. Separation of type III collagen from type I collagen and pepsin by differential denaturation and renaturation. Biochem Biophys Res Commun 1978; 83:180-86; PMID:358975; http://dx.doi.org/10.1016/0006-291X(78)90414-X

24. Barth D, Kyrieleis O, Frank S, Renner C, Moroder L. The role of cystine knots in collagen folding and stability, part II. Conformational properties of (Pro-Hyp-Gly)n model trimers with N- and C-terminal collagen type III cystine knots. Chemistry 2003; 9:3703-14; PMID:12898697; http://dx.doi.org/10.1002/chem.200304918

25. Tsung K, Inouye S, Inouye M. Factors affecting the efficiency of protein synthesis in Escherichia coli. Production of a polypeptide of more than 6000 amino acid residues. J Biol Chem 1989; 264:4428-33; PMID:2538444

26. Venugopal MG, Ramshaw JA, Braswell E, Zhu D, Brodsky B. Electrostatic interactions in collagen-like triple-helical peptides. Biochemistry 1994; 33:7948-56; PMID:8011657; http://dx.doi.org/10.1021/bi00191a023

27. Persikov AV, Ramshaw JAM, Brodsky B. Prediction of collagen stability from amino acid sequence. J Biol Chem 2005; 280:19343-49; PMID:15753081; http://dx.doi.org/10.1074/jbc.M501657200

28. Yu Z, Brodsky B, Inouye M. Dissecting a bacterial collagen domain from Streptococcus pyogenes: sequence and length-dependent variations in triple helix stability and folding. J Biol Chem 2011; 286:18960-68; PMID:21454494; http://dx.doi.org/10.1074/jbc.M110.217422

29. Bruckner P, Bächinger HP, Timpl R, Engel J. Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix ⇌ coil transition. Eur J Biochem 1978; 90:595-603; PMID:710449; http://dx.doi.org/10.1111/j.1432-1033.1978.tb12640.x

30. Boudko S, Frank S, Kammerer RA, Stetefeld J, Schulthess T, Landwehr R, Lustig A, Bächinger HP, Engel J. Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. J Mol Biol 2002; 317:459-70; PMID:11922677; http://dx.doi.org/10.1006/jmbi.2002.5439

31. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B. Triple-helix propensity of hydroxyproline and fluoroproline: comparison of host-guest and repeating tripeptide collagen models. J Am Chem Soc 2003; 125:11500-01; PMID:13129344; http://dx.doi.org/10.1021/ja036673+

32. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970; 227:680-85; PMID:5432063; http://dx.doi.org/10.1038/227680a0