Université de Montréal

Artificial Collagen for Cornea Repair

Par

Yasmina-Mia EL KHOURY

Département de pharmacologie et physiologie, Faculté de médecine

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès scientifiques appliquées (M. Sc. A.)

en génie biomédical

Mai 2020

Université de Montréal

Département de Pharmacologie et Physiologie, Faculté de Médecine

---

Ce mémoire intitulé

Artificial Collagen for Cornea Repair

Présenté par

Yasmina-Mia EL KHOURY

A été évalué par un jury composé des personnes suivantes

Alain Vinet
Président-rapporteur

May Griffith
Directeur de recherche

Marie-Claude Heuzey
Membre du jury

Résumé

Les patients atteints de cécité cornéenne résultant d'une maladie ou d'une blessure dans de nombreux pays ne seront probablement pas transplantés avec des cornées de donneurs humains en raison d'une grave pénurie mondiale de tissus de donneurs. Cependant, même si des cornées de donneurs étaient disponibles, les patients présentant une inflammation ou une maladie grave ne seraient pas aidés car ils courent un risque élevé de rejet des cornées de donneurs car celles-ci contiennent des cellules allogéniques. Les implants cornéens sans cellules qui ne déclenchent pas de rejet ont été développés comme alternatives à la transplantation de donneurs humains par le laboratoire Griffith, et ont montré dans un premier essai clinique chez l'homme qu'ils régénèrent de manière stable le tissu et les nerfs cornéens. Ces implants comprenaient du collagène humain recombinant, la principale protéine structurelle trouvée dans la cornée humaine. Cependant, les collagènes de pleine longueur sont difficiles et coûteux à produire et ne peuvent pas être personnalisés. Une grande variété de peptides plus courts qui imitent le collagène et d'autres molécules de la matrice extracellulaire ont été développés et testés. Cela comprend les peptides hybrides combinant le collagène et la soie (VBsilk).

Le but de ma thèse est de confirmer les simulations de VBsilk d'un peptide hybride collagène-soie produit au Griffith Lab. Un autre objectif est de déterminer les conditions de production et de purification pour montrer que le peptide simulé peut être converti en un peptide réel.

En bref, l'ADN codant pour une séquence de VBsilk a été cloné dans ClearColi, une souche d'*E. Coli* à faible endotoxine. Les bactéries ont été cultivées dans des cultures à grand volume. Le VBsilk a été extrait et purifié par FPLC. SDS-PAGE a montré que des bandes de protéines de taille appropriée étaient obtenues. Par conséquent, il est possible de produire le peptide VBsilk.

Mots-clés : cornée, collagène, soie, peptides, simulation, purification des protéines

Abstract

Patients with cornea blindness resulting from disease or injury in many countries are unlikely to be transplanted with human donor corneas due a worldwide severe shortage of donor tissues. However, even if donor corneas were available, patients with inflammation or severe disease would not be helped as they are at a high risk of rejecting donor corneas as these contain allogeneic cells. Cell-free corneal implants that do not trigger rejection were developed as alternatives to human donor transplantation by the Griffith lab, and shown in a first-in-human clinical trial to stably regenerate corneal tissue and nerves. These implants comprised recombinant human collagen, the main structural protein found in the human cornea. However, full-length collagens are difficult and expensive to produce, and cannot be customized. A wide variety of shorter peptides that mimic collagen and other extracellular matrix molecules have been developed and tested. This includes hybrid peptides combining collagen and silk (VBsilk).

The aim of my thesis is to is to confirm simulations of VBsilk, a hybrid collagen-silk peptide that was produced in the Griffith Lab. A further aim is to determine the conditions for the production and purification to show that simulated peptide can be converted into an actual peptide.

Briefly, the DNA coding for a VBsilk sequence was cloned into ClearColi, a strain of *E. coli* with low endotoxin. The bacteria were grown up in large volume cultures. The VBsilk was extracted and purified by FPLC. SDS-PAGE showed that appropriate-sized bands of protein were obtained. Hence, it is possible to produce VBsilk peptide.

Keywords: cornea, collagen, silk, peptides, simulation, protein purification.

Table of contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

Å: angstrom

ALK: anterior lamellar keratoplasty

B. mori: Bombyx mori

BSA: bovine serum albumin

CLPs: collagen like peptides

CMPs: collagen mimetic peptides

CVs: column volume

DALK: deep lamellar keratoplasty

dH2O: distilled water

DTT: Dithiothreitol

E. coli: Escherichia coli

ECL: enhanced chemiluminescence

ECM: extracellular matrix

EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide

EGF: Endothelial growth factor

ELPs: elastin like peptides

FPLC: Fast Protein Liquid Chromatography

GAGs: glycosaminoglycans

GMEC: Global Minimal Energy Conformation

HCl: Hydrochloric acid

HRP: horseradish peroxidase

IKVAV: isoleucine-lysine-valine-alanine

IpTG: Isopropyl ß-D-1-thiogalactopyranoside

KCl: Potassium Chloride

KPro: keratoprosthesis

LB: lysogeny broth

LPS: lipopolysaccharide

M: molar

mL/min: millilitre per miniute

mL: millilitres

mM: millimolar

MOPS: 3-(N-morpholino)propane sulfonic acid

MPC: 2-methacryloyloxyethyl phosphorylcholine

mS/cm: millisiemens per centimeter

ms: millisecond

Na-AC: sodium acetate

NaCl: sodium chloride

NaOH: sodium hydroxide

NHS: N-hydroxysuccinimide

$NiSO_4$: nickel sulfate

OD: optical density

PBS: Phosphate Buffer Saline

PEG: poly (ethylene glycol)

pI: isoelectric point

pKa: acid dissociation constant

PMSF: phenylmethanesulfonyl fluoride

PVDF: polyvinylidene fluoride

QSQE: quaternary structure quality estimate

RCF: relative centrifugal force

RGD: arginine-glycine-aspartic acid

RGDS: arginine-glycine-aspartic acid-serine

RHCIII: recombinant human collagen type III

RPM: revolutions per minute

S. pyrogenes: Streptococcus pyrogenes

SclB/Scl2: collagen-like protein 2

TBS: Tris-buffered saline

TBS-T: Tris-buffered saline with Tween 20

TEE: translation enhancing element

Tris-Cl: Tris (hydroxymethyl) aminomethane (THAM) hydrochloride

V: volt

YIGSR: tyrosine-isoleucine-glycine-serine-arginine

µg/mL: microgram per millilitre

To my parents, Zeina and Fady, and to my family

Acknowledgements

I would like to thank my colleagues from the Griffith Lab, especially Elle Edin and Fiona Simpson, with whom I have worked closely on several experiments. They helped me regain confidence in myself. Throughout the journey, I have learned to be independent and tougher.

I would like to thank my family and my close friends for encouraging me.

Principally, I would like to thank my supervisor, Dr. Griffith, for accepting me into her laboratory. She gave me the chance to work on this interesting research project.

Chapter 1 – Introduction

1.1 The human cornea

The cornea is the transparent outer layer of the eye. The human cornea has a diameter of approximately 11 mm. It is convex and aspherical, measuring about 500 µm in the center and gradually increasing to 550 µm in thickness at the periphery (Sridhar, 2018). The cornea serves as the main refractive surface of the eye, focusing light onto the retina for vision. Its ability to focus light is dependent upon its transparency.

The cornea is composed of three cell-containing layers, an outermost epithelium, middle stroma, and an innermost endothelium (Figure 1). There are two acellular layers as well. The Bowman's layer lies between the epithelium and the stroma, while the Descemet's membrane lies between the stroma and endothelium. These acellular layers are composed mainly of collagen.

The corneal epithelium is a stratified epithelium that is non-keratinized. It comprises five to six layers of epithelial cells overlain by a tear film. The tear film keeps the epithelium moist and the epithelial cells healthy. The corneal epithelial cells have tight junctions that form a barrier between the eye and the external environment and protecting it against any pathogens in the environment.

The innermost layer of the cornea comprises a single-celled layer called the endothelium. The endothelial cells serve a pumping function by regulating the movement of molecules and ions between the aqueous humor and stroma, thereby maintaining the state of hydration, and in turn, the optical clarity of the cornea (Bonanno, 2003; Nishida, 1997).

In between the epithelium and the endothelium lies the stroma, which makes up 85% of the corneal thickness. The stroma consists of a hydrated extracellular matrix (ECM) surrounding an interconnected network of cells known as keratocytes. The main component of the stromal

ECM is collagen, which is arranged in regularly spaced parallel lamellae. Other major components are proteoglycans, which are glycosaminoglycans (GAGs) with protein cores. The highly organized collagen and proteoglycans are responsible for the structural integrity and transparency of the cornea (DelMonte et al., 2011). The proteoglycan-rich hydrated matrix and their associated GAGs strengthen the corneas swelling and compressive material properties (Meek et al., 2015). The main collagen components in the stroma are type I and type V fibrillar collagen.

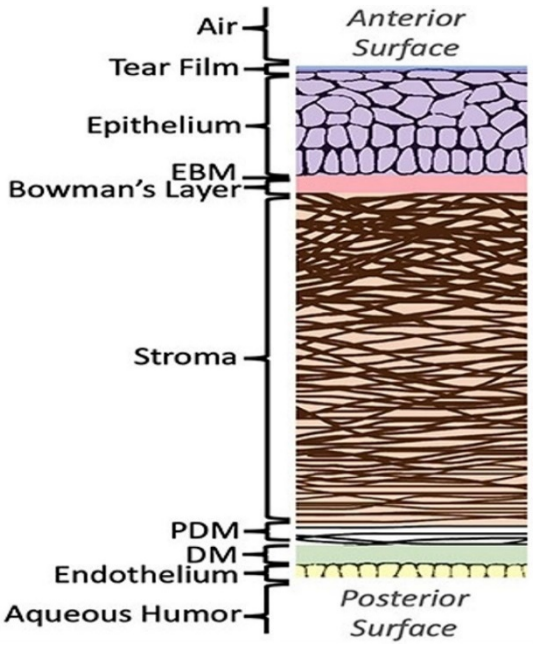


Figure 1. Structure of the human cornea. EBM, Epithelial basement membrane ; PDM, Pre-Descemet's membrane ; DM, Descemet's membrane.

Image from: Blackburn, B.J., Jenkins, M.W., Rollins, A.M. & Dupps, W. A review of structural and biomechanical changes in the cornea in aging, disease, and photochemical crosslinking J. Front. Bioeng. Biotechnol. doi.org/10.3389/fbioe.2019.00066

1.2 Corneal blindness and the need for implants

Any damage to the cornea caused by disease or injury that leads to permanent loss of transparency is likely to lead to vision impairment. There are many different causes of corneal

blindness, and it affects people of all ages, from the very young to the elderly. The causes can be unilateral (affects one eye) or bilateral (affects both eyes). Corneal blindness in high income (developed) countries is commonly due to aging and loss of endothelial cells and their pumping function. In low to middle income countries, infections or injuries (e.g. chemical injuries from the workplace) are more common. Globally, an estimated of 36 million people are blind, and 217 million have moderate to severe vision impairment (Bourne et al., 2017). The World Health Organization has named corneal vision impairment a priority eye disease (World Health Organization, 2018).

Transplantation with a human donor cornea from cadavers is the only widely accepted treatment worldwide for corneal blindness. There are an estimated 2000-3000 Canadians who are on waiting lists for up to two to three years in some provinces (Kramer, 2013). Globally, an estimated 12.7 million people are on waiting lists for corneal transplantation surgery, as there is only one donor cornea available for every seventy persons in need of transplantation (Gain et al., 2016). According to Gain et al. (2016), over 53% of the world's population live in countries with no access to corneal transplantation.

In order to solve the issues of donor cornea shortage, graft failure and rejection, researchers have started the splitting of donor corneas into two pieces for use in two different surgical procedures in two different patients (Heindl et al., 2011). Corneas that are not of sufficiently high quality for transplantation have also been irradiated to sterilize and then cut into lenticules for use as partial thickness grafts and patches (Mathews et al., 2019). Expansion of human corneal cells into large sheets is another way to deal with the donor tissue shortage but at the present time, the cultured cells only address single layer issues. They still need a donor transplant if more than one cornea layer is damaged or diseased (Rama et al., 2010).

In addition to the severe human donor cornea shortage, there is another problem. Even if donor corneas were readily available, those patients with inflammation or severe diseases will not benefit from the donor corneas. This is because immune cells react against allogeneic cells in the donor corneas, and in corneas that are already inflamed, the immune system is already

primed to act against the foreign cells. This leads to a high chance of graft rejection or failure (49% of high risk patients are expected to have graft failure) (Yu et al., 2016).

Artificial corneas have also been developed for clinical use. Corneal prostheses known as keratoprostheses have been developed for use in patients who are at high risk for rejecting human donor corneas (Avadhanam et al., 2015). However, even the Boston Keratoprosthesis that is the most widely used device has many side effects (Aref et al., 2015; Avadhanam & Liu, 2015; Avadhanam et al., 2015; Nouri et al., 2001; Talajic et al., 2012). Their use also necessitates lifetime of antibiotics to prevent infection (Nouri et al., 2001). As such, they are only being used for corneas at the end-stages of damage or disease (Avadhanam & Liu, 2015).

Decellularized xenogeneic and human donor organs and tissues have been studied, both as scaffolds or as reconstituted implants re-seeded with stem cells (Islam et al., 2019). Decellularized corneas, and in particular xenogeneic porcine corneas, have now been tested in clinical trials as alternatives to human corneas (Shi et al., 2017).

Cell-free implants made from recombinant human collagen have also been developed tested in clinical trials by our laboratory and shown to stably regenerate the corneas of patients in clinical trials or regular and high-risk patients (Fagerholm 2009, 2010, 2014).

1.3 Collagen

Collagen is the most abundant protein in mammals. It is found in bones, hair, and skin and in the cornea. The main collagen in the human body and the cornea is type I collagen. Type I collagen is a structural collagen that forms fibrils. Each fibril is 300 nm long and is composed of three left-hand helices twisted in a right-handed manner. Each collagen fibril is composed of chains of repeating amino acids with the motifs glycine (Gly)-Xaa-Yaa, where Xaa and Yaa are most often proline (Pro) and hydroxyproline (Hyp) respectively (Brazel et al., 1987). The repeating sequence of amino acids with glycine as the third amino acid in the triplet, gives each collagen fibril its helical structure (Figure 2). The helix has a 9.6 Å rise per turn, 3.3 Å residues per turn and 2.9 Å rise per residue (Branden et al., 2012). The chains can geometrically twist together by forming the

triple-helix with a peptidyl-proline bond. The presence of hydroxyproline residues acts to stabilize the triple helical collagen molecules through stereoelectronic effects (Shoulders et al., 2009). Mutations that result from substitution of Gly in particular can results in the destabilizing of collagen fibers, leading to genetic diseases like Osteogenesis Imperfecta, where individuals have brittle bones (Shoulders et al., 2009).

In the body, collagen has a hierarchical structure. Triple helical collagen molecules are bundled together to form fibrils, and these in turn form collagen-fibers, which in the cornea are highly aligned (An et al., 2014) (Figure 2).

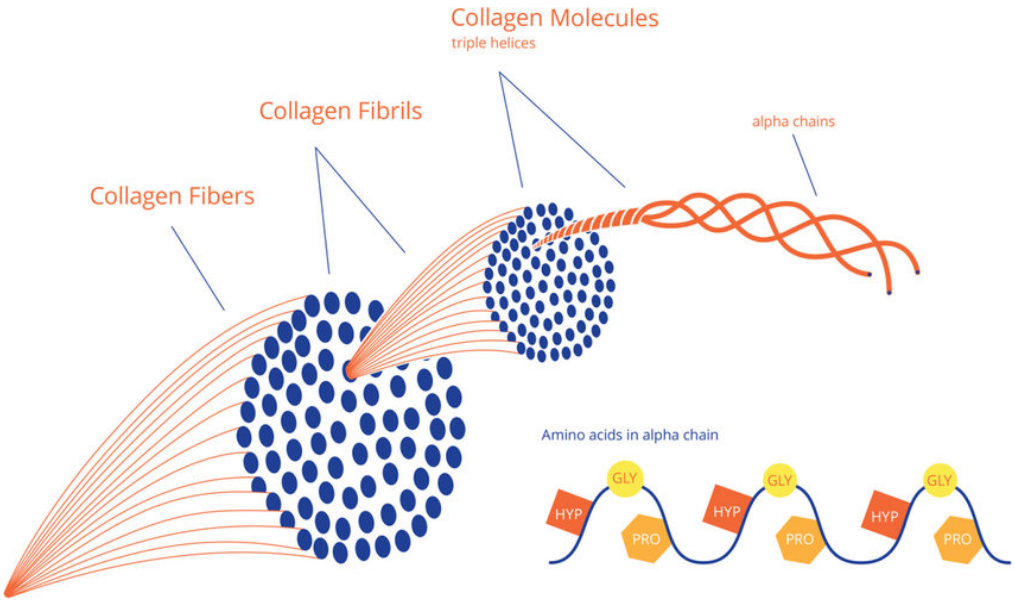


Figure 2. Structure of collagen from amino acid to fibers.

