UNIVERSITY OF CALIFORNIA,
IRVINE

Recombinant Collagen Variants for the Production of Mechanically and Biofunctionally Tunable Hydrogels

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Richard A. Que

Dissertation Committee:
Professor Szu-Wen Wang, Co-Chair
Professor Nancy Da Silva, Co-Chair
Assistant Professor Wendy Liu

2016

DEDICATION

To

my grandparents, Conrado and Francisca

my parents, Aristotle and Grace

my brother and sister, Dennison and Yi-Hsuan

and Dinh

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	two dimensional
3D	three dimensional
AcOH	acetic acid
AF532	Alexa Fluor 532 C5 maleimide
AFM	atomic force microscopy
ANOVA	analysis of variance
bCol	bovine collagen
bFGF	basic fibroblast growth factor
BIT9500	bovine serum albumin, insulin, and transferrin
BSA	bovine serum albumin
calcein-AM	calcein acetoxymethyl
cAMP	cyclic adenosine monophosphate
CD	circular dichroism
CEN/ARS	centromere sequence / autonomously replicating sequence
Cys	cysteine
ddH2O	double-distilled water
DDR	discordin domain receptor
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DPBS	Dulbecco's phosphate buffered saline
DTSSP	(3,3'-dithiobis(sulfosuccinimidyl propionate))

DTT	dithiolthreitol
EC	endothelial cell
ECFC-EC	endothelial colony forming cell-derived endothelial cell
ECM	extracellular matrix
EDC	N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide
EDTA	ethylenediaminetetraacetic
EGF	epidermal growth factor
EGM-2	endothelial cell growth medium 2
f	frequency
FBS	fetal bovine serum
G'	storage modulus
G''	loss modulus
GFAP	glial fibrillary acidic protein
GFOGER	hexapeptide glycine-phenylalanine-hydroxyproline-glycine-glutamic acid-arginine
GSPGGK	hexapeptide glycine-serine-proline-glycine-glycine-lysine
HA	hyaluronic acid
hCol	human collagen
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
IgG	immunoglobulin G
IKVAV	peptapeptide isoleucine-lysine-valine-alaline-valine
LN	laminin
Mal	maleimide

MAP2	microtubule-associated protein 2
MEM	minimal eagle's medium
MMP	matrix metalloproteinase
MSD	mean square displacement
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NHLF	normal human lung fibroblast
NSPC	neural stem/progenitor cell
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PDMS	polymethylsiloxane
PEG	polyethylene glycol
PEO	polyethylene oxide
PF	primary fragment
PHSRN	pentapeptide phenylalanine-histidine-serine-arginine-aspargine
PMSF	phenylmethanesulfonyl fluoride
rCol	recombinant collagen III
RGD	tripeptide arginine-glycine-aspartic acid
RNA	ribonucleic acid
RNAseq	RNA sequenceing
RT	room temperature
S.E.M.	standard error of the mean

SDM	site directed mutagenesis
SDS-PAGE	denaturing polyacrylamide gel electrophoresis
SF	secondary fragment
SMCC	succinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate
Sulfo-SMCC	(sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate)
TBS	tris-buffered saline
TCEP	tris(2-carboxyethyl)phosphine
TEM	transmission electron microscopy
TEMED	tetramethylethylenediamine
TGF-β	transforming growth factor beta 1
T_m	apparent melting temperature
VEGF	vascular endothelial growth factor
αSMA	alpha smooth muscle actin
γ	strain

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chairs, Dr. Szu-Wen Wang and Dr. Nancy Da Silva, for allowing me this opportunity and for their support and guidance. I also would like to express my thanks to Dr. Wendy Liu who graciously served on my committee. Without their guidance and support this dissertation would not have been possible.

I would also like to express my thanks to the members of the Wang Lab and Da Silva Lab for their thoughtful discussions, support and friendship. I thank Medea Neek and Tae Il Kim for the coffee break brainstorming sessions. I would like to additionally thank Dr. Michael Shen, Dr. Jin Choi, Dr. Abeer Jabaiah, and Dr. Sam Wei Polly Chan for their mentorship, training and brainstorming about puzzling experimental results.

I express my thanks to all my collaborators: Dr. Mohraz and Dr. Hubert Chan for support and guidance for sample characterization by rheology, Dr. Christopher Hughes and Duc Phan for providing cells and guidance for the endothelial cell and vascularization work, and Dr. Lisa Flanagan and Dr. Janahan Arulmoli for the neural stem/progenitor cell work. Financial support was provided by NSF Grants DMR-1006999 and CBET-1034566 and UCI's Graduate Division Bridge award.

Lastly, I would like to thank my family and Dinh Diep, who have always supported me throughout this journey.

Curriculum Viate

Richard Que

Education

2010-2016	Ph.D. Biomedical Engineering	University of California – Irvine
2010-2014	M.S. Biomedical Engineering	University of California – Irvine
2005-2009	B.S. Bioengineering (Biotechnology) Minor: Economics	University of California – San Diego

Teaching Experience

2013-2015	Teaching Assistant, Chemical Engineering	University of California – Irvine

Prior Employment

2011-2016 **Graduate Student Researcher, University of California - Irvine**
Engineering and expression of recombinant collagen in *S. cerevisiae* for modulating cell microenvironment

2009-2010 **Programmer / Analyst I**, **University of California - San Diego**
Developer for the COBRAToolbox http://opencobra.github.io/

2008-2009 **Lab Assistant I**, **University of California – San Diego**
In sillico modeling of Transcription/Translation and Metabolism of *E. coli*.
Expression and purification of recombinant protein in *E. coli.*

Presentations

4: "Engineering of Integrin Recognition Sites in Recombinant Human Collagen III to Control Cellular Response" 2015 Biomedical Engineering Society (BMES) Annual Meeting – Tampa, FL, October 2015, Richard Que, Sam Wei Polly Chan, Richard Lathrop, Nancy A. Da Silva, Szu-Wen Wang

3: "Recombinant Collagen Variants for the Production of Tunable Hydrogel Scaffolds" 2013 American Institute of Chemical Engineers Annual Meeting – San Francisco, CA, November 2013, Richard Que, Ali Mohraz, Nancy A. Da Silva, Szu-Wen Wang

2: "Endothelial Cell-Mimicking Surfaces to Mitigate the Host Response to Cardiovascular Implatns" ASM WEST / MATSCI UCI Symposium on Biomaterials, Medical Devices and Tools: Challenges with Design, Fabrication and Testing – Irvine, CA 2011, November 2011, Yoon Kyung Kim, Richard Que, Szu-Wen Wang, and Wendy Liu

1: "The COBRA Toolbox for Systems Biology v2.0." 110th General Meeting of the American Society for Microbiology – San Diego, CA, May 2010, Jan Schellenberger, Richard Que, Andrei Osterman, Bernhard Ø. Palsson, and Karsten Zengler

Peer-Reviewed Publications

6: Arulmoli, J., Wright, H. J., Phan, D. T. T., Sheth, U., Que, R. A., Botten, G. A., et al. (2016). Combination scaffolds of salmon fibrin, hyaluronic acid, and laminin for human neural stem cell and vascular tissue engineering. Acta Biomaterialia, 43, 122-138.

5: Que, R. A., Chan, S. W. P., Jabaiah, A. M., Lathrop, R. H., Da Silva, N. A., & Wang, S.-W. (2015). Tuning cellular response by modular design of bioactive domains in collagen. Biomaterials, 53, 309-317.

4: Que, R., Mohraz, A., Da Silva, N. A., & Wang, S.-W. (2014). Expanding Functionality of Recombinant Human Collagen Through Engineered Non-Native Cysteines. Biomacromolecules, 15(10), 3540-3549.

3: Kim, Y. K., Que, R., Wang, S. W., & Liu, W. F. (2014). Modification of Biomaterials with a Self-Protein Inhibits the Macrophage Response. Advanced healthcare materials, 3(7), 989-994.

2: Thiele, I., Fleming, R. M., Que, R., Bordbar, A., Diep, D., & Palsson, B. O. (2012). Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. PLoS One, 7(9), e45635.

1: Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature Protocols, 6(9), 1290-1307.

ABSTRACT OF THE DISSERTATION

Recombinant Collagen Variants
for the Production of Mechanically and Biofunctionally Tunable Hydrogels

By

Richard A. Que

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2016

Professor Szu-Wen Wang, Co-Chair

Professor Nancy Da Silva, Co-Chair

As the most abundant protein in humans, collagen has been utilized as a tissue engineering scaffold. However, it exhibits the same limitations as other natural substrates, such as the inability to decouple the modulation of mechanical and biofunctional properties. The properties of the cell microenvironment guide cell fate, making it advantageous to modulate the properties of the microenvironment. In this work, we employed a modular collagen platform to produce recombinant collagen variants with prescribed functionalities to independently tune the properties of collagen. Collagen variants with two to eight non-native cysteines were synthesized to form hydrogels of varying stiffness while holding protein and functional site concentration constant. The variants were characterized by circular dichroism and found to be triple-helical with melting temperatures near 37°C. Hydrogel gelation characteristics and stiffness were measured through microrheology and bulk rheology. Non-native cysteines were also utilized as specific covalent anchoring sites for TGF-β, which induced myofibroblast differentiation of fibroblasts that were cultured on these substrates.

We also manipulated integrin-binding sites, which are known to affect cell fate. Using a collagen variant with its integrin-binding sites removed, we tested the effects of the α2β1 integrin-binding sequence GFOGER from collagen I. By introducing GFOGER at up to four locations throughout the polymer, we demonstrated that the location at which GFOGER was introduced affected cellular adhesion. One location resulted in 39% of the cells adhered compared to native collagen III, while other locations along the polymer restored native levels of cellular adhesion. This variation hints at the importance of local context on the functionality of the inserted bioactive sequences and that additional modulation of properties can be achieved by placing the functional sequences at different locations throughout the polymer. Furthermore, modulation of cellular adhesion levels was achieved through mixing collagen substrates at different ratios.

We also demonstrated that the modular collagen platform also allows for the mixing-and-matching of fragments with distinct functionalities. Non-native cysteines and GFOGER were introduced within the same collagen protein to test the effect of GFOGER sites in 3D culture. NHLFs were successfully encapsulated within collagen-variant hydrogels. NHLFs remained spherical when encapsulated with a non-cell adhesion supporting collagen variant, while introduction of GFOGER into this variant resulted in NHLFs invading the gel. This research demonstrates the ability to produce full-length collagen variants with customizable functionalities to tune hydrogel stiffness and integrin-binding site presentation for potential tissue engineering applications such as vascularization.

Chapter 1: Introduction

1.1 Introduction

Collagen is a naturally occurring protein, and cells have the ability to degrade and remodel it as necessary [1]. As collagen is a major protein in the ECM, its function is to provide mechanical support for cells and it is widely used in tissue engineering applications. Defects in collagen can lead to dysfunctional connective tissues, such as in a severe form of Ehlers-Danlos syndrome where the body synthesizes a defective form of collagen III leading to brittle blood vessels and life-threatening health problems [2].

1.1.1 Properties of collagen

Collagen is a fibrous protein which provides structural support for resident cells and is a major component of the extracellular matrix (ECM). It is composed of three monomers which coil to form a triple helix with an end-to-end length of approximately 300 nm and diameter of approximately 1 nm (Figure 1.1) [3,4]. The triple-helical region of collagen is mainly composed of the repeated motif of Gly-X-Y, where X is typically proline, and Y is typically hydroxyproline [5]. The glycines are packed within the center of the helix while the prolines and hydroxyprolines help stabilize the protein through hydrogen bonding [6]. It is important to note that the post translational modification of proline to hydroxyproline is necessary to increase the thermostability of the collagen [7]. The thermostability of the collagen triple-helix is important, as the triple-helical structure provides resistance to most proteases, and is necessary for self-assembly into fibrils and for cellular adhesion [8–10].

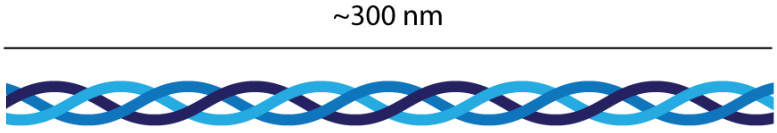


Figure 1-1: Triple-helical collagen. Collagen is composed of three identical monomers which coil to form a triple helix. The end to end length is approximately 300 nm.

Due to its abundance in the natural cell scaffold within humans as well as its natural cell binding capabilities, collagen has been utilized as a cell scaffold for tissue engineering purposes and

provides a strong base for the production of a customizable scaffold [11,12]. Within the extracellular matrix (ECM), collagen provides structural support for cells to adhere and migrate [13]. It has also been shown that the integrin binding site GFOGER from collagen I has pro-vasculogenic properties [14].

1.1.2 Angiogenesis and vasculogenesis

Angiogenesis and vasculogenesis, the formation of new vasculature from existing vasculature and *de novo*, respectively, play a significant role in morphogenesis, wound healing, and tumor formation [15]. Angiogenesis is broken up into a series of steps including migration, proliferation, alignment, tubule formation, branching, and anastomosis [16–18]. The process occurs in response to two sets of extracellular signals: soluble factors and ECM signals.

The intricate processes of angiogenesis and vasculogenesis are pushed forward by the presence of growth factors. Several growth factors have been implicated in promoting angiogenesis, with vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF) being two potent factors [17,19,20]. VEGF induces endothelial cell proliferation, promotes cell migration, and inhibits apoptosis [19]. bFGF is a potent mitogen leading to the migration and invasion of endothelial cells to form tubules [20]. Both VEGF and bFGF can induce vasculogenesis *in vivo* [19,21].

It has been shown that in addition to soluble factors, ECM proteins also affect angiogenesis. ECM proteins contain ligands which can bind to integrin receptors to promote adhesion, differentiation, and migration [14,22]. Substrate stiffness has also been shown to affect angiogenesis. Within flexible gels, the vascular networks formed were thin and dense with many sprouting cells and small vacuoles formed from single cells. Within ridged gels, deeper networks were formed (>30 μm) with multicellular lumens [23]. Another study showed that with increasing rigidity and integrin binding site concentration, total tube length increases [24].

1.1.3 Natural vs synthetic materials for tissue engineering scaffolds

Various materials have been utilized to produce tissue engineering scaffolds, ranging from natural to synthetic, each of which has their advantages and disadvantages. Synthetic materials, unlike natural materials, have the advantage of being highly tunable. Stiffness can be easily modified by incorporating crosslinking reagents. Peptides can be incorporated within the synthetic matrix to provide biofunctionality [25–28]. Peptides containing MMP cleavage sites or hydrolysable linkers can be added to modulate degradation rate [26]. Cell binding capability can be added through the integration of RGD or GFOGER peptides, and growth factors can be conjugated to modulate cells [24,26,28]. Others have incorporated collagen-based peptides within these synthetic scaffolds [27]. Although synthetic materials such as PEG provide a blank slate, allowing for complete control of the number of cell binding sites and degradation sites, this approach is limited by our understanding of the complex interactions between cells and their ECM [29].

Natural materials, such as collagen and fibrin, are biocompatible, degradable into non-toxic products, and contain innate cell adhesion sites and cell signaling capabilities. Natural materials interact better than synthetic materials with cells and can also enhance cell performance [29,30]. However, one drawback of natural materials is availability. Cadavers can be utilized, but they are also limited in supply [31]. The use of other mammals can solve issues of availability; however, the health risks increase with the use of allogenic or xenogenic sources. Materials from these sources would increase the risk of transmitting pathogens to the recipient as not all pathogens are able to be detected and removed [32,33].

1.1.3.1 Tuning the properties of natural scaffolds

Regardless of the source, natural materials have batch-to-batch variation and are not easily tunable mechanically [34]. Cells are modulated by the stiffness of their substrate environment which can help determine the physical properties of the resulting repaired tissue [35]. Scaffold stiffness can be increased using various methods of crosslinking [36–38]. Crosslinking limits the accessibility

of degradation sites, thereby decreasing degradation rate. Unfortunately, some crosslinking reagents can affect proliferation and biosynthesis. Glutaraldeyde introduces cytotoxic aldehyde molecules, and N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (EDC) introduces cytotoxic carbodiimide; even after extensive washing, these toxic components cannot be removed completely [36]. Although crosslinking provides a method to modulate substrate stiffness, it does not allow for increases in degradation rate. Furthermore, decreasing degradation rate by crosslinking cannot be decoupled from the stiffness of the scaffold.

Growth factors can also be introduced onto natural scaffolds utilizing adsorption or covalent attachment. EDC can be used to covalently link the carboxyl group on growth factors to amines in the scaffold [39]. As previously mentioned, the use of EDC is cytotoxic, and it can additionally allow growth factors to covalently link to one another instead of to the scaffold. To get around this issue, sulfhydryls can be introduced to the scaffold using Traut's reagent which binds to primary amines [12]. Growth factors can then be conjugated to the scaffold's introduced sulfhydryls using sulfo-SMCC [12] (Figure 1-2). As many proteins do not contain exposed sulfhydryls, utilizing this method would limit the dimerization of the growth factor. However, the locations of the introduced sulfhydryls cannot be controlled completely using this strategy due to the abundance of lysine within proteins. Both natural and synthetic scaffolds have their advantages and disadvantages. Thus, a scaffold which provides the advantages of both natural and synthetic materials would be ideal.

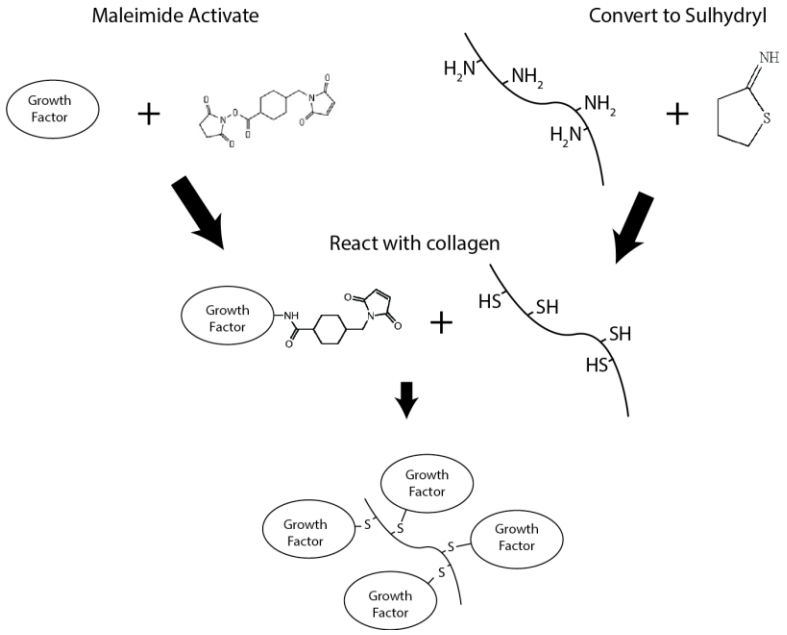


Figure 1-2: Schematic for immobilization of growth factors to collagen with SMCC. Growth factors are maleimide activated by reacting with SMCC. Sulfhydryls are introduced into collagen through reaction with Traut's reagent which reacts with primary amines. The maleimide activated growth factors can then react with the introduced sulfhydryls to covalently link the growth factors to collagen.

1.1.4 Modular collagen platform

Our lab (Chan et al. [40]) published methods to produce recombinant human collagen III *de novo* utilizing engineered *Saccharomyces cerevisiae* as a host, which can be used to address the issues of sourcing, inhomogeneity, and tuning of mechanical and biofunctional properties. The synthesis of a full-length gene encoding collagen is challenging, due to the dual requirement of repetitive amino acid sequences interspaced with unique sequences. By utilizing the degeneracy of the genetic code, oligonucleotide sequences were optimized to have a melting temperature of correct hybridization higher than the melting temperature of any incorrect hybridization [41]. These oligonucleotides can then be assembled by polymerase chain reaction (PCR) to form the collagen gene [42] (Figure 1-3). Oligonucleotides are PCR assembled into primary fragments. Primary fragments are PCR assembled into secondary fragments. Finally, secondary fragments are PCR assembled into the full-length gene. By changing the oligonucleotides used to assemble the gene, collagen variants containing specific mutations can be produced.

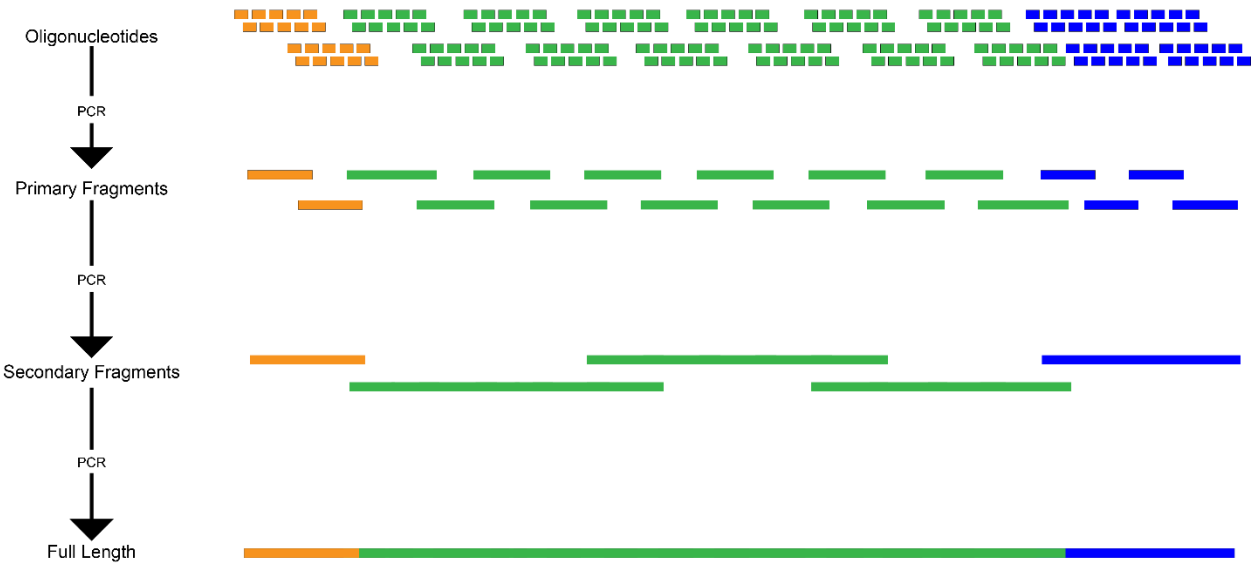


Figure 1-3: Schematic of modular collagen assembly. The collagen gene is divided into three sections, front (orange), middle (green), and back (blue). The middle region codes for the triple-helical section of collagen. The sections were split into overlapping oligonucleotides with lengths of approximately 60 nucleotides. The oligonucleotides were then assembled by PCR into primary fragments of approximately 300 bases. Primary fragments were assembled by PCR into secondary fragments. And secondary fragments were assembled by PCR into the full-length gene.

Hydroxylation of proline is necessary for the thermostability of the triple helix of collagen. To provide the proper hydroxylation of the collagen polymers, *S. cerevisiae* was engineered to express human prolyl-4-hydroxylase. Two copies of prolyl-4-hydroxylase α and β subunits under the *GAL1* promoter were integrated into the chromosomes of BY4741ΔTRP1 (*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 *trp1*::KanMX) to produce BYα2β2 [43]. This yeast system provides for the expression of recombinant collagen with triple-helical thermostability near physiological [44].

