

RESEARCH ARTICLE

Purification of recombinant bacterial collagens containing structural perturbations

Sonal Gahlawat¹, Vikas Nanda^{2,3}, David I. Shreiber^{1*}

1 Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, United States of America, **2** Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ, United States of America, **3** Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, NJ, United States of America

* shreiber@soe.rutgers.edu**OPEN ACCESS**

Citation: Gahlawat S, Nanda V, Shreiber DI (2023) Purification of recombinant bacterial collagens containing structural perturbations. PLoS ONE 18(5): e0285864. <https://doi.org/10.1371/journal.pone.0285864>

Editor: Yong Wang, Zhejiang University College of Life Sciences, CHINA

Received: March 7, 2023

Accepted: May 2, 2023

Published: May 17, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0285864>

Copyright: © 2023 Gahlawat et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This research was supported by the New Jersey Health Foundation (Grant/Award Number:

Abstract

Streptococcus pyogenes-derived recombinant bacterial collagen-like proteins (CLPs) are emerging as a potential biomaterial for biomedical research and applications. Bacterial CLPs form stable triple helices and lack specific interactions with human cell surface receptors, thus enabling the design of novel biomaterials with specific functional attributes. Bacterial collagens have been instrumental in understanding collagen structure and function in normal and pathological conditions. These proteins can be readily produced in *E. coli*, purified using affinity chromatography, and subsequently isolated after cleavage of the affinity tag. Trypsin is a widely used protease during this purification step since the triple helix structure is resistant to trypsin digestion. However, the introduction of Gly→X mutations or natural interruptions within CLPs can perturb the triple helix structure, making them susceptible to trypsin digestion. Consequently, removing the affinity tag and isolating collagen-like (CL) domains containing mutations is impossible without degradation of the product. We present an alternative method to isolate CL domains containing Gly→X mutations utilizing a TEV protease cleavage site. Protein expression and purification conditions were optimized for designed protein constructs to achieve high yield and purity. Enzymatic digestion assays demonstrated that CL domains from wild-type CLPs could be isolated by digestion with either trypsin or TEV protease. In contrast, CLPs containing Gly→Arg mutations are readily digested by trypsin while digestion with TEV protease cleaved the His6-tag, enabling the isolation of mutant CL domains. The developed method can be adapted to CLPs containing various new biological sequences to develop multifunctional biomaterials for tissue engineering applications.

Introduction

Over the past two decades, recombinant bacterial collagen from *Streptococcus pyogenes* has evolved into an indispensable biomaterial to gain insight into the sequence-structure-function relationship of normal and pathological collagens. *S. pyogenes* produces collagen-like

PC101-20, <https://www.njhealthfoundation.org>) and The Marfan Foundation (Grant/Award Number: 127611, <https://marfan.org>) awarded to D.I.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

proteins (also known as Scl2 or CLPs) that contain an N-terminal signal sequence, a variable globular domain (V-domain), a collagen-like (CL) triple helix domain consisting of repeating units of (Gly-X-Y)_n, and a C-terminal cell wall attachment domain [1]. The V-domain contains a coiled-coil motif and is essential for the trimerization and proper folding of the triple helix [2]. The recombinant CLPs can self-assemble into a stable triple helix structure even in the absence of hydroxyproline residues. The high content of proline residues along with charged and polar residues contribute to the stability of the triple helix via electrostatic interactions [3, 4]. As such, recombinant CLPs exhibit a high thermal stability of 35–39°C, comparable to animal collagen [5, 6]. Unlike animal collagens, CLPs do not inherently bind to human cell surface receptors and other components of the extracellular matrix (ECM) to elicit specific responses [7, 8]. Consequently, they can provide an effective biological control, allowing selective integration of specific human collagen ligand-binding sites to study specific protein-collagen interactions via recombinant DNA technology [9–13]. Hence, engineered CLPs can be used to probe the structure and function of animal collagen, including single amino acid substitutions and natural interruptions to mimic disease-causing mutations in collagen disorders.

Recombinant CLPs can be easily expressed in *E. coli* to achieve high protein expression using a cold-shock system that selectively induces expression of the target protein at low temperatures [14]. One of the most common methods to purify these bacterial CLPs is to include a polyhistidine (His6) tag at the N-terminal of the protein, followed by protein purification using immobilized metal affinity chromatography (IMAC) that utilizes the interaction between His6-tagged CLP and immobilized Ni²⁺ ions supplied by the chromatography matrix. Previous applications of CLPs to study triple helix structure and function all included a trypsin protease site to remove the His6-tag to obtain CL domains from the full-length designed construct [3, 10, 11, 13, 15–18]. This was motivated by the established proteolytic resistance of a native collagen triple helix to digestion by trypsin, pepsin, and chymotrypsin (Fig 1) [3, 5, 8, 16, 19]. However, a triple helix with disruptions and a partially folded structure is sensitive to digestion [20, 21]. Therefore, recombinant CLPs containing natural interruptions and mutations that could potentially disrupt the triple helical structure become susceptible to trypsin digestion (Fig 1) [22–27]. As a result, it is very challenging to remove the His6-tag or any other affinity tag without disrupting the triple helix structure of the proteins, and researchers who have used CLPs to study these kinds of interruptions forgo the cleavage step and leave the His6-tag intact.

We developed an efficient and reproducible purification protocol to isolate recombinant CLPs with structural perturbations by incorporating a Tobacco Etch Virus (TEV) protease cleavage site. Structurally distorted triple helices were generated through the introduction of Gly→Arg mutations within the integrin-binding site of CL domains. The Gly→Arg mutation would provide differential folding effects and has been associated with collagen disorders, including Osteogenesis Imperfecta and Ehlers-Danlos Syndrome and its subtypes (<https://eds.gene.le.ac.uk/>). The TEV protease is responsible for cleaving the His-tag from the designed full-length construct, while trypsin can be used to investigate the structural integrity of the triple helix (Fig 1). Optimization of protein expression and purification was performed to maximize bacterial CLPs production. Several parameters were examined that could affect protein expression, including bacterial host strain, culture media, and inducer concentration. The results indicate that this method can be applied to express and purify recombinant CLPs with sequences of varied molecular flexibility to obtain CL domains for downstream applications, including biophysical characterization of mutant proteins and study of cell function and behavior due to the included mutation(s).