Image from: Nijhuis, W. & Eastwood, D. & Allgrove, J. & Hvid, I. & Weinans, H. & Bank, R. & Sakkers, R (2019). Current concepts in osteogenesis imperfecta: Bone structure, biomechanics and medical management. J. Children's Orthopaedics 13. 1-11. 10.1302/1863-2548.13.180190.

1.4 Recombinant human collagen

To prevent the possibility of animal to human zoonotic transmission of disease, laboratory produced recombinant DNA techniques have been used to produce various proteins, including

collagens. Recombinantly produced human collagens are essentially the same as native collagen as DNA that codes for the collagen is used to reproduce the entire collagen molecule (Olsen et al., 2003). Full-length human and recombinant-human collagens have been made in transgenic expression systems with different species such as yeast (*Pichia pastoris*), plants (tobacco), silkworm and human fibroblasts.

Hydrated gels or hydrogels made from recombinant human collagen have now been tested as cell-free implants in clinical trials to regenerate diseased and damaged corneas. Hydrogels made from recombinant human collagen that was crosslinked with water-soluble N-Ethyl-N'-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and its co-reactant, N-Hydroxysuccinimide (NHS) were tested in an early feasibility study on 10 human subjects (Fagerholm et al., 2010). In this study, 10 patients who were on waiting list for corneal transplantation, aged from 18 to 75, who had vision loss from advanced keratoconus (9 patients) and scarring from bacterial keratitis (1 patient). It was shown that all patients were able to regenerate their corneal tissues and nerves, and the regenerated corneas have been stable for over four years (Fagerholm et al., 2014). This shows that collagen-based implants are very effective at promoting regeneration of corneas in patients.

However, like native extracted full-length collagens, recombinant human collagen is a large macromolecule and not easy to manipulate chemically. In addition, because recombinant collagens reproduce full-length native collagens, it also contains the numerous hydroxyproline amino acids that are characteristic of the protein. To produce recombinant collagen with hydroxyproline amino acids, it is not only necessary to produce the collagen but also to produce prolyl 4-hydroxylase, the enzyme that catalyzes the formation of 4-hydroxyproline from proline amino acids (Vuorela et al., 1997). In addition, recombinant pepsin is needed to cleave the telopeptides from the full-length protein prior to use (Yang et al., 2004). This means expression of three different complex proteins to make recombinant collagen. Therefore, to produce recombinant human collagen, it is necessary to express and produce three different complex proteins.

1.5 Peptide analogs of collagen and other extracellular matrix proteins for regenerative medicine

As discussed above, recombinant production of ECM proteins such as collagen is possible but can be complex. Many research groups including the Griffith laboratory has been developing short peptide equivalents to the much longer proteins of the ECM and many of these have been tested for different regenerative medicine applications (Rubert Perez et al., 2015). Among these are the collagen analogs.

Collagen-like peptides (CLPs) that are also known as collagen mimetic peptides (CMPs) have been developed as short peptide alternatives to native collagen. These range from a few amino acids in length to a few hundred amino acids. The shorter CLPs (40 amino acids or shorter) are usually synthesized on a peptide synthesizer using a solid support system. Longer CLPs, however, need to be produced either in segments that are then chemically linked together; or as a single long peptide chain using recombinant DNA technology (Edin et al., 2020).

Other ECM-derived peptides that have been used in place of much larger native proteins include cellular adhesion peptides such as arginine-glycine-aspartic acid-serine (RGDS) from fibronectin and isoleucine-lysine-valine-alanine (IKVAV) and containing tyrosine-isoleucine-glycine-serine-arginine (YIGSR) from laminin, which have been well-characterized. These have been used as integrin binding ligands in hydrogels. RGDS motifs have been used to improve keratinocytes and cell migration in skin (Bradshaw et al., 2014). The pentapeptide YIGSR promotes epithelial cell growth in cornea (Li et al., 2003), while IKVAV peptides and YIGSR stimulate neuraxial extension (Graf et al., 1987; Tashiro et al., 1989).

Another ECM macromolecule that has been widely tested for use as a biomaterial is silk. Silk is the fibrous secretion from the domestic silkworm *Bombyx mori*. Silk fibroin is the structural part of silk that is left after removal of the sticky sericin by degumming. Like collagen, it is made up of sequences of repeating amino acids sequence of (GAGAGS). However, silk shows better mechanical strength than collagen-based materials (Vepari et al., 2007). Silk fibroin also has controllable biodegradability (based on silk types and fabrication processes), can be transparent (with >95% transparency) while having high load bearing capacity (Vepari et al., 2007), and

tunable surface topography. In 1993, silk fibroin was recognized by the US Food and Drug Administration (FDA) as a biomaterial and has been used to make medical sutures (Melke et al., 2016). Silk can be easily and economically extracted using water-based mild processing and molded into various shapes, 3D constructs and micropatterns for use in regenerative medicine (Vepari et al., 2007), including corneal implants (Bray et al., 2011). However, unlike collagen which inherently has tripeptide arginine-glycine-aspartic acid (RGD) recognizable by the integrin receptors for cell adhesion, silk fibroin lacks specific cell adhesion motifs which limits its clinical application. One method to get around the inertness of silk is to hybridize it to collagen.

1.6 Modelling Protein structure

In silico methods are used to predict the 3D structures and behavior of proteins based on their primary amino acid sequence. Protein modelling relies mainly on two principles; homology-based predictions, and predictions based on physical interactions.

When predicting protein folding *in silico*, the Global Minimal Energy Conformation (GMEC) is the assumed actual protein conformation. Methods yielding GMEC follow some assumptions that need to be considered before running this kind of simulation. For large proteins, GMEC calculations are computationally heavy and might yield unrealistic predictions due to unaccounted for mechanistic properties. For designed proteins, a known structure often serves as the basis for the design. This allows for calculating GMEC for only shorter regions of the full strand. If computer-aided structure prediction is used for only parts of the protein, a good understanding of the folding mechanics is required. A change in a long sequence might destabilize the full motif, and even changes to flanking regions might inhibit the initial assembly of the structure. A large shortcoming of traditional computational prediction of folding is that in many proteins there are more than one individual strand involved in stabilizing a motif; many strategies utilize energy minima predictions for adjacent structures on a single strand but might fail when a structure can only be sustained in multimeric complexes. As the complexity of the assembly increases, the importance of combining design tools becomes more important. If assembly

dynamics for a given sequence or type of motif is known, these regions are often more easily and accurately designed by homology modeling compared to *de-novo* predictions (Samish, 2017).

There are a wide set of interconnected databases to turn to when performing homology-based predictions, the main ones being UniprotKB/SwissProt (going forward called by its original name; SwissProt), Uniprot/TrEMBL, RefSeq, NCBI Protein (includes RefSeq).

1.7 Structural motifs in proteins

The interactions within the amino acid chains of peptides result in a three-dimensional structure in the various segments of a protein. The most common secondary structures are alpha helices and beta-pleated sheets.

1.7.1 α-helixes

One of the most common and simplest protein secondary structures in naturally occurring proteins and peptides is the helix bundles. These are often found as coiled coils or as separate domains within larger, multi-domain proteins. Briefly, in an α-helix, the carbonyl (C=O) of one amino acid is hydrogen-bonded to the amine (Ludwiczak et al., 2019) of an amino acid that is located four places down the chain. This pattern of bonding pulls the polypeptide chain into a helical structure that resembles a curled ribbon. Each turn of the helix has 3.7 amino acids. The reactive organic or R groups of the amino acids stick outward from the α helix, where they are free to interact. α-helices are the only proteins that can exist as an individual strand with no internal strain (Berg et al., 2002). Examples of α-helix include short innate cationic peptides of the immune system like LL37, the cathelicidin produced by the human body.

Coiled coils

These are more complex, larger proteins. The sub-units of coiled coils are made up of alpha helixes with recognizable repeated motifs. Hydrophobic (H) and polar (P) residues are repeated in the heptad pattern HPPHPPP; where the individual positions are denoted abcdefg. Depending on

what position of the heptad starts the repeating pattern, the coiled-coil is designated an index; a pattern that starts with PPHPPPH would, for example, be a b-index coiled-coil. These motifs are common close to collagens in most organisms that produce them, and they act as assembly regions to initiate folding of collagens (Fletcher et al., 2012).

1.7.2 β-sheets

β-pleated sheets or β-sheets form when amino acid chains are stretched out and packed together to form sheet-like structures held together by hydrogen bonds. The hydrogen bonds form between carbonyl and amino groups of backbone, while the R groups extend above and below the plane of the sheet.

The peptide chains of a β-sheet may be parallel, pointing in the same direction (i.e., their N- termini are on one direction and their C-termini point the other way), or antiparallel, pointing in opposite directions (i.e., the N-terminus of one peptide chain is located next to the C-terminus of the other). They are often connected to each other by other structural elements such as loops, unstructured coils and helices. β-sheets are made up of regions rich in amino acids that are typically more hydrophobic, such as tyrosine, phenylalanine, tryptophan, threonine, valine, and isoleucine. Proline is usually only found in the unstructured regions between the secondary structures (Khakshoor et al., 2008; Nesloney et al., 1996).

1.8 VBsilk design

The overall three-dimensional structure or tertiary structure of a protein determines its final properties. In some cases, like collagen, multiple protein strands come together to form the quaternary structure of the final protein.

VBsilk was designed as a chimeric transgene protein. A series of “known” structures based on their amino acid sequence were combined to create a composite protein that in theory carries beneficial properties from several families of proteins that are regularly used for regenerative medicine applications. The focus in this design was to create a structure that had native folding

in each region were a derived sequence resided while also being appropriate to allow expression within a prokaryotic expression system such as *Escherichia coli* (*E. coli*).

VBsilk comprises a combination of sequences from a collagen (V-CL peptide), silk, elastin with RGD and YIGSR. The collagen region was selected due to collagens contribution to mechanical properties of ECM, and as collagens are recognizable to several cell surface receptors. Silk can confer high tensile strength. Elastin can both contribute elasticity that allows the material to retake its original form after repeated deformation, and while RGD and YIGSR are well-known fibronectin and laminin cell adhesion sites. During the design process, the focus was on feasibility of expression, and biological properties; advanced structural modelling was not performed. As discussed above, the combination of collagen and silk fibroin is considered to achieve improved mechanical and biological properties of hydrogels. Both collagen and silk are natural protein fibers that have been shown to be biocompatible and promote wound healing (Vepari et al., 2007; Fagerholm et al. 2010, 2014). VBsilk, which combines both collagen and silk amino acid sequences is also likely to be biocompatible. However, testing *in vitro* and *in vivo* is needed for confirmation.

1.9 Hypothesis and Objectives

This project focuses the translation of a novel collagen-silk-based hybrid protein, VBsilk, from an *in silico* design into a protein. The **hypothesis** I was testing is that through modelling and understanding folding, we can predict or explain behaviors in buffers and thereby optimize a purification strategy for VBsilk.

My **first objective** was to confirm simulations that a hybrid collagen-silk peptide that was produced in the Griffith Lab that we refer to as VBsilk, does contain collagenous helices in its predicted structure. My **second objective** was to determine the conditions for the production and purification of VBsilk, and to show that the design can be converted into actual peptides that can be purified and quantified.

Chapter 2 – Materials and Methods

2.1 Protein modelling of VBsilk

Protein BLAST searches were performed using BLASTp (NCBI) directly, or by proxy search in the SWISS-MODEL online tool (Bienert et al., 2017; Guex et al., 2009; Waterhouse et al., 2018). SWISS-MODEL was the primary tool used for general homology modelling in cases where full length templates were available. The given quaternary structure quality estimate (QSQE) score reflected the expected accuracy of the interchain contacts. SWISS-MODEL outputs verification data in form of Ramachandran plots as a complement to the similarity data acquired. In the parts of the sequence where templates were missing, or where the known structure was too short for efficient homology modelling, a combination of manual review of primary sequence for known common elements (common turn, coil, or sheet forming units) and *de novo*, energy minima based folding was utilized. Specialized protein prediction tools were used in conjunction with homology folding for suspected collagenous or coiled coil regions, this was performed to determine stability and type of oligomerization of these regions. For analysis of coiled coils "DeepCoil", a neural network-based tool for detecting coil-coiled domains in protein sequences (Ludwiczak et al., 2019; Zimmermann et al., 2018) was used. For analysis of collagenous regions Collagen Stability Calculator V 2.1 (https://compbio.cs.princeton.edu/csc/) was used (Persikov et al., 2005). The prediction of protein secondary structure was made with JPred v.4 Protein Secondary Structure Prediction Server (http://www.compbio.dundee.ac.uk/jpred/) that provided predicted Solvent Accessibility and Coiled-coil regions by Lupas method (Drozdetskiy et al., 2015).

The primary sequence of VBsilk was annotated, marking source sequences from which the protein was derived. The N-terminal region was an α-helix protein that was important for protein folding and its three-dimensional structure was predicted with SWISS-MODEL (https://swissmodel.expasy.org/repository). The study of coiled coils was used to predict the protein stability in the first 317 amino acids. The collagenous CL domain sequence was modelled

to analyse stability of its three identical sequence with the Collagen Stability Calculator. The C-terminal region after the collagenous CL domain was predicted with JPred. Likely enzyme cutting sites were predicted using ExPASy PeptideCutter (Gasteiger et al., 2005).

VBsilk was designed by Dr. Chyan-Jang Lee and modelling was performed by Dr. Elle Edin, post-doctoral fellow in the laboratory. What I did for this thesis was to confirm the remodelling by reproducing the simulations that were done.

2.2 Materials and equipment for VBsilk production

2.2.1 Materials

Synthetic DNA for VBsilk was prepared by GenScript (Leiden, Netherlands), expanded and stored as aliquots in a -80°C freezer until used. LB Broth Miller and ampicillin powders were purchased from BioShop Canada Inc. (Burlington, ON). Isopropyl ß-D-1-thiogalactopyranoside (IpTG) (Cat #BP1755-100), bovine serum albumin (Cat #AAJ6573118), thermo scientific gel code blue safe protein stain (Cat #PI24594), his-tagged protein standard (Cat #LC5606), secondary goat anti-mouse IgG horseradish peroxidase (HRP) antibody (Cat #PI32430), enhanced chemiluminescence (ECL) chemiluminescent substrate (Cat #WP20005) and pH buffer sachets (Cat #01-911-291) were bought from Thermo Fisher Scientific (St-Laurent, QC). Urea (Cat #800-140-DG) was purchased from Wisent Inc (St-Jean Baptiste, QC). The HisPrep FF 16/10 column (Cat #28-9365-52), was purchased from GE Life Sciences Canada (Mississauga, ON). SDS-PAGE and Western Blot kits, 2X Laemmli sample buffer (Cat #161-0737), 10X Tris/Glycine/SDS (Cat #1610772), mini Trans blot filter paper (Cat #1703932), immun-blot polyvinylidene fluoride (PVDF) membranes for protein blotting (Cat #1620177), 10% Mini-Protean precast gels (Cat #4561033) and roller for western blot (Cat# 1651279) were purchased from Bio-Rad (Mississauga, ON). HyClone HyPure Wifi Quality Water (Cat #SH30221.17) was bought from VWR (Mont-Royal, QC). Western blot box (Cat #B1200-7) was bought from Diamed Lab Supplies (Mississauga, ON). The color protein broad range (Cat #P7712S) was bought from New England Biolabs Ltd. (Ipswich, MA, USA). Recombinant protein A (Cat #AB52953) was purchased from AbCam (Toronto, ON). Anti-His-Tag antibody (Cat

#AM1010A) was purchased from FroggaBio (Toronto, ON). pH/Conductivity meter (Cat #RK-59200-04) was bought from Cole-Parmer Canada (Montreal, QC). Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich Canada Ltd (Oakville, ON). Toxic chemicals as imidazole, phenylmethanesulfonyl fluoride (PMSF), hydrochloric acid (HCl), sodium hydroxide (NaOH) were handled under the fume hood.

2.2.2 Equipment

The FPLC AktaPure 25 system (GE Life Sciences, Waltham, MA, USA) with Unicorn version 7.2 software was used for protein purification. Other specialized equipment included a TecanSpark (Tecan, Männedorf, Switzerland) microplate reader was used with SparkControl method editor version 2.2 to read OD of bacteria. The ImageQuant LAS 4000 (GE Life Sciences, Waltham, MA, USA) was used for quantitative imaging for SDS-PAGE gels and Western blots.

2.3 VBsilk producing plasmids

The pCold DNA family of vectors is a cold-shock recombinant protein expression system (Takara Bio Europe AB, Gothenberg, Sweden). A Cold Shock Protein A (cspA) promoter and related elements have been engineered into these vectors to increase target protein production at when these bacteria are given a cold shock (Qing et al. 2004). That is, they are switched from growing at 37 ℃ to 15 ℃. This cold shock suppresses the expression of bacterial cellular proteins and overall bacterial growth while allowing for the expression of target proteins at a high yield and with high purity as well as increased solubility as compared to conventional *E. coli* expression systems. The yield of target protein is expected to be 60% of all intracellular proteins produced.