This modular collagen system provided a bottom up approach to generating a full-length collagen III gene with the potential to produce custom tailored collagen with precise control of functional sequences. Thus, the collagen molecules can be modified at the DNA level. Recombinant collagen variants were produced in yeast, utilizing the methods outlined in Chan et al. [40], by combining the advantages of natural collagen with the customizability of synthetic scaffolds.

1.2 Motivation and Objectives

The knowledge and materials to control cellular behaviors is critical for endeavors in tissue engineering. It is known that cells respond to stimuli from their surroundings and that the ECM can guide cell fate [45]. ECM stiffness has been shown to have an effect on differentiation [46], while covalent conjugation of growth factors can increase the potency of the growth factors effect on cells [47]. Functional peptide sequences from ECM proteins have also been shown to help guide cellular processes towards desired phenotypes. Therefore, it is advantageous to produce an ECM which can be custom tailored with the required properties to direct cells towards the required phenotypes. Others have addressed this by utilizing short synthetic peptides and polymers [14,24,26,48], but, the context in which the short peptides are presented may have additional implications.

Utilizing this modular collagen system developed by our research group as a base, my research aimed to synthesize and investigate collagen variants which can be utilized as a customizable ECM for teasing out the contributions of individual ECM properties on cell fate processes. My Ph.D. work had three specific objectives:

(1) To construct and characterize recombinant collagen variants with varying numbers of non-native cysteines to serve as specific immobilization sites. These sites were used to immobilize growth factors and for cross-linking to control hydrogel stiffness.

(2) To construct and characterize recombinant collagen variants with varying numbers of the integrin binding site GFOGER. These variants included a non-cell supporting variant of collagen and variants with non-native GFOGER introduced to test the functionality of GFOGER in the context of full length collagen.

(3) To construct and characterize recombinant collagen variants combining both non-native cysteines and GFOGER to test the functionality of GFOGER in 3D cell cultures.

1.3 References

[1] A.Z. Eisen, J.J. Jeffrey, J. Gross, Human skin collagenase. Isolation and mechanism of attack on the collagen molecule, Biochim. Biophys. Acta. 151 (1968) 637–45.

[2] K.N. North, D.A. Whiteman, M.G. Pepin, P.H. Byers, Cerebrovascular complications in Ehlers-Danlos syndrome type IV, Ann Neurol. 38 (1995) 960–4. doi:10.1002/ana.410380620.

[3] K. Gelse, E. Poschl, T. Aigner, Collagens - structure, function, and biosynthesis, Adv Drug Deliv Rev. 55 (2003) 1531–1546.

[4] J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution, Science. 266 (1994) 75–81.

[5] P. Szpak, Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis, J. Archaeol. Sci. 38 (2011) 3358–3372. doi:10.1016/j.jas.2011.07.022.

[6] J. Bella, H.M. Berman, Crystallographic Evidence for Cα–H···O=C Hydrogen Bonds in a Collagen Triple Helix, J. Mol. Biol. 264 (1996) 734–742. doi:10.1006/jmbi.1996.0673.

[7] S. Jimenez, M. Harsch, J. Rosenbloom, Hydroxyproline stabilizes the triple helix of chick tendon collagen, Biochem. Biophys. Res. Commun. 52 (1973) 106–114. doi:10.1016/0006-291x(73)90960-1.

[8] L. Ryhänen, E.J. Zaragoza, J. Uitto, Conformational stability of type I collagen triple helix: Evidence for temporary and local relaxation of the protein conformation using a proteolytic probe, Arch. Biochem. Biophys. 223 (1983) 562–571. doi:10.1016/0003-9861(83)90621-5.

[9] J. Emsley, C.G. Knight, R.W. Farndale, M.J. Barnes, Structure of the Integrin α2β1-binding Collagen Peptide, J. Mol. Biol. 335 (2004) 1019–1028. doi:10.1016/j.jmb.2003.11.030.

[10] K.E. Kadler, D.F. Holmes, J.A. Trotter, J.A. Chapman, Collagen fibril formation., Biochem. J. 316 (1996) 1–11.

[11] F.J. O'Brien, Biomaterials & scaffolds for tissue engineering, Mater. Today. 14 (2011) 88–95. doi:10.1016/S1369-7021(11)70058-X.

[12] Q. He, Y. Zhao, B. Chen, Z. Xiao, J. Zhang, L. Chen, W. Chen, F. Deng, J. Dai, Improved cellularization and angiogenesis using collagen scaffolds chemically conjugated with vascular endothelial growth factor, Acta Biomater. 7 (2011) 1084–1093. doi:10.1016/j.actbio.2010.10.022.

[13] J. Emsley, C.G. Knight, R.W. Farndale, M.J. Barnes, R.C. Liddington, Structural Basis of Collagen Recognition by Integrin α2β1, Cell. 101 (2000) 47–56. doi:10.1016/S0092-8674(00)80622-4.

[14] J.R. García, A.Y. Clark, A.J. García, Integrin-specific hydrogels functionalized with VEGF for vascularization and bone regeneration of critical-size bone defects, J. Biomed. Mater. Res. A. 104 (2016) 889–900. doi:10.1002/jbm.a.35626.

[15] P. Martin, Wound Healing--Aiming for Perfect Skin Regeneration, Science. 276 (1997) 75–81. doi:10.1126/science.276.5309.75.

[16] J. Folkman, Tumor Angiogenesis: A Possible Control Point in Tumor Growth, Ann. Intern. Med. 82 (1975) 96.

[17] M.N. Nakatsu, C.C.W. Hughes, Chapter 4 An Optimized Three-Dimensional In Vitro Model for the Analysis of Angiogenesis, in: A.C. David (Ed.), Methods Enzymol., Academic Press, 2008: pp. 65–82. http://www.sciencedirect.com/science/article/pii/S0076687908020041.

[18] E.M. Conway, D. Collen, P. Carmeliet, Molecular mechanisms of blood vessel growth, Cardiovasc. Res. 49 (2001) 507–521. doi:10.1016/s0008-6363(00)00281-9.

[19] G. Neufeld, T. Cohen, S. Gengrinovitch, Z. Poltorak, Vascular endothelial growth factor (VEGF) and its receptors, FASEB J. 13 (1999) 9–22.

[20] R. Montesano, J.D. Vassalli, A. Baird, R. Guillemin, L. Orci, Basic fibroblast growth factor induces angiogenesis in vitro, Proc. Natl. Acad. Sci. 83 (1986) 7297–7301.

[21] A. Hayek, F.L. Culler, G.M. Beattie, A.D. Lopez, P. Cuevas, A. Baird, An in vivo model for study of the angiogenic effects of basic fibroblast growth factor, Biochem. Biophys. Res. Commun. 147 (1987) 876–880. doi:10.1016/0006-291X(87)91011-4.

[22] G.E. Davis, D.R. Senger, Endothelial Extracellular Matrix: Biosynthesis, Remodeling, and Functions During Vascular Morphogenesis and Neovessel Stabilization, Circ. Res. 97 (2005) 1093–1107. doi:10.1161/01.RES.0000191547.64391.e3.

[23] N. Yamamura, R. Sudo, M. Ikeda, K. Tanishita, Effects of the mechanical properties of collagen gel on the in vitro formation of microvessel networks by endothelial cells, Tissue Eng. 13 (2007) 1443–1453. doi:10.1089/ten.2006.0333.

[24] A.K. Jha, K.M. Tharp, J. Ye, J.L. Santiago-Ortiz, W.M. Jackson, A. Stahl, D.V. Schaffer, Y. Yeghiazarians, K.E. Healy, Enhanced survival and engraftment of transplanted stem cells using growth factor sequestering hydrogels, Biomaterials. 47 (2015) 1–12. doi:10.1016/j.biomaterials.2014.12.043.

[25] H.J. Lee, J.S. Lee, T. Chansakul, C. Yu, J.H. Elisseeff, S.M. Yu, Collagen mimetic peptide-conjugated photopolymerizable PEG hydrogel, Biomaterials. 27 (2006) 5268–76. doi:10.1016/j.biomaterials.2006.06.001.

[26] M.P. Lutolf, J.L. Lauer-Fields, H.G. Schmoekel, A.T. Metters, F.E. Weber, G.B. Fields, J.A. Hubbell, Synthetic matrix metalloproteinase-sensitive hydrogels for the conduction of tissue regeneration: Engineering cell-invasion characteristics, Proc. Natl. Acad. Sci. 100 (2003) 5413–5418. doi:10.1073/pnas.0737381100.

[27] C.M. Rubert Pérez, A. Panitch, J. Chmielewski, A Collagen Peptide-Based Physical Hydrogel for Cell Encapsulation, Macromol. Biosci. 11 (2011) 1426–1431. doi:10.1002/mabi.201100230.

[28] E.A. Phelps, N.O. Enemchukwu, V.F. Fiore, J.C. Sy, N. Murthy, T.A. Sulchek, T.H. Barker, A.J. García, Maleimide Cross-Linked Bioactive PEG Hydrogel Exhibits Improved Reaction Kinetics

and Cross-Linking for Cell Encapsulation and In Situ Delivery, Adv. Mater. 24 (2012) 64–70. doi:10.1002/adma.201103574.

[29] B. Dhandayuthapani, Y. Yoshida, T. Maekawa, D.S. Kumar, Polymeric Scaffolds in Tissue Engineering Application: A Review, Int. J. Polym. Sci. 2011 (2011). doi:10.1155/2011/290602.

[30] U.K. Saarialho-Kere, S.O. Kovacs, A.P. Pentland, J.E. Olerud, H.G. Welgus, W.C. Parks, Cell-matrix interactions modulate interstitial collagenase expression by human keratinocytes actively involved in wound healing, J. Clin. Invest. 92 (1993) 2858–2866. doi:10.1172/jci116906.

[31] S.L. Hansen, D.W. Voigt, P. Wiebelhaus, C.N. Paul, Using skin replacement products to treat burns and wounds, Adv Skin Wound Care. 14 (2001) 37-44–6.

[32] Z. Ruszczak, Effect of collagen matrices on dermal wound healing, Adv. Drug Deliv. Rev. 55 (2003) 1595–1611. doi:10.1016/j.addr.2003.08.003.

[33] D. Olsen, C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perälä, E.-R. Hämäläinen, M. Jarvinen, J. Polarek, Recombinant collagen and gelatin for drug delivery, Adv. Drug Deliv. Rev. 55 (2003) 1547–1567. doi:10.1016/j.addr.2003.08.008.

[34] C.A. Farrugia, I.V. Farrugia, M.J. Groves, Comparison of the Molecular Weight Distribution of Gelatin Fractions by Size-exclusion Chromatography and Light Scattering, Pharm. Pharmacol. Commun. 4 (1998) 559–562. doi:10.1111/j.2042-7158.1998.tb00675.x.

[35] R.J. Pelham, Y. Wang, Cell locomotion and focal adhesions are regulated by substrate flexibility, Proc. Natl. Acad. Sci. 94 (1997) 13661–13665.

[36] C.R. Lee, A.J. Grodzinsky, M. Spector, The effects of cross-linking of collagen-glycosaminoglycan scaffolds on compressive stiffness, chondrocyte-mediated contraction, proliferation and biosynthesis, Biomaterials. 22 (2001) 3145–54.

[37] L.H.H. Olde Damink, P.J. Dijkstra, M.J.A. Luyn, P.B. Wachem, P. Nieuwenhuis, J. Feijen, Glutaraldehyde as a crosslinking agent for collagen-based biomaterials, J. Mater. Sci. Mater. Med. 6 (1995) 460–472. doi:10.1007/bf00123371.

[38] C.N. Grover, J.H. Gwynne, N. Pugh, S. Hamaia, R.W. Farndale, S.M. Best, R.E. Cameron, Crosslinking and composition influence the surface properties, mechanical stiffness and cell reactivity of collagen-based films, Acta Biomater. 8 (2012) 3080–90. doi:10.1016/j.actbio.2012.05.006.

[39] Y.H. Shen, M.S. Shoichet, M. Radisic, Vascular endothelial growth factor immobilized in collagen scaffold promotes penetration and proliferation of endothelial cells, Acta Biomater. 4 (2008) 477–489. doi:10.1016/j.actbio.2007.12.011.

[40] S.W.P. Chan, S.-P. Hung, S.K. Raman, G.W. Hatfield, R.H. Lathrop, N.A. Da Silva, S.-W. Wang, Recombinant Human Collagen and Biomimetic Variants Using a De Novo Gene Optimized for Modular Assembly, Biomacromolecules. 11 (2010) 1460–1469. doi:10.1021/bm100052y.

[41] L.S.Z. Larsen, C.D. Wassman, G.W. Hatfield, R.H. Lathrop, Computationally Optimised DNA Assembly of synthetic genes, Int. J. Bioinforma. Res. Appl. 4 (2008) 324–336. doi:10.1504/ijbra.2008.019578.

[42] W.P.C. Stemmer, A. Crameri, K.D. Ha, T.M. Brennan, H.L. Heyneker, Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides, Gene. 164 (1995) 49–53. doi:10.1016/0378-1119(95)00511-4.

[43] A. Jabaiah, X. Wang, S.K. Raman, R. Ragan, N.A. Da Silva, S.-W. Wang, Nanoscale architecture and cellular adhesion of biomimetic collagen substrates, J. Biomater. Appl. 28 (2014) 1354–1365. doi:10.1177/0885328213508328.

[44] R.A. Que, S.W.P. Chan, A.M. Jabaiah, R.H. Lathrop, N.A. Da Silva, S.-W. Wang, Tuning cellular response by modular design of bioactive domains in collagen, Biomaterials. 53 (2015) 309–317. doi:10.1016/j.biomaterials.2015.02.074.

[45] D.E. Discher, D.J. Mooney, P.W. Zandstra, Growth Factors, Matrices, and Forces Combine and Control Stem Cells, Science. 324 (2009) 1673–1677. doi:10.1126/science.1171643.

[46] A.J. Engler, S. Sen, H.L. Sweeney, D.E. Discher, Matrix Elasticity Directs Stem Cell Lineage Specification, Cell. 126 (2006) 677–689. doi:10.1016/j.cell.2006.06.044.

[47] J.E. Saik, D.J. Gould, E.M. Watkins, M.E. Dickinson, J.L. West, Covalently immobilized platelet-derived growth factor-BB promotes angiogenesis in biomimetic poly(ethylene glycol) hydrogels, Acta Biomater. 7 (2011) 133–143. doi:10.1016/j.actbio.2010.08.018.

[48] X. Jia, K.L. Kiick, Hybrid Multicomponent Hydrogels for Tissue Engineering, Macromol. Biosci. 9 (2009) 140–156. doi:10.1002/mabi.200800284.

Chapter 2: Expanding Functionality of Recombinant Human Collagen Through Engineered Non-Native Cysteines

Adapted with permission from R. Que, A. Mohraz, N.A. Da Silva, S.-W. Wang, Expanding Functionality of Recombinant Human Collagen Through Engineered Non-Native Cysteines, Biomacromolecules. 15 (2014) 3540–3549. doi:10.1021/bm500735d.

2.1 Introduction

The environment that a cell inhabits plays a major role in determining cell fate. Mechanical properties, degradation characteristics, cell-interaction sites, and signaling factor densities of the extracellular matrix (ECM) environment have been shown to modulate proliferation, migration, and differentiation of the residing cells [1–4]. Therefore, major components of the ECM, such as the protein collagen are natural departure points for engineering cell substrates and microenvironments in tissue engineering and regenerative medicine. Ideally, biomimetic materials would have the additional advantage over native ECM materials of being amenable to independent control over properties such as protein concentrations, mechanical properties, and densities of cell-interaction sites [3]. Furthermore, engineered materials would not suffer from the typical concerns for materials from animal sources (e.g., batch-to-batch variability, immunogenicity, pathogen transmission) [5–7]. Examples of peptides or polymers used as artificial matrices have included hyaluronic acid, fibrinogen, elastin, and the entirely synthetic polymer polyethylene glycol [8–12], but none of the peptides are as prevalent in native ECM as the protein collagen.

However, despite collagen being the major component of the ECM, and despite the corresponding high interest to engineer it, efforts to produce collagen synthetically as a matrix backbone have been complicated by challenging issues[13,14]. Collagen's glycine-X-Y tripeptide repeating sequence, together with the unique sequences embedded within the tripeptide repeats required for cell interaction sites, and the large overall number of amino acids, result in difficulties in generating an encoding synthetic gene due to a high propensity for oligonucleotide mishybridizations [15]. Furthermore, prolines in the Y position must be post-translationally hydroxylated for sufficient thermostability of the collagen triple helix [16]. These difficulties have limited most collagen-mimetic materials to short peptides [13,17–21] or repeating collagen-like domains [22].

Our research group has developed a bottom-up strategy to produce full-length, hydroxylated recombinant collagen III with the native human sequence (rCol) or with alternatively-defined sequences. Collagen III, one of the fibular collagens, is located in elastic tissues such as the skin and vasculature as well as other tissues alongside collagen I [23,24]. It is a homotrimer, expressed from a single gene, reducing the number of genes requiring modification to introduce non-native elements into the resulting protein. To overcome the mishybridization problem between oligonucleotides during gene synthesis, the DNA sequence was optimized utilizing an algorithm that considered the degeneracy of the genetic code and oligonucleotide hybridization melting temperatures [15]. Gene synthesis was then performed in modules via PCR assembly. The protein is produced in *Saccharomyces cerevisiae* that has been genetically engineered with the ability to hydroxylate prolines in the correct positions within collagen [25]. This strategy provides complete control of the amino acid sequence and allows for the tailoring of sequence, location, and frequency of functional sites within the protein. Our previous work has shown that recombinant human collagen III produced by this platform yields stable triple-helical collagen with correct structure, and this collagen interacts favorably with mammalian cells [25].

In this study, we examined the feasibility of using our platform to fabricate custom-designed collagen variants. We introduced non-native elements into the human collagen III scaffold, and then probed the structural effects and the functionality of these modified sites. Varying numbers of cysteines (Cys), which are not present in the triple-helical regions of any fibrillar collagen, were inserted at specific sites to be used as crosslinking or attachment sites for bioactive proteins. Although lysines are often used to chemically functionalize collagen from native sources, there are 38 lysines present in the triple-helical region of collagen III [26], resulting in difficulty in controlling the location and quantity of any lysine-based reactions. In contrast, the exact numbers and locations of Cys sites can be precisely controlled using our platform.

We then examined whether the specifically-designed collagens are functional for crosslinking and attachment of bioactive molecules. The sulfhydryl group on Cys is available for reacting with maleimide to form a thioether linkage. One advantage of this approach is to avoid the cytotoxicity of crosslinking lysine amines using either glutaraldehyde or N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (EDC) [27,28]. Furthermore, the use of lysines as crosslinking or attachment sites may alter ligand availability as they are contained in binding sequences [29]. These collagen cysteine variants allow for the formation of crosslinked collagen hydrogels with the potential of decoupling stiffness from collagen concentration through the control of cysteine density, and thus crosslinking density, per collagen. To examine the feasibility of modulating cell fate through differentiation of fibroblasts to myofibroblasts [30–32], we immobilized transforming growth factor beta 1 (TGF-β) onto the non-native Cys residues of recombinant collagen variants.

The results of our study demonstrate the ability to produce recombinant collagen with tunable physical and biological properties. Triple helical collagens with precise numbers of cysteines were produced using novel modular genes. The collagens successfully formed hydrogels via crosslinking, and promoted the differentiation of fibroblasts to myofibrolasts through covalently tethered growth factors.

2.2 Materials and Methods

2.2.1 Cells and reagents

Escherichia coli DH5α was purchased from Strategene (La Jolla, CA). Restriction endonucleases, DNAse I, and RNase I_f were acquired from New England Biolabs (Ipswitch, MA). Phenylmethanesulfonyl fluoride (PMSF), tris(2-carboxyethyl)phosphine (TCEP), bovine serum albumin (BSA), galactose, 40 kDa MWCO desalting columns, methyl-PEG24-NHS [ms(PEG)24], and succinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate (SMCC) were obtained from

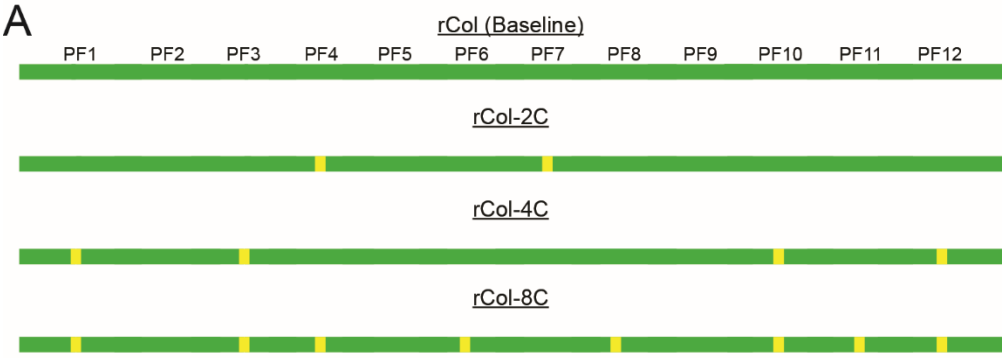
Thermo Fisher (Waltham, MA). Pepsin was purchased from MP Biomedicals (Santa Ana, CA). TGF-β1 was obtained from R&D Systems (Minneapolis, MN). The crosslinking reagent 20 kDa 4-arm PEG maleimide was purchased from NANOCS (Boston, MA). Acrylamide solution was acquired from Bio-Rad (Hercules, CA). Native collagen proteins used as controls were obtained from Millipore (Billerica, MA). Dulbecco's modified Eagle's medium (DMEM), CelLytic M, and monocolonal antibodies against α-smooth muscle actin were purchased from Sigma (St. Louis, MO). Fetal bovine serum (FBS), penicillin, streptomycin, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT), Alexa Fluor 532 C5 maleimide (AF532), and secondary antibody alkaline phosphatase conjugated anti-mouse IgG antibodies were obtained from Life Technologies (Carlsbad, CA). Casamino acids, yeast nitrogen base, and tissue culture plates were purchased from BD Biosciences (Franklin Lakes, NJ). NIH/3T3 mouse fibroblast cells and HT-1080 human fibrosarcoma cells were obtained from ATCC (Manassas, VA). Cells were maintained in DMEM supplemented with 10% FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin at 37 °C and 5% CO_2.

2.2.2 Fabrication of genes encoding collagen variants

Full length collagen-mimetic proteins were fabricated using the modular gene assembly strategy described in Chan et al. to produce recombinant human collagen [15]. The triple-helical region of collagen, for which the gene has been conventionally difficult to assemble, was divided into 12 modules to facilitate synthesis and sequence design. By utilizing the degeneracy of the genetic code, oligonucleotide sequences were optimized to yield higher melting temperatures for correct hybridizations than for incorrect hybridizations [33]. These oligonucleotides were assembled into 12 primary fragments by polymerase chain reaction (PCR). These primary fragments were then PCR-assembled into secondary fragments, and the secondary fragments were PCR-assembled into the full length gene. We showed that this process enables the fabrication of genes encoding the amino acid sequence of human collagen III (which is defined here as "baseline" collagen, rCol) and variants of collagen with desired modulated sequences [15].