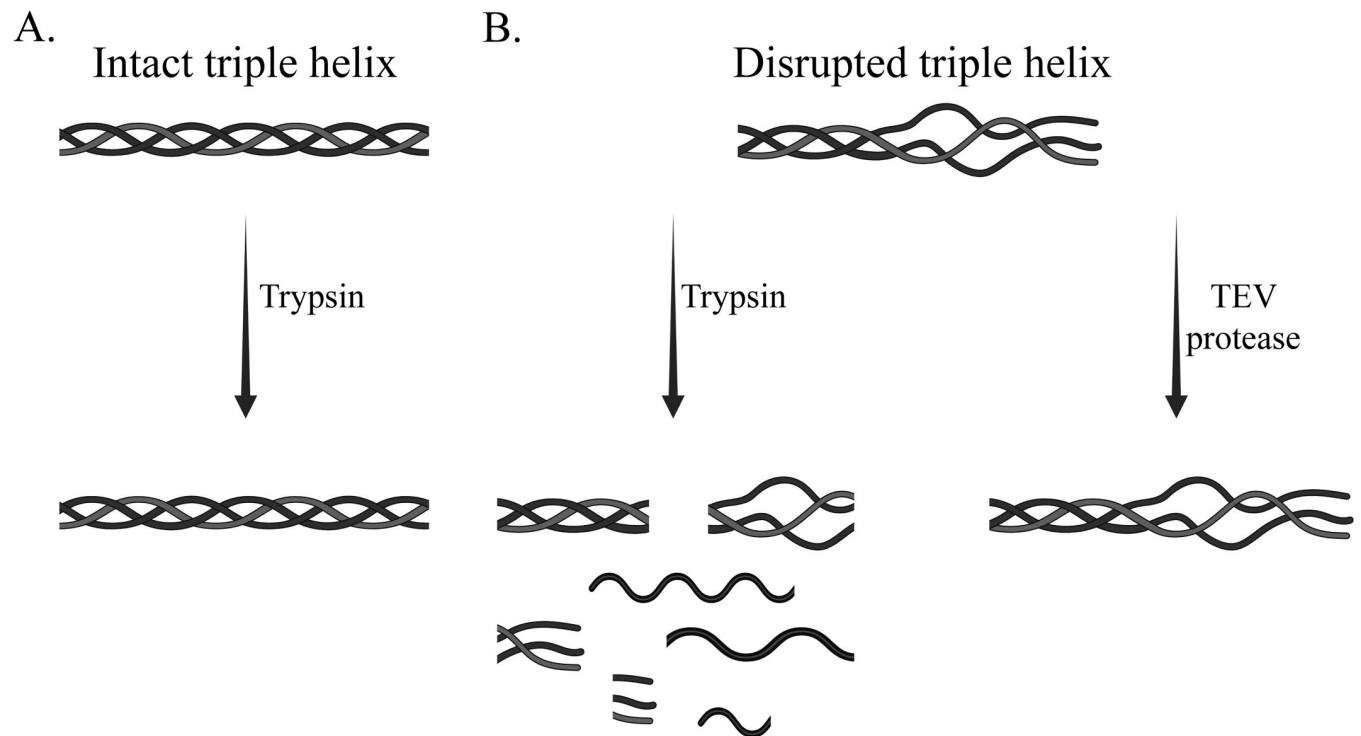


Fig 1. Schematic illustration of disrupted triple helix digestion by trypsin and TEV protease. (A) An intact triple helix is resistant to digestion by trypsin. (B) Due to mutations and natural interruptions, the triple helix become susceptible to trypsin digestion, which might result in protein degradation. In contrast, TEV protease does not degrade disrupted triple helix, enabling the isolation of mutant triple helices for more precise characterization and functional studies and for use as a potential biomaterial.

<https://doi.org/10.1371/journal.pone.0285864.g001>

Materials and methods

Chemicals and reagents

Sodium phosphate monobasic (AC389870025), sodium phosphate dibasic (S374-500), imidazole (AC122025000), carbenicillin (BP26485), tryptone (BP1421-2), yeast extract (BP9727-2), and Luria Broth (LB) Agar (BP9724-500) were purchased from Fisher Scientific. Dialysis tubing cellulose membrane (D9527-100FT), acetic acid (695092-2.5L), trypsin (T1426-50MG), and Amicon® Ultra-15 Centrifugal Filter Unit (UFC901024) were purchased from Millipore Sigma. Sodium Chloride (0241-2.5KG) was purchased from VWR. Ni-NTA Agarose (30210) was purchased from Qiagen. 4X Laemmli sample buffer (1610737), Precision Plus protein dual color standard (1610374), 4-15% Mini-PROTEAN® TGX™ Precast Protein Gels (4561086), and 2-mercaptoethanol (1610710) were purchased from Bio-Rad Laboratories. BL21-DE3 competent cells (C25271) were purchased from New England BioLabs. Isopropylthio- β -galactoside, IPTG, (I2481C50) was purchased from Gold Biotechnology. Collagen-I (C857) was purchased from Elastin Products Company, Inc. pCold-I expression vector was purchased Takara Bio Inc.

Design of recombinant bacterial CLPs

The gene constructs used for bacterial CLPs are based on the DNA sequence for the fragment of the Scl2.28 allele of *S. pyogenes* encoding for the combined globular and collagen-like proteins but lacking the C-terminal domain, as previously described [1, 3, 5, 8, 24]. In brief, constructs included an N-terminal globular domain (V) followed by two CL domains. An enzymatic cleavage site was introduced between the V-domain and the two following CL

domains to facilitate isolation of the CL domains. Constructs contained either the trypsin cleavage sequence (LVPR↓GSP) or the TEV protease sequence (ENLYFQ↓G). This DNA sequence, termed VCL₂, was synthesized commercially with codon optimization for expression in *E. coli* (GenScript, Piscataway, New Jersey), and served as the template for creating the mutant construct with a Gly-to-Arg mutation at residue #201 and #444 within the IBS of each CL domain, which was denoted VCL₂(G→R). The sequences of the final constructs were confirmed by sequencing prior to transformation and protein expression. The final DNA sequences of all recombinant bacterial collagens were cloned into the pCold-I vector systems (GenScript, Piscataway, New Jersey) for expression in *E. coli*. The constructs also included an N-terminal His6-tag that was provided by the pCold-I vector for protein purification.

Expression and purification of the proteins

All constructs were transformed into *E. coli* BL21-DE3 strain. For protein expression, a positive clone was selected for the primary culture and was grown in LB media with 100 µg/mL ampicillin at 37°C with shaking at 250 RPM for overnight. The following day, the primary culture was used to inoculate 1L of 2X YT medium (10 gm/L yeast extract, 16 gm/L tryptone, and 5 gm/L NaCl, pH 7.4) with ampicillin (100 µg/mL). The cells were grown at 37°C with shaking until the A600 absorbance reading reached an optical density in the range of 0.4 to 0.6 A.U. Subsequently, the cells were cooled down to 25°C for 1 hour, and 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added for protein induction. After 10 hours of protein induction, the cells were further cooled to 15°C for 14 hours, after which the cells were harvested by centrifugation (4,000 RPM, 20 mins, 4°C). The cell pellet was stored at -80°C until further use.