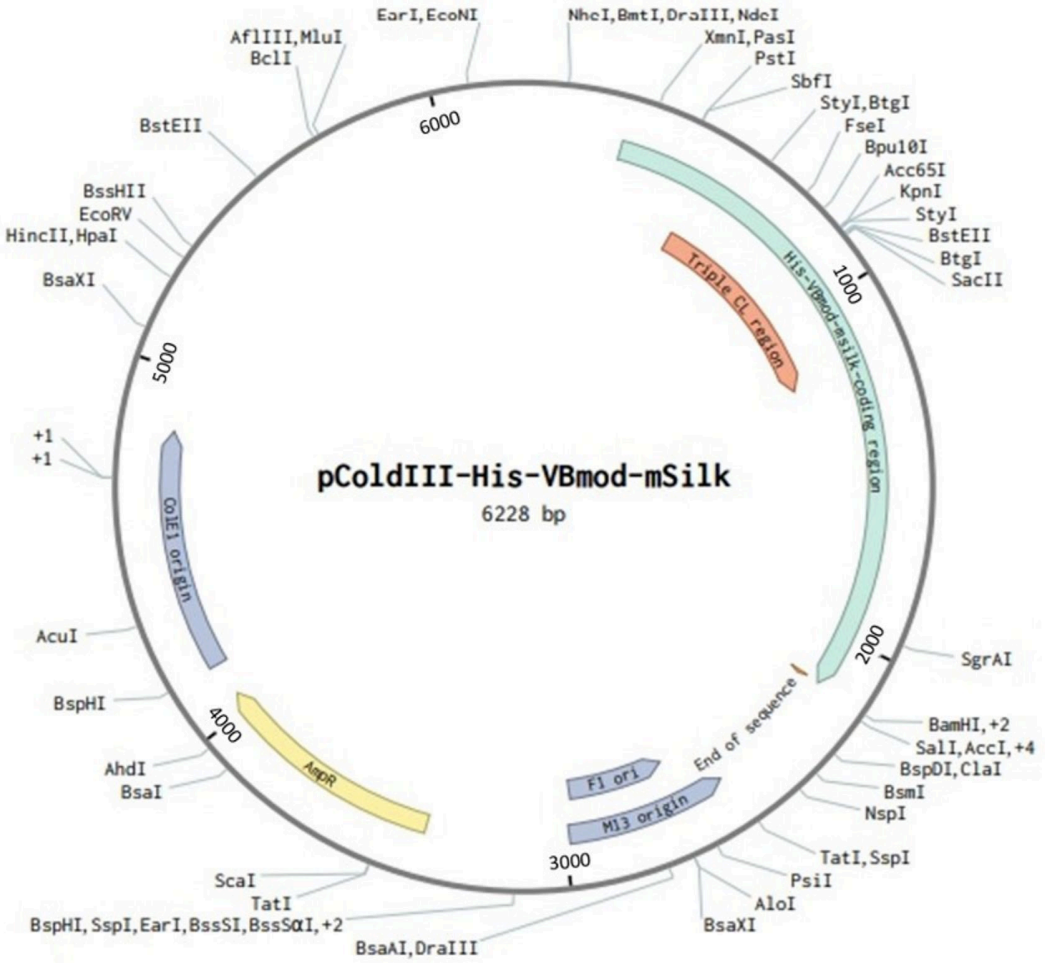


Figure 3. Plasmid map of pColdIII that was used to clone VBsilk. The multiple cloning sites start at the NdeI site.

2.4 Growing up VBsilk producing bacteria

2.4.1 Solutions

For 4 L of LB Miller broth, 100 g of LB Miller was weighed on a precision balance and added to 4 L of dH_2O in a 6 L sterilized glass flask. For 2.5 L, 62.5 g of LB Miller broth was weighed and added to 2.5 L of dH_2O. For 2 L, 50 g of LB Miller broth was weighed and added to 2 L of dH_2O.

Aluminium foil was added to cover the flask and sealed with autoclaving tape to autoclave the solution at 121 °C for 30 mins.

1 M IpTG stock solution for 5 tubes of 15 mL was made was made by weighing on a precision balance 11.9 g of IpTG in 50 mL tube containing already distilled water (dH_2O) and vortexed to dissolve. The solution was moved into a 60 mL syringe and was filtered using a 0.22 µm syringe filter (Corning®) into 10 mL of 5 tubes of 15 mL each; The solution was stored in -20 °C.

2 M imidazole stock solution was prepared by adding 200 mL of dH_2O in a 1L bottle. 27.23 g of imidazole was weighed on precision balance and combined to the 200 mL and topped 1 L of dH_2O. Stock solution was filtered with 1 L filter bottle 0.2 µm then stored in 4 °C.

0.2 M PMSF stock solution was prepared by weighing 348.4 mg of PMSF on precision balance and added into a 50 mL tube where 10 mL of 100% ethanol was added then vortexed to dissolve. Solution was stored in -20 °C.

Ampicillin stock solution for 5 tubes of 15 mL was made by weighing on a precision balance 5 g of ampicillin in 50 mL distilled water (dH_2O) and vortexed to dissolve. The solution was moved into a 60 mL syringe and was filtered using a 0.22 µm syringe filter (Corning®) into 10 mL of 5 tubes of 15 mL each; The solution was stored in -20 °C.

2.4.2 Growing ClearColi©

The protocol used is a laboratory protocol that is described in Edin et al. (2020). ClearColi© are *E. coli* that have been genetically modified to expressed significantly reduced amounts of lipopolysaccharide (LPS) to not cause an endotoxic response in human cells. These cells were used to express VBsilk.

Flasks of 4 L or 2.5 L or 2 L LB Miller broth were sterilized and cooled to room temperature, 4 mL or 2.5 mL or 2 mL ampicillin was added since ampicillin has a final concentration of 100 µg/mL. With a 50 mL serological glass pipet, 40 mL of the LB miller and ampicillin prepared was pipetted in a 125 mL sterilized glass flask. The frozen aliquot tube of bacteria ClearColi pColdIII-VBsilk in -80 °C freezer was scratched with a 10 µm tip and placed in the mix then incubated in a shaking incubator with an aluminium foil at 250 rpm at 37 °C for, care was taken to not thaw the

bacteria pellet above -80 °C. 1 mL was measured with a serological glass pipet to load in a cuvette, the optical density (OD) was read with cuvette application on Tecan Spark, with the Spark control method editor version 2.2. The absorbance of bacterial growth was read at 600 nm since large particles, in the range of several microns, have high absorbance at 600 nm. When the smaller feeder culture reached OD 1, it was added to the 4 L LB Miller and ampicillin or 2.5 L or 2 L LB Miller and ampicillin then incubated at 250 rpm at 37 °C until OD_{600} of 1 was reached.

To induce VBsilk protein expression, 1 mL/L of IpTG stock was added to the bacterial cultures, yielding a final concentration of 1 M. A magnet stirrer was added to the flask and this was placed on a stirring plate and set to 300 rpm in a 4 °C cold room for 24 hours. After 24 hours, a 1 L phosphate lysis buffer prepared by combining 50 mM phosphate buffer, 690 mM NaCl, 13.5 mM KCl, pH 7.4, 20 mM imidazole from the 2 M stock, 5 M urea, 0.5% Triton X-100, 10% v/v glycerol and 1 mM PMSF in ethanol (added separately) then stored in 4 °C.

PMSF was added to inhibit serine peptidases that cleave peptide bonds and prevent denaturation. After ~24 hours, the culture was divided into 1 L centrifugal tubes and centrifuged at 4000 relative centrifugal force (rcf) for 10 mins at 4 °C. Supernatant was removed and a volume of phosphate lysis buffer that was five times greater than the amount of pellet collected was added to re-suspend the pellet. Re-suspended crude lysates were transferred in 50 mL tubes and frozen at -80 °C for 24 hours. The crude lysate was then thawed and sonicated at 60% amplitude; 1 s on, 200 ms off; 4 × 2 mins with 4 mins of cooling on ice between each cycle to further break down any bacterial cell membranes. Tubes were centrifuged at 15000 rcf for 30 mins at 4 °C and supernatant was collected in a sterile glass bottle.

All bacterial pellets were lysed as described above or stored frozen in a -20 °C freezer until used.

2.5 Preparation of the FPLC for peptide purification and Nickel Sulfate ($NiSO_4$) Loading

20% ethanol was made in a 1 L cylinder by diluting 20 mL of 100% ethanol to 1 L dH_2O then filtered with 0.2 µm filter bottle of 1 L. The AktaPure 25 FPLC system (Figure 4) was cleaned according to the manufacturer's directions as per the user's manual.

Histidine increased affinity of nickel ions and made his-tag proteins to bind further and so column was loaded with $NiSO_4$. The column was cleaned with 1 column volumes (CVs) of 20% ethanol with a flow rate of 10 mL/min, with 5 CVs of distilled water with 5 mL/min of flow. 0.4CVs of $NiSO_4$ was passed through the column at a flow rate of 1 mL/min.

Column cleaning was repeated after $NiSO_4$ loading using 5 CVs of distilled water and 20% ethanol were repeated. When the run was ended, the column cleaning was repeated with a further 2 CVs of 20% ethanol from tube A_1 (shown in Figure 4) at 5 mL/min of flow and 10% concentration of tube B_1. Concentration of tube B_1 was set at 0% and switched to tube A_2 and run with 2 CVs of 20% ethanol. The column was equilibrated with 20% ethanol to be stored for a longer period.

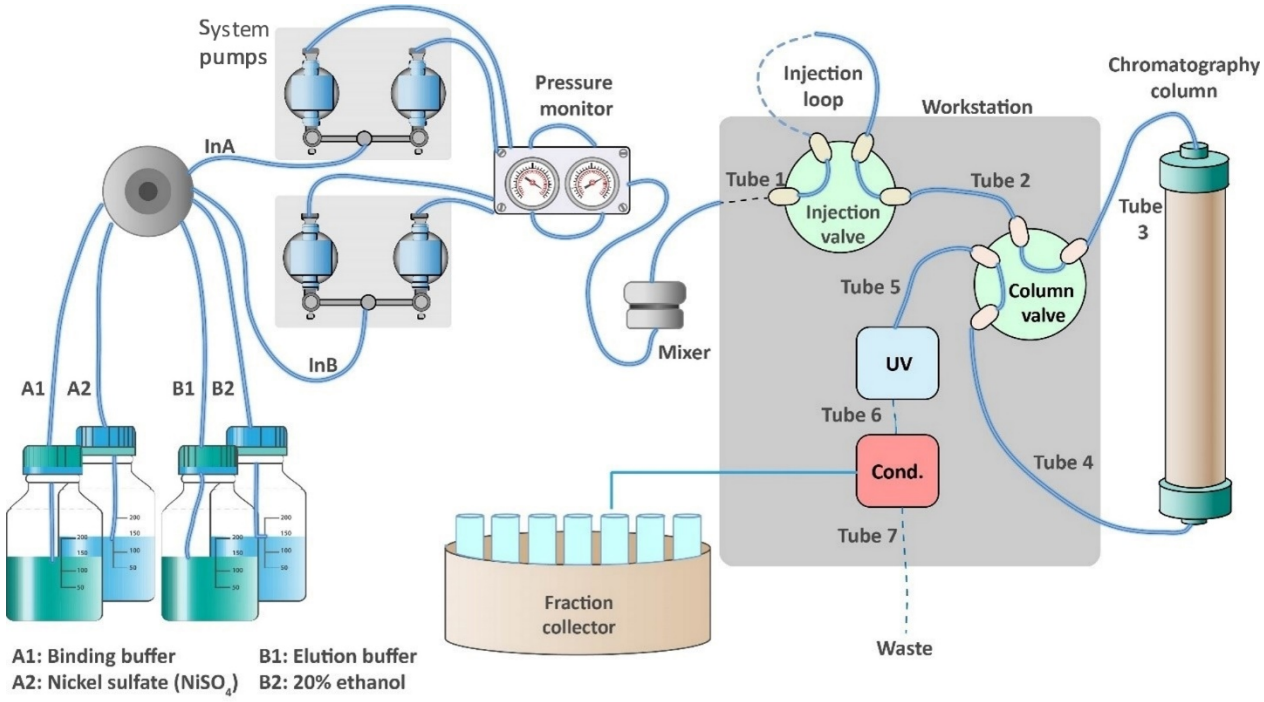


Figure 4. Diagram showing the flow path within the AktaPure 25 FPLC system.

Inlet valve: Two A and two B inlet ports. No integrated air sensor. InA/InB: Inlets delivering buffers or sample to system pumps. System pumps: Delivery of buffers or sample in purification runs. Pressure monitor: System pressure reading. Mixer: Mixing buffer composition homogeneously. Injection valve: Sample directing onto the column. Column valve: Column connection to instrument. UV: Measuring UV absorbance at a fixed wavelength of 280 nm. Cond.: Measuring conductivity of buffers and proteins. Fraction collector: Collection of multiple fraction tubes (GE Health Care Life Sciences, 2012b).

2.6 FPLC Purification of VBsilk

A range of buffers and different buffering conditions e.g. type of buffer (components of buffers), concentrations of buffers, detergents and other additives were tested. The buffers are important to prevent sudden pH changes that could irreversibly affect their folding, solubility, and future function. In general, the pH of the buffer solution should be within 1.0 pH unit of the pKa to provide appropriate buffering capacity, i.e. there is sufficient amount of the molecule in both its acidic and basic forms to neutralize the solution in case of a H^+ or OH^- influx. Common protein

purification buffers include sodium or Tris phosphate. Tris-based buffers could not be used because they contain primary amines that could interfere with future amide crosslinking of the purified VBsilk to produce hydrogels. Hence, a sodium phosphate-based buffering system was selected. Table 1 shows the buffers that were tested.

Table 1. Range and components of buffering systems tested in order to optimize yields of purified VBsilk protein.

Buffers		pH	Phosphate Buffer	Na-AC	MOPS	Urea	Imidazole	NaCl	Tween20	TritonX-100
Buffer 1	Binding buffer	7.4	10 mM	0	0	5 M	20 mM	138 mM	0	0
	Elution buffer						500 mM			
Buffer 2	Binding buffer	7.4	10 mM	0	0	3 M	15 mM	138 mM	0	0
	Elution buffer					5 M	500 mM			
Buffer 3	Binding buffer	4	0	50 Mm	0	5 M	10 Mm	276 Mm	0.1%	0
	Elution buffer						500 Mm			
Buffer 4	Binding buffer	4	0	50 mM	0	5 M	10 mM	276 mM	1%	0
	Elution buffer						500 mM			
Buffer 5	Binding buffer	4	0	50 mM	0	5 M	10 mM	276 mM	0	0.1%
	Elution buffer						500 mM			
Buffer 6	Binding buffer	4	0	50 mM	0	5 M	10 mM	276 mM	0	1%
	Elution buffer									

Buffer 7	Elution buffer	7.9	0	0	50 mM	5 M	10 mM	276 mM	0.1%	0
	Elution buffer						500 mM			
Buffer 8	Elution buffer	7.9	0	0	50 mM	5 M	10 mM	276 mM	1%	0
	Elution buffer						500 mM			

FPLC is used to separate macromolecules for purification of large amounts of protein mixtures. The present aim was to get rid of all undesired impurities. Briefly, 1 L of binding buffer and elution buffer were prepared for buffer 1 and 1 L of buffer 2 was made and filtered with 0.2 µm filter bottle of 1 L. For buffers 3 to 8, 3 L of binding buffers were prepared for sample dialysis (appendices A.1), 700 mL was filtered with 0.2 µm filter bottle of 1 L while 300 mL of elution buffer was made and was then filtered with 0.2 µm filter bottle of 250 mL.

VBsilk contains a his-tag for purification. Urea was added to provide denaturation conditions in which his-tag is completely exposed so it can bind to the nickel columns. Imidazole was used in binding buffer to protect the column resin and avoid excessive binding of non-target proteins. Imidazole is used in the elution buffer to be able to extract protein. System was cleaned according to FPLC system manufacturer handbook and many FPLC runs were made for optimization to obtain a purified VBsilk in a 4 °C fridge.