To produce genes coding for collagen containing varying numbers of non-native cysteines (Cys), site-directed mutagenesis (SDM) was performed on eight of the twelve primary fragments of the baseline collagen gene to build a library of primary fragments containing one Cys per fragment (see Figure 2-1). Complementary mutagenic primers used in SDM were designed to change an amino acid of similar size to cysteine in the X or Y position of the Gly-X-Y tripeptide near the center of each primary fragment (Table 2-1). This assisted the assembly of primary fragments with and without these non-native cysteines. The mutagenesis procedure was based on protocols described previously [15].

A



B

Variant	Amino Acid Location Number of Cysteines
rCol	none
rCol-2C	285 (PF4), 731 (PF7)
rCol-4C	42 (PF1), 224 (PF3), 803 (PF10), 984 (PF12)
rCol-8C	42 (PF1), 224 (PF3), 285 (PF4), 465 (PF6), 748(PF8), 803 (PF10), 888 (PF11), 984 (PF12)

Figure 2-1: Summary of recombinant collagen variants containing non-native cysteines (Cys) in the triple helical domain. (A) Variants were produced containing 2 to 8 Cys (yellow) spread throughout the triple helical domain. Green regions denote native amino acid sequences of human collagen III. "PF" indicates the primary fragment (one of 12 modules used in gene synthesis). (B) Collagen variants with the respective amino acid site number of the introduced cysteines.

Table 2-1: List of oligonucleotides used to introduce non-native cysteines for each primary fragment.

Primary fragment number	Oligonucleotides sequences (5' → 3'; forward and reverse strands)
PF1	CACCTGGTGAGCCCGGACAATGTGGTCCAAGTGGTCCTCCTGGTC GACCAGGAGGACCACTTGGACCACATTGTCCGGGCTCACCAGGTG
PF2	CCGTGGTAATGATGGTGCTAGAGGATGTGACGGACAACCTGGCCCACCTGGAC GTCCAGGTGGGCCAGGTTGTCCGTCACATCCTCTAGCACCATCATTACCACGG
PF3	CGGTCCACCCGGAATTAACGGTTGTCCAGGAGGTAAAGGCGAAATGGGTCCT AGGACCCATTTCGCCTTTACCTCCTGGACAACCGTTAATTCCGGGTGGACCG
PF4	GACCAAGGGGAGAACGTGGAGAATGTGGCATACCCGGAGTACCCGGTGCTAAAG CTTTAGCACCGGGTACTCCGGGTATGCCACATTCTCCACGTTCTCCCCTTGGTC
PF5	GACCAAGGGGAGAACGTGGAGAATGTGGCATACCCGGAGTACCCGGTGCTAAAG CTTTAGCACCGGGTACTCCGGGTATGCCACATTCTCCACGTTCTCCCCTTGGTC
PF6	CAGGTGGAGACAAAGGTGATTGTGGACCACCTGGACCCCAAGGATTGC GCAATCCTTGGGGTCCAGGTGGTCCACAATCACCTTTGTCTCCACCTG
PF7	CAAAAGGTGATAAGGGTGAACCTGGTTGTCCTGGCGCAGACGGTGTTCCCGG CCGGGAACACCGTCTGCGCCAGGACAACCAGGTTCACCCTTATCACCTTTTG
PF8	CCAGGAATTGCTGGCCCAAGGGGTTGTCCAGGTGAAAGGGGAGAAACAGG CCTGTTTCTCCCCTTTCACCTGGACAACCCCTTGGGCCAGCAATTCCTGG
PF9	CAAAAGGTGATAAGGGTGAACCTGGTTGTCCTGGCGCAGACGGTGTTCCCGG CAAAAGGTGATAAGGGTGAACCTGGTTGTCCTGGCGCAGACGGTGTTCCCGG
PF10	CAGGCATGCCAGGTCCTAGGGGCTGTCCAGGTCCACAAGGCGTTAAAG CTTTAACGCCTTGTGGACCTGGACAGCCCCTAGGACCTGGCATGCCTG
PF11	GGTCCTTGTGGACCAGCCGGCAAATCTGGTGATAGAGGAGAGTCAGGTC GACCTGACTCTCCTCTATCACCAGATTTGCCGGCTGGTCCACAAGGACC
PF12	GACCAAGCGGACCACCTGGTAAATGTGGAACTTCTGGTCATCCAGGTCC GGACCTGGATGACCAGAAGTTCCACATTTACCAGGTGGTCCGCTTGGTC

2.2.2.1 Primary fragments

For each baseline primary fragment in a final volume of 50 µL, 50 ng of plasmid DNA containing the primary fragment was mixed with 300 nM of each mutagenic primer, 200 µM dNTPs, 2.5U PfuUltraII and 1x PfuUltraII reaction buffer. The mixture was cycled in a thermocycler with the following program: 10 minutes at 95 °C, followed by 16 cycles of 20 seconds at 95 °C, 30 seconds at 55 °C, 1 minute at 72 °C. The sample was incubated with 20 U of DpnI for 2 hours. The product band of expected size was extracted from an agarose gel, ligated into the pCR-BluntII TOPO vector (Invitrogen), transformed into *Escherichia coli* DH5α (Stratagene), and plated on LB-Kan. DNA from colonies were sequenced to screen for mutations.

2.2.2.2 Secondary fragments

Primary fragments with and without cysteines were mixed-and-matched to assemble secondary fragments with cysteines in the desired locations by mixing 2 nM of each primary fragment

with 200 nM of each secondary fragment primer, 200 μM dNTPs, 1.5U PfuUltraII, and 1x PfuUltraII reaction buffer in a final volume of 50 μL. The mixture was cycled in a thermocycler with the following program: 10 minutes at 95 °C, followed by 16 cycles of 20 seconds at 95 °C, 30 seconds at 68 °C, 45 seconds at 72 °C, and a final step of 5 minutes at 72 °C. The sample was run on an agarose gel and the product band of expected size was extracted and ligated into the pCR-BluntII TOPO vector (Invitrogen), transformed into *Escherichia coli* DH5α, and plated on LB-Kan. DNA from colonies were sequenced to screen for the correct sequence.

2.2.2.3 Full length collagen genes

Genes encoding collagen variants were assembled from the secondary fragments by mixing 2 nM of each secondary fragment with 200 nM of each cloning primer, 200 μM dNTPs, 1 mM $MgSO_4$, 1 U KOD Hot Start DNA Polymerase (EMD Millipore) and 1x KOD reaction buffer in a final volume of 50 μL. The mixture was cycled in a thermocycler with the following program: 10 minutes at 95 °C, followed by 16 cycles of 20 seconds at 95 °C, 30 seconds at 68 °C, 1.25 minutes at 72 °C, and a final step of 5 minutes at 72 °C. The product band was extracted from an agarose gel, ligated into the pCR-BluntII TOPO vector, transformed into *E. coli* DH5α, and plated on LB-Kan. Colonies were sequenced to screen for the correct sequence. Full-length collagen variant genes were then transferred to the CEN/ARS plasmid as previously described [34] to generate plasmids YCpMCOL-2C, YCpMCOL-4C, and YCpMCOL-8C, which contain 2, 4, and 8 non-native cysteines, respectively. Figure 2-1 summarizes the collagen variants and their corresponding Cys mutations that are encoded by these genes.

2.2.3 Protein expression and purification

Full length recombinant biomimetic collagens were expressed in *S. cerevisiae* BYα2β2, which was engineered to contain two integrated copies each of the human prolyl-4-hydroxylase α-subunit and β-subunit genes to achieve hydroxylation of prolines in the collagen biopolymers [25]. BYα2β2 was transformed with plasmids YCpMCOL-2C, YCpMCOL-4C, and YCpMCOL-8C to express collagen

containing 2 Cys [named rCol-2C(1)], 4 Cys [rCol-4C(1)], and 8 Cys [rCol-8C(1)], respectively. In this manuscript, these mutants will be abbreviated as rCol-2C, rCol-4C, and rCol-8C.

To express the recombinant collagen, the yeast were cultured in selective media as previously described [25], resuspended to 0.1 g/L in ice-cold buffer containing 0.1 M Tris-HCl, 0.4 M NaCl, 2 U/ml DNase, 1 U/ml RNase, and 1 mM PMSF (pH 7.5), and mechanically lysed using a French press cell disruptor (Thermo Fisher). EDTA and TCEP were added to the solutions to 1 mM each, and the solutions were incubated for 1 hour at 4 °C. The pH was lowered to 2, and the samples were digested with pepsin at a concentration of 0.2 mg/ml for 12 hours at 4 °C, conditions which degrade endogenous yeast proteins while leaving hydroxylated triple-helical collagen intact. The digested lysate was cleared by centrifugation. Protein was precipitated from the cleared supernatant by adding acetic acid (AcOH) and NaCl to 0.5 M and 3M, respectively, and centrifuged. The precipitate was dissolved in 0.1 N HCl, the solution was raised to pH 7.4 with 200 mM Tris-HCl (pH 8.6), and TCEP was added to 1 mM. After incubation at 4 °C for 1 hour, the collagen was precipitated by adding NaCl to 3 M. The pellet was dissolved in 0.1 N HCl and dialyzed against 0.05 M AcOH. Protein concentration was determined by BCA assay (Pierce, Rockford, IL) using bovine collagen type III as a standard. SDS-PAGE confirmed size and purity of proteins.

2.2.4 Characterization of recombinant collagen variants

2.2.4.1 Circular dichroism.

To determine the structure and melting temperatures of the collagen variants, circular dichroism (CD) spectroscopy was performed using a Jasco J-810 spectropolarimeter [35]. The purified recombinant collagen samples (100 - 300 µg/ml) in 50 mM AcOH were placed in a 1 mm quartz cell and scanned at 10 nm/minute. To determine the apparent melting temperature (T_m), the temperature was increased at a rate of 1 °C/minute while measuring ellipticity at 221 nm. The molar ellipticity vs. temperature was fit to the Gibbs-Helmholtz equation [36]. Apparent T_m values are reported as mean ± standard deviation and are replicates of at least 3 independent measurements.

2.2.4.2 Cellular adhesion.

To assess whether the introduced cysteines affected cell adhesion, an adhesion assay was performed as previously described with the recombinant collagens and controls (human collagen III; BSA) [25]. Briefly, wells in non-tissue culture treated 96-well plates were coated with 40 µL of 20 µg/mL protein in phosphate buffered saline (PBS, pH 7.4) overnight at 4 °C, and then blocked with 1% BSA in DMEM for 1 hour at room temperature (RT, approximately 22 °C). To verify that comparable amounts of protein were deposited between different samples, an alkaline phosphatase assay was performed as previously described [25]. HT-1080 cells were seeded at 1.5 x 10^5 cells/cm^2 in FBS-free DMEM supplemented with 0.1% BSA and incubated at 30 °C for 4 hours. Cells were washed with DMEM, incubated with DMEM + 10% FBS for 1 hour at 37 °C, and imaged. To quantify the relative number of cells adhered, an MTT assay was used; we incubated cells with 12 mM MTT in DMEM at 37 °C for 2 hours, and lysed cells by incubating with 70 µL of 20% w/v sodium dodecyl sulfate, 2 mM HCl, 400 mM AcOH in dimethylformamide/water (50:50 v/v) overnight at 37 °C. Absorbance at 570 nm was measured using a SpectraMax M2 plate reader (Molecular Devices, Sunnyvale, CA).

2.2.5 Functionalization of collagen through non-native cysteines (Cys)

2.2.5.1 Conjugation of non-native cysteines with fluorescent markers

To determine the degree of accessibility and chemical functionality of the non-native Cys, the thiol side groups of the Cys were labeled with AF532. Recombinant collagen (1 mg/ml) was incubated with TCEP at a ratio of 2.5 TCEP to 1 cysteine in 50 mM phosphate buffer (pH 7) for 30 minutes at RT, and was then reacted with AF532 for 2 hours at RT. As a control, native human collagen was treated with TCEP and AF532 under the same conditions. Unreacted AF532 was removed with a 40 kD desalting column and absorbance of the samples was measured at 532 nm. The amount of AF532 within each sample was determined by a standard curve for free AF532. The non-specific binding of AF532 to the native collagen control was subtracted as the background.

2.2.5.2 Crosslinking of non-native cysteines to generate collagen hydrogels

The buffer of the different variants of purified collagen proteins (1 mg/ml) was exchanged into 50 mM HEPES (pH 5). TCEP was then added at a ratio of 2.5 TCEP to 1 cysteine and incubated at RT for 30 minutes to reduce any existing disulfide bonds. A 20 kDa, four-arm PEG crosslinker, derivatized with maleimide, was added and mixed. Gel formation was characterized by rheological measurements as described below.

2.2.5.3 Conjugation of non-native cysteines with growth factor

To examine the effect on cell growth of an immobilized growth factor, TGF-β was attached to rCol-4C. TGF-β was activated with maleimide by reacting it with SMCC (80 SMCC to 1 TGF-β) at RT for 30 minutes, and excess SMCC was removed using a desalting column. Non-tissue culture treated 6-well plates were incubated overnight at 4 °C with 20 μg/mL rCol-4C in PBS. Maleimide-activated TGF-β (50 ng in 1 ml PBS) was reacted with the surface-adsorbed rCol-4C for 2 hours at 4 °C. Wells were washed with PBS and incubated with DMEM + 1% BSA for 1 hour at RT. These surfaces were then used for subsequent cellular assays.

2.2.6 Characterization of crosslinked hydrogels

2.2.6.1 Passive microrheology

To monitor the formation of a hydrogel from rCol-4C, we used passive multiple particle tracking microrheology [37]. Amine-functionalized polystyrene fluorescent tracer particles of 1 μm diameter were PEGylated with ms$(PEG)_{24}$ to reduce microparticle aggregation. Samples containing collagen protein mixed with crosslinker (at a 1 maleimide to 1 Cys ratio) were prepared as described above at a volume of approximately 10 μL. Fluorescent tracer particles were immediately mixed into the sample after the addition of crosslinker, pipetted between a microscope slide and coverslip separated by 70 μm, and sealed (Norland optical adhesive 81, Norland Products, Cranbury, NJ). Passive microrheology was performed using an Axio Observer inverted microscope (Carl Zeiss

Microimaging, Inc., Jena, Germany) attached to a Vt-Eye confocal scanner (VisiTech International, Sunderland, UK). Images were taken at 16 fps for 5 minutes every hour with a 20x (numerical aperture = 0.4) objective and 2x digital zoom. Samples were monitored over time at RT and the mean square displacement (MSD) of particles was measured by tracking particle centers of mass [38]. Following the method initially outlined by Larsen and Furst, we scaled each individual MSD curve to construct pre-gel and post-gel master curves, and identified the gel point as the transition time between the two curves [39]. To determine the upper limit of elastic modulus measureable with our passive microrheology setup, a control sample of 20% polyacrylamide was prepared from 30% acrylamide/bis-acrylamide (29:1) in a final concentration of 375 mM Tris (pH 8.8), 0.1% w/v APS, 0.1% v/v TEMED.

2.2.6.2 Bulk rheology

To confirm the formation of a crosslinked hydrogel using a secondary method and to determine bulk viscoelastic properties of the crosslinked samples, bulk rheology was performed on a stress-controlled rheometer (Anton Paar MCR 301, Graz, Austria) equipped with a 25 mm parallel plate geometry and Peltier temperature control system. Samples containing 2 mg/ml collagen mixed with crosslinker (at a 2 maleimide to 1 Cys ratio) were prepared as described above, substituting PBS for HEPES, at a volume of approximately 150 µL and immediately loaded onto the rheometer. A solvent trap filled with ddH_2O was utilized to provide a humidified environment and prevent evaporation from the gel. Oscillatory measurements were performed at a strain of γ = 0.01 and frequency of f = 1 Hz, with a gap height of 200 µm at 22.5 °C. The storage (G') and loss (G'') moduli were recorded over time until G' showed no change over several hours. A frequency sweep was subsequently performed from f = 0.01 to 100 Hz, with all other settings remaining the same.

2.2.7 Cell response assay to growth factors immobilized on engineered collagen substrates

We explored the ability to modulate cellular activity through covalently immobilizing growth factors to recombinant collagen variants. Fibroblast differentiation into myofibroblasts was

evaluated by detection of α smooth muscle actin (αSMA) via Western blot [40]. Collagen surfaces (rCol, rCol-4C, and native human collagen) incubated with maleimide-functionalized TGF-β in 6-well plates were prepared as described in section 2.5. NIH/3T3 cells were seeded at a density of 6.3 x 10^4 cells per cm^2 in DMEM + 10% FBS and incubated at 30 °C and 5% CO_2 for 6 hours to allow cells to adhere. Wells were washed with PBS, and media was replaced with DMEM + 0.1% FBS. For the controls, collagens that were not mixed with maleimide-functionalized TGF-β were adsorbed to the surfaces, and soluble TGF-β (not maleimide-functionalized) was added to the cell culture media to a final solution concentration of 5 ng/mL (positive controls) and 0 ng/mL (negative controls), respectively. Cells were incubated at 30 °C, 5% CO_2 for 48 hours then lysed with CelLytic M. Cell lysate was analyzed by Western blot using anti-αSMA primary antibodies raised in mouse and alkaline phosphatase-conjugated rabbit anti-mouse secondary antibodies. Intensities of bands were quantified using Image J software and normalized to the respective positive control (the collagen variant mixed with 5 ng/ml soluble TGF- β) within each blot. Intensity results are reported as mean ± standard deviation and are replicates of 4 independent experiments.

2.2.8 Statistical analysis

For cell adhesion, results are reported as mean ± standard deviation of at least 3 independent experiments. Statistical significance was determined using the software program R[41]. We used one-way analysis of variance (ANOVA) followed by the Bonferroni post-test to evaluate pairwise comparisons.

2.3 Results

2.3.1 Fabrication and expression of cysteine-containing collagen variants

Non-native Cys codons were successfully introduced into primary gene fragments through site-directed mutagenesis, and these gene fragment modules were utilized to assemble the genes encoding full-length collagen variants (Figure 2-1). Using these genes, the collagen-based proteins

rCol-2C, rCol-4C, and rCol-8C, with 2, 4, and 8 Cys per collagen, respectively (corresponding to 6, 12, and 24 Cys per triple-helical polymeric strand), were successfully expressed and purified from engineered yeast strain BYα2β2. Approximately 0.5 mg purified protein was recovered per L of culture for rCol-2C and rCol-4C and 0.05 mg purified protein was recovered per L of culture for rCol-8C. SDS-PAGE analysis showed that the purified rCol-2C and rCol-4C are the expected size, which is consistent with the rCol control (baseline; recombinant human collagen III) and the hCol control (native human collagen III) (Figure 2-2). These proteins are also as pure and pepsin-resistant as the rCol baseline human collagen. Although rCol-8C was expressed in yeast, purification yields were inconsistent and suggested disulfide crosslinking of the Cys thiols. Therefore, subsequent investigations focused primarily on rCol-2C and rCol-4C.

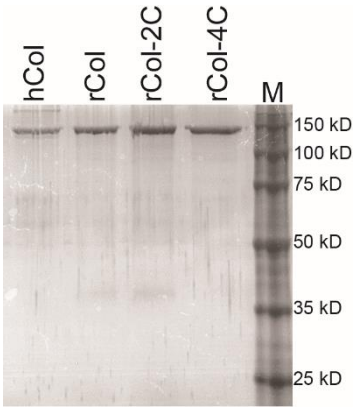


Figure 2-2: SDS-PAGE of purified collagens and cysteine variants of collagen. Lanes: (1) native human collagen III (hCol); (2) recombinant human collagen III (rCol); (3) recombinant collagen with 2 Cys (rCol-2C); (4) recombinant collagen with 4 Cys (rCol-4C); (5) molecular weight standard (M). The purified recombinant collagens are comparable in size and purity to native human collagen III.

2.3.2 Confirmation of structure, thermostability, and cell adhesion capability

The structure of the recombinant collagen variants was examined by CD. Variants rCol-2C and rCol-4C were both confirmed to be triple helical, with the characteristic CD spectra exhibiting a positive peak at 221 nm and negative peak at 198 nm (Figures 2-3A & 2-3B). The thermostability of the triple helix of rCol-2C and rCol-4C were interrogated by CD temperature scans, and yielded

apparent melting temperatures (T_m) of 35.9 ± 0.1 °C for rCol-2C and 37.2 ± 0.5 °C for rCol-4C (Figures 2-3A & 2-3B, insets). Although these values are lower than the apparent T_m of 39.1 ± 0.1 °C for native human collagen III, the T_m for rCol-4C is higher than that of recombinant baseline collagen (rCol, T_m = 35.8 ± 0.5 °C) [25].

We examined whether the introduction of Cys would affect cellular adhesion to the collagen variants. Human collagen III, bovine serum albumin (BSA), rCol, rCol-2C, and rCol-4C were deposited onto surfaces of 96-well plates. Surface protein densities were verified to be equivalent by an ALP-streptavidin-biotin assay. At these experimental conditions, HT-1080 cells adhered to and spread on rCol-2C and rCol-4C substrates at comparable levels as on rCol (baseline) and human collagen III (Figure 2-4A). In contrast, cells did not adhere to the BSA negative control. These observations were also supported by the quantitative results of the MTT assay (Figure 2-4B) (n=3). Cell viability of cells seeded on any of the collagen surfaces were not significantly different from one another, while cell viability of cells seeded on all of the collagen surfaces were significantly different from cells seeded on surfaces coated with BSA (p<0.001, ANOVA). Thus, cell adhesion onto recombinant collagen is not inhibited by the introduction of non-native cysteines.

Characterization of rCol-2C and rCol-4C shows that the structure, stability, and cellular adhesion to these variants are similar to those of rCol and to each other. Therefore, we selected the recombinant collagen with the greater number of changes from the baseline, rCol-4C, as the representative variant for subsequent functional characterization.

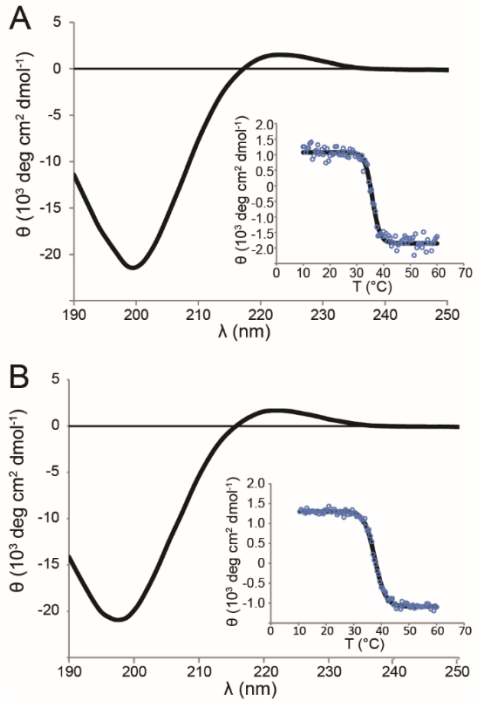


Figure 2-3: Representative circular dichroism (CD) data for Cys variants of recombinant collagen. (A) CD wavelength spectra and thermostability scan (inset) for rCol-2C. (B) CD wavelength spectra and thermostability scan (inset) for rCol-4C. Both variants display the characteristic peaks for triple helical collagen and apparent melting temperatures of 35.9 ± 0.1 °C (for rCol-2C) and 37.2 ± 0.5 °C (for rCol-4C).