For protein extraction, the cell pellet was thawed and resuspended in lysis buffer, and the cells were ruptured by sonication in ice water for 10 mins (30 sec ON and 30 sec OFF cycle at 40% amplitude). Clarified lysate was obtained by centrifugation at 18,000 RPM for 30 mins at 4°C. The His-tagged recombinant proteins were purified by immobilized metal affinity chromatography (IMAC) using Ni²⁺ charged affinity resin (Ni-NTA, Qiagen, Germantown, Maryland). To aid in the binding of Ni²⁺ with His-tagged protein, the clarified lysate was incubated with Ni-NTA, pre-equilibrated with binding buffer with gentle shaking at 4°C. (NOTE: for a cell pellet obtained from 500 mL bacterial culture, 2 gravity columns were used for the purification with each column containing approximately 5 mL of Ni-NTA resin.) After an hour of incubation, the column was washed with (i) 20 column volumes (CV) of binding buffer (50 mM sodium phosphate buffer, 500 mM of sodium chloride, and 10 mM of imidazole, pH 8), (ii) 20 CV of high-salt buffer (50 mM sodium phosphate buffer, 1 M of sodium chloride, and 10 mM of imidazole, pH 8), and (iii) 15–20 CV of low imidazole buffer (50 mM sodium phosphate buffer, 500 mM of sodium chloride, and 50 mM of imidazole, pH 8) to remove the non-specific proteins bound to the resin. Finally, the protein was eluted in a step gradient of imidazole concentration (100 mM, 250 mM, and 500 mM and pH 8) with three fractions (10 mL each) collected for each concentration. The purification buffers contained β-mercaptoethanol (BME) as the reducing agent to prevent any disulfide bonds formation. All fractions containing the protein were collected and verified for purity level with SDS-PAGE. Eluted fractions containing the protein were pooled and dialyzed against 5 mM sodium phosphate buffer, pH 7.4 for 48 hours with three buffer changes. The dialyzed protein was lyophilized and stored at -80°C.

Enzymatic digestion assay

Enzymatic digestion assays were performed to examine the structural stability of triple helix. All recombinant proteins were digested with either trypsin or TEV protease at a ratio of 1:25

and 1:10, respectively, at 20°C for 2 mins, 5 mins, 15 mins, or overnight (16 hours). After digestion, the reaction was stopped by adding 4X SDS sample buffer (3:1 ratio) and heating the samples to 95°C for 5 mins. The digestion profile of the constructs was analyzed by SDS-PAGE.

Results

Design and expression of recombinant constructs with trypsin and TEV protease sites

The control protein, referred to as VCL₂, includes an N-terminal V-domain and two tandem CL domains (CL₂), which are separated by the linker sequence GAAGVM (Fig 2A). Each of the two CL domains included an integrin binding site (IBS), GFPGER, integrated within the triple helix structure for cell binding. GGPCPPC sequences were inserted at the N- and C-terminal of the CL domains to aid in stability and allow subsequent functionalization for downstream applications. The VCL₂(G→R) mutant protein includes a single Gly→Arg mutation (GFPRER) near the middle of the CL domain within the IBS sequence to have minimum impact on the folding and assembly of the triple helix (Fig 2B).

Lastly, the introduction of two different enzymatic cleavage sites between the V-domain and CL₂ domains allowed us to remove His6-tag and isolate CL domains from the full-length

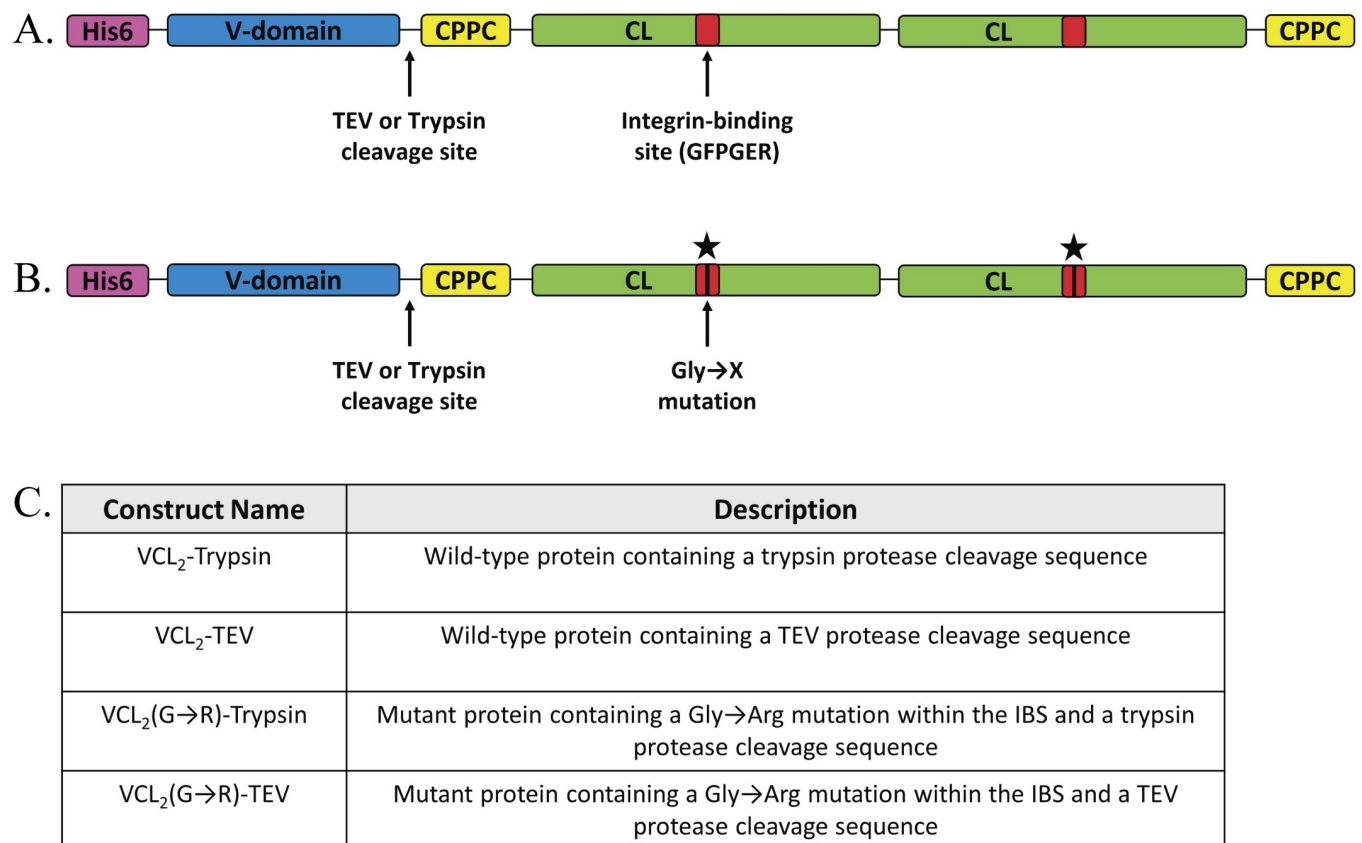


Fig 2. Schematic illustration of the design of recombinant bacterial collagens. (A) VCL₂ with a protease cleavage site (highlighted by “↑”) and an IBS within each CL domain. (B) Version of the VCL₂ construct with Gly→Arg within the IBS, where the second Gly residue of GFPGER was mutated to Arg (highlighted by “★”). The ‘CPPC’ domains represent the amino acid sequences inserted at the N- and C- terminal of the CL domains to aid in stability. (C) Description of the designed protein constructs.