All buffers were used at the same temperature of 4 °C fridge. 1 L of buffer 1 was prepared respectively and filtered with 0.2 µm filter bottle of 1 L. As first run, 85 mL of the collected VBsilk supernatant was applied to the column containing $NiSO_4$, pre-equilibrated with 5 CVs of binding buffer 1. After washing the column with 5 CVs of loading buffer 1, the protein was linearly eluted in elution buffer 1 and 25 mL was collected in fraction collection tubes. The unpurified sample was collected in the flow through, in a clean collection flask, by running into an outlet valve. Column got cracked due to its clogging from previous runs made in the laboratory. A new column was charged with $NiSO_4$ as method 2.5. Another run was made with a new method by pre-equilibrating the column with 5 CVs of binding buffer 1, applying 75 mL of VBsilk supernatant into the column that was then washed with 5 CVs of binding buffer 1 and the protein was eluted with

4 different gradient elution steps of 3 CVs each. Fraction volume of collected tubes was set at 10 mL. With these parameters buffer 2 was tested using a sample volume of 200 mL. 10 mL of sample was used with the same previous buffers in another run to further purify the protein by changing the methods with 4 different gradient elution steps where first step was at 4 CVs and others of 3 CVs each. 20 mL of the resulted flow through was loaded and tested with buffer 2 in a new run using this new method. To differentiate sample purification, two same runs were made with this method before and after loading the column freshly with $NiSO_4$: by charging the column with 80 mL and 100 mL of sample supernatant respectively. Since column was recently filled with $NiSO_4$, another attempt of undiluted 50 mL protein supernatant was applied to the column with buffer 1.

A fresh sample batch was produced and was performed in two different runs of buffer 1 with the new method, as sample volumes were 72 mL and 60 mL respectively. In run no. 14, the usual automated run was replaced by manual sample application (due to an issue with automation in the instrument). Manual loading and elution steps were reset to 3 CVs each. Manual sample application was used until run no. 17, using buffer 1 (shown in Table 1) with sample volumes at 80 mL, 45 mL, 70 mL and 7 mL each.

Different buffers with the addition of detergents to promote protein denaturation were tested. 3 L of binding buffers 3 to 8 were made and 2 L were used to dialyze sample against those buffers respectively. The remaining, 700 mL was filtered with 0.2 µm filter bottle of 1 L while 300 mL of elution buffer was made and was then filtered with 0.2 µm filter bottle of 250 mL.

As a final run, production of VBsilk was made with a different bacterium aliquot and was loaded in column using the last method with buffer 1. In all the runs the protein fractions containing peaks of purified protein in chromatogram were pooled and tested on gels.

2.7 SDS-Polyacrylamide Gel Electrophoresis

To check the size of the proteins produced by ClearColi to determine if VBsilk was being produced, Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate out proteins to check for VBsilk bands in the bacteria lysate

2.7.1 Solutions

Laemmli sample buffer (2X) used consisted of 4% SDS (w/v), 20% glycerol, 0.004% bromophenol blue (colour dye), 0.125 M Tris-Cl, pH 6.8 and 10% dithiothreitol (DTT).

1 M DTT was made by weighing 1.54 g of DTT in a precision balance. In a 15 mL tube, 4 mL of dH_2O was added then the weighed DTT was poured in and topped to 10 mL with dH_2O. The stock solution was vortexed until dissolved. The solution was moved with a 10 mL syringe and filtered using a 0.22 µm syringe filter into a 10 mL tube then stored in -20 °C prior to use.

1 L 1X running buffer was prepared by diluting 1:10 10X Tris/Glycine/SDS in dH_2O.

20 mg/mL bovine serum albumin (BSA) stock was made by weighing 20 mg of BSA in a 1.5 mL sterile tube with an analytical balance and added 1 mL of dH_2O then vortexed.

0.5 µg/mL BSA was mixed in a 1.5 mL sterile tube by adding 39 µL of dH_2O and 1 µL of 20 mg/mL BSA stock.

2.7.2 SDS-PAGE gel electrophoresis

A 1/10 dilution of the protein fractions with identifiable peaks on the FPLC chromatogram, as well as a sample from the flow through were prepared along with an undiluted flow through, and 0.5 µg BSA protein benchmark were separated out on SDS-PAGE gels. DTT was added to the samples to denature the proteins for faster separation. Glycerol was added to each sample to allow for easier loading. Sample dilutions and 0.5 µg BSA were prepared in a 1.5 µL sterilized Eppendorf tubes by adding 80 µL of dH_2O, 10 µL of glycerol and 10 µL of sample. In a new sterile tube, 20 µL of the dilution and undiluted flow through were added to 20 µL of 2X Laemmli sample buffer. Samples were boiled at 98 °C at 500 rpm for 10 mins with thermomixer.

Pre-cast “Mini-protean” 8% acrylamide gels were placed into the electrophoresis tank assembly. 1X running buffer was placed in the upper chamber and lower chambers. The combs were removed from the gels and samples were loaded into the wells. 20 µL samples was loaded in each well, apart from 3 µL ladder of color protein standard loaded in well 1. After connecting the tanks via electrodes to the power supply, the gels were run at 300 V for approximately 20 mins or until dye front of the samples was approximately 1 cm from the bottom of the gel. When run was done, the gel cassette assembly was opened, the precast gel got open with key according to arrows respectively and wells were cut with razor blade. Gel was put in western blot box, washed with 100 mL dH_2O for 5 mins 3 times, with 20 mL blue stain and with 100 mL dH_2O for 1h.

Densitometry was performed to quantify the proteins in the different lanes. This was done by placing the gel on the center of orange plate for colorimetry where digitalization and epi-illumination of the gel were captured by using azure c200 or Fuji image QuantLAS 4000 (GE Life Sciences). BSA weighed 66.5 kDa and VBsilk around 80 kDa. Bands were selected and analyzed for quantification on Fiji with ImageJ application: To outline the first lane, the rectangular selection tool was used. On the application status bar to analyze gels, the first lane was selected. The rectangular selection tool was moved to the next lane and to analyze gels on the status bar the next lane was selected. This step was repeated for each remaining lane. To generate the lane profile plots, analyze, gels then plot lanes were selected. The straight-line selection tool to draw base lines was used so that each peak of interest determined a closed area. Size was measured by clicking inside each peak with the wand tool and was copied on an Excel sheet. Protein concentrations were estimated by calculating the ratio of absorbance between the target fraction band and the BSA band; the BSA had a known loading amount, and thus enabled us to back calculate the amounts that had been in the fraction samples.

2.8 Western Blots

VBsilk produced in ClearColi© was tagged with a 6-His tag to allow the identification of target proteins in the subsequent bacterial lysate. The presence of the 6-His Tag allowed for the use of

an anti-His Tag antibody to localize and confirm using Western or immune-blotting that bands of protein of interest had been obtained in the SDS-PAGE separated protein samples.

2.8.1 Solutions

1 L of 1X standard transfer buffer or Towbin buffer was made by adding 25 mM Tris, 192 mM glycine, 20% (v/v) methanol (pH 8.3).

2 L of 1X tris-buffered saline (TBS) was composed of 20 mM Tris-HCl and 500 mM NaCl, with the pH meter the pH was adjusted to 7.5.

For 1 L tris-buffered saline with Tween20 (TBS-T), 1 L of 1X TBS was combined in a 1 L bottle with 0.05% Tween 20.

Blocking solution was made by weighing in a 50 mL tube 2.5 g BSA to 50 mL of 1X TBS-T. Solution was then vortexed until dissolved.

0.1 M NaOH was prepared by dissolving 400 µL of 12.5 M NaOH in a 50 mL tube of dH_2O.

Primary antibody was 5 mL of blocking solution with 5 µL of His-Tag antibody.

Secondary antibody was 5 mL of blocking solution with 10 µL HRP.

2.8.2 Western blotting of separated proteins

SDS-PAGE separated proteins were first transferred from the gels onto polyvinylidene fluoride (PVDF) membranes that bind proteins with high efficacy through hydrophobic interactions. The PVDF membrane was activated at room temperature with 20 mL of methanol in a western blot box for 5 mins. The submerged protein transfer was prepared on the gel holder cassette by preventing air bubbles on superposing respectively a black foam pad, 2 blotting filter papers, the SDS-PAGE gel, the activated PVDF membrane, 2 blotting filer papers and a second black foam pad. Each step was rolled on with the roller for western blot to prevent air bubbles. Gel holder cassette was locked and closed then put in the electrode assembly in the same way than another gel holder cassette that contained 2 black foam pads (cathode side was faced the cathode side of blotting module).

The electrode assembly was put in mini tank according to electrodes colors where transfer buffer was filled until top to the mark of 2 gels of the mini tank and a small stir bar in was added. In a medium plastic bucket, ice to cover bottom was put and the mini tank placed in. The bucket was settled on mini stir plate and the stirring was turned to 2 until magnet was spinning in the middle of the mini tank. Transfer buffer was filled to 4 gels of mini tank and closed electrode lid respectively to colors and connected in voltage power supply. Crushed ice was inserted in medium plastic bucket and mini tank and lid were covered. Power supply voltage was set to 100 V and run for 1 hour. When run was done, gel holder cassette containing the protein transfer was opened and membrane was taken out. If ladder was showing on membrane and gel was clear, transfer was correctly done. If not, it was run for further timing.

For immunodetection of the His-tagged VBsilk bands, PVDF membranes with transferred protein bands were washed with 10 mL of TBS-T to cover membrane for 5 mins on rotating table. With a 25 mL glass serological pipet 20 mL of Ponceau S was added for 5 mins on rotating table, then washed with 10 mL of dH_2O for 5 mins on rotating table and bands were identified. Membrane was washed with 5 mL of 0.1 M NaOH for 5 mins on rotating table. In order to avoid membrane-protein interaction, membrane was blocked with 5 mL of blocking solution to cover membrane for 1 hour on rotating table in room temperature. 5 mL of primary antibody His-Tag was added to detect the 6 x His VBsilk on a rotating table overnight in cold room. After 24h primary antibody was discarded, and the membrane washed 2 x 5 mins with 5 mL TBS-T. The membrane was then overlain with the secondary HRP antibody consisting of affinity-purified antibody that detected specified target, for one hour in a cold room, in an aluminium-covered western blot box due to the light sensitivity of the reagents. The PVDF membrane was then washed 2 x 5 mins with 5 mL TBS-T with a final wash of 5 mL TBS for 5 mins. An ECL high sensitivity chemiluminescent substrate detection kit was used for imaging the His-tagged VBsilk bands on a Fuji image QuantLAS 4000 densitometer. The densitometry was made by placing the membrane on top of a zip lock on a black tray for chemiluminescence by eliminating air bubbles. In a 50 mL tube with 1000 µL pipet, 2 mL of Reagent A was added of the ECL Chemiluminescent Substrate kit and 2 mL of Reagent B then mixed with pipet and spilled on surface of membrane to cover it eventually for 1 min (equal parts of reagents). Excess of substrate was absorbed with kimwipes.

The membrane was read on chemiluminescence with Fuji image QuantLAS 4000. Recombinant protein A weighed 39 kDa and bands were shown with a known size of 10 mg. Bands were selected and analyzed for quantification on Fiji with ImageJ application.

Chapter 3 – Results

3.1 VB silk modelling

A high-quality template for the V-region of VBsilk was found using the SWISS-MODEL software, based on the *Streptococcus pyrogenes (S. pyrogenes)* protein *Scl 2* (also referred to as "bacterial collagen"). Both the protein design source and the template were derived from the same sequence, but VBsilk had been "humanized" by substituting amino acids found in human collagen, so they were not identical (54.69% identity). The SWISS-MODEL template search is summarised in Table 2.

Table 2. SWISS-MODEL template search results for V-region of VBsilk. The source structure for the model was an x-ray crystallography model of SclB (/Scl2).

Template	Seq Identity	Oligo-state	QSQE	Method	Resolution	Seq Similarity	Range	Description
4nsm.1.A	54.69	homo-trimer	0.16	X-ray	1.60Å	0.44	44 - 106	Collagen-like protein SclB

The resulting model for this region is a coiled coil homotrimer with an outer alpha helical bundle surrounding the coiled coil (Figure 5). The feasibility of this folding is validated by a Ramachandran plot (Figure 6). The Ramachandran plot represented torsional angles phi (Φ) and psi (Ψ) of the peptide backbone with residues that are likely to reside in α-helices or β-structures. The β-structures are few and likely interrupted as they reside in linker regions between α-helixes. The stability of the inner coiled coil is described by the second peak in (Figure 7).

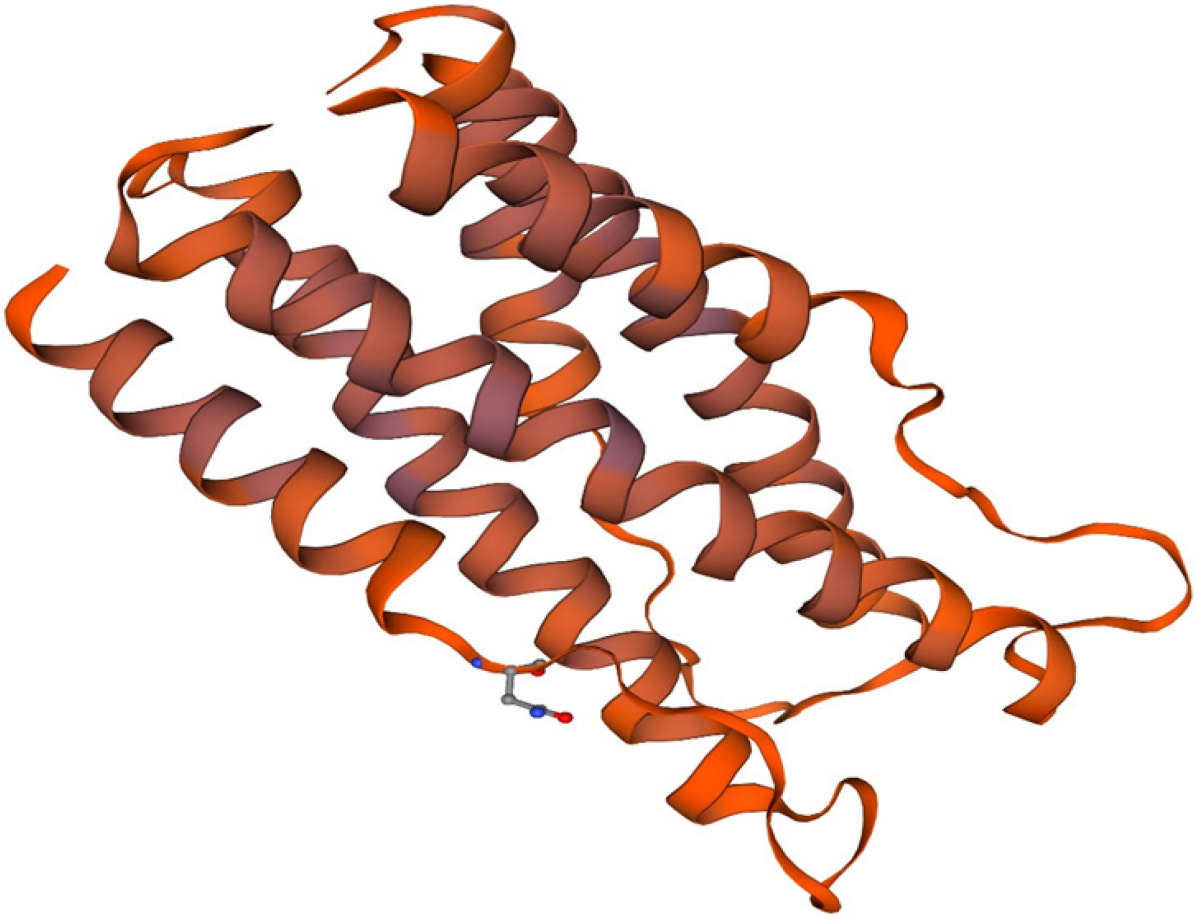


Figure 5. N-terminal region of VBsilk modelled using SWISS-MODEL based on SclB Protein from *Streptococcus pyrogenes*. An inner tri-helical coiled coil is surrounded by three shorter coils, forming an oval globule.