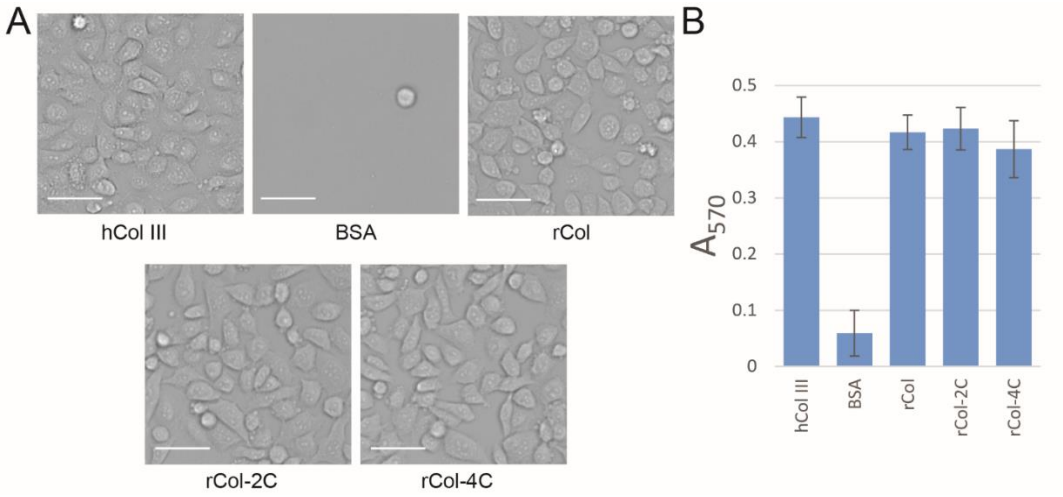


Figure 2-4: Cell adhesion assays. (A) Images of HT-1080 cells, seeded at 30 °C, on surfaces adsorbed with native human collagen III (hCol III, positive control), bovine serum albumin (BSA, negative control), baseline recombinant human collagen (rCol), rCol-2C, and rCol-4C. (B) MTT assay to evaluate the relative number of viable cells on these substrates. Cells are able to bind to recombinant collagen containing non-native cysteines.

2.3.3 Accessibility and functionality of non-native cysteines

2.3.3.1 Cysteine reactions with small molecules

To determine the accessibility and functionality of Cys within our collagen variants, we mixed rCol-4C with a dye molecule that can couple to the Cys thiols. Our results demonstrate that the Cys within rCol-4C is solvent accessible and chemically reactive. The conjugation ratio was determined to be 3.1 ± 0.2 molecules of AF532 to 1 collagen monomer for the given reaction conditions. This corresponds to approximately 77% of the cysteines being labeled.

2.3.3.2 Formation of crosslinked collagen hydrogels and evaluation by microrheology

To examine whether Cys could be crosslinked for hydrogel formation, rCol-4C was reacted with a maleimide-functionalized PEG crosslinker. Using passive microrheology, we observed that this protein and crosslinker mixture gradually progresses from a viscoelastic liquid to a viscoelastic solid. Tracer particles were tracked over time and the particle MSDs were calculated at various time points. Each MSD curve was scaled utilizing time cure superposition to form the pre-gel and post-gel master curves shown in Figure 2-5 [39]. The MSD of the tracer particles initially progressed over time from a slope of near 1 (diffusive dynamics indicative of a liquid-like viscous medium) to the gel point with a slope near 0.5 [42]. The post-gel master curve shows that samples continued to evolve as the slope of the MSD plot decreases to near 0 (localized dynamics indicative of a solid-like elastic medium) after approximately 50 h. At this point, particles became arrested at the upper limit of stiffness (evaluated here as G' or storage/elastic modulus) distinguishable by our microrheology setup; using a 20% poly-acrylamide gel as a standard, this maximum stiffness was determined to be close to 0.3 Pa. Time to gelation varied from approximately 30 to 70 hours for identical experimental conditions, demonstrating the kinetics of gelation can vary between replicates. However, the consistent decrease of the MSD slope over time to nearly zero indicates the formation of a crosslinked matrix. The viscoelastic properties of the samples were further evaluated by conventional bulk rheology (see results below). As expected, the recombinant baseline collagen control (no Cys) mixed

with crosslinker demonstrated only diffusive behavior with slope near unity in the MSD plot and did not exhibit signatures of gelation up to 15 hours; beyond that timepoint, aggregation of microparticles prevented further measurements.

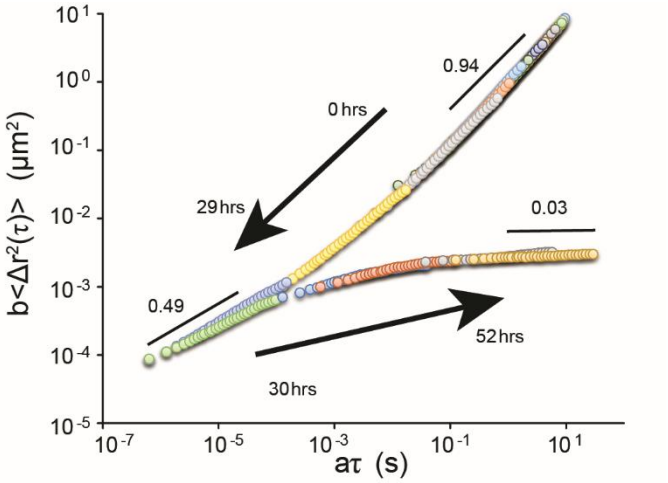


Figure 2-5: Representative pre-gel and post-gel master curves of crosslinked rCol-4C. Gelation of rCol-4C, crosslinked with maleimide-functionalized PEG, was tracked with passive multiple particle tracking microrheology. Scaled mean squared displacement (MSD) is on the vertical axis and scaled time is on the horizontal axis. Each individual MSD curve taken at different times during gelation is shifted to account for changes in MSD and the relaxation time as crosslinks percolate through the sample. Data shows this sample transitioned from a viscoelastic liquid to a viscoelastic solid over time, with gel point time at approximately 29 hours.

2.3.3.3 Formation of crosslinked collagen hydrogels and evaluation by bulk rheology.

After detecting gelation through microrheology, the viscoelastic properties of hydrogels formed from the collagen variants were independently characterized using a conventional rheometer until the storage (G') and loss (G") moduli remained unchanged over several hours (Figure 2-6A & 2-6B). Samples initially showed more viscous than elastic behavior (G" > G'), but over time became more solid-like (G' > G"). The rCol-2C and rCol-4C samples reached the gel point (defined here as the time to reach G' = G"), at approximately 2.6 ± 0.4 h and 1.1 ± 0.1 h, respectively. G' values plateaued near 5-10 Pa for rCol-4C after approximately 11.5 ± 1.4 h (Figure 2-6A). For rCol-2C, G' was 3-4 Pa after approximately 7.6 ± 0.6 h (Figure 2-6B). At this point, the storage and loss moduli remained relatively independent of frequency (Figure 2-6C & 2-6D); this is characteristic of viscoelastic solids

and gel-like materials [43,44]. Recombinant collagen without the introduced non-native cysteines (rCol) was also tested under the same crosslinking conditions as used for rCol-2C and rCol-4C. The sample remained more viscous than elastic (G" > G') over the testing period, never reaching the gel point. Collectively, these observations support the formation of a collagen gel through Cys crosslinking.

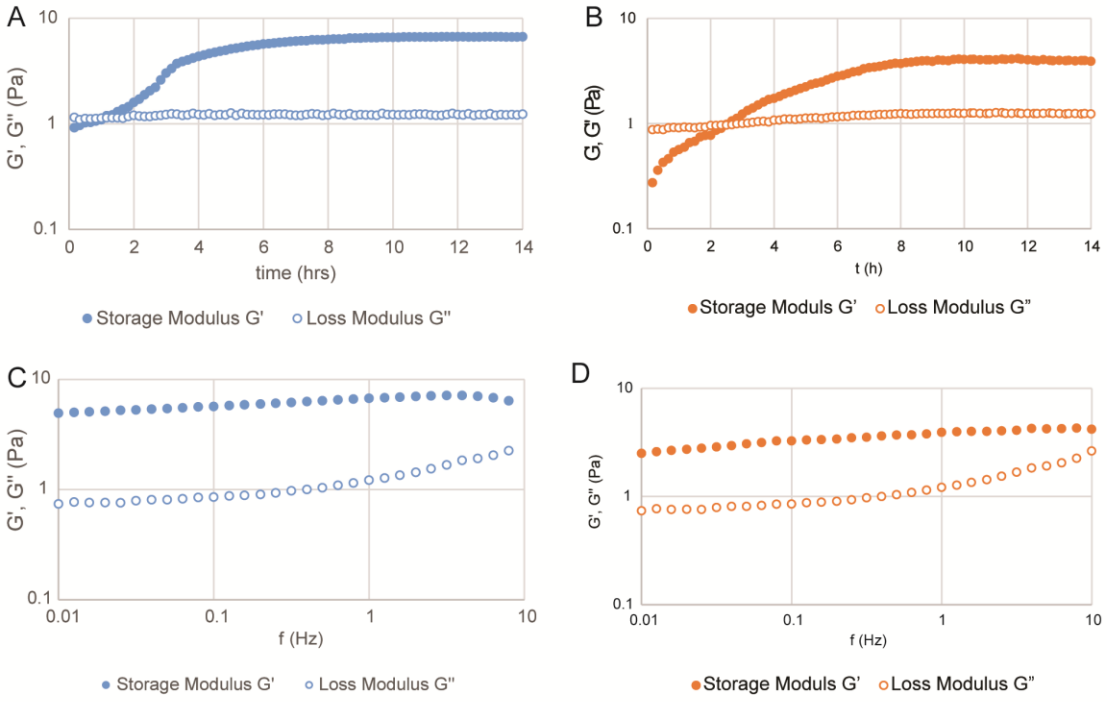


Figure 2-6: Representative oscillatory rheology data of crosslinked rCol-4C and rCol-2C. (A) Time sweep of rCol-4C. (B) Time sweep of rCol-2C. Storage modulus G' (closed symbols) and loss modulus G" (open symbols) over time shows transition from liquid-like to solid-like material. (C) Frequency sweep of rCol-4C. (D) Frequency sweep of rCol-2C. Storage modulus G' (closed symbols) is greater than the loss modulus G" (open symbols) over all frequencies measured for the final solid-like material.

2.3.4 Cellular response to biomimetic collagen conjugated with TGF-β

To investigate the feasibility of modulating cellular activity through the covalent conjugation of growth factors, collagen-coated surfaces were reacted with maleimide-activated TGF-β, and NIH/3T3 cells incubated on the coated surfaces were assayed by Western blot for expression of αSMA. The myofibroblast marker, αSMA, was detected in all samples to varying degrees. As expected,

the control samples with 5 ng/ml TGF-β in solution (positive controls) exhibited the highest amounts of αSMA, while the samples without TGF-β (negative controls) contained the lowest amounts of αSMA (Figure 2-7). Importantly, we observed that the band intensity of rCol-4C with TGF-β-maleimide was 1.8 ± 0.1 times the intensity of rCol (no Cys) with TGF-β-maleimide, highlighting the effect of covalently immobilizing TGF-β to the Cys collagen variant for promoting myofibroblast differentiation.

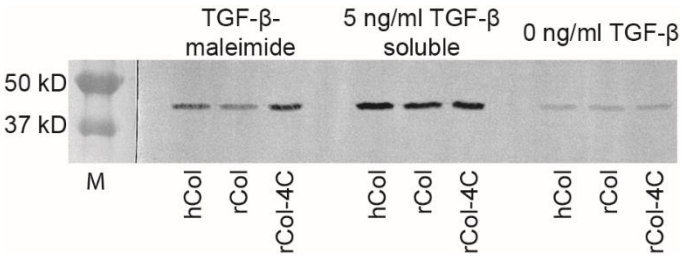


Figure 2-7: Representative anti-αSMA Western blot of NIH/3T3 cells seeded on different collagen substrates. Lanes: (M) Molecular weight markers; (hCol) Native human collagen; (rCol) Baseline recombinant human collagen; (rCol-4C) Recombinant collagen with four Cys per collagen. The first group was treated with maleimide-activated TGF-β and washed with PBS, the second group was treated with TGF-β in solution as a positive control, and the third group was treated with PBS-alone as a negative control.

2.4 Discussion

The requirements of the Gly-X-Y peptide repetition, together with the need for unique local sequences embedded within the repeats for cell-recognition sites and a relatively large protein length, makes the collagen gene difficult to synthesize via *de novo* strategies. This synthesis challenge poses a great hurdle to tailoring collagen's protein properties (e.g., tuning scaffold stiffness to guide cell fate or reduce compliance mismatch) to better suit bioengineering applications. Our modular collagen platform is able to address these challenges to yield prescribed, full-length collagen variants. By introducing mutations into the nucleotide sequence of primary fragments, genes encoding different numbers of non-native cysteines were produced by mixing-and-matching primary fragments for PCR assembly. The successful assembly of full-length collagen genes with multiple non-native functional mutations demonstrates the flexibility of the modular collagen platform. If

significant changes to the nucleotide sequence are desired, the method of site directed mutagenesis may not suffice, and optimization of the DNA sequence for the desired amino acid sequence and assembly of primary fragments from oligonucleotides may be necessary [15].

The introduction of four non-native cysteines per molecule, placed in the X or Y position of the Gly-X-Y repeat, did not affect the formation of full-length triple helical collagen. Although rCol-4C has an approximate 2 °C lower apparent T_m than the value for native human collagen III, it is approximately 1.5 °C higher than that for recombinant baseline collagen III (rCol) [25]. Furthermore, rCol-2C retains a similar T_m as rCol. Thus cell studies were performed at 30 °C, below the collagens T_m. Our yeast expression system has genetically introduced human prolyl-4-hydroxylase to hydroxylate the prolines in collagen. Without the hydroxylase enzymes, the resulting collagen was too unstable to enable purification. The lower apparent T_m of collagen produced by the recombinant yeast system, compared to native human collagen, may be due to differences in proline hydroxylation [25] or lysine hydroxylation, the latter of which increases collagen stability by promoting intramolecular crosslinking between collagen [45,46]. Interestingly, we consistently observe that collagen variants with four Cys yield higher T_m values than recombinant rCol without Cys, potentially due to stabilization from low levels of disulfide crosslinking. The thermostability of our recombinant collagen variants may limit their use to temperatures below 37 °C at this time. To address these stability questions, we are currently investigating strategies to increase thermostability, including strain engineering (e.g., increasing intracellular ratios of proline hydroxylase to collagen in the yeast) or substrate engineering (e.g., methods to stabilize collagen on solid substrates).

We showed that the introduced cysteines can function as unique and specific crosslinking or attachment sites. Our examination of protein sequences did not identify any cysteines embedded in the triple-helical regions of any collagen from native mammalian sources. Thus, introduction of Cys into the polymeric chain through genetic engineering allows for unique control over the location and the number of the chemically-functional thiols in each collagen molecule. Previous studies have

introduced sulfhydryls to collagen using Trauts reagent, which reacts with primary amines [47]; however this strategy does not enable control over the number or locations of reactive sites. This lack of control could potentially lead to heterogeneity of crosslinks or immobilized factors, resulting in heterogeneous local properties (e.g., irregular porosity or mechanics, local gradients of growth factor) within the scaffold.

Passive multiparticle microrheology confirmed that our Cys in rCol-4C can be crosslinked to form hydrogels; the material transitioned from a liquid-like to a solid-like character over time as the crosslinking reaction of introduced cysteines with malemide-functionalized PEG proceeded. The resulting collagen reached stiffness values that exceeded the upper sensitivity limit of our passive microrheology setup. Therefore, to gain additional information regarding the viscoelastic properties, we utilized traditional oscillatory rheology. Gelation was also observed, and differences in gel point times between microrheology and bulk rheology may be related to the differences in experimental protocols needed to perform the respective techniques (e.g., pH, presence of PEGylated microparticles), the size of the sample, and variations obtained from probing local versus bulk regions. Particle aggregation that occurred when monitoring rCol samples may have been due to salt bridging between charges on the microparticles and collagen.

One key advantage of using passive microrheology over traditional bulk rheological measurements for artificial ECM studies is the usage of very small sample sizes; however, as our current study has also shown, passive microrheology is limited in the range of measurable parameters, because tracer particle movements are solely driven by thermal energy. A possible alternative technique is active microrheology, which utilizes an external force (e.g., optical or magnetic) to move tracer particles within the material. Active microrheology also requires only small sample quantities, yet potentially provides access to a broader range of measurable rheological parameters compared to passive microrheology [48].

The storage moduli measured for our crosslinked samples were comparable to those of collagen gels of the same concentration formed through fibril formation [49,50]. Utilizing the equation $G_0 = vkT$, we can roughly estimate the distance between crosslinks for our hydrogels [51]. The calculated distances between crosslinks are 100 to 110 nm for rCol-2C and 70 to 90 nm for rCol-4C. These numbers are close to the calculated average distance between the introduced non-native cysteines within each collagen strand (approximately 120 nm and 90 nm for rCol-2C and rCol-4C, respectively). Although the storage modulus or stiffness of the formed gel was relatively low for typical tissue engineering applications, experimental conditions such as the protein concentration, crosslinking ratios, and the length of the crosslinkers could be optimized to obtain the desired viscoelastic properties. For example, studies with recombinant elastin utilize the relatively high concentration of 100 mg/mL to obtain desired mechanical properties [8]. Another study utilizing high molecular weight heparin displayed how varying crosslinker length as well as crosslinker ratios can modulate hydrogel stiffness [42].

We have also shown the ability of the introduced Cys thiols to function as anchoring sites for biofunctional molecules. Tethering of growth factors or other biologically-active molecules increases the number and type of tunable parameters in artificial ECM scaffolds. We demonstrated that collagen variants with TGF-β attached to non-native cysteines induced αSMA expression of NIH/3T3 fibroblasts to a greater degree than collagen without the cysteines, supporting a greater extent of myofibroblast differentiation. Other bioactive molecules, such as antibodies for cell capture [47] or VEGF to promote vascularization [52], could be utilized in lieu of TGF-β to customize the cellular microenvironment.

One limiting constraint of using native ECM sources is the inability to independently control parameters such as protein concentrations, mechanical properties, and densities of cell-reactive sites within collagen [53]. Our platform enables such decoupling in the context of collagen, allowing the

determination of the degree to which concentration of collagen, cell signaling site densities, or stiffness of the substrate modulates specific cell behavior. Other hydrogel systems exist that enable this decoupling, but are comprised primarily of the less-abundant ECM proteins (e.g., elastin [54]) or synthetic polymers (e.g., PEG [9]). Furthermore, collagen mimetic peptides are short (10 nm vs 300 nm), thus, reducing versatility[55]. Though advances have enabled collagen mimetic peptides to form fibers of 3 to 4 μm in length[56], they have not been shown to support cellular activity, nor have the functional sites of collagen been introduced. These alternative strategies rely on identifying specific cell-ECM interactions to engineer the respective known functions. In contrast, our platform utilizes the entire full-length sequence of collagen, which may include uncharacterized functionality and be beneficial for clinical applications.

2.5 Conclusions

In this study, we successfully produced full-length recombinant collagen variants containing non-native cysteines. We demonstrated the functionality of these engineered Cys residues by utilizing them for crosslinking into hydrogels and for growth factor immobilization. The introduction of two or four non-native Cys per collagen monomer (6 or 12 per trimer, respectively) resulted in stable triple helices and supported cell adhesion, both at comparable degrees relative to recombinant collagen without Cys. Thus, this recombinant collagen platform produced unique, specifically-prescribed mutants of collagen-mimetic biopolymers, demonstrating proof-of-concept to introduce non-native functional properties into a physiologically-important biopolymer.

2.6 References

[1] G.E. Davis, D.R. Senger, Endothelial Extracellular Matrix: Biosynthesis, Remodeling, and Functions During Vascular Morphogenesis and Neovessel Stabilization, Circ. Res. 97 (2005) 1093–1107. doi:10.1161/01.RES.0000191547.64391.e3.

[2] N. Yamamura, R. Sudo, M. Ikeda, K. Tanishita, Effects of the mechanical properties of collagen gel on the in vitro formation of microvessel networks by endothelial cells, Tissue Eng. 13 (2007) 1443–1453. doi:10.1089/ten.2006.0333.

[3] D.E. Discher, D.J. Mooney, P.W. Zandstra, Growth Factors, Matrices, and Forces Combine and Control Stem Cells, Science. 324 (2009) 1673–1677. doi:10.1126/science.1171643.

[4] M.P. Lutolf, J.A. Hubbell, Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering, Nat Biotech. 23 (2005) 47–55. doi:10.1038/nbt1055.

[5] Z. Ruszczak, Effect of collagen matrices on dermal wound healing, Adv. Drug Deliv. Rev. 55 (2003) 1595–1611. doi:10.1016/j.addr.2003.08.003.

[6] D. Olsen, C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perälä, E.-R. Hämäläinen, M. Jarvinen, J. Polarek, Recombinant collagen and gelatin for drug delivery, Adv. Drug Deliv. Rev. 55 (2003) 1547–1567. doi:10.1016/j.addr.2003.08.008.

[7] D.M. Asher, The transmissible spongiform encephalopathy agents: concerns and responses of United States regulatory agencies in maintaining the safety of biologics., Dev. Biol. Stand. 100 (1999) 103–118.

[8] K.S. Straley, S.C. Heilshorn, Independent tuning of multiple biomaterial properties using protein engineering, Soft Matter. 5 (2009) 114–124.

[9] M.P. Lutolf, J.L. Lauer-Fields, H.G. Schmoekel, A.T. Metters, F.E. Weber, G.B. Fields, J.A. Hubbell, Synthetic matrix metalloproteinase-sensitive hydrogels for the conduction of tissue regeneration: Engineering cell-invasion characteristics, Proc. Natl. Acad. Sci. 100 (2003) 5413–5418. doi:10.1073/pnas.0737381100.

[10] G.P. Raeber, M.P. Lutolf, J.A. Hubbell, Molecularly Engineered PEG Hydrogels: A Novel Model System for Proteolytically Mediated Cell Migration, Biophys. J. 89 (2005) 1374–1388. doi:10.1529/biophysj.104.050682.

[11] D. Sengupta, S.C. Heilshorn, Protein-engineered biomaterials: highly tunable tissue engineering scaffolds, Tissue Eng. Part B Rev. 16 (2010) 285–293.

[12] K. Shapira-Schweitzer, D. Seliktar, Matrix stiffness affects spontaneous contraction of cardiomyocytes cultured within a PEGylated fibrinogen biomaterial, Acta Biomater. 3 (2007) 33–41. doi:10.1016/j.actbio.2006.09.003.

[13] M.D. Shoulders, R.T. Raines, COLLAGEN STRUCTURE AND STABILITY, Annu. Rev. Biochem. 78 (2009) 929–958. doi:10.1146/annurev.biochem.77.032207.120833.

[14] S. Chattopadhyay, R.T. Raines, Review collagen-based biomaterials for wound healing, Biopolymers. 101 (2014) 821–833. doi:10.1002/bip.22486.

[15] S.W.P. Chan, S.-P. Hung, S.K. Raman, G.W. Hatfield, R.H. Lathrop, N.A. Da Silva, S.-W. Wang, Recombinant Human Collagen and Biomimetic Variants Using a De Novo Gene Optimized for Modular Assembly, Biomacromolecules. 11 (2010) 1460–1469. doi:10.1021/bm100052y.