<https://doi.org/10.1371/journal.pone.0285864.g002>

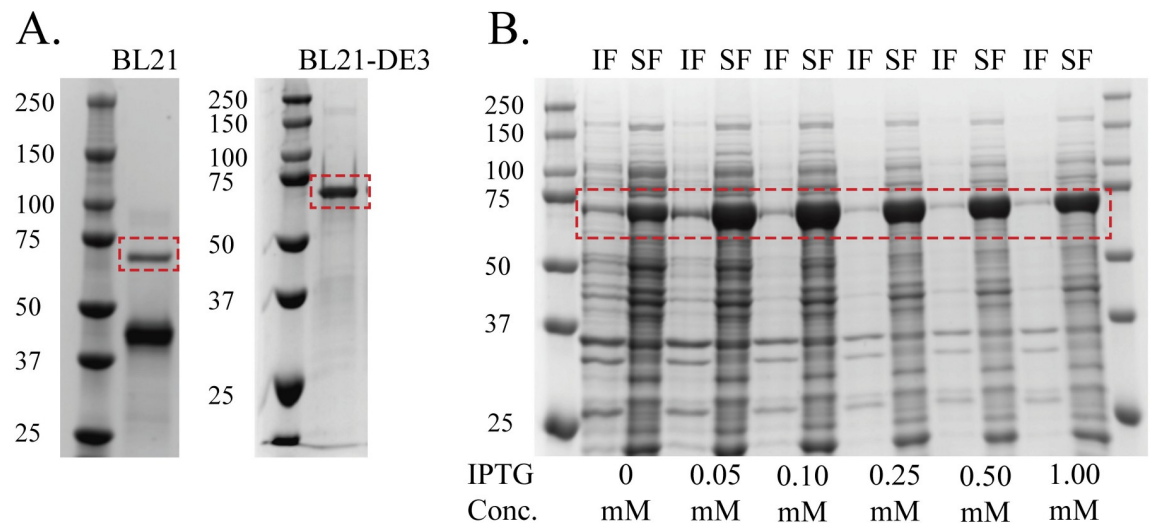


Fig 3. Protein expression of VCL₂ was optimized to achieve high expression and protein solubility. (A) SDS-PAGE of the VCL₂ purified proteins expressed in *E. coli* BL21 and BL21-DE3. Protein expression in BL21-DE3 cells resulted in purification of VCL₂ with high purity, without any degradation products. (B) SDS-PAGE of insoluble fractions (IF) and soluble fractions (SF) of VCL₂ after induction with varying IPTG concentrations, where lower IPTG concentrations induced high protein expression. Molecular Weight standards are in kDa. Conc. = Concentration.

<https://doi.org/10.1371/journal.pone.0285864.g003>

constructs. Four different types of constructs were designed for protein expression and purification: VCL₂-Trypsin, VCL₂-TEV, VCL₂(G→R)-Trypsin, and VCL₂(G→R)-TEV (Fig 2C).

We used the cold-shock pCold-I vector system to express and purify recombinant CLPs with and without specific Gly→X mutations. Recombinant protein overexpression at low temperatures often improves protein solubility and stability [14]. We optimized the protein expression by comparing different *E. coli* strains (BL21 vs. BL21-DE3), culture media (TB vs. 2X YT), and inducer concentration (0–1 mM). We initially utilized the BL21 bacterial strain for the expression of the VCL₂ protein, since protein expression was driven by the cold-shock promoter, unlike a traditional T7 RNA polymerase-based vector, where protein expression is strictly controlled by the inducer. Purification of BL21-expressed protein resulted in target protein and a dominant lower molecular weight band that is likely a CL₂ degradation product (Fig 3A). By switching to BL21-DE3 bacterial strain, we nearly eliminated the second band (Fig 3A).

To improve expression and solubility, we tested different IPTG concentrations. High protein expression was obtained using lower IPTG concentrations (0.05 and 0.01 mM, Fig 3B) and inducing protein expression at the mid-exponential growth phase using 2X YT media vs. TB media. Protein expression using TB media led to high protein loss in inclusion bodies, which could not be recovered using detergents. To identify the optimal conditions for protein expression, we selected conditions that led to high protein expression in the soluble fraction and a low expression in the insoluble fraction. Testing various IPTG concentrations allowed us to identify an optimal level of inducer for protein expression with minimal protein loss in inclusion bodies. The combination of low temperatures and lower IPTG concentrations enhanced both protein expression and solubility.

A bench-scale purification of all constructs was carried out at 4°C. The His6-tag at the N-terminal allowed protein purification by IMAC using a Ni-NTA resin, resulting in pure samples for each protein construct as verified by SDS-PAGE (Fig 4). We also utilized an imidazole gradient to obtain a high yield with maximum purity. We used three concentrations of imidazole (100 mM, 250 mM, and 500 mM) to elute the protein and determine the optimal