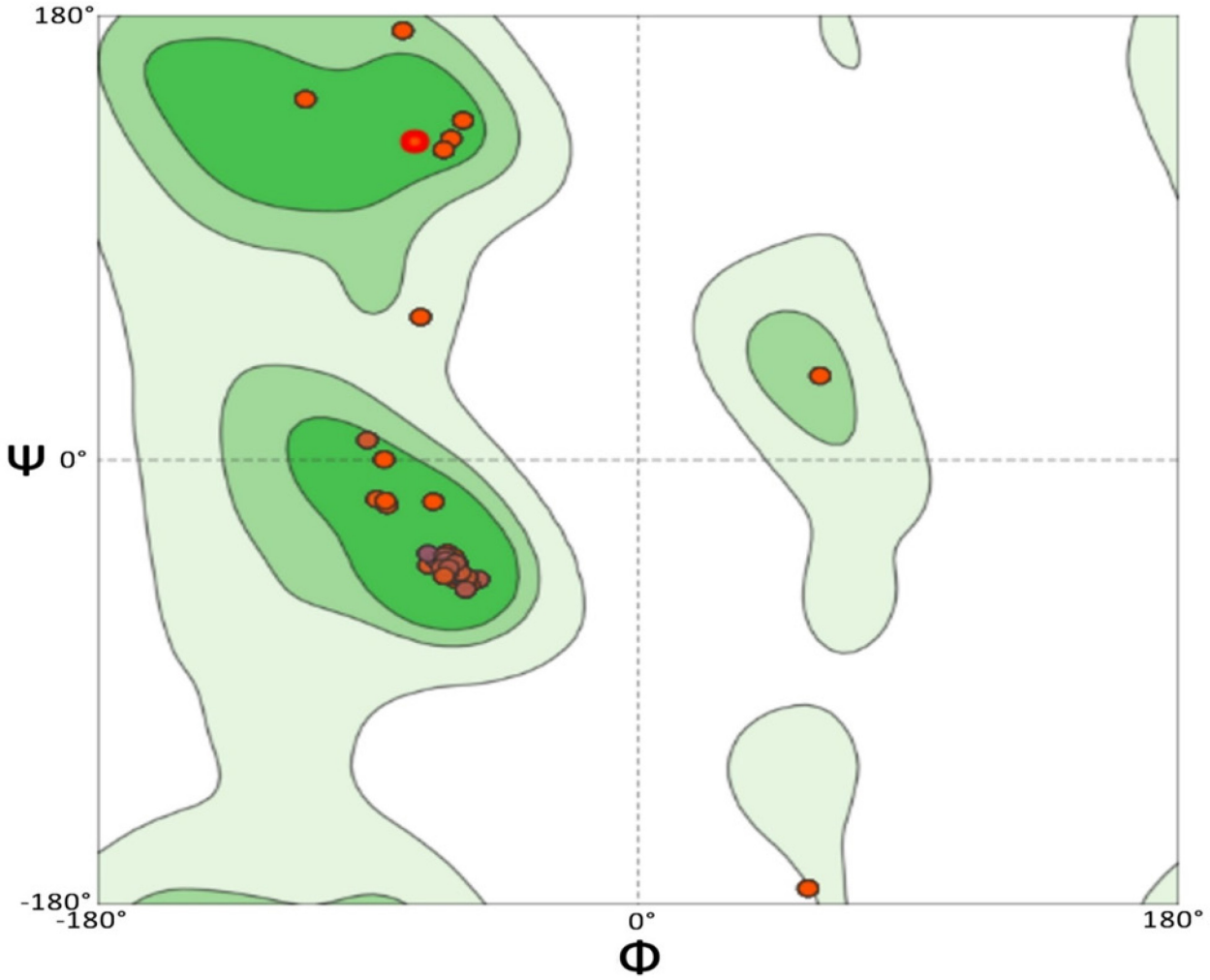


Figure 6. Ramachandran Plot of VBsilk, representing torsional angles of the peptide backbone. The dark region in the top left indicates residues likely to reside in β-structures; the dark region intersecting the 0 on the Y-axis indicates residues likely to reside in α-helices. Residues residing in the upper right quadrant indicates turns. The residue found in the lower right quadrant is in a sterically un-favoured position.

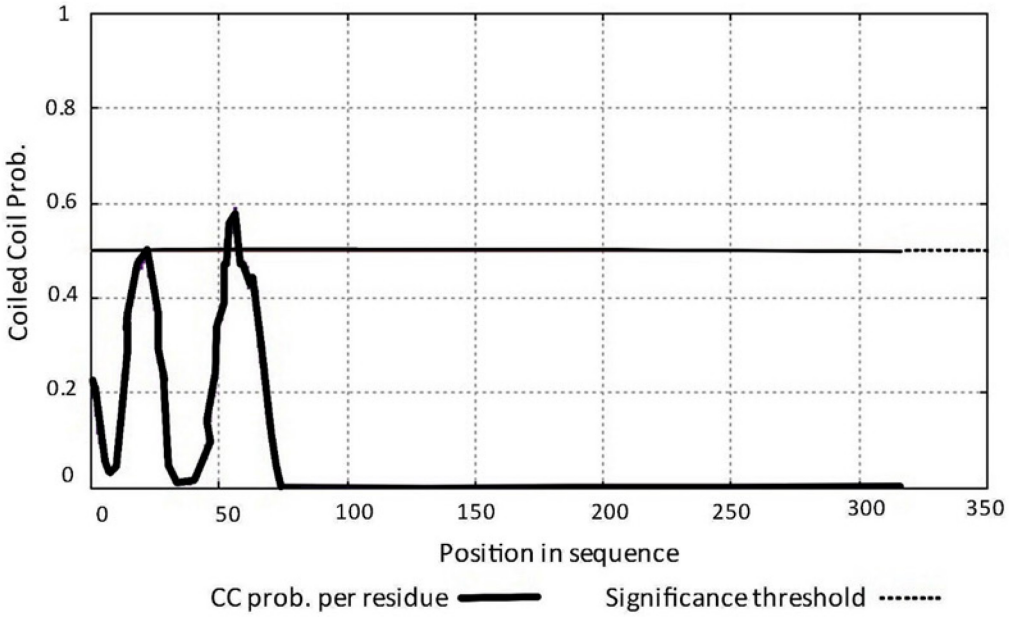


Figure 7. The first 317 amino acids of VBsilk graphed for likelihood for presence of a coiled coil using as predicted using the "DeepCoil" software (Ludwiczak et al., 2019; Zimmermann et al., 2018)

P-BLAST performed on the first third of the collagenous region of VBsilk yielded a close match to LPXTG cell wall anchor domain-containing protein (88% identity, 92% positives, E_{value}: $2e^{-30}$) and to collagen-like protein 2 (also known as SclB/Scl2) (88% identity, 92% positives, E_{value}: $1e^{-29}$) both of these sequences are *S. pyrogenes* proteins. The partial search was due to the VBsilk sequence containing three identical repeats of this region. The collagenous region downstream of the V-region was modelled for stability. Peak stability was found in the central regions of each of the three identical repeated collagen sequences (Figure 8).

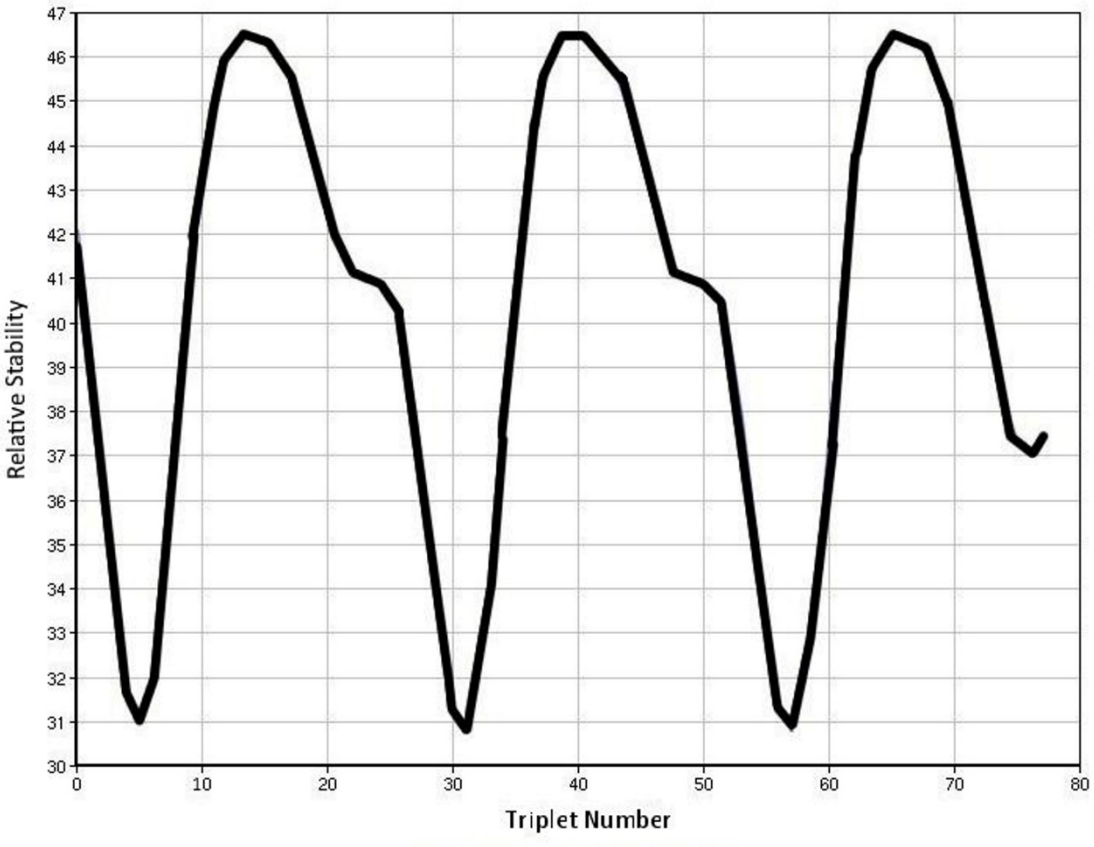
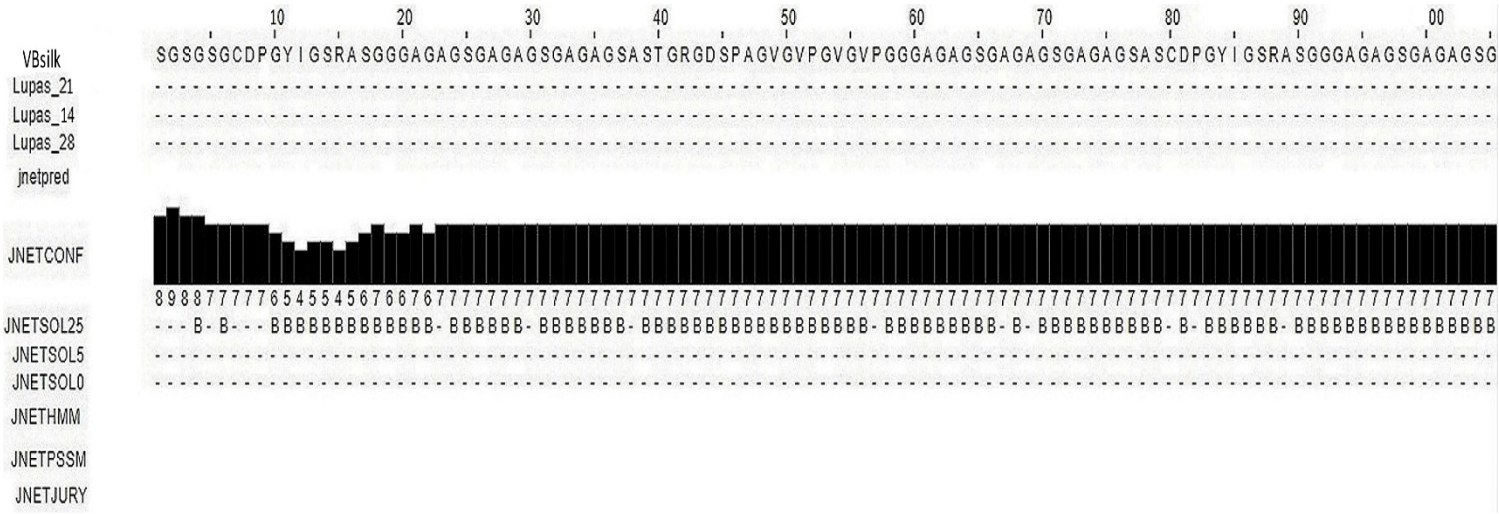
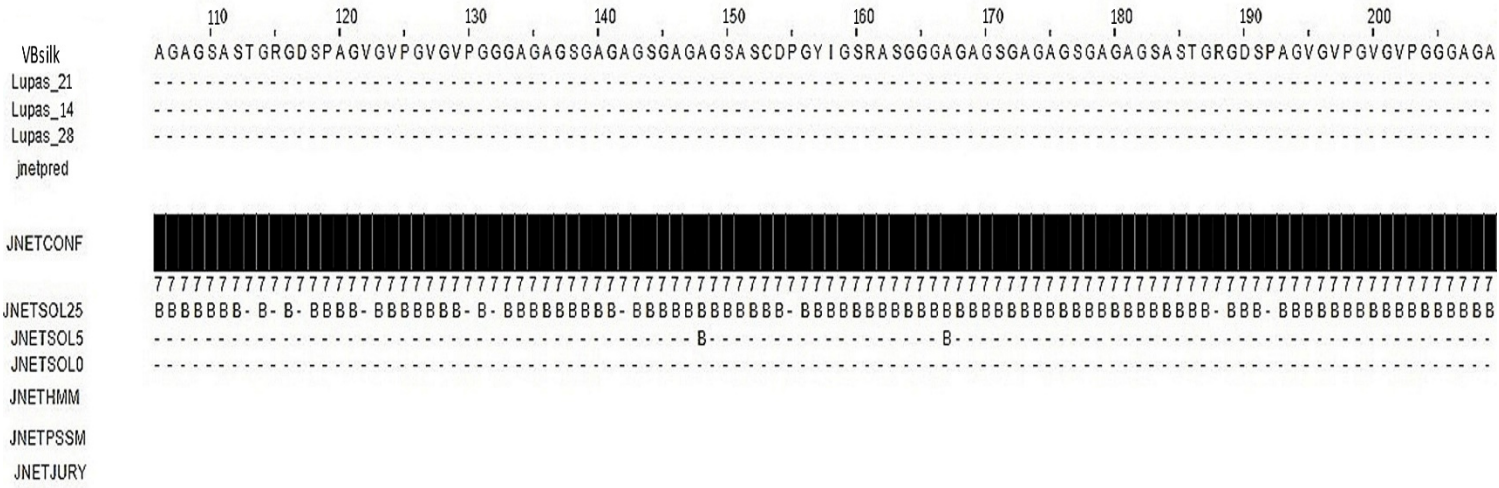


Figure 8. Cumulative gaussian curve of VBsilk CL domain (residue 96-317) stability profile as determined using the Collagen Stability Calculator V2. (Persikov et al., 2005).

The C-terminal domain of VBsilk was difficult to model using traditional methods. P-BLAST searches could not find similar proteins due to its short repeated, non-naturally occurring motifs. Jpred prediction did not find any likely beta or alpha structures in this region but did yield a high confidence prediction for accessibility of peptide side chains (Figure 9). A high degree of buried residues were predicted across this span of the protein, only very short sequences were predicted as entirely exposed, the largest part of this part of the sequence has a partially buried nature, but with > 25 % of side chains being exposed, small regions (4 positions) formed peaks of extremely likely buried residues; GSGAGAGSASC, SGGGAGAGSGAGAG, SGGGAGAGSG, and VGVPGVGVPGGGA were the most likely to be buried.


10 20 30 40 50 60 70 80 90 00
VBsilk
SGSGSGCDPGYIGSRASGGGAGAGSGAGAGSGAGAGSASTGRGDSPAGVGVPGVGVPGGGAGAGSGAGAGSGAGAGSASCDPGYIGSRASGGGAGAGSGAGAGSG
Lupas_21
Lupas_14
Lupas_28
jnetpred
JNETCONF
8988777776545545676676777777777777777777777777777777777777777777777777777777777777777777777777777777777777
JNETSOL25
---B-B---BBBBBBBBBBBBB-BBBBBB-BBBBBBB-BBBBBBBBBBBBBBBBBB-BBBBBBBBB-B-BBBBBBBBBB-B-BBBBBB-BBBBBBBBBBBBBBBB
JNETSOL5
JNETSOL0
JNETHMM
JNETPSSM
JNETJURY


110 120 130 140 150 160 170 180 190 200
VBsilk
AGAGSASTGRGDSPAGVGVPGVGVPGGGAGAGSGAGAGSGAGAGSASCDPGYIGSRASGGGAGAGSGAGAGSGAGAGSASTGRGDSPAGVGVPGVGVPGGGAGA
Lupas_21
Lupas_14
Lupas_28
jnetpred
JNETCONF
7777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777777
JNETSOL25
BBBBBBB-B-B-BBBB-BBBBBBB-B-BBBBBBBBB-BBBBBBBBBBBB-BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB-BBB-BBBBBBBBBBBBBBBBB
JNETSOL5
JNETSOL0
JNETHMM
JNETPSSM
JNETJURY