[16] P. Szpak, Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis, J. Archaeol. Sci. 38 (2011) 3358–3372. doi:10.1016/j.jas.2011.07.022.

[17] C.M. Rubert Pérez, A. Panitch, J. Chmielewski, A Collagen Peptide-Based Physical Hydrogel for Cell Encapsulation, Macromol. Biosci. 11 (2011) 1426–1431. doi:10.1002/mabi.201100230.

[18] D.E. Przybyla, J. Chmielewski, Higher-Order Assembly of Collagen Peptides into Nano- and Microscale Materials, Biochemistry (Mosc.). 49 (2010) 4411–4419. doi:10.1021/bi902129p.

[19] C.D. Reyes, A.J. García, Engineering integrin-specific surfaces with a triple-helical collagen-mimetic peptide, J. Biomed. Mater. Res. A. 65A (2003) 511–523. doi:10.1002/jbm.a.10550.

[20] B. Brodsky, G. Thiagarajan, B. Madhan, K. Kar, Triple-helical peptides: An approach to collagen conformation, stability, and self-association, Biopolymers. 89 (2008) 345–353. doi:10.1002/bip.20958.

[21] S.M. Yu, Y. Li, D. Kim, Collagen mimetic peptides: progress towards functional applications, Soft Matter. 7 (2011) 7927–7938. doi:10.1039/C1SM05329A.

[22] I. Majsterek, E. McAdams, E. Adachi, S.T. Dhume, A. Fertala, Prospects and limitations of the rational engineering of fibrillar collagens, Protein Sci. 12 (2003) 2063–2072. doi:10.1110/ps.0385103.

[23] P.M. Royce, B. Steinmann, Connective tissue and its heritable disorders: molecular, genetic, and medical aspects, John Wiley & Sons, 2003. http://books.google.com/books?hl=en&lr=&id=x-Z-cXUGlL8C&oi=fnd&pg=PR5&dq=In+Connective+tissue+and+its+heritable+disorders:+molecular,+genetic+and+medical+aspects,+2nd+ed.%3B+Royce&ots=FysP6xccm4&sig=AgNVsbPuhWif7vCYuPLQ6RV8n5w (accessed July 9, 2014).

[24] J.-M. Bourhis, N. Mariano, Y. Zhao, K. Harlos, J.-Y. Exposito, E.Y. Jones, C. Moali, N. Aghajari, D.J.S. Hulmes, Structural basis of fibrillar collagen trimerization and related genetic disorders, Nat. Struct. Mol. Biol. 19 (2012) 1031–1036. doi:10.1038/nsmb.2389.

[25] A. Jabaiah, X. Wang, S.K. Raman, R. Ragan, N.A. Da Silva, S.-W. Wang, Nanoscale architecture and cellular adhesion of biomimetic collagen substrates, J. Biomater. Appl. 28 (2014) 1354–1365. doi:10.1177/0885328213508328.

[26] R.L. Strausberg, E.A. Feingold, L.H. Grouse, J.G. Derge, R.D. Klausner, F.S. Collins, L. Wagner, C.M. Shenmen, G.D. Schuler, S.F. Altschul, B. Zeeberg, K.H. Buetow, C.F. Schaefer, N.K. Bhat, R.F. Hopkins, H. Jordan, T. Moore, S.I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A.A. Farmer, G.M. Rubin, L. Hong, M. Stapleton, M.B. Soares, M.F. Bonaldo, T.L. Casavant, T.E. Scheetz, M.J. Brownstein, T.B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S.S. Raha, N.A. Loquellano, G.J. Peters, R.D. Abramson, S.J. Mullahy, S.A. Bosak, P.J. McEwan, K.J. McKernan, J.A.

Malek, P.H. Gunaratne, S. Richards, K.C. Worley, S. Hale, A.M. Garcia, L.J. Gay, S.W. Hulyk, D.K. Villalon, D.M. Muzny, E.J. Sodergren, X. Lu, R.A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madan, A.C. Young, Y. Shevchenko, G.G. Bouffard, R.W. Blakesley, J.W. Touchman, E.D. Green, M.C. Dickson, A.C. Rodriguez, J. Grimwood, J. Schmutz, R.M. Myers, Y.S.N. Butterfield, M.I. Krzywinski, U. Skalska, D.E. Smailus, A. Schnerch, J.E. Schein, S.J.M. Jones, M.A. Marra, Mammalian Gene Collection Program Team, Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 16899–16903. doi:10.1073/pnas.242603899.

[27] A.B. Moshnikova, V.N. Afanasyev, O.V. Proussakova, S. Chernyshov, V. Gogvadze, I.P. Beletsky, Cytotoxic activity of 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide is underlain by DNA interchain cross-linking, Cell. Mol. Life Sci. CMLS. 63 (2006) 229–234. doi:10.1007/s00018-005-5383-x.

[28] C.R. Lee, A.J. Grodzinsky, M. Spector, The effects of cross-linking of collagen-glycosaminoglycan scaffolds on compressive stiffness, chondrocyte-mediated contraction, proliferation and biosynthesis, Biomaterials. 22 (2001) 3145–54.

[29] B. Trappmann, C.S. Chen, How cells sense extracellular matrix stiffness: a material's perspective, Curr. Opin. Biotechnol. 24 (2013) 948–953. doi:10.1016/j.copbio.2013.03.020.

[30] S. Barrientos, O. Stojadinovic, M.S. Golinko, H. Brem, M. Tomic-Canic, PERSPECTIVE ARTICLE: Growth factors and cytokines in wound healing, Wound Repair Regen. 16 (2008) 585–601. doi:10.1111/j.1524-475X.2008.00410.x.

[31] J.P. Grande, Role of transforming growth factor-beta in tissue injury and repair, Proc. Soc. Exp. Biol. Med. 214 (1997) 27–40.

[32] R. Montesano, L. Orci, Transforming growth factor beta stimulates collagen-matrix contraction by fibroblasts: implications for wound healing, Proc. Natl. Acad. Sci. 85 (1988) 4894–4897.

[33] L.S.Z. Larsen, C.D. Wassman, G.W. Hatfield, R.H. Lathrop, Computationally Optimised DNA Assembly of synthetic genes, Int. J. Bioinforma. Res. Appl. 4 (2008) 324–336. doi:10.1504/ijbra.2008.019578.

[34] S.W. Chan, J. Greaves, N. Da Silva, S.-W. Wang, Assaying proline hydroxylation in recombinant collagen variants by liquid chromatography-mass spectrometry, BMC Biotechnol. 12 (2012) 51.

[35] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, Nat Protoc. 1 (2007) 2876–2890.

[36] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nat Protoc. 1 (2007) 2527–2535.

[37] T.G. Mason, D.A. Weitz, Optical Measurements of Frequency-Dependent Linear Viscoelastic Moduli of Complex Fluids, Phys. Rev. Lett. 74 (1995) 1250–1253.

[38] J.C. Crocker, D.G. Grier, Methods of Digital Video Microscopy for Colloidal Studies, J. Colloid Interface Sci. 179 (1996) 298–310. doi:10.1006/jcis.1996.0217.

[39] T.H. Larsen, E.M. Furst, Microrheology of the Liquid-Solid Transition during Gelation, Phys. Rev. Lett. 100 (2008) 146001.

[40] R.A. Evans, Y.. C. Tian, R. Steadman, A.O. Phillips, TGF-β1-mediated fibroblast–myofibroblast terminal differentiation—the role of smad proteins, Exp. Cell Res. 282 (2003) 90–100. doi:10.1016/S0014-4827(02)00015-0.

[41] R Core Team, R: A Language and Environment for Statistical Computing, Vienna, Austria, 2013. http://www.R-project.org/.

[42] K.M. Schultz, A.D. Baldwin, K.L. Kiick, E.M. Furst, Gelation of Covalently Cross-Linked PEG–Heparin Hydrogels, Macromolecules. 42 (2009) 5310–5316. doi:10.1021/ma900766u.

[43] H.H. Winter, F. Chambon, Analysis of Linear Viscoelasticity of a Crosslinking Polymer at the Gel Point, J. Rheol. 1978-Present. 30 (1986) 367–382. doi:http://dx.doi.org/10.1122/1.549853.

[44] D. Calvet, J.Y. Wong, S. Giasson, Rheological Monitoring of Polyacrylamide Gelation: Importance of Cross-Link Density and Temperature, Macromolecules. 37 (2004) 7762–7771. doi:10.1021/ma049072r.

[45] S. Jimenez, M. Harsch, J. Rosenbloom, Hydroxyproline stabilizes the triple helix of chick tendon collagen, Biochem. Biophys. Res. Commun. 52 (1973) 106–114. doi:10.1016/0006-291x(73)90960-1.

[46] M. Yamauchi, M. Shiiba, Lysine Hydroxylation and Cross-linking of Collagen, in: C. Kannicht (Ed.), Post-Transl. Modif. Proteins, Humana Press, 2008: pp. 95–108. http://dx.doi.org/10.1007/978-1-60327-084-7_7.

[47] C. Shi, Q. Li, Y. Zhao, W. Chen, B. Chen, Z. Xiao, H. Lin, L. Nie, D. Wang, J. Dai, Stem-cell-capturing collagen scaffold promotes cardiac tissue regeneration, Biomaterials. 32 (2011) 2508–2515. doi:10.1016/j.biomaterials.2010.12.026.

[48] I. Sriram, E.M. Furst, R.J. DePuit, T.M. Squires, Small amplitude active oscillatory microrheology of a colloidal suspension, J. Rheol. 1978-Present. 53 (2009) 357–381. doi:http://dx.doi.org/10.1122/1.3058438.

[49] R.K. Willits, S.L. Skornia, Effect of collagen gel stiffness on neurite extension, J. Biomater. Sci. Polym. Ed. 15 (2004) 1521–1531. doi:10.1163/1568562042459698.

[50] Y. Yang, L.M. Leone, L.J. Kaufman, Elastic Moduli of Collagen Gels Can Be Predicted from Two-Dimensional Confocal Microscopy, Biophys. J. 97 (2009) 2051–2060.

[51] R.G. Larson, The Structure and Rheology of Complex fluids, (1999). http://www.matdl.org/matdlwiki/index.php?title=softmatter:The_Structure_and_Rheology_of_Complex_Fluids&redirect=no (accessed July 9, 2014).

[52] Y.H. Shen, M.S. Shoichet, M. Radisic, Vascular endothelial growth factor immobilized in collagen scaffold promotes penetration and proliferation of endothelial cells, Acta Biomater. 4 (2008) 477–489. doi:10.1016/j.actbio.2007.12.011.

[53] K.J. Lampe, S.C. Heilshorn, Building stem cell niches from the molecule up through engineered peptide materials, Bioeng. Approaches Nerv. Syst. Repair. 519 (2012) 138–146. doi:10.1016/j.neulet.2012.01.042.

[54] K.J. Lampe, A.L. Antaris, S.C. Heilshorn, Design of three-dimensional engineered protein hydrogels for tailored control of neurite growth, Acta Biomater. 9 (2013) 5590–5599. doi:10.1016/j.actbio.2012.10.033.

[55] F.W. Kotch, R.T. Raines, Self-assembly of synthetic collagen triple helices, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 3028–3033. doi:10.1073/pnas.0508783103.

[56] S. Rele, Y. Song, R.P. Apkarian, Z. Qu, V.P. Conticello, E.L. Chaikof, D-periodic collagen-mimetic microfibers, J. Am. Chem. Soc. 129 (2007) 14780–14787. doi:10.1021/ja0758990.

Chapter 3: Tuning Cellular Interaction through Introduction and Removal of Integrin Binding Sites

Parts of this chapter are reprinted from Biomaterials, Vol 53, Richard A. Que, Sam Wei Polly Chan, Abeer M. Jabaiah, Richard H. Lathrop, Nancy A. Da Silva, Szu-Wen Wang, Tuning cellular response by modular design of bioactive domains in collagen, 309-317, 2015, with permission from Elsevier

3.1 Introduction

The importance of the microenvironment in cellular development, self-renewal, and differentiation has propelled efforts to create biomimetic materials that can direct cellular response for therapeutic applications and regenerative medicine. Characteristics of the local extracellular matrix (ECM), including nanoscale architecture, mechanical properties, cell-interaction site presentation, and degradation rates, have been shown to regulate cellular behavior [1-3]. However, limitations of using naturally-derived ECM components have included the inability to decouple such properties for independent assessment [2], and the use of animal sources, which give materials that are often poorly-defined and immunogenic in clinical applications [4, 5]. To address these limitations, strategies to generate cell-responsive synthetic materials have included the chimeric integration of bioactive ECM sites into synthetic peptides or polymers (e.g., polyethylene glycol, hyaluronic acid) [6-8] and the use of recombinant systems for molecular-level control (e.g., elastin) [9].

Given that > 90% of natural ECM tissue and approximately 30% of all mammalian proteins comprise the protein collagen [10, 11], it follows that collagen is extensively used in regenerative medicine investigations as a scaffold on which to grow cells. However, current tissue engineering efforts rely on "off-the-shelf" native collagen obtained from animal sources, such as bovine collagen or Matrigel, so there is little or no control over molecular-level parameters, such as protein sequence, cell-collagen interactions, and residual bioactive impurities [8]. Furthermore, a synthetic bottom-up approach has been elusive, due to the difficulties in generating synthetic genes encoding the $(\text{glycine-X-Y})_N$ backbone of collagen and the stringent need to post-translationally hydroxylate prolines in the Y-position for stability [12]. Consequently, synthesis of collagen-mimetic material has primarily focused on peptides [13-17], unhydroxylated collagen-like polymers from bacteria [18, 19], or tandem repeating collagen domains [20].

Our research group has developed a platform for creating recombinant human collagen III in

which we can specifically tailor the identity, location, and frequency of functional sites within the biopolymer [12, 21]. Collagen III was selected as the molecular scaffold because it is a homotrimer; experimentally, this requires only one gene to be synthesized with no need to separate the heterotrimeric populations of product. Synthetic genes encoding the collagen-mimetic biopolymers are fabricated from oligonucleotides in which DNA sequences are optimized for gene assembly and yeast expression using a biocomputational strategy. This optimization is necessary to minimize the mishybridization propensities due to the repetitive glycine-X-Y amino acid sequences and G-C rich sequences [12]. Twelve gene modules (primary fragments, PF) were designed which span the entire triple-helical region of collagen III, and the baseline DNA modules encode for the human amino acid protein sequence. The protein is synthesized in a recombinant *Saccharomyces cerevisiae* yeast system that has been genetically altered to express human α- and β-prolyl hydroxylase, which impart stability to the collagen triple-helix by post-translational hydroxylation [22]. In this investigation, we demonstrate the modularity of this platform and its ability to create defined, non-native variants of human collagen III in the context of integrin binding sites.

Integrins are the primary receptors that mediate cell adhesion and mechanical interactions with the extracellular matrix and are important in processes such as cell adhesion, migration, and differentiation [23, 24]. Modulation of the cellular adhesion motif concentration has led to modulation of the robustness of vasculogenesis, neuronal differentiation, and neurite extension [25, 26]. Genetic mutations in these receptors can result in pathologies such as tumor growth and metastasis, muscular dystrophy, and thrombosis [24]. While biomaterials engineering has focused on manipulating the fibronectin-based sequence RGD [7], only approximately one-third of all integrins bind to this sequence [24]. Thus, there is untapped potential to engage alternative integrins for biomaterials design, thereby expanding the potential to modulate alternative types of cells and their corresponding processes.

In this investigation, multiple native integrin binding sites were removed from the natural

amino acid sequence of collagen III [27, 28], and a site from collagen I but nonexistent in collagen III (GFOGER) [29] was introduced at various locations and frequencies. Prior examples of biomaterials engineering with GFOGER have been limited to short peptides, which are able to support cellular adhesion, promote osteoblast-specific gene expression and vascularization [13, 15], but do not present the peptides in native context. Our bottom-up, modular strategy enables the re-design of full-length collagen at the molecular level to tailor collagen for tuning cellular microenvironments and response. This alternative strategy could expand the level of control over cellular behavior in therapeutics and regenerative medicine.

3.2 Materials and methods

3.2.1 Materials

E. coli strain DH5α (Zymo Research) was used for plasmid maintenance and amplification. Haploid *S. cerevisiae* strain BYα2β2 [22] was used to express the collagen baseline and all mutants. TOPO vector used to hold PCR generated gene fragments, fetal bovine serum (FBS), penicillin, and streptomycin were purchased from Life Technologies. Restriction enzymes, DNA ligase, DNase and RNase were from New England Biolabs. KOD Hot Start DNA Polymerase, human collagen III, bovine collagen I, anti-integrin α2 (MAB1950Z), and anti-integrin β1 (MAB2253Z) were purchased from EMD Millipore. PfuUltra II DNA polymerase was from Agilent Technologies. DNA fragments were purified using ZymoClean Gel DNA Recovery Kit (Zymo Research) after gel electrophoresis. Phenylmethylsulfonyl fluoride (PMSF) and BCA assay were from Pierce. Pepsin and phosphate buffer saline (PBS) were from MP Biomedicals. Human collagen I, bovine collagen III, and calcein-AM were purchased from BD Bioscience. Streptavidin-alkaline phosphatase was purchased from GE Healthcare. All *E. coli* and yeast growth media were from Difco. DMEM, dibutyryl cyclic AMP (cAMP), mouse anti-MAP2 (M9942), rabbit anti-GFAP (G9269), and monoclonal anti-vinculin antibodies were purchased from Sigma. Epidermal growth factor (EGF), basic fibroblast growth factor (bFGF), and platelet-derived growth factor (PDGF) were purchased from PeproTech. Rabbit anti-Ki67 (KI67P-

CE) was purchased from Leica Biosystems. Mouse anti-phosphohistone H3 (9706S) was purchased from Cell Signaling Technologies. Unless otherwise noted, all other chemicals were purchased from Fisher Scientific.

3.2.2 Construction of genes encoding collagen-mimetic mutants

3.2.2.1 Integrin binding site removal and insertions.

Our goal was to develop collagen substrates with varying frequencies, locations, and combinations of α2β1 binding sites and to examine their structure, stability, and biological effect on interacting cells. We used our modular human collagen gene, which has been described elsewhere [12], as the underlying backbone upon which desired variants were made. The design of full-length human collagen III includes twelve gene modules (or "primary fragments", PF), which span and encode for the entire triple-helical region (See Figure 1 for "Baseline" collagen, rCol) and additionally includes the N- and C-terminal propeptides and telopeptides. These ends were included to promote correct triple-helical formation. Presence of the triple-helix initiating C-propeptide was confirmed by Western blot after protein expression.

Mutants of human collagen were designed as described in Figure 3-1 and Table 3-1. One of these variants included a collagen (rCol-0G) with the native α2β1 binding sites removed (GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER hexamer motifs [27, 28]) and replaced with a non-binding sequence. To identify a replacement non-binding hexameric sequence, we calculated the thermodynamic melting temperatures (T_m) for oligonucleotides encoding an exhaustive list of hexamer peptides to identify sequences that would minimally interfere with PCR-based gene synthesis. The top-ten hexamer candidates were then cross-checked with their respective Toolkit III peptides [28] to determine the sequences likely to yield a well-defined, thermostable triple helix with low cellular binding. This analysis yielded the non-binding hexamer GSPGGK, which replaced the native α2β1 binding sites. Although the melting temperature of short peptides is lower when glycines are introduced into the X-position of Gly-X-Y [30], work by Raynal et al. reported that the

region with the native GSPGGK site had a T_m which was approximately average for all the 27-mer peptides spanning the entire triple helical region of collagen III [28]. Therefore, we did not expect the introduction of this sequence to be disruptive to the overall melting temperature of rCol-0G.

An integrin binding site from collagen I (GFOGER [29]) was systematically introduced into the rCol-0G scaffold. To preserve the immediate local context and stability of the GFOGER sequence, we included the three native flanking amino acids on both sides of the collagen I sequence. Therefore, to create the GFOGER variants, the amino acid sequence GER-GFOGER-GVQ was substituted for twelve amino acids in each of the respective replacement sites within the rCol-0G scaffold. Three factors went into the selection of the new locations for introducing the non-native GFOGER sequences. The first criterion was to choose primary fragments (PF) that yielded relatively uniform spacing between the GFOGERs. Secondly, the proximity to the center of the selected PFs was preferred, since mutagenesis in locations close to the fragment edges would overlap with the adjacent fragment and require additional oligonucleotide and PCR fragment synthesis. Finally, we favored regions that minimized the number of amino acid changes between the native and GFOGER sequences. For example, the GFOGER of PF1 was introduced into the same location as the native GROGER sequence.

3.2.2.2 Mutagenesis of primary fragments in collagen variants

For introducing GFOGER sequences into PFs 1, 4, 8, and 11, we used previously described strategies of either PCR-assembly using oligonucleotides encoding the PF1 [12] or site-directed mutagenesis using PF 4, 8, and 11 for rCol-0G as the template DNA. Oligonucleotide sequences are listed in Tables A3-1 (for PCR assembly of rCol-0G and rCol-1G-1) and A3-2 (for mutagenesis of PF 4, 8, 11). After defining the amino acid sequences of the different variants (Table 3-1), the DNA sequences were optimized by a computational algorithm to favor correct hybridizations between synthetic oligonucleotides for PCR assembly and to disrupt incorrect hybridizations, and for expression in *S. cerevisiae* [12, 31].

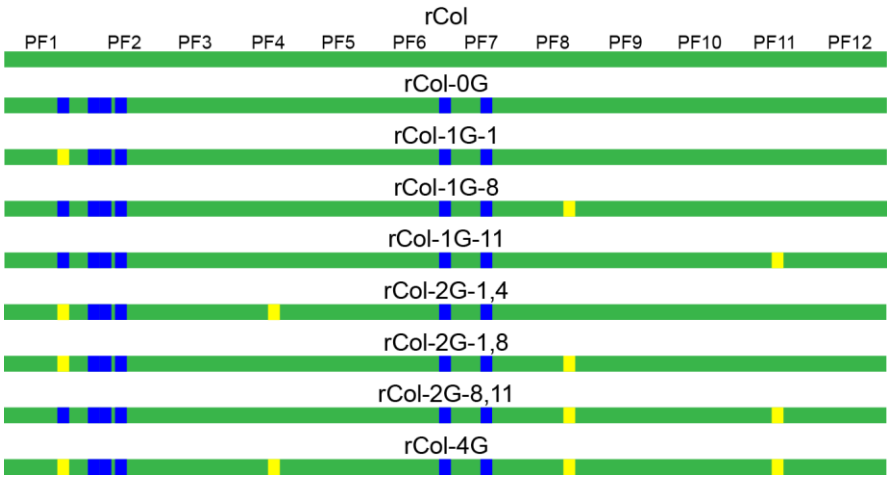


Figure 3-1: Summary of recombinant collagen variants containing integrin-binding sites in the triple helical domain. Variants were produced by removing all native integrin sites from human collagen III (rCol-0G) and systematically introducing non-native integrin-binding sites (GFOGER) in different frequencies and locations. Green regions denote amino acid sequences native to human collagen III, blue regions are those where native integrin sites have been replaced with non-binding sites, and yellow regions indicate non-native GFOGER site from collagen I. "PF" indicates the primary fragment of the triple-helical region (one of 12 modules used in gene synthesis). The N- and C-propeptide regions are not pictured; while they are expressed, they are removed during purification.