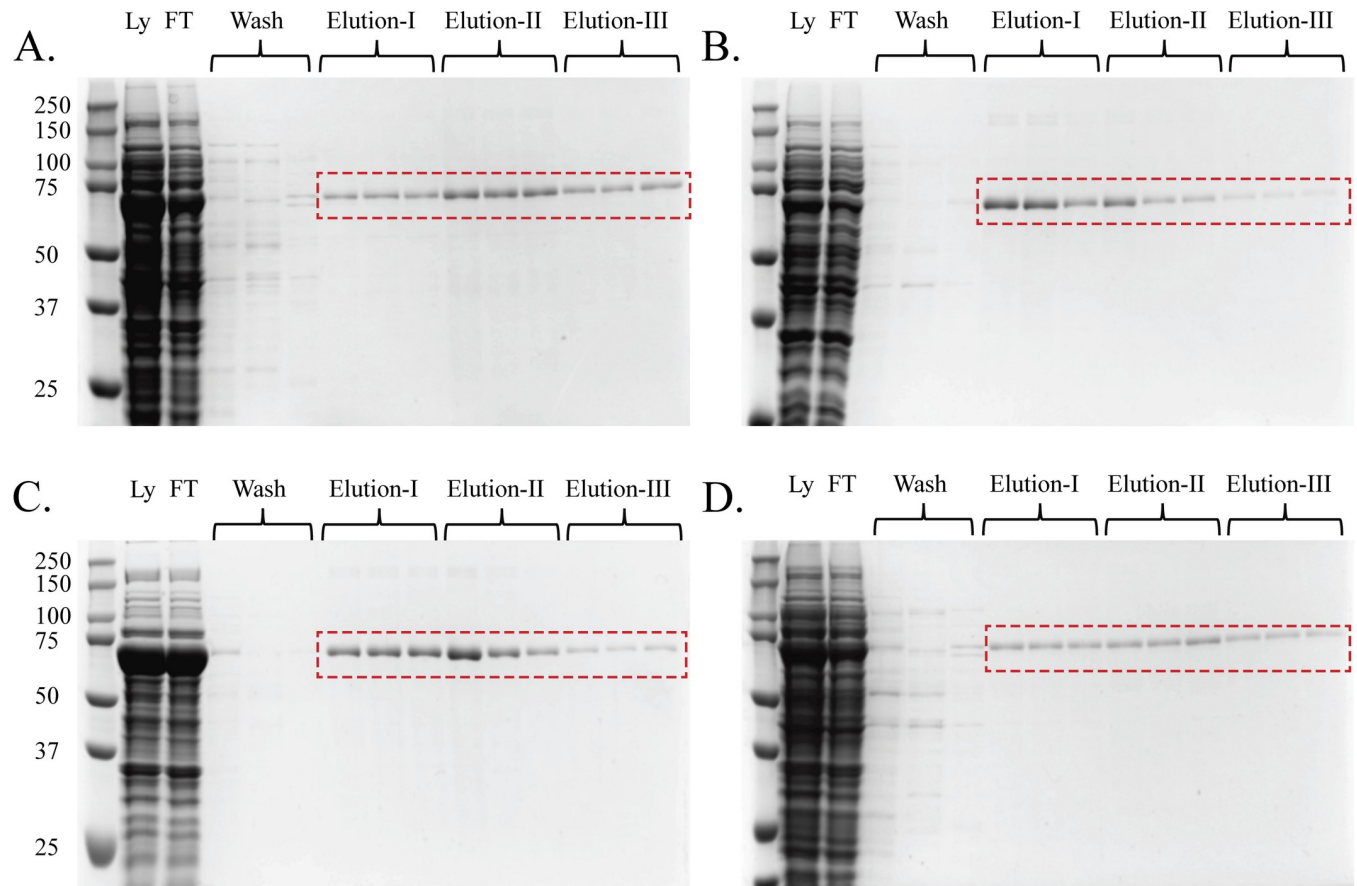


Fig 4. Ni-NTA purification of expressed proteins lead to high purity. SDS-PAGE of the purified proteins, (A) VCL₂-Trypsin, (B) VCL₂-TEV, (C) VCL₂(G→R)-Trypsin, and (D) VCL₂(G→R)-TEV. Cell lysate was incubated with Ni-NTA resin and FT was collected. The column was washed with binding buffer, high salt buffer, and low imidazole buffer to remove non-specific bound proteins. His-tagged protein was eluted with buffer containing 100 mM (Elution-I), 250 mM (Elution-II), and 500 mM (Elution-III) imidazole with three fractions collected per concentration. Molecular weight standards in kDa, Ly = lysate, and FT = Flowthrough (containing the unbound protein).

<https://doi.org/10.1371/journal.pone.0285864.g004>

concentration for elution of the protein. Three different fractions were collected for each imidazole concentration to separate contaminants not removed during washing and to achieve full recovery of the target protein.

The monomeric molecular weight of VCL₂-Trypsin, VCL₂-TEV, VCL₂(G→R)-Trypsin, and VCL₂(G→R)-TEV are 57.24 kDa, 57.41 kDa, 57.55 kDa, and 57.61 kDa, respectively. On SDS-PAGE, the protein bands for monomeric chains were higher than expected, which has previously been shown for proteins with rod-like structures [3]. The protein purity was sufficient to examine the effectiveness of each protease to isolate CL domains with and without Gly→X mutations.

High protein expression in all constructs was seen, as demonstrated by the high-intensity band in loaded lysates on the SDS-PAGE. Such high protein expression led to inefficient binding of the protein with Ni-NTA resin, resulting in protein elution during sample loading, despite a variety of attempts to improve the binding efficiency, including optimization of the binding buffer with varying concentrations of imidazole and detergents, purification under denaturing conditions, changing the incubation time between the protein and resin, and increasing the volume of the resin. Recovery of the protein lost in the flow-through was easily achieved by performing multiple rounds of purification using the collected flow-through.

Alternatively, larger-sized gravity columns or high-performance liquid chromatography (HPLC) instruments can be utilized to attain high yields in one-step purification.

Structural perturbation in triple helix increases susceptibility to trypsin digestion

Trypsin digestion assays are one of the most common techniques employed to examine the structural stability of triple helices. Mutations that disrupt the triple helix conformation can increase proteolytic susceptibility. As expected, trypsin digestion of VCL₂-Trypsin cleaves the His6-tag and V-domain, allowing the isolation of CL₂ protein (Fig 5A). The presence of an extra band is observed, which suggests digestion of a small fraction of unfolded CL₂. However, the introduction of Gly→Arg within the IBS of bacterial collagen, VCL₂(G→R)-Trypsin, increased sensitivity to trypsin digestion with detectable CL₂ cleavage products after only 2 minutes (Fig 5B). This is consistent with the highly disruptive nature of Gly→Arg substitutions within the triple helix [22, 28–30]. After overnight digestion with trypsin, the VCL₂(G→R)-Trypsin protein is entirely degraded, hampering purification and significantly reducing the final yield of CL₂(G→R).

TEV protease sites facilitate purification of CL domains containing structural perturbations

To minimize proteolysis of CL₂(G→R) during purification, we synthesized VCL₂ and VCL₂(G→R) constructs containing the TEV protease cleavage site. The digestion profile of VCL₂-TEV with TEV protease was similar to that of VCL₂-Trypsin digested with trypsin (Fig 5C). While trypsin digestion of VCL₂(G→R)-Trypsin produced multiple degradation products, TEV protease digestion of VCL₂(G→R)-TEV only produced the His6-tag + V-domain and the target CL₂(G→R) (Fig 5D). As expected, trypsin digestion of VCL₂(G→R)-TEV produced multiple degradation products (Fig 5E). These results confirm that the purification of intact CL domains with G→X mutations can be achieved by replacing trypsin with TEV protease. Replacing the trypsin cleavage site with TEV offers a reliable method to isolate CL domains containing structural perturbations irrespective of the type of substituted amino acid, the location of the mutation, and the surrounding local sequence, all of which play a decisive role in determining the conformation and stability of the triple helix.

Discussion

Given that CLPs are increasingly used in collagen-based biomaterials and as tools for understanding collagen structure and function [31–36], it is critical to develop effective methods for purifying and studying them. The inclusion of the His6-tag at the N-terminal sequence allowed the purification of expressed CLPs using IMAC, which can be easily scaled up. However, isolating CL domains and removing the His6-tag (or any other affinity tag) using current protocols depends on the formation of a fully-folded, triple helix structure.