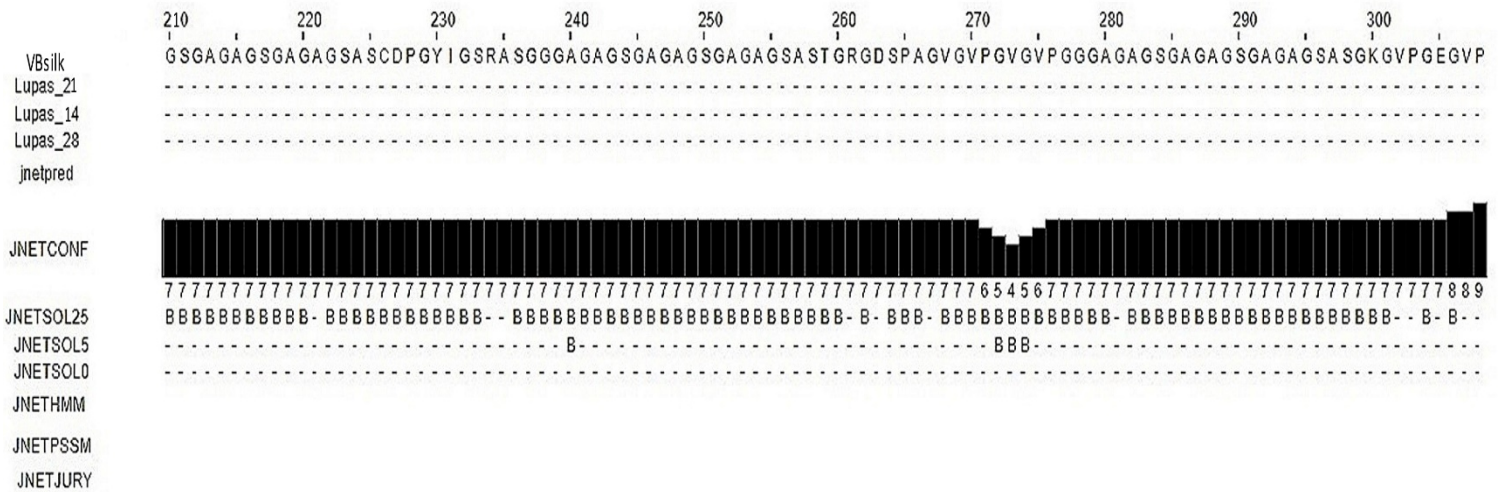

210 220 230 240 250 260 270 280 290 300
VBsilk
GSGAGAGSGAGAGSASCDPGYIGSRASGGGAGAGSGAGAGSGAGAGSASTGRGDSPAGVGVPGVGVPGGGAGAGSGAGAGSGAGAGSASGKGVPGEGVP
Lupas_21
Lupas_14
Lupas_28
jnetpred
JNETCONF
7777777777777777777777777777777777777777777777777777777777777777777765456777777777777777777777777778889
JNETSOL25
BBBBBBBBBBB-BBBBBBBBBBBB--BBBBBBBBBBBBBBBBBBBBBBBB-B-BBB-BBBBBBBBBBBB-BBBBBBBBBBBBBBBBBB--B-B--
JNETSOL5
JNETSOL0
JNETHMM
JNETPSSM
JNETJURY

Figure 9. Jpred prediction of VBsilk C-terminal prediction, sequence starting at residue 318 (Drozdetskiy et al., 2015). Lupas_21, Lupas_14 and Lupas_28 did not find any predicted α-helices or β-sheets. The JNETSOL25, 5 and 0, represents prediction of Solvent Accessibility. The levels are: Exposed, 25% or more Solvent Accessibility accessible, 5% or more Solvent Accessibility accessible and Buried (< 5% exposed). The JNetPRED is the consensus prediction. JNetCONF identified the confidence estimated for the prediction, high values mean high confidence prediction. The JNetHMM-HMM is the profile-based prediction and JNETPSSM-PSSM is the based prediction, no α-helices or β-sheets were found.

The ExPASy PeptideCutter was used to create a list of any likely enzyme targets since the aim is to produce biomaterials for use in the human cornea as implants. The results are summarized in Table 3 (Gasteiger et al., 2005). As seen, many potential degradation sites were found in the VBsilk sequence. Most of these enzymes are ones that are produced by *E. coli*, or other bacteria including skin dwelling bacteria. Examples include the Arg-C and Asp-N endopeptidases, and Staphylococcal peptidase I. Enzymes that are found in human tissue include thrombin, trypsin and pepsin.

Table 3. List of enzymes that could cut VBsilk and the amino acid residue positions affected.

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	21	18 60 68 70 129 141 153 207 219 231 285 297 309 330 357 403 430 476 503 549 576
Asp-N endopeptidase	23	8 41 43 45 76 80 95 127 149 173 205 227 251 283 305 322 358 395 431 468 504 541 577
Asp-N endopeptidase + N-terminal Glu	61	8 9 11 12 19 23 32 41 42 43 45 58 60 63 76 80 85 95 97 112 115 118 127 130 139 142 149 151 163 173 175 190 193 196 205 208 217 220 227 229 241 251 253 268 271 274 283 286 295 298 305 307 322 358 395 431 468 504 541 577 619
BNPS-Skatole	1	67
CNBr	5	1 50 108 186 264
Chymotrypsin-high specificity (C-term to [FYW], not before P)	7	49 56 67 326 399 472 545
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	31	1 2 3 4 5 6 7 21 25 29 40 45 47 49 50 53 54 56 57 67 71 72 78 80 108 186 264 326 399 472 545
Clostripain	21	18 60 68 70 129 141 153 207 219 231 285 297 309 330 357 403 430 476 503 549 576
Formic acid	23	9 42 44 46 77 81 96 128 150 174 206 228 252 284 306 323 359 396 432 469 505 542 578
Glutamyl endopeptidase	38	10 12 13 20 24 33 43 59 61 64 86 98 113 116 119 131 140 143 152 164 176 191 194 197 209 218 221 230 242 254 269 272 275 287 296 299 308 620
Iodosobenzoic acid	1	67
LysC	29	14 16 34 35 52 69 73 84 95 117 126 135 138 149 162 173 195 204 213 216 227 240 251 273 282 291 294 305 615

LysN	29	13 15 33 34 51 68 72 83 94 116 125 134 137 148 161 172 194 203 212 215 226 239 250 272 281 290 293 304 614
NTCB (2-nitro-5-thiocyanobenzoic acid)	4	321 394 467 540
Pepsin (pH1.3)	16	21 24 25 28 29 40 44 45 52 53 56 57 79 92 170 248
Pepsin (pH>2)	24	21 24 25 28 29 40 44 45 48 52 53 55 56 57 66 67 79 92 170 248 326 399 472 545
Proteinase K	197	8 10 12 13 15 17 19 20 21 22 24 25 26 29 32 33 37 39 40 43 45 48 49 51 53 54 55 56 57 59 61 63 64 67 71 72 75 79 80 86 92 98 99 101 105 111 113 116 119 122 125 131 132 140 143 144 147 152 156 159 164 170 176 177 179 183 189 191 194 197 200 203 209 210 218 221 222 225 230 234 237 242 248 254 255 257 261 267 269 272 275 278 281 287 288 296 299 300 303 308 312 315 326 327 331 336 338 342 344 348 350 353 355 362 364 366 369 371 376 378 382 384 388 390 393 399 400 404 409 411 415 417 421 423 426 428 435 437 439 442 444 449 451 455 457 461 463 466 472 473 477 482 484 488 490 494 496 499 501 508 510 512 515 517 522 524 528 530 534 536 539 545 546 550 555 557 561 563 567 569 572 574 581 583 585 588 590 595 597 601 603 607 609 612 617 620 622
Staphylococcal peptidase I	37	10 12 20 24 33 43 59 61 64 86 98 113 116 119 131 140 143 152 164 176 191 194 197 209 218 221 230 242 254 269 272 275 287 296 299 308 620
Thermolysin	118	7 14 16 21 25 28 31 39 49 52 53 56 62 70 71 74 78 79 100 104 107 110 124 146 155 158 178 182 185 188 202 224 233 236 256 260 263 266 280 302 311 314 326 330 335 337 341 343 347 349 352 361 363 368 375 377 381 383 387 389 392 399 403 408 410 414 416 420 422 425 434 436 441 448 450 454 456 460 462 465 472 476 481 483 487 489 493 495 498 507 509 514 521 523 527 529 533 535 538 545 549 554 556 560 562 566 568 571 580 582 587 594 596 600 602 606 608 611
Thrombin	4	357 430 503 576
Trypsin	50	14 16 18 34 35 52 60 68 69 70 73 84 95 117 126 129 135 138 141 149 153 162 173 195 204 207 213 216 219 227 231 240 251 273 282 285 291 294 297 305 309 330 357 403 430 476 503 549 576 615

These chosen enzymes do not cut:

Caspase1

Caspase10

Caspase2

Caspase3

Caspase4

Caspase5
Caspase6
Caspase7
Caspase8
Caspase9
Enterokinase
Factor Xa
GranzymeB
Hydroxylamine
Proline-endopeptidase
Tobacco etch virus protease

3.2 Growth Curve of VBsilk producing E. coli

A growth curve of ClearColi© was produced to determine the optimal time to stop the growth and proceed with the cold shock to induce VBsilk protein production. The optimal time is when the bacteria are still within exponential growth, but with maximal numbers of cells. Hence, the best time is while the bacteria are still growing exponentially before the growth stops and the curve reaches a plateau. Figure 10 shows the growth curve obtained for VBsilk-transfected ClearColi©.

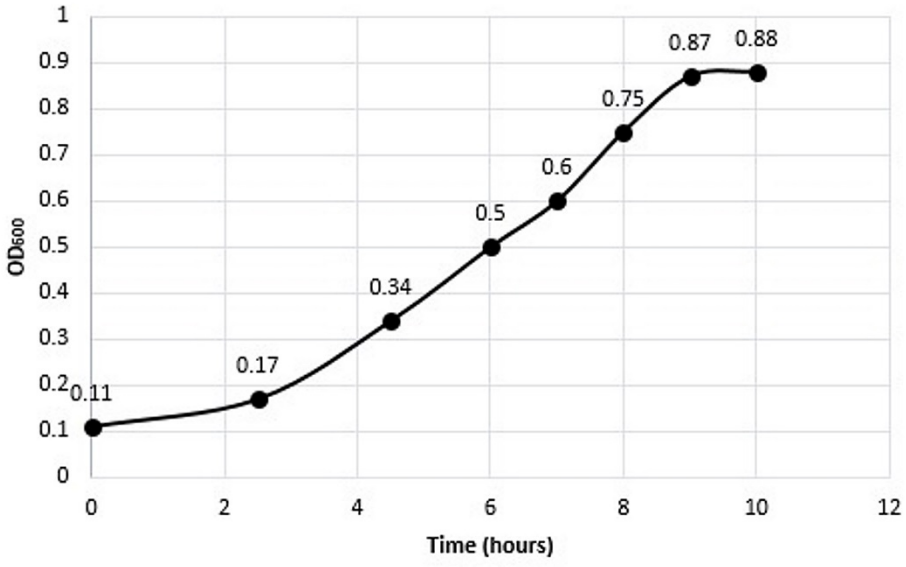


Figure 10. Growth curve of ClearColi transfected with pColdIII-VBsilk. The growth starts slower and becomes exponential between 4 to 8 hours with a plateau showing at hours 9 and 10.

3.3 FPLC Purification of VBsilk

FPLC purification of VBsilk is optimized with different buffers having different pH and salt concentrations. The buffers were chosen to have pH that was either above or below the pI of VBsilk which was calculated to be 5.41. UV absorbance indicated protein content.

Buffers 1 and 2 had a neutral pH of 7.41, although they were two pH units above the pI of VBsilk. In the chromatogram obtained with buffers 1 and 2 (Figure 11), identifiable protein peaks were obtained, which when investigated by SDS-PAGE and western blotting, corresponded to VBsilk. Buffers 3 to 6 that were tested had a lower pH of 4, and high salt content is used from buffer 3 to 8. Buffer 3 through 8 did not show chromatogram peaks during fraction collection.

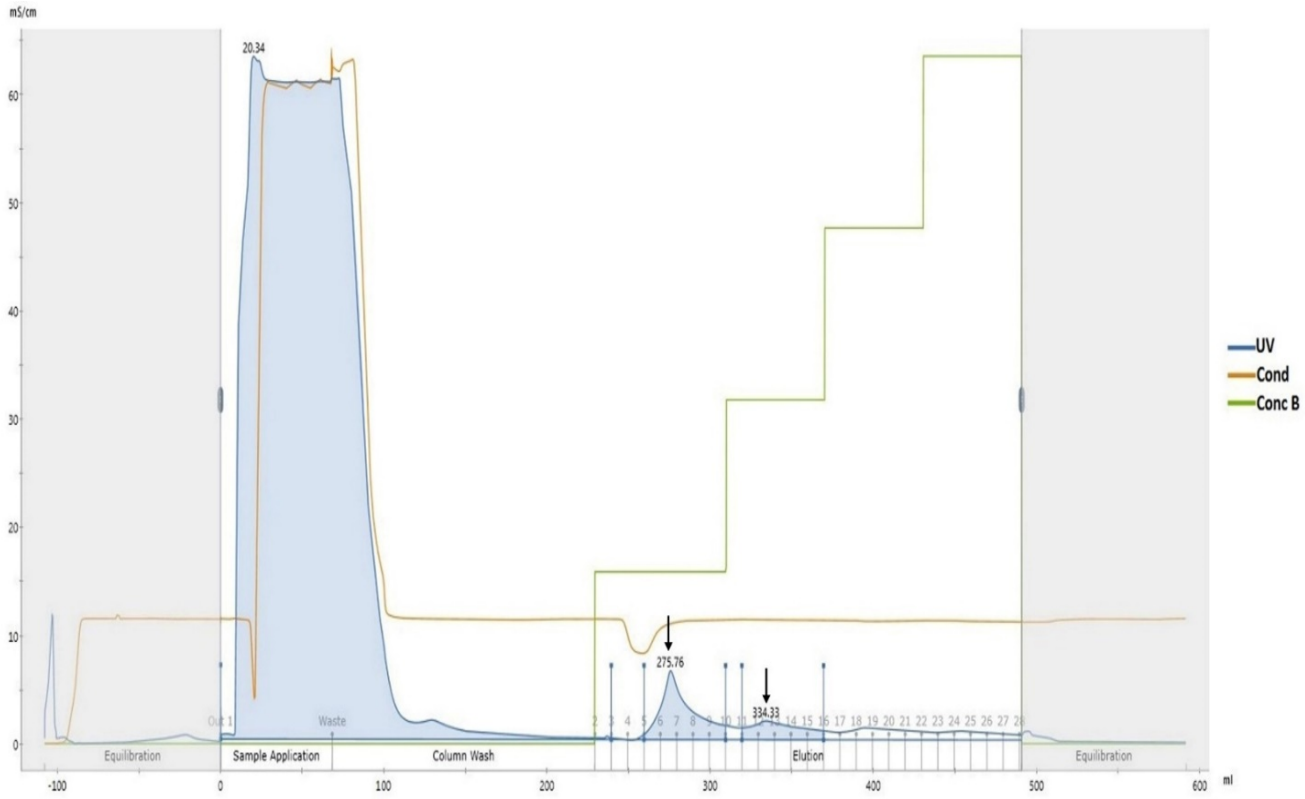


Figure 11. FPLC chromatogram describing VBsilk loading on a HisPrep FF 16/10 column. This chromatogram represented mS/cm of conductivity in function of mL protein solutions pumped. Protein peaks are shown in fraction 6 and 12, as indicated by the arrows.

3.4 SDS-PAGE

The BSA is served as a positive control for VBsilk quantification. Contaminant proteins can however be seen in fraction 6 (two bands of lower molecular weight) using buffer 1 (Figure 12). A clear band is seen in fraction 12 with a lower yield but with a higher purity. A larger amount of VBsilk band is seen in the flow through which contain many contaminants. Calculated yields from each buffer condition is summarized in Table 4. Protein quantification of SDS-PAGE gels was performed using Fiji with ImageJ application. Densitometry scans obtained with Fiji and ImageJ software give a two-dimensional graph representing x-axis as horizontal distance from peak. Peak areas represent protein bands, which were used to calculate VBsilk yield (Figure 13).

There was a large amount of protein in the flow through, including bands in the position of VBsilk. As the flow through volume was about 10 times the volume of the fractions collected, it can be estimated that five to ten times the concentration of pure VBsilk was being lost during the FPLC purification process.

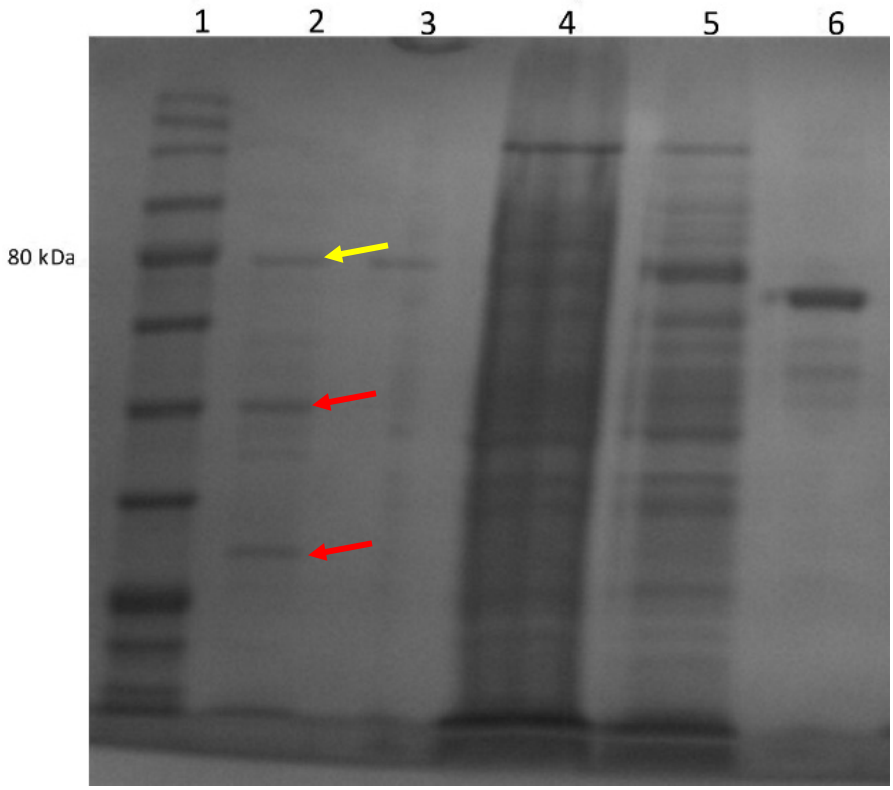


Figure 12. SDS-PAGE gel showing separation of fractions of VBsilk obtained after FPLC. Lane 1, 3 µL color protein broad range ladder; Lane 2, 1 µL fraction 6; Lane 3, 1 µL fraction 12; Lane 4, 10 µL flow through; Lane 5, 1 µL flow through; Lane 6: 0.5 µg BSA. Three prominent bands are marked by arrows. The yellow arrowed band is approximately 80 kDa, corresponding most likely to VBsilk. The smaller red arrowed bands could be degradation products of VB silk or contamination from bacterial proteins.