Table 3-1: Summary of recombinant collagen variants and native human collagen controls. Listed are the abbreviated name of the variant and the number and location of GFOGER integrin binding sites. Primary fragment (PF) indicates the region in which the GFOGER sequence is located.

Name of collagen	Number of GFOGER sites	Description of collagen variant
hCol I	1	Human collagen I from native source; 1 native GFOGER site
hCol III	--	Human collagen III from native source (no GFOGER, but contains GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER binding sites)
rCol	--	Human collagen III amino acid sequence (no GFOGER, but contains GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER binding sites)
rCol-0G	0	Human collagen III with GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER integrin binding hexamers replaced with non-binding GSPGGK
rCol-1G-1	1	rCol-0G with GFOGER in PF1
rCol-1G-8	1	rCol-0G with GFOGER in PF8
rCol-1G-11	1	rCol-0G with GFOGER in PF11
rCol-2G-1,4	2	rCol-0G with GFOGERs in PF1 and PF4
rCol-2G-1,8	2	rCol-0G with GFOGERs in PF1 and PF8
rCol-2G-8,11	2	rCol-0G with GFOGERs in PF8 and PF11
rCol-4G	4	rCol-0G with GFOGERs in PF1, PF4, PF8, and PF11

For rCol-0G, the PFs containing the non-binding hexamer GSPGGK (replacing the native hexameric integrin binding sites) were PCR assembled from oligonucleotides using previously-described conditions [12]. To introduce the non-native GFOGER sequences into PFs, we used previously-described strategies of PCR-assembly using oligonucleotides for encoding GFOGER into PF1 [12] and site-directed mutagenesis using PFs 4, 8, and 11 for rCol-0G as the template DNA [32]. Mutagenesis primers were designed based on Liu et al. [32] which allows for the relatively large length of the mutation. Mutation primers were designed to overlap only on their 5' ends, centered to span across the mutation site, and designed to have higher annealing temperatures for the non-overlapping region than the overlapping region. The mutagenesis was performed in a reaction of 1 μM mutation primer pairs mixed with 10 ng of their respective DNA template, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. PCR was performed with 5 min at 95°C followed by 12 cycles of 1 min at 95°C, 1 min at 55 – 65°C, and 2 min at 72°C, followed by 1 min at 50 – 60°C, and 10 min at 72°C. Methylated template DNA was digested with 20 U DpnI for 2 hr at 37°C, gel purified, and transformed into *Escherichia coli* DH5α. Colonies with plasmids containing the desired PF sequences were identified. The PFs containing the mutated sequences were ligated into pCR-BluntII-TOPO vectors, and the correct final sequences were confirmed by DNA sequencing (Genewiz). The correct final sequences for PFs were confirmed by DNA sequencing.

3.2.2.3 Fabrication of genes encoding full-length collagen variants

Full-length genes were assembled based on our previously-reported methods [12, 21], in which secondary fragments were PCR-assembled from PFs, and full-length genes for collagen variants were assembled from the secondary fragments. Full-length collagen mutants containing 1 to 4 introduced GFOGER sites, with the rCol-0G as the underlying background scaffold, were assembled from primary fragments following protocols detailed in Chan et al. [12].

PFs were amplified by PCR by combining 200 nM of each oligo flanking each primary fragment, 50 ng of PF template DNA, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. Reactions were performed using a 10 min at 95°C followed by 25 cycles of 20 s at 95°C, 30 s at 62°C, and 15 s at 72°C, followed by 5 min at 72°C. PCR product was then gel purified for use in assembly of secondary fragments.

To assemble secondary fragments, 2 nM of each required primary fragment was mixed together with 200 nM each oligo flanking each secondary fragment, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. Reactions were performed using a 10 min denaturing step at 95°C, followed by 30 cycles of 20 s at 95°C, 30 s at 68°C, and 45 s at 72°C, followed by 5 min at 72°C. Reaction product was agarose gel purified and ligated into pCR-BluntII-TOPO vectors. Colonies with plasmids containing the correct sequence were identified by DNA sequencing.

Secondary fragments were amplified by PCR by combining 200 nM of each oligo flanking each secondary fragment, 50 ng of secondary fragment template DNA, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, 1 x PfuUltra II reaction buffer in a final volume of 50 μL. PCR was performed using a 10 min denaturing step at 95°C followed by 25 cycles of 20 s at 95°C, 30 s at 62°C, and 45 s at 72°C, followed by 5 min at 72°C. PCR product was then gel purified for use in assembly of full length collagen.

To assemble full length collagen mutants, 2 nM of each secondary fragment was mixed together with 200 nM primers containing a PmeI site upstream of the gene and a NotI site downstream. Full length collagen genes were assembled using 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. Reactions were performed using a 10 min denaturing step at 95°C followed by 30 cycles of 20 s at 95°C, 30 s at 68°C, and 1.5 min at 72°C, followed by 5 min at 72°C. The reaction product was agarose gel purified and ligated into pCR-BluntII-TOPO vectors. Full-length collagen genes with GFOGER mutations were

excised with PmeI and NotI and ligated into the CEN/ARS plasmid. Final sequences for the collagen variants were confirmed by DNA sequencing.

3.2.3 Expression and purification of recombinant collagen variants

Expression and purification were performed using previously-described protocols as a departure point [21, 22]. The full-length collagen genes were cloned in the CEN/ARS plasmid and transformed in *S. cerevisiae* strain BYα2β2 in which we had integrated two copies each of α- and β-prolyl hydroxylase genes into the genome [22]. Selective SDC(A) and SGC(A) media contain casamino acids (5 g/L), yeast nitrogen base without amino acids (6.7 g/L), adenine sulfate (20 mg/L), and glucose (20 g/L, in SDC(A)) or galactose (20 g/L, in SGC(A)), respectively. Cells were inoculated from stock cells, grown overnight in 16 × 125 mm culture tubes containing 5 mL SDC(A) medium, and used to inoculate (1% v/v) 2.8 L Fernbach flasks containing 1 L of SGC(A) induction medium. The yeast were cultivated at 250 rpm and 30°C (Series 25, New Brunswick), and OD_{600} was monitored. The cells were harvested at an OD_{600} of 2.5 to 3.5 during late exponential phase by centrifugation at 4000 ×g, and then frozen at -80°C.

To purify the collagen product from yeast, the yeast cells were thawed and resuspended at 0.1-0.2 g wet cell/mL in cold Tris-buffered saline (TBS, 100 mM Tris, 400 mM NaCl, pH 7.4) with 1 mM PMSF, 2 U/mL DNase, and 1 U/mL RNase. The cells were lysed in a pre-cooled French Press (Thermo Scientific) using 25000 psi for two cycles. Ethylenediaminetetraacetic acid (EDTA; 2 mM) was added to the lysate on ice and hydrochloric acid was added to lower the pH to 2. Pepsin (20 mg/mL in 0.1M HCl) was added to a final concentration of 0.2 mg/mL and the solution was incubated at 4°C overnight to digest the background protein and collagen propeptides, leaving the triple-helical component intact. The supernatant was collected by ultracentrifugation (Optima LE80K, Beckman) for 30 min at 70,000 × g and 4°C, and all subsequent steps were performed on ice or in the cold room. Glacial acetic acid was added to a final concentration of 0.5 M and NaCl was added to a final concentration of 3 M with stirring. Precipitated protein was collected by ultracentrifugation for 30

min at 70,000 ×g and 4°C, and resuspended in cold 0.1 N HCl. The pH of the resuspension was raised to 7.4 with 200 mM Tris pH 8.6, and NaCl solution was added to a final concentration of 3 M. Precipitated protein was collected by ultracentrifugation again and resuspended in cold 0.1 N HCl. Protein solutions were dialyzed against 50 mM acetic acid. Concentration of collagen protein was determined using BCA assay, with bovine collagen type III (BD Biosciences) as a standard. Typical yields ranged from 300 – 500 μg of purified protein per L of culture.

3.2.4 Stability and structural analysis

Confirmation of triple-helix formation and thermostability were performed via circular dichroism (CD), and atomic force microscopy was used to image proteins and measure structural characteristics. These methods are based on previously published methods [21, 22].

CD scans were performed on a spectropolarimeter equipped with a Peltier temperature controller (Jasco). The protein samples (collagen variants and native collagen controls) in 50 mM AcOH were scanned between 190–260 nm at 10°C at a scanning speed of 50 nm/min in 0.1 cm path length quartz cells. For thermal unfolding, we monitored the ellipticity at 221 nm from 15 to 65°C at a heating rate of 1°C/min. The apparent melting temperature (T_m) was defined as the temperature at which 50% of the protein is unfolded. This was determined by fitting the experimental thermal profile to a thermodynamic model via a multi-parameter, non-linear regression using the Levenberg-Marquardt algorithm within MATLAB [33, 34]. CD wavelength scans were performed on at least 3 independent batches of protein, and T_m measurements were performed on at least 2 independent batches.

AFM was used to characterize the structure of the recombinant collagen variants. Freshly cleaved mica surfaces were incubated with 1 ng/μl of protein in 0.5 M acetic acid for 3 min at room temperature. Surfaces were then washed with ddH_2O, dried with nitrogen, and imaged with AFM.

AFM imaging was performed in air with a Bruker MultiMode 8 instrument in PeakForce Tapping mode. Tips were ScanAsyst-Air-HR probes with a resonance frequency of 130 kHz, nominal

tip radius of 2 nm, and spring constant of 0.3 N/m. AFM images were obtained with scan speeds of 1.95 or 3.91 Hz and rendered with background slopes corrected using NanoScope software. The length and height of each collagen strand was estimated using the "section" function of NanoScope Analysis. Measurements are reported as mean ± standard deviation of at least 8 independent polymer strands from at least 2 different images.

3.2.5 Cell adhesion on collagen substrates

3.2.5.1 Substrate deposition and cell adhesion imaging

HT-1080 cells were grown in DMEM (with 10% FBS, 100 units/ml penicillin, and 100 µg/ml streptomycin) and subcultured every 3 days. Cells were trypsinized, washed, and resuspended in DMEM (with 0.1% BSA) before seeding onto the protein-adsorbed surfaces. To prepare these surfaces, 20 µg/mL native and recombinant collagens (in PBS) were incubated on 96-well non-tissue culture treated polystyrene plates (Corning) for 24 h at 4°C. Our prior investigations showed that protein adsorption onto non-tissue culture treated wells was saturated when surfaces were incubated with ≥10 µg/ml collagen [22]. Relative surface protein concentrations were quantified as described in SI.

Protein solutions were removed, and surfaces were incubated with DMEM with 0.5% BSA for 1 h. HT-1080 cells suspended in DMEM containing 0.1% BSA were then seeded at 3.15×10^4 cells/cm^2 on the protein-adsorbed surfaces and incubated for 4 h at 30°C. The wells were washed with serum-free DMEM and imaged (Olympus IX-51). To quantify cell spreading, we measured cell areas. The boundaries for individual cells were manually defined using Image J (v 1.48), and the areas from fifty adherent cells (N=50) on each substrate were used to obtain statistical information.

3.2.5.2 Quantifying number of cells adhered to surfaces

HT-1080 cells were incubated on the different protein-adsorbed surfaces for 4 h at 30°C. To remove non-specifically bound cells, wells were filled with DMEM containing 0.1% BSA, and the plate was sealed with Microseal tape (Bio-Rad). The plates were inverted, centrifuged at 900 x g, and

washed with PBS. Adhered cells were incubated with 4 μM of calcein-AM (in PBS and 0.1% BSA) at 37°C for 1 h. Solutions were removed, and cells were lysed with CelLytic M (Sigma). Fluorescence was measured at 494/517 nm (Ex/Em) (SpectraMax M2) and normalized to values for hCol I.

3.2.5.3 Cellular adhesion modulation assays

To evaluate whether modulation of adhesion could be achieved through mixtures of binding and non-binding variants, native hCol I, rCol-1G-8, and rCol-4G were diluted with rCol-0G, yielding solutions containing 0.3 to 20 μg/ml of the GFOGER-containing collagen, but with 20 μg/ml total protein concentration. As described above, protein mixtures were adsorbed and incubated with cells, and cell adhesion was quantified.

3.2.5.4 Integrin inhibition assays

We examined the specificity of HT1080 cell binding via α2β1 integrins to the collagen substrates. HT-1080 cells were resuspended in DMEM supplemented with 0.1% BSA at 3.15×10^4 cells/cm^2. Anti-α1, anti-α2, anti-α3, anti-β1, or non-specific IgG antibodies (at 10 μg/mL) were incubated with the cell suspension for 15 min at 37°C. These cell/antibody solutions were incubated with collagen-adsorbed surfaces for 4 h at 30°C. Cells were quantified with calcein-AM as described above.

We examined the specificity of human neural stem/progenitor cell (hNSPC) binding via α1, α2, α3, and β1 integrins to collagen substrates. HNSPCs were resuspended in hNSPC basal media (DMEM/F12, 20% BIT 9500 (bovine serum albumin, insulin, and transferrin), 1 % penicillin/streptomycin/amphotericin) supplemented with 40 ng/mL of bFGF, EGF, and PDGF. Anti-α1, anti-α2, anti-α3, anti-β1, or non-specific IgG antibodies (at 20 μg/mL) were incubated with the cell suspension for 15 min at 37°C. Cells were seeded onto collagen-adsorbed surfaces at 3.15×10^4 cells/cm^2 and incubated for 24 h at 37°C. Cells were washed with MEM thrice and incubated with 4 μM calcein-AM in PBS for 1 h. Cells were lysed with Cellytic M and fluorescence was read at 494/517 excitation/emission.

3.2.5.5 Quantification of focal adhesion complexes

To investigate the effect of varying integrin binding sites within collagen substrates on the formation of focal adhesion complexes, we measured the relative amount of vinculin based on a previous protocol [35]. Wells of a non-tissue culture treated 6-well plate were coated with collagen-variant proteins. HT-1080 cells were seeded at 6.3 x 10^4 cells/cm^2 in DMEM supplemented with 10 % FBS and incubated for 4 h at 30°C. The solution was gently removed and replaced with DPBS, and bound integrins were crosslinked with 1 mM DTSSP in PBS for 30 min. Unreacted DTSSP was quenched with 50 mM Tris-HCl in DPBS for 5 min. Uncrosslinked cellular components were extracted with 0.1% SDS in DPBS supplemented with 1 mM PMSF for 5 min. Wells were gently washed with DPBS, and crosslinked proteins were released by the addition of 50 mM DTT and 0.1% SDS in DPBS (without Ca^{2+} and Mg^{2+}) at 37°C for 30 min. Vinculin was detected through Western blot using monoclonal anti-vinculin antibodies. Band intensities were quantified using ImageJ and normalized to the hCol I band intensity for each experiment. Statistical analysis was performed on three independent experiments (N=3).

3.2.6 Human NSPC proliferation and differentiation

We examined the ability for hNSPCs seeded on collagen adsorbed surfaces to support differentiation into astrocytes and neurons. Non-tissue culture treated surfaces were coated with collagen at 20 µg/mL for 24 h at 4°C or 20 µg/mL of laminin at 37°C for 24 h. hNSPCs in hNSPC basal media supplemented with 40 ng/mL of bFGF, EGF, and PDGF were seeded at 3.15 x 10^4 cells/cm^2 and incubated at 37°C. Media was changed every other day. On day 1, wells for proliferation were fixed, stained and imaged as described below. On day 4 media was exchanged for differentiation media (96% Neurobasal, 2% B-27, 1% GlutaMax, and 1% penicillin/streptomycin, with 20 ng/mL brain-derived neurotrophic factor, 20 ng/mL glial-derived nerotrophic factor, and 0.5 µM dibutyryl cyclic AMP). Cells were incubated for 14 days, refreshing media every other day, then fixed, stained and imaged as described below.

3.2.6.1 Immunostaining.

Cells were fixed with 4% paraformaldehyde for 10 min as described previously [36]. Fixed cells were treated with 0.3% Triton X-100 in PBS for 5 min and blocked for 1 h with 5% BSA in PBS. Cells were incubated with primary antibody in 1% BSA in PBS overnight at 4°C. For differentiation experiments, neurons were stained with 1:200 mouse anti-MAP2, and astrocytes were stained with 1:200 rabbit anti-GFAP. For proliferation experiments, cells were stained with 1:500 rabbit anti-Ki67 to indicate cells not in G_0, and cells undergoing mitosis were stained with 1:200 mouse anti-phosphohistone H3. Cells were washed thrice with PBS for 5 min then incubated with secondary antibodies, donkey anti-mouse 488 (A21202) and donkey anti-rabbit 555 (A31572) at 1:200 dilution in 1% BSA in PBS at room temperature for 2 h protected from light. Cells were washed thrice with PBS and counterstained with 1:500 Hoechst 33342 in PBS. Wells were covered with VECTASHIELD (Vector Labs) and imaged with a Nikon Eclipse Ti microscope with a 20X objective.

3.2.6.2 Quantification of proliferation

To quantify the proliferation rate of hNSPCs on the various protein substrates, Image J was utilized. Total cell number was counted using Hoechst labeled nuclei. Few cells were labeled with anti-phosphohistone H3 that were also labeled with anti-Ki67, so only cells labeled with anti-Ki67 were utilized to determine the number of proliferating cells. Proliferating percentage was defined as (# of nuclei Ki67 positive) / (total number of nuclei).

3.2.7 Statistical analyses

For cell adhesion and focal adhesion quantification assays, statistical analyses were performed using at least three independent experiments. Each independent cell adhesion and antibody inhibition experiment contained at least three replicates (total N=9); cell area analyses contained at least 50 cells (N=50). Statistical significance was determined using the software program R [37]. We performed a one-way, analysis of variance (ANOVA) analysis, followed by Tukey's HSD post-test comparison. Data is reported as mean ± standard error of the mean (S.E.M.)

unless otherwise stated, and values were considered statistically significant when p-values were < 0.05.

3.3 Results and Discussion

3.3.1 Design strategy and synthesis of collagen genes encoding different combinations of non-native integrin binding sites.

To examine the feasibility of introducing non-native cell-responsive sites in a collagen scaffold and to study their respective bioactive properties, we designed a set of collagen variants containing the collagen I GFOGER $\alpha 2\beta 1$ integrin binding sequence in different locations, frequencies, and combinations within a collagen III scaffold (summarized in Table 3-1, Figure 3-1). The gene synthesis for this set of collagen variants was performed using our modular human collagen III gene, described previously [12, 22], coding for the biopolymer backbone.

Since collagen III naturally exhibits strong adhesion to cells, it was necessary first to design a full-length collagen variant in which native integrin interaction sites were removed to serve as a non-binding scaffold. Studies have identified the main $\alpha 2\beta 1$ integrin binding amino acid sequences in human collagen III to be the peptide hexamers GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER [27, 28] (single amino acid-abbreviations used). Based on the overlap between calculated optimized oligonucleotide melting temperatures (T_m) for hexameric sequences in gene synthesis [12], together with experimentally-determined triple helical sequences that are both thermostable and exhibit minimal cell adhesion properties [28], we identified the hexameric sequence GSPGGK to replace each of the five native integrin binding sites in collagen III. This yielded the DNA encoding the protein sequences for the collagen variant "rCol-0G," and this gene assembled correctly as confirmed by DNA sequencing.

The integrin binding sequence GFOGER was then introduced into the rCol-0G scaffold in primary fragments 1, 4, 8, and 11. GFOGER (Gly-Phe-Hyp-Gly-Glu-Arg) is found once within natural

collagen I but never in collagen III, and it is specifically recognized by α1β1, α2β1, α10β1, and α11β1 integrins [24, 29]. As with rCol-0G, all genes encoding these variants correctly assembled, demonstrating the modularity of the optimized collagen III gene and the ability to fabricate the genes encoding both single-GFOGER variants and those with a combination of GFOGER sites (Table 3-1 and Figure 3-1). These full-length genes were then placed into the yeast expression system for expression, purification, and characterization of each collagen variant.

3.3.2 The collagen scaffold was tolerant of numerous changes while retaining triple-helical architecture.

All collagen variants were successfully expressed and purified (Figure A3-1). Circular dichroism (CD) showed that all variants yielded spectra similar to data for rCol [22], and results confirmed their triple-helical structure (Figure 3-2; Figure A3-2). Figure 3-2a shows that the variant with the maximum number of changes (rCol-4G) still remains triple-helical. The linear, triple-helical structure is confirmed by AFM imaging (Figure 3-2c). Height measurements through the centers of single polymeric strands yielded 1.1 ± 0.3 nm, 0.9 ± 0.2 nm, and 0.9 ± 0.2 nm for rCol-0G, rCol-4G, and the hCol III control, respectively, consistent with previously-reported values for collagen assembled in a triple helix [22, 38].

Thermostability scans were performed to determine apparent melting temperatures T_m (Figure 3-2b), and results for all variants are summarized in Table 3-2. Recombinant collagen III (with native human collagen III sequence; rCol) gave an average T_m of 35.8 ± 0.5°C, and all recombinant GFOGER variants yielded comparable T_m values. We note that all of the recombinant proteins consistently yielded T_m values approximately 3.5°C lower than human collagen from native tissue sources. This slightly lower stability has also been observed for fibrillar-type collagens expressed in other *S. cerevisiae* systems [39] and is likely due to the heterologous expression in yeast [12, 22]. Because yeast does not natively hydroxylate proline, which imparts stability to collagen, we have designed the strain to contain two copies of the α- and β-prolyl hydroxylase genes integrated

into the yeast genome (BYα2β2), with the collagen gene placed on a CEN/ARS plasmid (YCpMCOL) to facilitate variant manipulation [12, 22]. The resulting proline hydroxylation level of our yeast system using amino acid analysis yields 31.1 ± 2.4%, lower than the 45 ± 1.4% measured for native hCol III. Ongoing investigations continue to optimize the yeast expression system further to increase hydroxylation levels and thermostability of the collagen trimer.

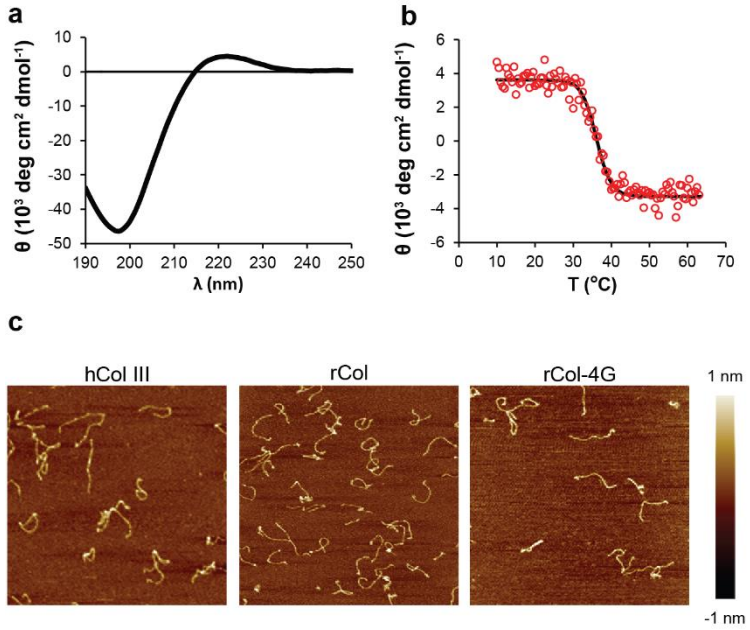


Figure 3-2: Structural characterization of collagen variant with 4 GFOGER sites introduced into a non-binding collagen scaffold (rCol-4G) and controls. (a) Representative circular dichroism (CD) wavelength spectra for rCol-4G. (b) Representative thermostability scan measuring ellipticity at 221 nm vs. temperature for rCol-4G. (c) Representative AFM images for native human collagen III (hCol III), recombinant human collagen (rCol), and rCol-4G. Scanned areas are 1 μm × 1 μm.