We have developed a robust protocol for isolating full-length recombinant CLPs containing structural perturbations due to Gly→X mutations. Bacterial CLPs are often engineered to incorporate specific sequences of interest, such as Gly→X mutations and other natural interruptions, to study collagen's structure and function, especially in inherited collagen disorders. Such modifications could also be used to alter biomaterial mechanics or tailor cell-biomaterial interactions. While existing protocols based on trypsin digestion provide a reliable tool to investigate the impact of Gly substitutions on triple helix structure, the use of the same enzyme for purification restricts our ability to obtain full-length recombinant collagens containing Gly substitutions.

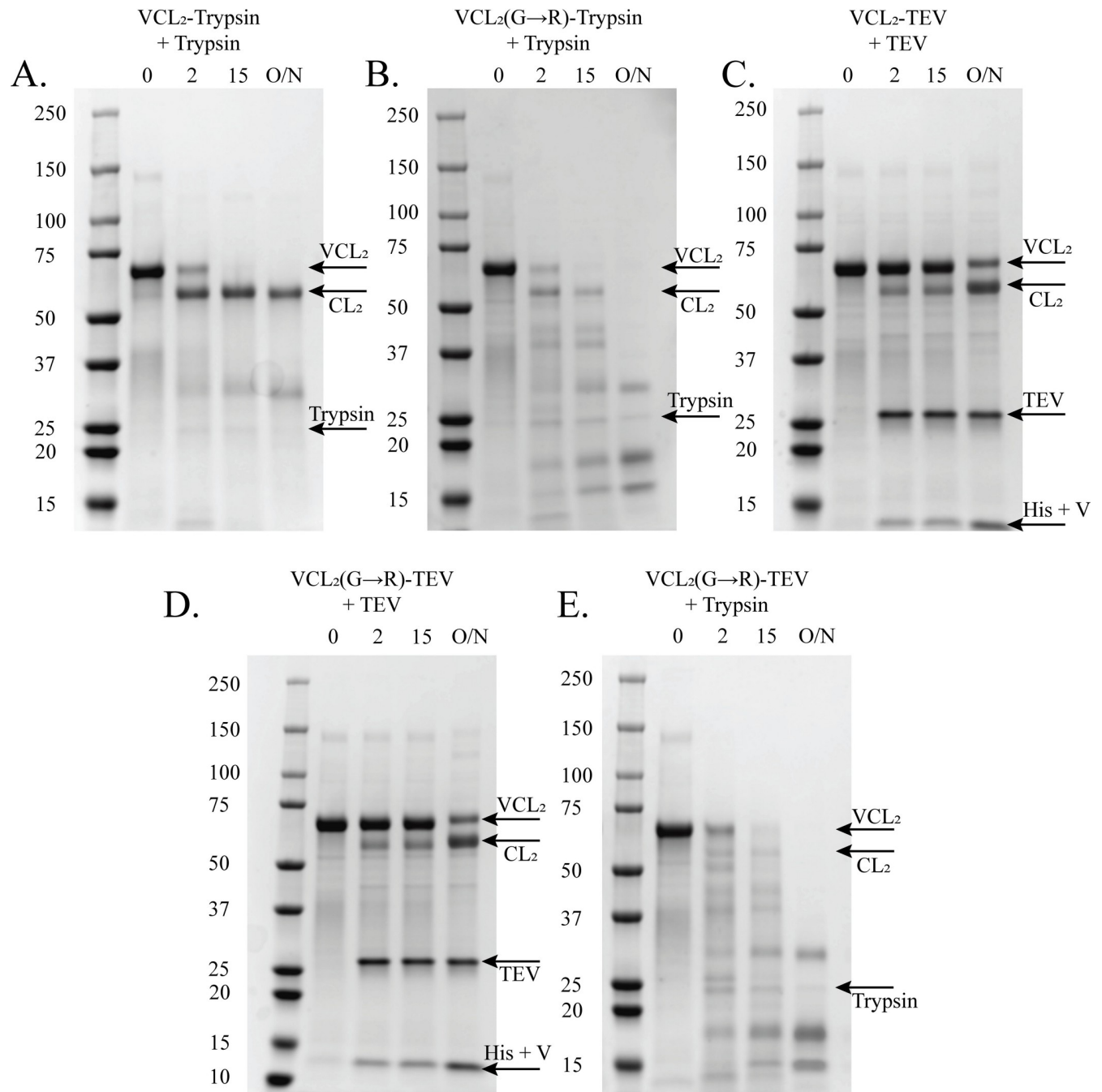


Fig 5. Incubation of distorted triple helices with trypsin demonstrated protein digestion while incubation with TEV protease allowed isolation of CL domains containing Gly→X mutations. SDS-PAGE of (A) VCL₂-Trypsin, (B) VCL₂(G→R)-Trypsin, (C) VCL₂-TEV, (D) VCL₂(G→R)-TEV, and (E) VCL₂(G→R)-TEV purified proteins digested with either trypsin or TEV protease for different time points (0 mins, 2 mins, 15 mins, and O/N). Digestion of VCL₂ with trypsin and TEV protease resulted in removal of His₆-tag and V-domain and isolation of CL₂. However, digestion of VCL₂(G→R) with trypsin led to protein digestion starting at 2 mins only. Digestion of VCL₂(G→R)-TEV with TEV protease resulted in the isolation of CL₂(G→R) protein without degrading the protein. Molecular weight standards in kDa, O/N = Overnight digestion with the protease.

<https://doi.org/10.1371/journal.pone.0285864.g005>

In this study, we identified shake flask culture conditions that maximize bacterial CLP expression. Selecting a suitable *E. coli* strain proved an essential step since the type of strain had a major impact on protein expression, solubility, and yield. Moreover, culture media

composition and conditions affect bacterial growth, which will ultimately impact protein expression and solubility. By sampling culture media and inducer concentration, we were able to obtain high protein yields. Furthermore, to prevent unwanted degradation products, we incorporated a TEV protease cleavage site instead of the widely-used trypsin site to remove the His6-tag and V-domain. This enabled the isolation of the CL domains containing Gly→Arg mutation within the IBS of CL domain. This method can be further adapted to engineering CLPs containing multiple G→X mutations, ligand-binding sites, natural interruptions, or other sequences of interest that are found in animal collagens that are anticipated to compromise triple helix integrity.

Conclusions

While trypsin is a widely used protease to extract and purify animal collagens for commercial applications, low-substrate specificity and the promiscuous nature of the protease suggest a high potential for mis-cleavages at undesired locations, resulting in irreproducible and unpredictable results. This is particularly important when designing a tailored biomaterial using recombinant bacterial collagens for biomedical applications. Human collagen consists of regions with variable flexibility, which play a crucial role in imparting structural and functional properties in their native environment. Engineering CLPs with specific biological functionality, including cell-cell and cell-ECM interactions or cell-guided degradation, requires incorporating human collagen sequences that can alter the triple helix structure of these fabricated biomaterials. Thus, our method offers one approach to introduce a wide range of sequences to develop advanced biomaterials for tissue engineering applications. This technology provides a versatile platform to design novel biomaterials, where the introduction of multiple biological motifs could potentially impact the triple helix structure.