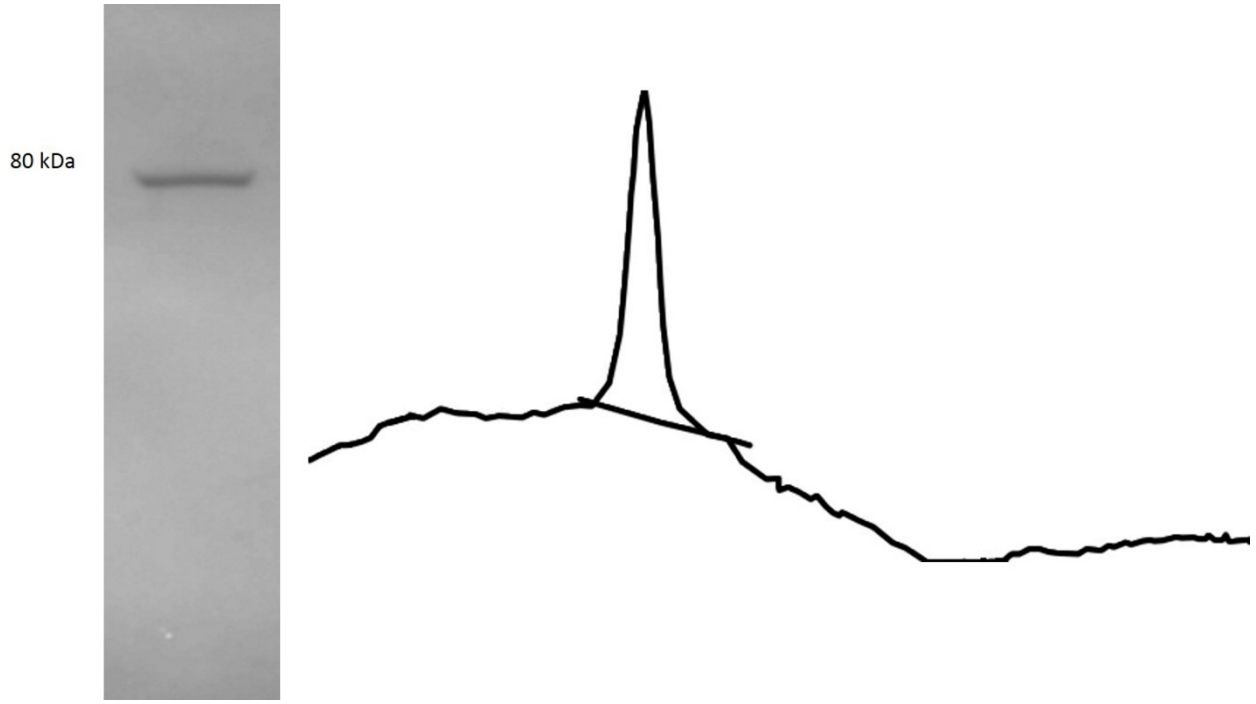


Figure 13. Protein quantification of SDS-PAGE gels using Fiji with ImageJ application. Plot profile displayed a two-dimensional graph representing x-axis as horizontal distance from peak. Peak areas represent protein bands, which were used to calculate VBsilk yield. In this case, the striking peak seen corresponds to the 80 kDa band seen that has the expected molecular weight for VBsilk.

Table 4. Effect of different buffering conditions on VBsilk yield.

Buffer	Run Number	Culture Volume (L)	Fraction	VBsilk Yield (mg/L)
Buffer 1	No. 1	2.5	4	1.07
Buffer 2	No. 3	2.5	10	1.32
Buffer 2	No. 4	2.5	10	1.95
Buffer 1	No. 10	2.5	12	0.69
Buffer 1	No. 12	2	12	0.88
Buffer 1	No. 14	4	10	0.42
Buffer 1	No. 15	2	10	0.84

3.5 Western Blotting

His-tagged recombinant protein A served as a positive staining control. The presence of few minor secondary bands showed that the VBsilk can be obtained with very limited degradation (Figure 14). The absence of the two contaminant bands of lower molecular weight that were marked by the red arrows in Figure 12 of the SDS-PAGE separated proteins showed that those bands were most likely bacterial contaminants rather than degraded VBsilk.

The high amounts of VBsilk protein that was detected in the flow through showed that there was inefficient protein binding to FPLC column.

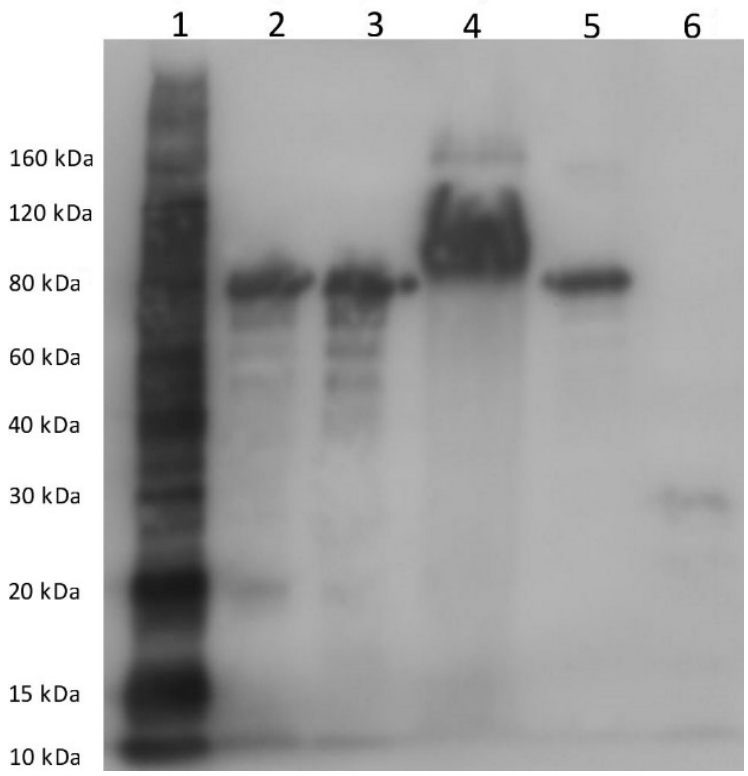


Figure 14. Western blot of proteins obtained after FPLC purification using anti-His6 antibody. Lane 1, 2.5 µL his-tagged protein standard; Lane 2, 1 µL fraction 6; Lane 3, 1 µL fraction 12; Lane 4, 10 µL flow through; Lane 5, 1 µL flow through; Lane 6: 0.2 µg Recombinant Protein A. Positive staining confirmed the presence of VBsilk in the fractions collected. However, VBsilk is also seen in the flow through lanes.

Chapter 4 – Discussion

4.1. Protein modelling of VBsilk

VBsilk is a chimeric protein of 625 amino acids. Modelling results from the SWISS-MODEL showed that the coiled coil region of VBsilk has a 54.69% identity to three identical units of polypeptide of SclB, a "bacterial collagen" with a QSQE of 0.16. The tertiary structure of the region immediately preceding the collagenous region is a coiled coil homotrimer that can initiate collagenous triple-helix folding (Yu et al., 2010). The modelling results also showed that VBsilk contains minimal α-helical and β-sheet portions; α-helical and β-sheet structures are mainly found in the N-terminal region proceeding the collagenous region. The C-terminal region consists of small motifs with a high tendency to form tracts with amorphous/variable folding, most of the side chain in this region will form complexes with each other via electrostatic interactions.

Coiled coils are stabilized by hydrophobic amino acids and electrostatic contacts that might bind to the FPLC column. However, analysis of the coiled-coils structures could depend on the solvent, temperature, pH and salt interactions. The 2D Ramachandran plot modelled with the same application represented α-helices and β-sheet that could saturate hydrogen bonds in the peptide backbone and insure stability. The torsion angles could predict the folding. Moreover, the highest density of dots is seen in the right-handed α-helix. The residues in the left-handed α-helix are glycine or asparagine or aspartate, however, glycine does not have side chains and can be found in all four sections of the Ramachandran plot. Further applications are used for this prediction. Following an algorithm, DeepCoil application calculated the probability of prediction of the coiled coils stability that determined two peaks of α-helical regions above the significant threshold.

With the use of the Collagen Stability Calculator, the downstream CL domain was shown to have the stability of the triple-helical structure of the collagen. This is achieved by combination of stabilizing molecular mechanisms. The design choices involving repeated inclusion of the sequence GVGVP is likely prohibiting definite determination of secondary and tertiary of this

region, as this motif is known to form a metastable amorphous network of hydrogen bonds (Rodriguez-Cabello et al., 1999). The full secondary sequence is made with the prediction of the C-terminal of VBsilk with Jpred by the JNet algorithm. JNet algorithm predicted the buried or exposed amino acids in the folded structure of the protein. It also predicted Solvent Accessibility and Coiled-coils regions. A range of simulations was performed and basically, the model predicted that the C-terminal of VBsilk is largely unstructured and amorphous.

Many potential enzyme degradation sites were found in the VBsilk sequence using the ExPASy PeptideCutter software. As the aim is to make implants from VB silk, it would be important to know what enzymes could cut and destabilize the implants. Knowing the degradation sites could also allow modification of the protein by removing sequences or adding to it for future customization. The results showed that most of the enzymes that would affect VB silk are from *E. coli* or skin dwelling bacteria. Three human enzymes – pepsin, trypsin and thrombin could also degrade VBsilk. However, these are found in the stomach and blood, respectively. So, there is not much concern that either pepsin, trypsin or thrombin will be a problem for implants in the eye. However, it would be important to ensure that VBsilk is produced under very clean or sterile conditions to circumvent bacterial attack.

The molecular weight of VBsilk was predicted from its amino acid sequence as well as the N- and C-terminal modifications. Several molecular weight calculators are available on-line (https://www.selleckchem.com/peptide-calculator.html; https://www.peptide2.com/peptide_molecular_weight_calculator.php). The actual protein obtained was 80 kDa, which corresponded to the predicted molecular weight.

4.2 VBsilk producing E. coli and induction of protein production

Growth of ClearColi© bacteria producing VBsilk followed a sigmoid growth pattern. Once initiated, the growth became exponential over several hours until a plateau was reached. As *E. coli* is a facultative anaerobe, i.e., it can grow in both aerobic and anaerobic conditions. When oxygen is present, it will use oxygen to grow. However, if oxygen is absent, it can switch to

anaerobic respiration. However, under anaerobic conditions, bacterial growth slows down. Additionally, during production process, certain cultures might have grown faster than others due to the culture volume made of 4 L or 2 L and the amount of bacteria added (Kram et al., 2014). VBsilk production is maximal during the exponential phase. Hence, induction to produce VBsilk is optimal and best when there are a maximum number of bacterial cells present to make protein.

The pColdIII vector was chosen to express VBsilk because the protein is a large one and difficult to express, and therefore was not amenable to expression using the more common T7 system. The pCold vector system was engineered to specifically express difficult proteins, repressing the production of bacterial proteins allowing for expression of the target protein. VBsilk was expressed but the purification results showed that bacterial proteins were also expressed. The cold shock protocol used that was optimized for other collagen-like peptides (Edin et al. 2020) differed from that of the manufacturer. It is possible that temperature for the cold shock will need to be optimized for better VBsilk yield.

The modelling which showed the structure of VBsilk is fairly complex with both coiled coils and beta-sheets had suggested the use of the pColdIII system. The results confirmed that the use of this expression vector was able to guide the vector selection. The pI can also be calculated from the sequence which allowed for buffer selection for purification.

4.3 FPLC Purification of VBsilk

The modelling showed that VBsilk was a relatively complex peptide with different secondary structures. Even though the pCold system was used, it was still expected that expressed protein would remain intracellular as opposed to excreted, so the bacteria had to be lysed to release the protein for purification.

For ion exchange purification, FPLC buffers used for binding and elution could be of different pH. The pH is used was designed to be more acidic or basic than the pI of the protein to allow VBsilk to carry a net positive or negative charge to bind to the column. When pH is below the pI of VBsilk, the carboxylic group of the protein is hydrogenated and VBsilk is positively

charged. Conversely, when pH is above the pI, VBsilk is negatively charged. The pH can be changed to elute the protein off the column. More commonly, however, pH remains the same, but the elution is done with a buffer with higher ionic strength, as in the present case.

The overall pI is relevant for folding and quaternary structure, which can affect binding. This could be explained by the binding of other amino acids that have affinity for metal binding at neutral pH such as cysteine, lysine, arginine, glutamine that are protonated (Sigel et al., 2007). The FPLC buffers used had different concentrations of urea and imidazole to aid purification. Salt concentrations are important due to the increasing ionic strength that competed with the bound components for charges on the surface of the medium and for protein elution as a monomer. Detergents are often used for cell lysis to lease proteins, and they also break up protein-protein and protein-lipid interaction. There are several classes of detergents but only non-ionic detergents such as Triton X-100 and Tween20, and a chaotropic detergent, urea, were tested in this thesis project.

Overall, in buffers 1 and 2, low concentration of salt with a basic pH led to solubility of the protein by competing for electrostatic interactions even if buffer 1 has low yield, gels showed high purity of VBsilk. Buffer 1 had similar urea concentrations in both binding and elution solutions. Buffer 2 was made with lower urea and imidazole concentrations for column binding. A higher urea and imidazole concentration were used for elution. The amounts of urea used were not highly denaturing but appeared to be sufficient to prevent unwanted folding of the VBsilk and to keep the protein solubilized, since we were able to obtain fractions of the protein. The higher imidazole competes with the his-tag that is bound to the column and helps in the elution. While there was insufficient data for statistical analysis, buffer 2 did result in higher yields of VBsilk of 0.43 mg/L and 0.65 mg/L respectively. No non-ionic detergents were added to these buffers.

The phosphate-based buffers used did not have a wide buffering range and might have contributed to a lower yield due to low VBsilk binding to the column. However, to allow better binding, sample could be loaded repeatedly on column (GE Health Care Life Sciences, 2012a). Other buffers were excluded since they were not recommended buffering for the column or cannot be used downstream in chemistry which demands excessive purification. For example,

Tris-based buffers were not used because of possible interactions with downstream chemistry. An extensive dialysis step would need to be added between FPLC and lyophilization at a minimum, to allow for Tris to be used as a buffer. Additionally, the column used may not have been optimal. Even if nickel is the better ion to purify his-tag, other metal ions could be tested on the column with HiTrap™ IMACHP 1 ml column (GE Health Care Life Sciences, 2004).

4.4 SDS-PAGE and Western Blot

The production of recombinant proteins in bacteria must have high yields and purity to be useful for tissue engineering. The SDS-PAGE showed that some FPLC purification conditions yielded VBsilk with different purities. The SDS-PAGE sample buffers used were selected to break all non-covalent interactions between proteins by reducing the disulfide bridges to dissociate the protein. Some secondary bands were seen in protein fractions collected. Western blotting showed that most of the bands of lower molecular weight were not stained for the His-tag. This suggested that the bands resulting more likely from contamination with undesired bacterial proteins rather than denaturation of the VBsilk during collection and processing.

The large amount of VBsilk in in the flow through suggests that the bacteria lysis and solubilization of the proteins process were not efficient. A large amount of VBsilk had most likely remained within inclusion bodies that were not lysed and therefore did not get purified. The presence of contaminating protein coupled with the presence of VBsilk in the flow through or waste buffer shows that the purification process has yet to be optimized. The fractions of protein collected however do show that although small, pure amounts of VBsilk can be obtained.

A total of 7.17 mg of VBsilk was obtained. To make a single corneal implant, however, 10 to 20 mg of peptide will be needed. This means that gram-scale amounts of purified protein will be needed. The results indicate that while purification could be optimized, the major issues were in the lysis of the bacteria and solubilization of the inclusion bodies. This means an extra step is needed to extract the insoluble inclusion bodies and several protocols are available (Palmer et al., 2004; Singh et al., 2015).