Table 3-2: Apparent melting temperature (T_m) of recombinant collagen baseline and mutants (from N=2 independent batches) and human collagen controls.

Collagen variant	T_m (°C)
human collagen III (native source)	39.1 ± 0.1
human collagen I (native source)	40.8 ± 0.1
rCol (human collagen III; recombinant source)	35.8 ± 0.5
rCol-0G	35.3 ± 0.5
rCol-1G-1	35.6 ± 0.3
rCol-1G-8	35.4 ± 0.1
rCol-1G-11	35.8 ± 0.4
rCol-2G-1,4	35.8 ± 0.5
rCol-2G-1,8	35.7 ± 0.3
rCol-2G-8,11	36.1 ± 0.4
rCol-4G	35.6 ± 0.8

The high tolerance for mutations in the collagen scaffold shows the robustness of the modular platform and the scaffold's structural backbone. The trimeric assembly and thermostability results were not obvious *a priori*, as these variants contained significantly more amino acid changes than had been reported previously. For example, rCol-0G and rCol-4G contained 19 and 40 amino acid changes from the native human collagen III sequence, respectively. In contrast, the maximum number of designed mutations (of which we are aware) was five amino acids in a localized cluster [40], and even single amino acid changes in the Gly-position of fibrillar collagen can cause severe phenotypes [41]. This demonstrates that the collagen backbone can indeed serve as an underlying scaffold for modular assembly, as long as the Gly-X-Y sequence pattern is retained and the sequence motifs that are substituted into the scaffold form stable trimers.

The retention of triple-helical structure after bioactive site incorporation is particularly important for the functional properties of the collagen-mimetic polymer, since integrin receptors for collagen simultaneously interact with multiple strands within intact triple-helical collagen [42]. This result also suggests that prior challenges to create extensively-mutated collagen *in vitro* was likely due to limitations of recombinant DNA technology, rather than the inherent triple-helical stability of collagen variants, and codon optimization for gene synthesis was necessary to minimize oligonucleotide mishybridizations [12]. It is unlikely that these triple-helical recombinant collagens

will form higher-order structures, such as fibrils. Telopeptides are necessary for fibril assembly [43], and our purification protocol uses pepsin to remove endogenous yeast proteins, which likely also cleaves the telopeptide region. Alternative avenues to produce collagen variants with intact telopeptides are currently being explored. Although fibril assembly to obtain higher-order three-dimensional structures is an important aspect of collagen's use as a tissue engineering hydrogel scaffold, hydrogel formation can also be accomplished with other strategies, such as utilizing introduced non-native cysteines [21].

3.3.3 Cellular adhesion to collagen III is abrogated upon removal of integrin binding sites.

Differences between the adhesion of HT-1080 cells to the variants of collagen were observed in cell microscopy images (Figures 3-3a – 3-3l), and relative cell numbers were quantified by calcein-AM (Figure 3-3m). As expected, cells adhered to native human collagen I and III controls from tissue sources (Figures 3-3l, 3-3k), but did not attach to bovine serum albumin (BSA) (Figure 3-3a). The baseline recombinant collagen (rCol), produced in yeast and comprising the same amino acid sequence as native human collagen III, bound as many cells as the native collagens ($p > 0.05$) (Figure 3-3j). Accessible surface concentrations of collagens were confirmed to be at saturated and comparable levels between all samples, so any observed differences in cell binding are due to molecular-level differences in the collagen substrates rather than surface protein concentrations (Figure A3-4).

Replacement of the five integrin sites from collagen III [27, 28] with the non-binding sequence GSPGGK [28] yielded the variant rCol-0G, on which no adhesion of HT-1080 cells was observed (Figure 3-3b). Since this collagen variant remained triple-helical and showed no loss of stability relative to the recombinant baseline protein (Figure A3-2 and Table 3-2), the lack of cell adhesion is specific to the loss of interactions between the integrin receptors on the cell surface and the hexameric peptide sequences within collagen. The variant rCol-0G, therefore, can serve as a non-binding collagen III scaffold, allowing for the evaluation of introduced integrin binding sites.

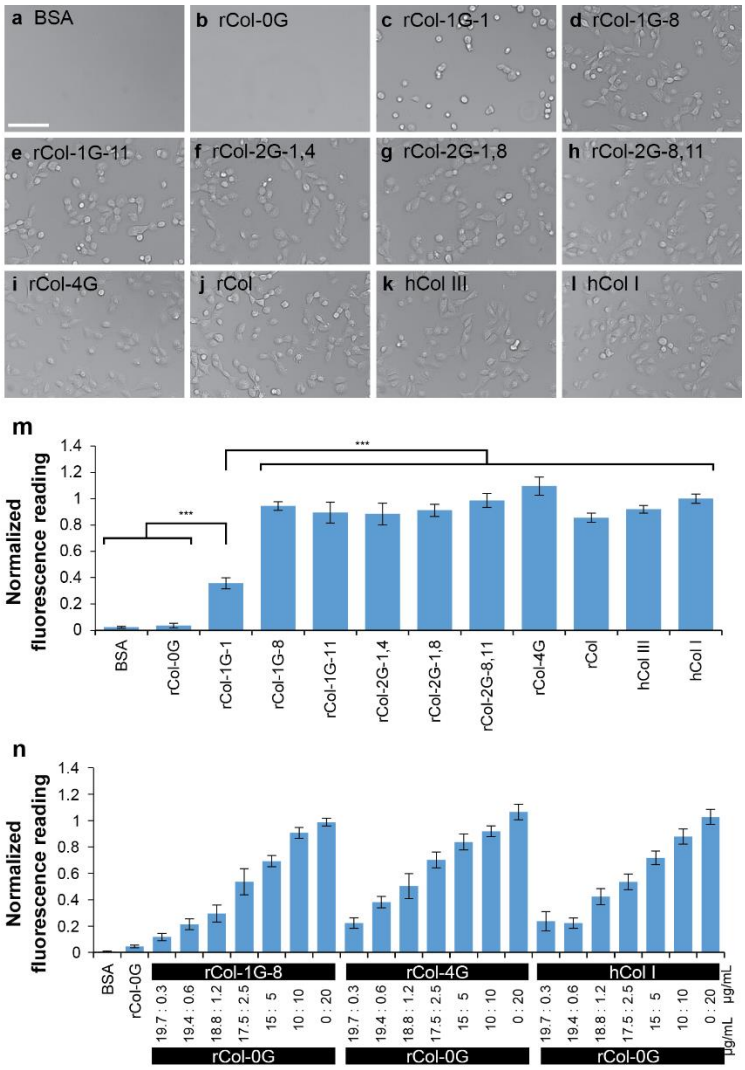


Figure 3-3: Adhesion of HT-1080 cells to collagen and collagen-mimetic substrates. (a-l) Representative cell images for adhesion to collagen variants and controls. Scale bar = 100 μm. Substrates are: (a) bovine serum albumin (BSA); (b) rCol-0G; (c) rCol-1G-1; (d) rCol-1G-8; (e) rCol-1G-11; (f) rCol-2G-1,4; (g) rCol-2G-1,8; (h) rCol-2G-8,11; (i) rCol-4G; (j) recombinant human collagen III (rCol); (k) native human collagen III (hCol III); and (l) native human collagen I (hCol I). (m) Fluorescence measurements quantifying the relative number of cells adhered to collagen substrates using a calcein-AM assay. Values are normalized to average for hCol I. Cell adhesion on BSA and rCol-0G are significantly lower than on all other variants ($p < 0.001$), and the adhesion on rCol-1G-1 is also significantly different than all other groups ($p < 0.001$). Comparisons between substrates showing high adhesion (substrates shown in panels 3d through 3l) indicate no significant differences ($p > 0.05$). ***$p < 0.001$. Mean ± S.E.M. (n) Fluorescence measurements quantifying the relative number of cells on surfaces coated with solutions of 0.3 – 20 μg/ml binding collagen (rCol-1G-8, rCol-4G, and hCol I) diluted in non-binding collagen (rCol-0G). Total protein concentrations (binding + non-binding collagens) are 20 μg/ml. Fluorescence values are normalized to average for 20 μg/ml hCol I. Mean ± S.E.M.

3.3.4 Introduction of the collagen I integrin binding site (GFOGER) at different frequencies and locations altered cellular adhesion.

Crystal structures of a GFOGER-containing peptide bound by α2β1 integrin showed that integrin interacts with two of the three triple-helical collagen strands [42, 44], thereby confirming that the triple-helical tertiary structure of collagen is critical for recognition. We observed that the introduction of only a single GFOGER motif into the ~1000-amino acid triple-helical region of the rCol-0G backbone recovered cellular adhesion (Figures 3-3c – 3-3e). Interestingly, we observed a marked difference in adhesion, depending on the location of the GFOGER placement site. Single-GFOGER substitutions, placed in primary fragment (PF) 8 or 11 (rCol-1G-8; rCol-1G-11) increased the number of adhered cells to values similar to those for native human collagen I and III (Figure 3-3m), and differences between cell spreading areas (per cell) for these two variants and native human collagen I and III were not significant ($p > 0.05$). Taken together, this demonstrates similar cell numbers and quality of attachment. That a single hexamer sequence from collagen I can promote this degree of adhesion was surprising, considering that native collagens contain not only one but multiple integrin binding sites. In contrast, cell adhesion to collagen with a single-GFOGER introduced into PF1 (rCol-1G-1) was weaker than on other GFOGER-containing collagen substrates (rCol, rCol-1G-8, rCol-1G-11, rCol-2G variants, rCol-4G) and native controls, with only approximately one-third the number of cells attached to the rCol-1G-1 substrate (Figure 3-3m). Furthermore, cells on rCol-1G-1 did not spread as well as on the other surfaces (Figure 3-3c), with the average surface area per cell on rCol-1G-1 being 50% and 54% of areas on hCol I and hCol III, respectively ($p < 0.001$).

The marked difference in binding between the rCol-1G mutation locations (in PF 1, 8, or 11) demonstrates the importance of the local collagen context for integrin-binding interactions. The results of the AFM analysis (Figure 3-2d) support a possible mechanism for this difference. Although images show that all recombinant collagens are linear and triple-helical, the apparent lengths for the imaged GFOGER mutants (224 ± 12 nm for rCol-0G, 210 ± 21 nm for rCol-1G-1, and 235 ± 27 nm for

rCol-4G) were 10% shorter than for the controls (256 ± 10 nm for native hCol III and 251 ± 13 nm for rCol). Closer examination of individual GFOGER-variant trimers indicate that one end of the trimeric polymer often has a greater height or aggregate for a length ranging approximately 15-30 nm. This suggests that the decreased overall length could be due to one end of the GFOGER-collagens folding back on itself. If the folding propensity is in the N-terminus, then this could occlude GFOGER binding in PF1. Our data does not support the alternative scenario of proteolytic truncation; SDS-PAGE analysis shows all collagens in this investigation exhibit comparable molecular weights (Figure A3-1). Furthermore, trimer misalignment as a mechanism for decreased adhesion in rCol-1G-1 is also unlikely, since the N-and C-terminal propeptide domains are expressed and therefore present to promote correct triple-helical assembly before their cleavage during purification.

3.3.5 Modulation of adhesion could be achieved by mixing different ratios of the binding and non-binding variants.

Binding variants rCol-1G-8 and rCol-4G were diluted with rCol-0G, producing various concentrations of the GFOGER-containing proteins while keeping total protein amount at 20 μg/ml. We observed that the number of adhered cells could be tuned as rCol-1G-8 or rCol-4G was increased between 0.3 and 20 ug/ml (Figure 3-3n). Variant rCol-4G, containing four GFOGERs, did not significantly bind additional cells over the collagen with only one GFOGER (rCol-1G-8), hinting that the distance between GFOGER sites (intra-trimeric vs. inter-trimeric GFOGERs) may also be as important as the number of sites.

3.3.6 Inhibition of cell adhesion by anti-α2 and anti-β1 integrin antibodies.

We examined the subunit specificity of integrin-based cell adherence to GFOGER. Integrins that bind to collagens are α1β1, α2β1, α10β1, and α11β1 [24, 29], with the most broadly expressed and well-studied being α1β1 and α2β1 which are involved in biological functions that include immunity, angiogenesis, and matrix remodeling [45]. In HT-1080 cells, the primary integrin that binds to triple-helical collagen (including the GFOGER sequence) is α2β1 [15, 46]. We therefore

evaluated the specificity of cell binding to our GFOGER variants by inhibiting interactions with the α2 and β1 subunits using antibodies.

Antibodies for the β1 subunit strongly inhibited cell adhesion to all modular-GFOGER substrates (Figure 3-4). All of the native human collagen controls and recombinant collagen variants containing GFOGER sites exhibited a significant decrease in cell adhesion when cells were first incubated with an antibody specific for the β1 subunit, relative to media-alone. Overall, cell adhesion was reduced by an average 86% with anti-β1 antibody for the variants which had demonstrated binding. This result is consistent with investigations showing that all collagen-binding integrins contain the β1 subunit [24].

Incubation of cells with α2-binding antibodies only partially inhibited cell adhesion to the GFOGER-containing substrates, with an average 38% decrease for the variants that showed binding without inhibition, and the degree of inhibition varied with the underlying substrate. Incubation of cells with antibody concentrations higher than that of 10 μg/ml yielded no additional decrease in cellular adhesion. In rCol-1G-1, which exhibited weaker binding compared to native collagens, the α2 antibody inhibited nearly all binding. However, for variants demonstrating stronger binding (*e.g.,* rCol-2G-8,11; rCol-4G), cell binding was not significantly inhibited by anti-α2, suggesting that the increased number of GFOGER sites can sufficiently bind to non-α2β1 collagen-binding integrins (*e.g.*, α1β1, α10β1, α11β1). This partial inhibition agrees with prior results for native collagen I [15] and is consistent with α2β1 being the primary, but not sole, integrin receptor expressed by HT-1080 cells.

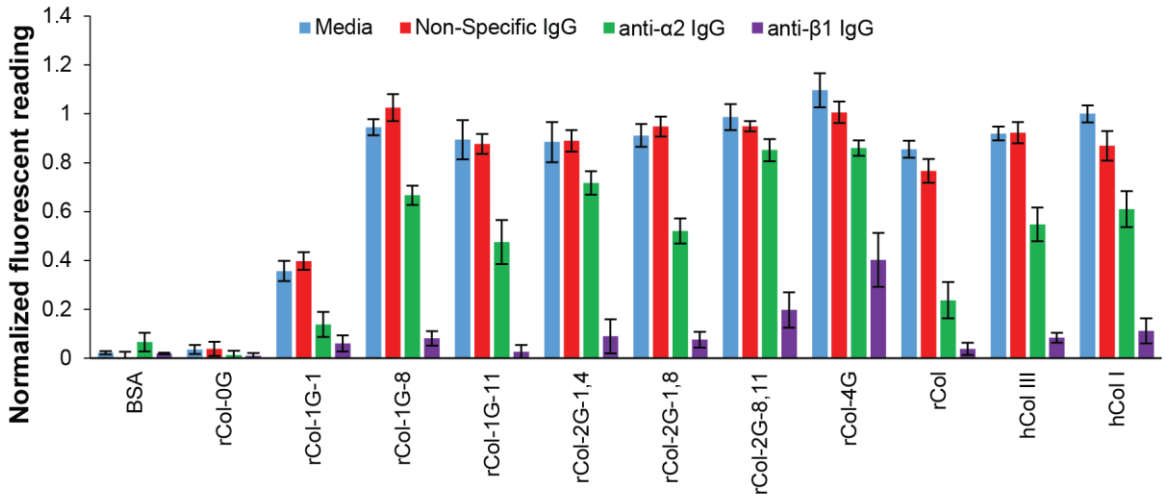


Figure 3-4: Cell adhesion after inhibition of binding to α2 and β1 integrin subunits. HT-1080 cells were incubated with no antibodies (media alone), non-specific IgG, anti-α2 IgG, and anti-β1 IgG prior to incubation with different collagen substrates. Data quantifies the relative number of cells adhered using a calcein-AM assay. Fluorescence values are normalized to average for hCol I with no antibodies. A significant decrease in adhesion due to anti-α2 antibody, relative to media-only conditions, is observed for hCol I, hCol III, rCol, rCol-1G-8, rCol-2G-1,8 ($p < 0.001$) and rCol-1G-1, rCol-1G-11 ($p < 0.01$). A significant decrease in adhesion due to anti-β1 antibody, relative to media-only conditions, is observed for hCol I, hCol III, rCol, rCol-1G-1, rCol-1G-8, rCol-1G-11, rCol-2G-1,4, rCol-2G-1,8, rCol-2G-8,11, and rCol-4G ($p < 0.001$). Significant differences between α2- and β1-subunit antibody inhibition for a given substrate are observed for hCol I, hCol III, rCol-1G-8, rCol-1G-11, rCol-2G-1,4, rCol-2G-1,8, rCol-2G-8,11 ($p < 0.001$) and rCol-4G ($p < 0.01$). Mean ± S.E.M.

For studies utilizing short GFOGER-peptide substrates, however, the addition of α2-binding antibodies to cells completely blocked all cell binding [15, 47]. This apparent discrepancy suggests the importance of the full-length collagen in its native structure for integrin-mediated adhesion. This importance of the context of epitope presentation is also supported by prior work in which 30 times more copies of GFOGER was needed when the motif was surface-adsorbed and displayed in a peptide than when presented in full-length collagen I (by molar concentrations) [47]. While other recognition sites such as GLOGER and GASGER in collagen I have also been reported as binding to α2β1 integrin [48], each site only occurs once in human collagen I [49] and at similar or weaker binding than GFOGER [27]. Therefore, their presence alone does not account for the significantly higher binding of full-length collagen I relative to a GFOGER-peptide, and suggests that the full structure of collagen in presenting the sequence plays a significant role in cell adhesion.

One possible explanation for the increased adhesion to full-length collagen relative to peptide may be that collagen provides numerous binding domains for other cell surface receptors, some of which support cooperative and synergistic interactions with integrin-mediated effects [21]. For example, several discoidin domain receptor (DDR) binding domains have been identified in human collagen III near the integrin binding domains [21], and DDRs have been shown to enhance integrin-mediated cell adhesion [50]. Such effects are not recapitulated in systems that use only GFOGER peptides. Synergy between neighboring sequences has been reported in other protein systems. For example, cell binding to the RGD motif from fibronectin is significantly increased when presented concurrently with fibronectin's PHSRN peptide sequence in native spatial context [51, 52]. Furthermore, different integrin binding specificities are elicited for linear RGD and RGD-PHSRN peptides. Understanding the mechanisms of synergistic epitope effects can be important for regulating cell proliferation and differentiation [21, 52, 53].

Another factor which could potentially contribute to the observed increased binding to full-length collagen is the accessibility of the binding motifs. Immobilization of a GFOGER peptide utilizing a PEO linker required approximately three times less peptide than peptide adsorbed to the surface (for the same number of cells adhered), suggesting that increased accessibility of GFOGER due to the linker could be important for adhesion [47]. We may also be seeing this effect for full length collagen if only short regions of the collagen are required for adsorption to the surface. The total increased adhesion is likely a combination of multiple effects which includes both site accessibility and synergistic receptor interactions.

3.3.7 The number of focal adhesion complexes is modulated by GFOGER-collagen substrates.

We examined the presence of vinculin, which is present in focal adhesion complexes, in cells seeded on varying collagen substrates [54, 55]. Our results demonstrate that the number of GFOGER sites on an external substrate can affect intracellular processes (Figure 3-5). In Western blot analysis, the average vinculin band intensity for rCol-4G is greater than those for all other samples and

controls, with cells seeded on rCol-4G exhibiting approximately 56% higher levels of vinculin than on the native hCol I. Furthermore, cells incubated with the collagen variant containing no integrin binding sites (rCol-0G) shows significantly less vinculin than with hCol I, with intensities averaging only 4% of native and close to that of the negative BSA control. The values for variant rCol-1G-1 averaged an intermediate intensity that was 12% of hCol I. These results are consistent with the cell adhesion data and show the potential to manipulate intracellular biochemical activity through integrin binding sites in the underlying substrates [13].

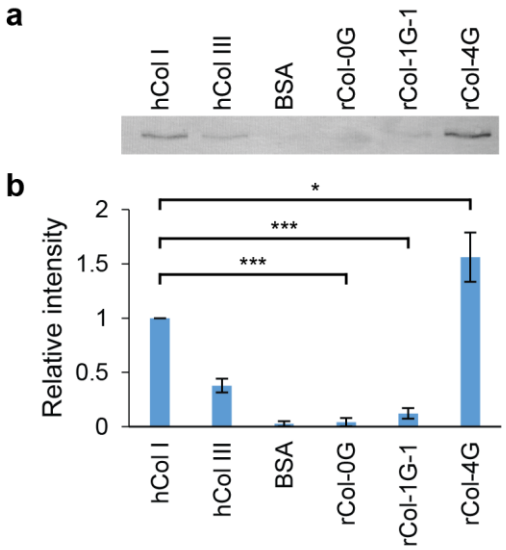


Figure 3-5: Western blot analyses for vinculin in focal adhesion complexes. (a) Representative Western blot. (b) Band intensities, with values normalized to hCol I in each independent experiment. Intensity for rCol-4G is greater than those for all other samples and controls (with $p = 0.018$ compared to hCol I). $^{*}p < 0.05$, $^{***}p < 0.001$. Mean ± S.E.M.

3.3.8 HNSPCs adhesion is β1 mediated but not α2 mediated

Previous tests utilized the HT1080 human fibrosarcoma cell line. In order to test a more clinically relevant cell line, we investigated the effects of GFOGER on hNSPCs. Initial tests showed adhesion of hNSPCs to collagen I and III (Figure 3-6b,c). We also tested adhesion onto collagen variants containing 0 to 4 GFOGER sites. Little to no adhesion was observed on rCol-0G with the cells adopting circular morphology (Figure 3-6h). On variants containing GFOGER sites, adhesion was observed with cells spreading across the surface (Figure 3-6b,i). GFOGER is often mentioned as an

α2 β1 integrin binding motif; however, RNAseq and flow cytometry data showed low α2 integrin transcript number or expression (<2% cells displayed α2 integrin) [56, 57]. HNSPCs express high numbers of β1, α3 and α7 integrin which are laminin and fibronectin binding [57]. Integrin α3β1 has also been implicated with collagen I binding; however, the amino acid sequence which it recognizes has not been determined [23, 58]. To determine which integrins were responsible for adhesion, we incubated the hNSPCs with antibodies to block specific integrins before seeding the cells onto the protein coated surfaces.

Adhesion of hNSPCs to collagen was inhibited by β1 integrin blocking antibodies (Figure 3-6e, f, j-l). Cells incubated with anti-integrin β1 antibodies were unable to adhere to human collagen I and III, rCol, and its GFOGER variants. Adhesion of hNSPCs to laminin was not inhibited (Figure 3-6d) as adhesion could be achieved through binding of β4 integrin [57, 58].