Supporting information

S1 File.
(PDF)

Acknowledgments

The authors acknowledge Natalie Losada and Dr. Eddy Arnold (Center for Advanced Biotechnology and Medicine, Rutgers University) for providing access to protein purification resources and Jennifer Timm (Department of Marine and Coastal Sciences and Center for Advanced Biotechnology and Medicine, Rutgers University) for providing valuable expertise in protein purification methodology and sharing the TEV protease.

Author Contributions

Conceptualization: Sonal Gahlawat, Vikas Nanda, David I. Shreiber.

Data curation: Sonal Gahlawat.

Formal analysis: Sonal Gahlawat.

Funding acquisition: David I. Shreiber.

Investigation: Sonal Gahlawat.

Methodology: Sonal Gahlawat, Vikas Nanda, David I. Shreiber.

Resources: Vikas Nanda, David I. Shreiber.

Supervision: Vikas Nanda, David I. Shreiber.

Writing – original draft: Sonal Gahlawat, David I. Shreiber.

Writing – review & editing: Sonal Gahlawat, Vikas Nanda, David I. Shreiber.

References

1. Xu Y, Keene DR, Bujnicki JM, Höök M, Lukomski S. Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. *J Biol Chem.* 2002 Jul 26; 277(30):27312–8. <https://doi.org/10.1074/jbc.M201163200> PMID: 11976327
2. Squeglia F, Bachert B, De Simone A, Lukomski S, Berisio R. The Crystal Structure of the Streptococcal Collagen-like Protein 2 Globular Domain from Invasive M3-type Group A Streptococcus Shows Significant Similarity to Immunomodulatory HIV Protein gp41. *J Biol Chem.* 2014 Feb 21; 289(8):5122–33. <https://doi.org/10.1074/jbc.M113.523597> PMID: 24356966
3. Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JAM, Inouye M, et al. Self-association of streptococcus pyogenes collagen-like constructs into higher order structures. *Protein Sci.* 2009 Jun; 18(6):1241–51. <https://doi.org/10.1002/pro.134> PMID: 19472339
4. Privalov PL, Tiktopulo EI, Tischenko VM. Stability and mobility of the collagen structure. *J Mol Biol.* 1979 Jan 15; 127(2):203–16. [https://doi.org/10.1016/0022-2836\(79\)90240-7](https://doi.org/10.1016/0022-2836(79)90240-7) PMID: 430563
5. Yu Z, An B, Ramshaw JAM, Brodsky B. Bacterial collagen-like proteins that form triple-helical structures. *J Struct Biol.* 2014 Jun; 186(3):451–61. <https://doi.org/10.1016/j.jsb.2014.01.003> PMID: 24434612
6. Han R, Zwiefka A, Caswell CC, Xu Y, Keene DR, Lukomska E, et al. Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. *Appl Microbiol Biotechnol.* 2006 Aug; 72(1):109–15. <https://doi.org/10.1007/s00253-006-0387-5> PMID: 16552563
7. Cosgriff-Hernandez E, Hahn MS, Russell B, Wilems T, Munoz-Pinto D, Browning MB, et al. Bioactive hydrogels based on Designer Collagens. *Acta Biomater.* 2010 Oct; 6(10):3969–77. <https://doi.org/10.1016/j.actbio.2010.05.002> PMID: 20466083
8. An B, Kaplan DL, Brodsky B. Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Front Chem.* 2014; 2:40. <https://doi.org/10.3389/fchem.2014.00040> PMID: 25003103
9. Diaz Quiroz JF, Rodriguez PD, Erndt-Marino JD, Guiza V, Balouch B, Graf T, et al. Collagen-mimetic proteins with tunable integrin binding sites for vascular graft coatings. *ACS Biomater Sci Eng.* 2018 Aug 13; 4(8):2934–42. <https://doi.org/10.1021/acsbiomaterials.8b00070> PMID: 33435015
10. Parmar PA, Chow LW, St-Pierre JP, Horejs CM, Peng YY, Werkmeister JA, et al. Collagen-mimetic peptide-modifiable hydrogels for articular cartilage regeneration. *Biomaterials.* 2015 Jun; 54:213–25. <https://doi.org/10.1016/j.biomaterials.2015.02.079> PMID: 25907054
11. Parmar PA, St-Pierre JP, Chow LW, Spicer CD, Stoichevska V, Peng YY, et al. Enhanced articular cartilage by human mesenchymal stem cells in enzymatically mediated transiently RGDS-functionalized collagen-mimetic hydrogels. *Acta Biomater.* 2017 Mar 15; 51:75–88. <https://doi.org/10.1016/j.actbio.2017.01.028> PMID: 28087486
12. Xu K, Nowak I, Kirchner M, Xu Y. Recombinant Collagen Studies Link the Severe Conformational Changes Induced by Osteogenesis Imperfecta Mutations to the Disruption of a Set of Interchain Salt Bridges. *J Biol Chem.* 2008 Dec 5; 283(49):34337–44. <https://doi.org/10.1074/jbc.M805485200> PMID: 18845533
13. Peng YY, Stoichevska V, Schacht K, Werkmeister JA, Ramshaw JAM. Engineering multiple biological functional motifs into a blank collagen-like protein template from *Streptococcus pyogenes*. *J Biomed Mater Res A.* 2014 Jul; 102(7):2189–96. <https://doi.org/10.1002/jbm.a.34898> PMID: 23913780
14. Qing G, Ma LC, Khorchid A, Swapna GVT, Mal TK, Takayama MM, et al. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat Biotechnol.* 2004 Jul; 22(7):877–82. <https://doi.org/10.1038/nbt984> PMID: 15195104
15. Peng YY, Yoshizumi A, Danon SJ, Glattauer V, Prokopenko O, Mirochnitchenko O, et al. A *Streptococcus pyogenes* derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. *Biomaterials.* 2010 Apr; 31(10):2755–61. <https://doi.org/10.1016/j.biomaterials.2009.12.040> PMID: 20056274
16. Peng YY, Stoichevska V, Madsen S, Howell L, Dumsday GJ, Werkmeister JA, et al. A simple cost effective methodology for large scale purification of recombinant non-animal collagens. *Appl Microbiol Biotechnol.* 2014 Feb; 98(4):1807–15. <https://doi.org/10.1007/s00253-013-5475-8> PMID: 24402415