4.5. Conclusions and potential impact of results

Protein simulations predicted the structure of VBsilk based on the modelling done providing information on how to produce protein in this thesis. The modelling allowed for prediction of the protein structure that allowed for planning for the expression and subsequent purification strategy for VBsilk.

VBsilk DNA was expressed in pColdIII and transfected into low endotoxin producing ClearColi©. The growth of the ClearColi© followed a standard sigmoid curve with an exponential growth rate. The ClearColi© produced proteins that were lysed to give rise to a crude protein lysate that could be purified by FPLC. SDS-PAGE and Western blotting confirmed that the proteins contained VBsilk.

VBsilk was produced in milligram quantities in this thesis and there was a lot of loss of proteins and contamination with bacterial proteins was noted. This shows that the purification process by FPLC has yet to be optimized. However, despite the inefficiencies, fractions of VBsilk were collected. Moreover, in the future, with optimization of the purification, the production could be scaled up to produce sufficient amounts for use to produce biomaterials for making implants, including those that can be used as corneal implants.

Bibliographic references

An, B., Kaplan, D. L., & Brodsky, B. (2014). Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Front Chem, 2*, 40. doi:10.3389/fchem.2014.00040

Aref, A. A., & Vajaranant, T. S. (2015). Management of Glaucoma Associated with Boston KPro Type I Implantation. In *Keratoprostheses and Artificial Corneas* (pp. 107-115).

Avadhanam, V. S., & Liu, C. S. (2015). A brief review of Boston type-1 and osteo-odonto keratoprostheses. *Br J Ophthalmol, 99*(7), 878-887. doi:10.1136/bjophthalmol-2014-305359

Avadhanam, V. S., Smith, H. E., & Liu, C. (2015). Keratoprostheses for corneal blindness: a review of contemporary devices. *Clin Ophthalmol, 9*, 697-720. doi:10.2147/OPTH.S27083

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). *Biochemistry. 5th edition.*

Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res, 45*(D1), D313-D319. doi:10.1093/nar/gkw1132

Bonanno, J. (2003). Identity and regulation of ion transport mechanisms in the corneal endothelium. *Progress in Retinal and Eye Research, 22*(1), 69-94. doi:10.1016/s1350-9462(02)00059-9

Bourne, R. R. A., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., . . . Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health, 5*(9), e888-e897. doi:10.1016/s2214-109x(17)30293-0

Bradshaw, M., Ho, D., Fear, M. W., Gelain, F., Wood, F. M., & Iyer, K. S. (2014). Designer self-assembling hydrogel scaffolds can impact skin cell proliferation and migration. *Sci Rep, 4*, 6903. doi:10.1038/srep06903

Branden, C. I., & Tooze, J. (2012). *Introduction to Protein Structure*: Garland Science, Taylor & Francis Group, LLC, an informa business.

Bray, L. J., George, K. A., Ainscough, S. L., Hutmacher, D. W., Chirila, T. V., & Harkin, D. G. (2011). Human corneal epithelial equivalents constructed on Bombyx mori silk fibroin membranes. *Biomaterials, 32*(22), 5086-5091. doi:10.1016/j.biomaterials.2011.03.068

Brazel, D., Oberbäumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R., & Kühn, K. (1987). Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 nontriplet interruptions located within the collagenous domain. . *Eur. J. Biochem, 168*(3), 529-536.

DelMonte, D. W., & Kim, T. (2011). Anatomy and physiology of the cornea. *J Cataract Refract Surg, 37*(3), 588-598. doi:10.1016/j.jcrs.2010.12.037

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Res, 43*(W1), W389-394. doi:10.1093/nar/gkv332

Edin, E., Simpson, F., & Griffith, M. (2020). Synthesis and Application of Collagens for Assembling a Corneal Implant. *Springer*.

Fagerholm, P., Lagali, N. S., Merrett, K., Jackson, W. B., Munger, R., Liu, Y., . . . Griffith, M. (2010). A biosynthetic alternative to human donor tissue for inducing corneal regeneration: 24-month follow-up of a phase 1 clinical study. *Science Translational Medicine, 2*(46), 46ra61.

Fagerholm, P., Lagali, N. S., Ong, J. A., Merrett, K., Jackson, W. B., Polarek, J. W., . . . Griffith, M. (2014). Stable corneal regeneration four years after implantation of a cell-free recombinant human collagen scaffold. *Biomaterials, 35*(8), 2420-2427. doi:10.1016/j.biomaterials.2013.11.079

Fletcher, J. M., Boyle, A. L., Bruning, M., Bartlett, G. J., Vincent, T. L., Zaccai, N. R., . . . Woolfson, D. N. (2012). A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth Biol, 1*(6), 240-250. doi:10.1021/sb300028q

Gain, P., Jullienne, R., He, Z., Aldossary, M., Acquart, S., Cognasse, F., & Thuret, G. (2016). Global Survey of Corneal Transplantation and Eye Banking. *JAMA Ophthalmol, 134*(2), 167-173. doi:10.1001/jamaophthalmol.2015.4776

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571-607.

GE Health Care Life Sciences. (2004). *HisPrep™ FF 16/10*.

GE Health Care Life Sciences. (2012a). *ÄKTA™ Laboratory-scale Chromatography Systems*.

GE Health Care Life Sciences. (2012b). *ÄKTA™ pure*.

Graf, J., Iwamoto, Y., Sasaki, M., Martin, G. R., Kleinman, H. K., Robey, F. A., & Yamada, Y. (1987). Identification of an Amino Acid Sequence in Laminin Mediating Cell Attachment, Chemotaxis, and Receptor Binding *Cell Press, 48*(6), 989-996.

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis, 30 Suppl 1*, S162-173. doi:10.1002/elps.200900140

Heindl, L. M., Riss, S., Bachmann, B. O., Laaser, K., Kruse, F. E., & Cursiefen, C. (2011). Split cornea transplantation for 2 recipients: a new strategy to reduce corneal tissue cost and shortage. *Ophthalmology, 118*(2), 294-301. doi:10.1016/j.ophtha.2010.05.025

Islam, M. M., Sharifi, R., Mamodaly, S., Islam, R., Nahra, D., Abusamra, D. B., . . . Gonzalez-Andrades, M. (2019). Effects of gamma radiation sterilization on the structural and biological properties of decellularized corneal xenografts. *Acta Biomater, 96*, 330-344. doi:10.1016/j.actbio.2019.07.002

Khakshoor, O., & Nowick, J. S. (2008). Artificial beta-sheets: chemical models of beta-sheets. *Curr Opin Chem Biol, 12*(6), 722-729. doi:10.1016/j.cbpa.2008.08.009

Kramer, L. (2013). Corneal transplant wait list varies across Canada. *CMAJ, 185*(11), E511-512. doi:10.1503/cmaj.109-4517

Li, F., Carlsson, D., Lohmann, C., Suuronen, E., Vascotto, S., Kobuch, K., . . . Griffith, M. (2003). Cellular and nerve regeneration within a biosynthetic extracellular matrix for corneal transplantation. *Proceedings of the National Academy of Sciences of the United States of America, 100*(26), 15346-15351.

Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V., & Dunin-Horkawicz, S. (2019). DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics, 35*(16), 2790-2795. doi:10.1093/bioinformatics/bty1062

Mathews, P. M., Fogla, R., Samayoa, E., VanCourt, S., & Akpek, E. K. (2019). Long-term clinical outcomes of keratoplasty using gamma-irradiated corneal lenticules. *BMJ Open Ophthalmol, 4*(1), e000396. doi:10.1136/bmjophth-2019-000396

Meek, K. M., & Knupp, C. (2015). Corneal structure and transparency. *Prog Retin Eye Res, 49*, 1-16. doi:10.1016/j.preteyeres.2015.07.001

Melke, J., Midha, S., Ghosh, S., Ito, K., & Hofmann, S. (2016). Silk fibroin as biomaterial for bone tissue engineering. *Acta Biomater, 31*, 1-16. doi:10.1016/j.actbio.2015.09.005

Nesloney, C. L., & Kelly, J. W. (1996). Progress Towards Understanding β-Sheet Structure *Bioorganic & Medicinal Chernistry, 4*(6), 739-766.

Nishida, T. (1997). *Cornea: Fundamentals of cornea and external disease in Cornea*. Mosby-Year Book Inc.

Nouri, M., Terada, H., Alfonso, E. C., Foster, S., Durand, M. L., & Dohlman, C. H. (2001). Endophtalmitis After KeratoprosthesIs. *Archives of Ophtalmology, 119*(4), 484.

Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., . . . Polarek, J. (2003). Recombinant collagen and gelatin for drug delivery. *Adv Drug Deliv Rev, 55*(12), 1547-1567. doi:10.1016/j.addr.2003.08.008

Palmer, I., & Wingfield, P. T. (2004). Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. Current protocols in protein science, Chapter 6, Unit–6.3. Retrieved from: https://doi.org/10.1002/0471140864.ps0603s38

Persikov, A. V., Ramshaw, J. A., & Brodsky, B. (2005). Prediction of collagen stability from amino acid sequence. *J Biol Chem, 280*(19), 19343-19349. doi:10.1074/jbc.M501657200

Qing, G., Ma, L. C., Khorchid, A., Swapna, G. V., Mal, T. K., Takayama, M. M., . . . Inouye, M. (2004). Cold-shock induced high-yield protein production in Escherichia coli. *Nat Biotechnol, 22*(7), 877-882. doi:10.1038/nbt984

Rama, P., Matuska, S., Paganoni, G., Spinelli, A., De Luca, M., & Pellegrini, G. (2010). Limbal Stem-Cell Therapy and Long-Term Corneal Regeneration *The New England Journal of Medicine, 363*(2), 147-155.

Rubert Perez, C. M., Stephanopoulos, N., Sur, S., Lee, S. S., Newcomb, C., & Stupp, S. I. (2015). The powerful functions of peptide-based bioactive matrices for regenerative medicine. *Ann Biomed Eng, 43*(3), 501-514. doi:10.1007/s10439-014-1166-6

Samish, I. (2017). *Computational Protein Design* Springer Protocols.

Shi, Y., Bikkuzin, T., Song, Z., Jin, X., Jin, H., Li, X., & Zhang, H. (2017). Comprehensive evaluation of decellularized porcine corneal after clinical transplantation. *Xenotransplantation, 24*(6), 11.

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M., & Panda, A.K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Fact 14, 41. Retrieved from: https://doi.org/10.1186/s12934-015-0222-8

Shoulders, M. D., & Raines, R. T. (2009). Collagen structure and stability. *Annu Rev Biochem, 78*, 929-958. doi:10.1146/annurev.biochem.77.032207.120833

Sigel, A., Sigel, H., & Sigel, R. K. O. (2007). *Metal Ions in Life Sciences* (Vol. 2). John Wiley & Sons, Ltd.

Sridhar, M. S. (2018). Anatomy of cornea and ocular surface. *Indian J Ophthalmol, 66*(2), 190-194. doi:10.4103/ijo.IJO_646_17

Talajic, J. C., Agoumi, Y., Gagne, S., Moussally, K., & Harissi-Dagher, M. (2012). Prevalence, progression, and impact of glaucoma on vision after Boston type 1 keratoprosthesis surgery. *Am J Ophthalmol, 153*(2), 267-274 e261. doi:10.1016/j.ajo.2011.07.022

Tashiro, K., Sephel, G. C., Weeks, B., Sasaki, M., Martin, G. R., Kleinman, H. K., & Yamada, Y. A. (1989). A Synthetic Peptide Containing the IKVAV Sequence from the A Chain of Laminin Mediates Cell Attachment, Migration, and Neurite Outgrowth. *THE JOURNAL OF BIOLOGICAL CHEMISTRY, 264*(27), 16174-16182.

Vepari, C., & Kaplan, D. L. (2007). Silk as a Biomaterial. *Prog Polym Sci, 32*(8-9), 991-1007. doi:10.1016/j.progpolymsci.2007.05.013

Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., & Kivirikko, K. I. (1997). Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast Pichia pastoris: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *The EMBO Journal, 16*(22), 6702–6712.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res, 46*(W1), W296-W303. doi:10.1093/nar/gky427

World Health Organisation. (2018). Blindness and vision impairment prevention.

Yang, C., Hillas, P. J., Baez, J. A., Nokelainen, M., Balan, J., Tang, J., . . . Polarek, J. W. (2004). The Application of Recombinant Human Collagen in Tissue Engineering. *BioDrugs, 18*(2), 103-119.

Yu, T., Rajendran, V., Griffith, M., Forrester, J. V., & Kuffova, L. (2016). High-risk corneal allografts: A therapeutic challenge. *World J Transplant, 6*(1), 10-27. doi:10.5500/wjt.v6.i1.10

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., . . . Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol, 430*(15), 2237-2243. doi:10.1016/j.jmb.2017.12.007

Appendices

A1. Dialysis for optimization of FPLC Buffers

Spectra/Por molecular porous membrane tubing (Cat #25218-468) and Spectra/Por® Closures for Dialysis Membrane Tubing (Cat #97059-180) were purchased from VWR (Mont-Royal, QC). MOPS (Cat# BP308-500) were bought from Thermo Fisher Scientific (St-Laurent, QC). All other chemicals were purchased from Sigma-Aldrich Canada Ltd (Oakville, ON).

3 liters of buffers 3 to 8 from table 3 were made in a 5 L beaker each. Dialysis membrane tubing was cut to the correct length for the sample volume by adding 5 to 10 cm to have space to attach the dialysis tubing closures. The dialysis tubing was put in a 500 mL beaker of dH_2O for 5 mins to remove the residual preservatives from the tubing before its use. The bottom end of the dialysis tube was clamped by folding over the end of the tubing and attaching the weighted closure on the folded edge. The tubing was fully open with a 50 mL serological pipette at the top end to allow air into the tubing, then sample was carefully pipetted in. The end of the tubing was folded over and clamped at the top ensuring that the air bubble remained between the closure and the sample. In a 1 L beaker, 1 L of the 3 L buffers 3 to 8 were added respectively with an appropriately sized stir bar and sample was put in. Beakers were put on a mini stir plate at 200 rpm and buffers were changed after 6 hours then after ~12 hours to further separate molecules.

A2. FPLC Preparation and $NiSO_4$ Loading

Correct charging of the sepharose column was verified by visual inspection of the coloration of the column, and an extinguished UV and conductivity signal. A shift in column color from white to blue validates binding of $NiSO_4$, and an extinguished UV and conductivity signal validates that there is no further leaching of Ni from the column. Figure A1 shows the chromatogram obtained from the loading.

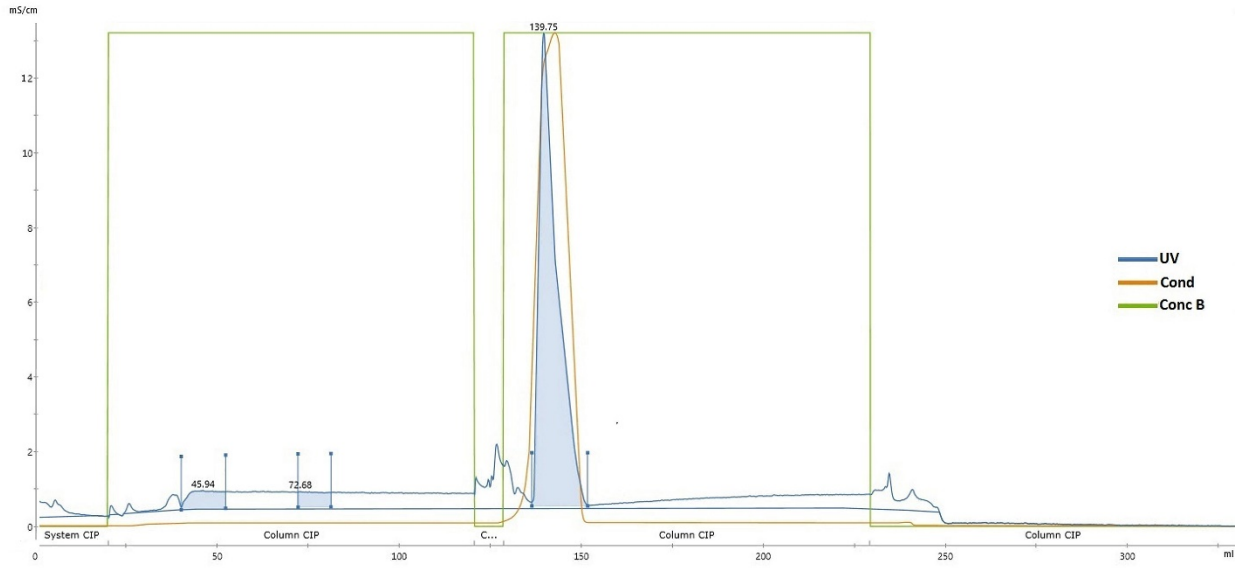


Figure A1. FPLC chromatogram describing $NiSO_4$ loading on HisPrep FF 16/10 column. This chromatogram is represented mS/cm of conductivity in function of mL pumped. System CIP, Column CIPs and Column loading are displayed.