17. Stoichevska V, Peng YY, Vashi AV, Werkmeister JA, Dumsday GJ, Ramshaw JAM. Engineering specific chemical modification sites into a collagen-like protein from *Streptococcus pyogenes*. *J Biomed Mater Res A*. 2017 Mar; 105(3):806–13. <https://doi.org/10.1002/jbm.a.35957> PMID: 27806444
18. Parmar PA, Skaalure SC, Chow LW, St-Pierre JP, Stoichevska V, Peng YY, et al. Temporally degradable collagen-mimetic hydrogels tuned to chondrogenesis of human mesenchymal stem cells. *Biomaterials*. 2016 Aug; 99:56–71. <https://doi.org/10.1016/j.biomaterials.2016.05.011> PMID: 27214650
19. Peng YY, Howell L, Stoichevska V, Werkmeister JA, Dumsday GJ, Ramshaw JAM. Towards scalable production of a collagen-like protein from *Streptococcus pyogenes* for biomedical applications. *Microbial Cell Factories*. 2012 Nov 5; 11(1):146. <https://doi.org/10.1186/1475-2859-11-146> PMID: 23126526
20. Bruckner P, Prockop DJ. Proteolytic enzymes as probes for the triple-helical conformation of procollagen. *Anal Biochem*. 1981 Jan 15; 110(2):360–8. [https://doi.org/10.1016/0003-2697\(81\)90204-9](https://doi.org/10.1016/0003-2697(81)90204-9) PMID: 7015914
21. Ryhänen L, Zaragoza EJ, Uitto J. Conformational stability of type I collagen triple helix: evidence for temporary and local relaxation of the protein conformation using a proteolytic probe. *Arch Biochem Biophys*. 1983 Jun; 223(2):562–71. [https://doi.org/10.1016/0003-9861\(83\)90621-5](https://doi.org/10.1016/0003-9861(83)90621-5) PMID: 6344798
22. Cheng H, Rashid S, Yu Z, Yoshizumi A, Hwang E, Brodsky B. Location of Glycine Mutations within a Bacterial Collagen Protein Affects Degree of Disruption of Triple-helix Folding and Conformation. *J Biol Chem*. 2011 Jan 21; 286(3):2041–6. <https://doi.org/10.1074/jbc.M110.153965> PMID: 21071452
23. Hwang ES, Brodsky B. Folding delay and structural perturbations caused by type IV collagen natural interruptions and nearby Gly missense mutations. *J Biol Chem*. 2012 Feb 3; 287(6):4368–75. <https://doi.org/10.1074/jbc.M111.269084> PMID: 22179614
24. Yigit S, Yu H, An B, Hamaia S, Farndale RW, Kaplan DL, et al. Mapping the Effect of Gly Mutations in Collagen on $\alpha 2\beta 1$ Integrin Binding. *J Biol Chem*. 2016 02; 291(36):19196–207.
25. Chhum P, Yu H, An B, Doyon BR, Lin YS, Brodsky B. Consequences of Glycine Mutations in the Fibronectin-binding Sequence of Collagen. *J Biol Chem*. 2016 Dec 30; 291(53):27073–86. <https://doi.org/10.1074/jbc.M116.753566> PMID: 27799304
26. Qiu Y, Mekkat A, Yu H, Yigit S, Hamaia S, Farndale RW, et al. Collagen Gly missense mutations: Effect of residue identity on collagen structure and integrin binding. *J Struct Biol*. 2018; 203(3):255–62. <https://doi.org/10.1016/j.jsb.2018.05.003> PMID: 29758270
27. Yeo J, Qiu Y, Jung GS, Zhang YW, Buehler MJ, Kaplan DL. Adverse effects of Alport syndrome-related Gly missense mutations on collagen type IV: Insights from molecular simulations and experiments. *Biomaterials*. 2020 May; 240:119857. <https://doi.org/10.1016/j.biomaterials.2020.119857> PMID: 32085975
28. Hoop CL, Kemraj AP, Wang B, Gahlawat S, Godesky M, Zhu J, et al. Molecular underpinnings of integrin binding to collagen mimetic peptides containing vascular Ehlers-Danlos syndrome-associated substitutions. *J Biol Chem*. 2019 Aug 12; jbc.RA119.009685. <https://doi.org/10.1074/jbc.RA119.009685> PMID: 31406019
29. Xiao J, Madhan B, Li Y, Brodsky B, Baum J. Osteogenesis Imperfecta Model Peptides: Incorporation of Residues Replacing Gly within a Triple Helix Achieved by Renucleation and Local Flexibility. *Biophysical Journal*. 2011 Jul 20; 101(2):449–58. <https://doi.org/10.1016/j.bpj.2011.06.017> PMID: 21767498
30. Persikov AV, Ramshaw JAM, Brodsky B. Prediction of Collagen Stability from Amino Acid Sequence*. *Journal of Biological Chemistry*. 2005 May 13; 280(19):19343–9. <https://doi.org/10.1074/jbc.M501657200> PMID: 15753081
31. Picker J, Lan Z, Arora S, Green M, Hahn M, Cosgriff-Hernandez E, et al. Prokaryotic Collagen-Like Proteins as Novel Biomaterials. *Front Bioeng Biotechnol*. 2022 Mar 17; 10:840939. <https://doi.org/10.3389/fbioe.2022.840939> PMID: 35372322
32. Meganathan I, Pachaiyappan M, Aarthy M, Radhakrishnan J, Mukherjee S, Shanmugam G, et al. Recombinant and genetic code expanded collagen-like protein as a tailorable biomaterial. *Mater Horiz*. 2022 Oct 31; 9(11):2698–721. <https://doi.org/10.1039/d2mh00652a> PMID: 36189465
33. Browning MB, Dempsey D, Guiza V, Becerra S, Rivera J, Russell B, et al. Multilayer vascular grafts based on collagen-mimetic proteins. *Acta Biomater*. 2012 Mar; 8(3):1010–21. <https://doi.org/10.1016/j.actbio.2011.11.015> PMID: 22142564
34. Merrett K, Wan F, Lee CJ, Harden JL. Enhanced Collagen-like Protein for Facile Biomaterial Fabrication. *ACS Biomater Sci Eng*. 2021 Apr 12; 7(4):1414–27. <https://doi.org/10.1021/acsbomaterials.1c00069> PMID: 33733733
35. Cereceres S, Touchet T, Browning MB, Smith C, Rivera J, Höök M, et al. Chronic Wound Dressings Based on Collagen-Mimetic Proteins. *Adv Wound Care (New Rochelle)*. 2015 Aug 1; 4(8):444–56. <https://doi.org/10.1089/wound.2014.0614> PMID: 26244101

36. Parmar PA, St-Pierre JP, Chow LW, Puetzer JL, Stoichevska V, Peng YY, et al. Harnessing the Versatility of Bacterial Collagen to Improve the Chondrogenic Potential of Porous Collagen Scaffolds. *Adv Healthc Mater.* 2016 Jul; 5(13):1656–66. <https://doi.org/10.1002/adhm.201600136> PMID: 27219220