After determining that hNSPCs are adhering to the collagens through β1 integrin, we continued to test α integrin blocking antibodies to determine which heterodimer pairs were responsible for adhesion. Adhesion of hNSPCs to collagen surfaces was not inhibited by α2, α3, nor α7 integrin blocking antibodies. Incubation of α1 integrin blocking antibodies reduced cell binding to collagen surfaces compared to a no treatment control (Figure 3-6k). The integrin binding motif GFOGER is normally represented as a high affinity α2β1 integrin binding site, but, α1 β1 is also able to bind to this sequence. It was surprising to see that incubation of hNSPCs with anti-integrin α3 antibodies did not reduce adhesion as α3β1 integrin was previously identified as a collagen I binding pair and both integrins are highly expressed on hNSPCs [57-59]. Furthermore, previously published flow cytometry data showed that less than 2% of the hNSPCs displayed integrin α1 or α2 [57]. We observed much greater than 2% adhesion of hNSPCs to the surfaces compared to laminin and fibronectin although there were significantly less cells bound to the collagen surfaces.

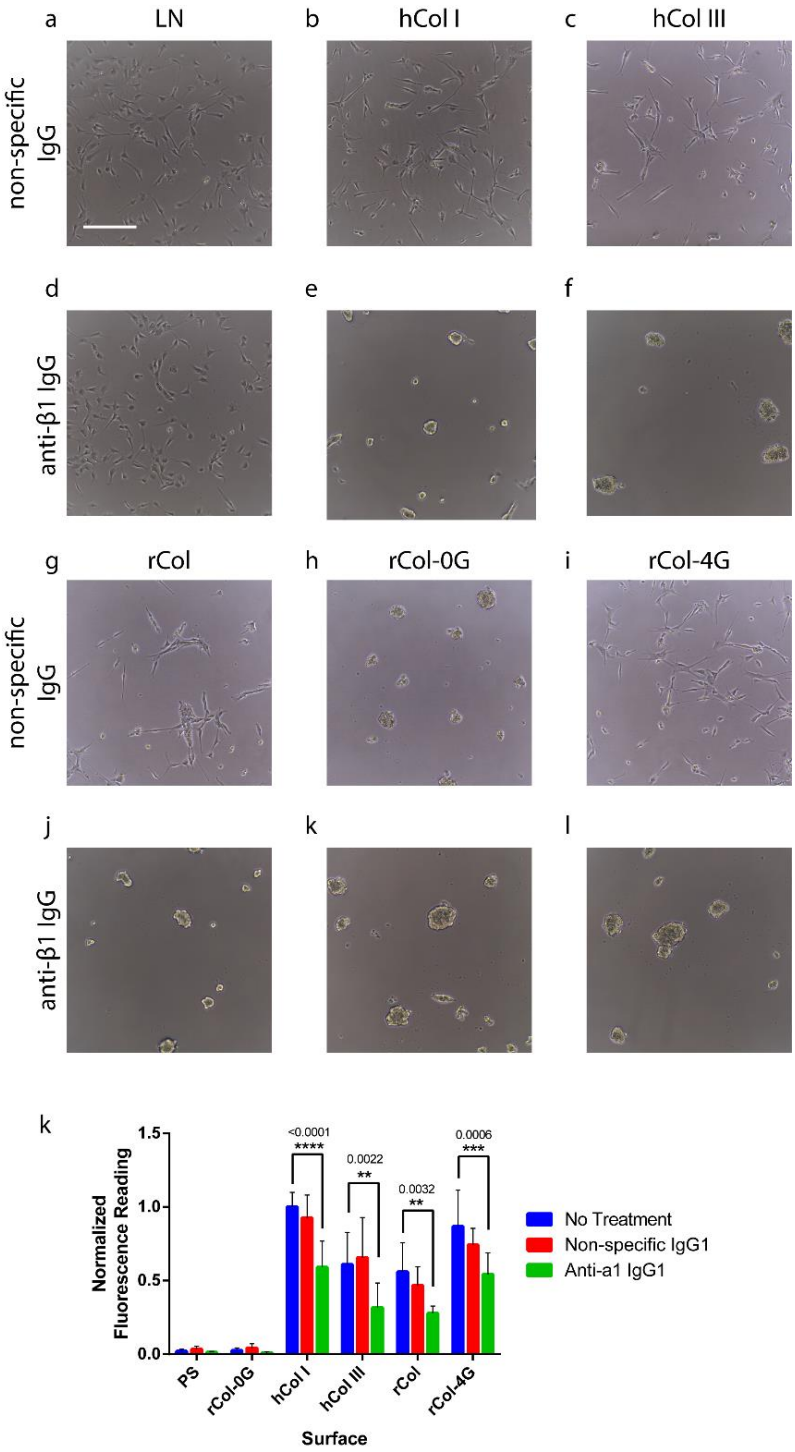


Figure 3-6: Adhesion of hNSPCs onto collagen variants and controls. HNSPCs adhered to (a) laminin, (b) hCol I, (c) hCol III, (f) rCol, and (h) rCol-4G. (g) rCol-0G did no support hNSPC adhesion and the cells formed clusters, preferring to adhere to one another instead of to the protein. Blocking β1 integrin prior to seeding abrogated cell adhesion to the collagen surfaces (e, f, j-l) but had no effect on laminin (d). (k) quantification of hNSPCs adhered to the surfaces with media only, non-specific IgG, and anti-α1 IgG. Incubation with anti-α1 IgG resulted in significant decrease in the number of cells adhered vs no treatment for each surface. Scale bar is 200 μm.

3.3.9 HNSPC proliferation and differentiation on collagen with varying GFOGER

With the verification that hNSPCs were able to adhere to the collagen surfaces and to further investigate the effects of GFOGER on adhered cells, we quantified the ability of hNSPCs to proliferate and differentiate on the GFOGER variants. Cells that are not in the G_0 phase of the cell cycle express Ki67 while Phospohohistone H3 (PH3) is present when cells are undergoing mitosis. Ki67 is a nuclear protein which is associated with cellular proliferation [60]. Phosphorylation of histone H3, is tightly correlated with chromosome condensation during mitosis [61]. We observed similar percentages of proliferating cells on the collagen surfaces as with laminin control (Figure 3-7). As observed during the adhesion assays, less cells were adhered to the collagen surfaces. However, the fact that the same percentage of cells were proliferating shows that the cells that are able to adhere are viable and able to proliferate similarly to laminin. The collagen surfaces were also able to support differentiation of hNSPCs to neurons and astrocytes as cells were stained with neuronal marker MAP2 and astrocyte marker GFAP (Figure 3-8). The support of differentiation is a critical attribute of any hNSPC substrate.

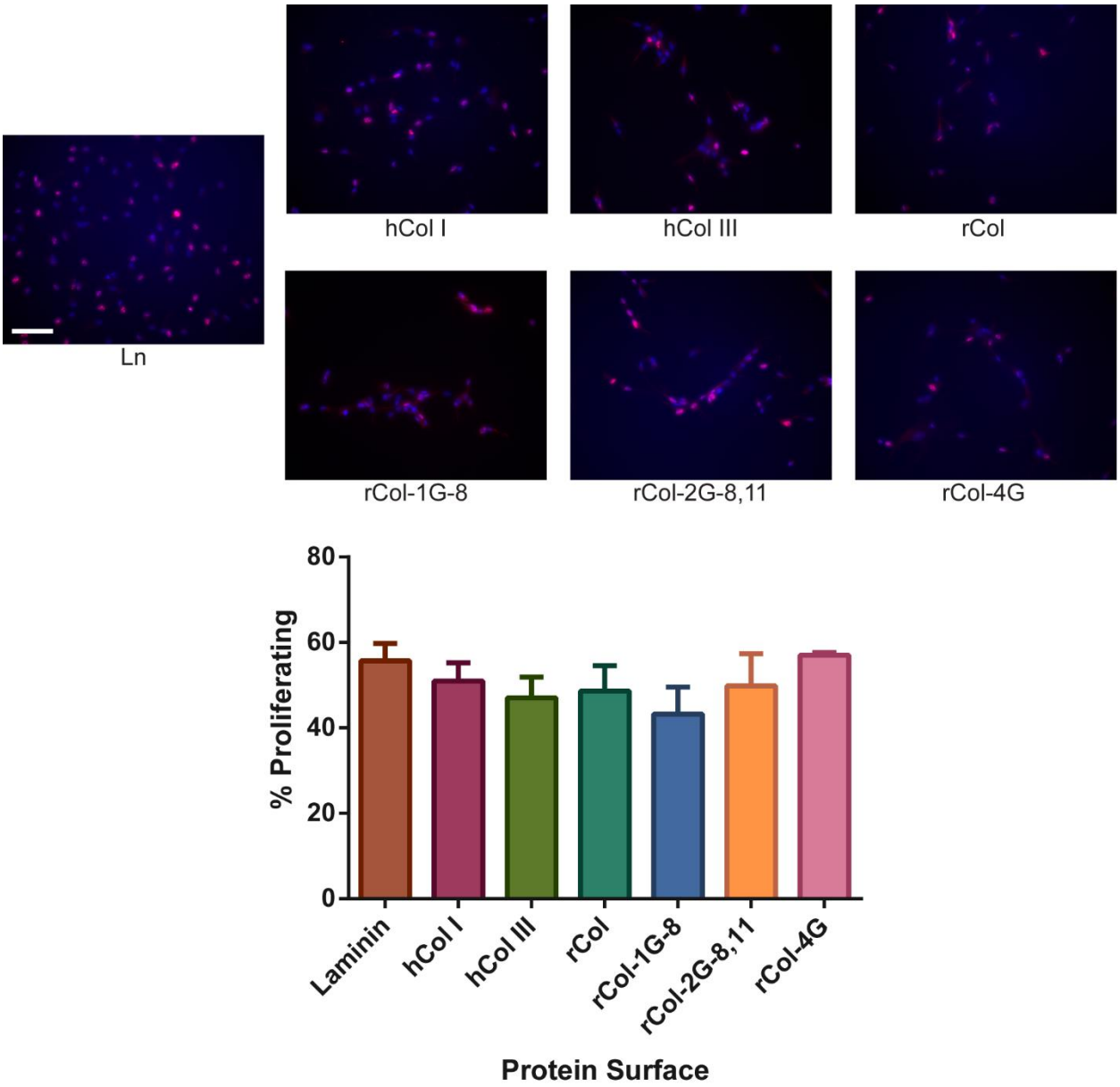


Figure 3-7: Proliferation of hNSPCs on collagen variants and controls. HNSPCs were cultured in proliferation media for 1 day and stained to assess proliferation using the marker Ki67 (red). Cell nuclei were stained with Hoechst (blue). Cells differentiated into both lineages for all surfaces. Cells did appear more clumped on collagen surfaces compared to laminin. Scale bar is 100 μm.

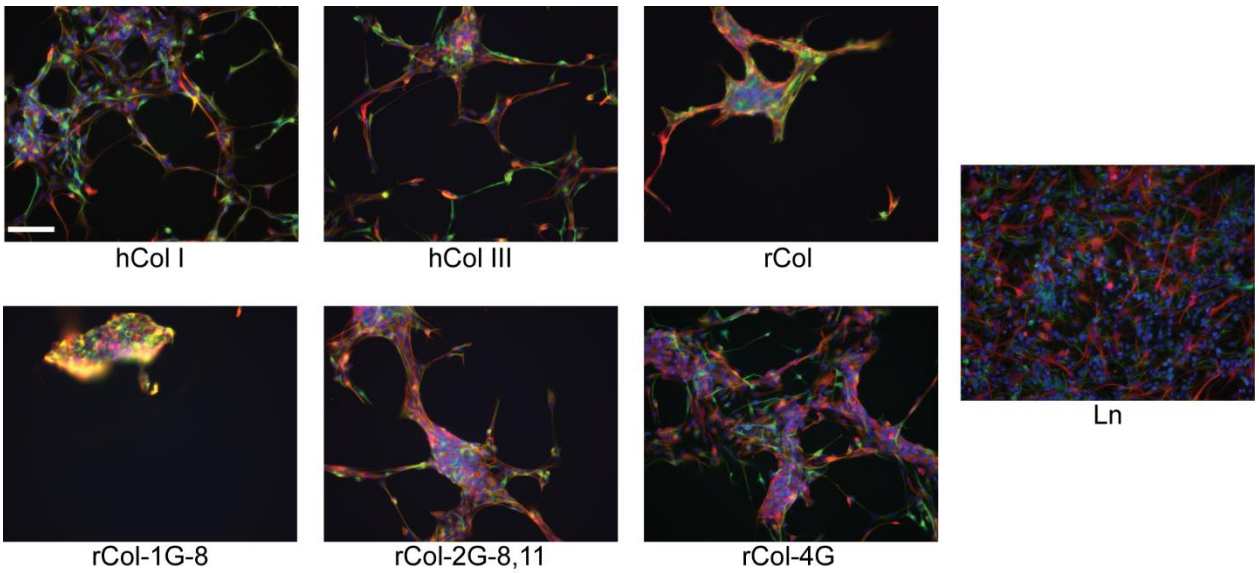


Figure 3-8: Differentiation of hNSPCs on collagen variants and controls. HNSPCs were cultured in differentiation media for 14 days and stained to assess differentiation into neurons using the markers MAP2 (red) and into astrocytes using the marker GFAP (green). Cell nuclei were stained with Hoechst (blue). Cells differentiated into both lineages for all surfaces. Cells did appear more clumped on collagen surfaces compared to laminin. Scale bar is 100 μm.

3.4 Conclusion

In this study, we used a bottom-up, modular strategy that enabled systematic placement of integrin binding sites in full-length collagen and allowed for the complete control of the number and location of these sites. Using the biopolymers created from this strategy, we found that the insertion of only a single GFOGER site into a non-binding collagen variant was sufficient to restore full cellular adhesion. However, the location of the introduced site affected the degree of recovered adhesion and hints at the importance of epitope context. Cellular activity could also be modulated by mixing distinctly different collagen variants at tunable ratios. Adhesion, proliferation, and differentiation of hNSPCs on the collagen variants was also achievable, making it a potential substrate for hNSPCs. All variants were shown to be triple-helical, linear polymers, comparable to native human collagen. This investigation shows that one can re-design native collagen at the molecular level to tailor collagen-mimetic materials for manipulating cellular microenvironments and biochemical responses.

3.5 References

[1] M. P. Lutolf, Hubbell J. A. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. Nat Biotech. 2005;23:47-55.

[2] D. E. Discher, Mooney D. J., Zandstra P. W. Growth Factors, Matrices, and Forces Combine and Control Stem Cells. Science. 2009;324:1673-7.

[3] K. von der Mark, Park J., Bauer S., Schmuki P. Nanoscale engineering of biomimetic surfaces: cues from the extracellular matrix. Cell Tissue Res. 2010;339:131-53.

[4] M. Uemura, Refaat M. M., Shinoyama M., Hayashi H., Hashimoto N., Takahashi J. Matrigel supports survival and neuronal differentiation of grafted embryonic stem cell-derived neural precursor cells. Journal of Neuroscience Research. 2010;88:542-51.

[5] A. k. Lynn, Yannas I. v., Bonfield W. Antigenicity and immunogenicity of collagen. Journal of Biomedical Materials Research Part B: Applied Biomaterials. 2004;71B:343-54.

[6] M. P. Lutolf, Weber F. E., Schmoekel H. G., Schense J. C., Kohler T., Muller R., et al. Repair of bone defects using synthetic mimetics of collagenous extracellular matrices. Nature Biotechnology. 2003;21:513-8.

[7] L. Perlin, MacNeil S., Rimmer S. Production and performance of biomaterials containing RGD peptides. Soft Matter. 2008;4:2331.

[8] X. Q. Jia, Kiick K. L. Hybrid Multicomponent Hydrogels for Tissue Engineering. Macromolecular Bioscience. 2009;9:140-56.

[9] R. L. DiMarco, Heilshorn S. C. Multifunctional Materials through Modular Protein Engineering. Advanced Materials. 2012;24:3923-40.

[10] L. C. Abraham, Zuena E., Perez-Ramirez B., Kaplan D. L. Guide to collagen characterization for biomaterial studies. Journal of Biomedical Materials Research Part B: Applied Biomaterials. 2008;87B:264-85.

[11] D. E. Birk, Silver F. H., Trelstad R. L. Matrix Assembly. In: Hay ED, editor. Cell Biology of Extracellular Matrix. second ed. New York: Plenum Press; 1991. p. 221-54.

[12] S. W. P. Chan, Hung S. P., Raman S. K., Hatfield G. W., Lathrop R. H., Da Silva N. A., et al. Recombinant Human Collagen and Biomimetic Variants Using a De Novo Gene Optimized for Modular Assembly. Biomacromolecules. 2010;11:1460-9.

[13] C. D. Reyes, Garcia A. J. Alpha2beta1 integrin-specific collagen-mimetic surfaces supporting osteoblastic differentiation. J Biomed Mater Res A. 2004;69:591-600.

[14] F. W. Kotch, Raines R. T. Self-assembly of synthetic collagen triple helices. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:3028-33.

[15] C. M. Yamazaki, Kadoya Y., Hozumi K., Okano-Kosugi H., Asada S., Kitagawa K., et al. A collagen-mimetic triple helical supramolecule that evokes integrin-dependent cell responses. Biomaterials. 2010;31:1925-34.

[16] O. D. Krishna, Jha A. K., Jia X., Kiick K. L. Integrin-mediated adhesion and proliferation of human MSCs elicited by a hydroxyproline-lacking, collagen-like peptide. Biomaterials. 2011;32:6412-24.

[17] V. Hernandez-Gordillo, Chmielewski J. Mimicking the extracellular matrix with functionalized, metal-assembled collagen peptide scaffolds. Biomaterials. 2014;35:7363-73.

[18] Y. Y. Peng, Yoshizumi A., Danon S. J., Glattauer V., Prokopenko O., Mirochnitchenko O., et al. A Streptococcus pyogenes derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. Biomaterials. 2010;31:2755-61.

[19] B. An, Kaplan D. L., Brodsky B. Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. Frontiers in Chemistry. 2014;2.

[20] I. Majsterek, McAdams E., Adachi E., Dhume S. T., Fertala A. Prospects and limitations of the rational engineering of fibrillar collagens. Protein Science. 2003;12:2063-72.

[21] R. Que, Mohraz A., Da Silva N. A., Wang S. W. Expanding functionality of recombinant human collagen through engineered non-native cysteines. Biomacromolecules. 2014;15:3540-9.

[22] A. Jabaiah, Wang X., Raman S. K., Ragan R., Da Silva N. A., Wang S. W. Nanoscale architecture and cellular adhesion of biomimetic collagen substrates. Journal of biomaterials applications. 2014;28:1354-65.

[23] R. O. Hynes. Integrins - Versatility, Modulation, and Signaling in Cell-Adhesion. Cell. 1992;69:11-25.

[24] M. Barczyk, Carracedo S., Gullberg D. Integrins. Cell Tissue Res. 2010;339:269-80.

[25] J. R. García, Clark A. Y., García A. J. Integrin-specific hydrogels functionalized with VEGF for vascularization and bone regeneration of critical-size bone defects. Journal of Biomedical Materials Research Part A. 2016;104:889-900.

[26] K. Saha, Irwin E. F., Kozhukh J., Schaffer D. V., Healy K. E. Biomimetic interfacial interpenetrating polymer networks control neural stem cell behavior. Journal of Biomedical Materials Research Part A. 2007;81A:240-9.

[27] J. K. Kim, Xu Y., Xu X., Keene D. R., Gurusiddappa S., Liang X., et al. A novel binding site in collagen type III for integrins alpha1beta1 and alpha2beta1. J Biol Chem. 2005;280:32512-20.

[28] N. Raynal, Hamaia S. W., Siljander P. R. M., Maddox B., Peachey A. R., Fernandez R., et al. Use of synthetic peptides to locate novel integrin alpha(2)beta(1)-binding motifs in human collagen III. Journal of Biological Chemistry. 2006;281:3821-31.

[29] C. G. Knight, Morton L. F., Peachey A. R., Tuckwell D. S., Farndale R. W., Barnes M. J. The collagen-binding A-domains of integrins alpha(1)beta(1) and alpha(2)beta(1) recognize the same

specific amino acid sequence, GFOGER, in native (triple-helical) collagens. Journal of Biological Chemistry. 2000;275:35-40.

[30] A. V. Persikov, Ramshaw J. A. M., Kirkpatrick A., Brodsky B. Amino Acid Propensities for the Collagen Triple-Helix†. Biochemistry. 2000;39:14960-7.

[31] L. S. Z. Larsen, Wassman C. D., Hatfield G. W., Lathrop R. H. Computationally Optimised DNA Assembly of synthetic genes. International Journal of Bioinformatics Research and Applications. 2008;4:324-36.

[32] H. Liu, Naismith J. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnology. 2008;8:91.

[33] D. Ren, Dalmau M., Randall A., Shindel M. M., Baldi P., Wang S. W. Biomimetic Design of Protein Nanomaterials for Hydrophobic Molecular Transport. Adv Funct Mater. 2012;22:3170-80.

[34] N. J. Greenfield. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc. 2006;1:2527-35.

[35] N. Gallant, García A. Quantitative Analyses of Cell Adhesion Strength. In: Coutts A, editor. Adhesion Protein Protocols: Humana Press; 2007. p. 83-95.

[36] L. A. Flanagan, Lu J., Wang L., Marchenko S. A., Jeon N. L., Lee A. P., et al. Unique Dielectric Properties Distinguish Stem Cells and Their Differentiated Progeny. STEM CELLS. 2008;26:656-65.

[37] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria2013.

[38] J. Bella, Eaton M., Brodsky B., Berman H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. Science. 1994;266:75-81.

[39] D. R. Olsen, Leigh S. D., Chang R., McMullin H., Ong W., Tai E., et al. Production of Human Type I Collagen in Yeast Reveals Unexpected New Insights into the Molecular Assembly of Collagen Trimers. Journal of Biological Chemistry. 2001;276:24038-43.

[40] K. E. Williams, Olsen D. R. Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen. Matrix Biology. 2009;28:373-9.

[41] H. Kuivaniemi, Tromp G., Prockop D. J. Mutations in Collagen Genes - Causes of Rare and Some Common Diseases in Humans. Faseb J. 1991;5:2052-60.

[42] J. Emsley, Knight C. G., Farndale R. W., Barnes M. J. Structure of the integrin alpha2beta1-binding collagen peptide. J Mol Biol. 2004;335:1019-28.

[43] K. Kadler, Holmes D., Trotter J., Chapman J. Collagen fibril formation. Biochem J. 1996;316:1-11.

[44] J. Emsley, Knight C. G., Farndale R. W., Barnes M. J., Liddington R. C. Structural basis of collagen recognition by integrin alpha2beta1. Cell. 2000;101:47-56.

[45] J. Heino. The collagen family members as cell adhesion proteins. Bioessays. 2007;29:1001-10.

[46] M. Pfaff, Aumailley M., Specks U., Knolle J., Zerwes H. G., Timpl R. Integrin and Arg Gly Asp Dependence of Cell-Adhesion to the Native and Unfolded Triple-Helix of Collagen Type-Vi. Experimental Cell Research. 1993;206:167-76.

[47] C. D. Reyes, Garcia A. J. Engineering integrin-specific surfaces with a triple-helical collagen-mimetic peptide. Journal of Biomedical Materials Research Part A. 2003;65A:511-23.

[48] P. R. Siljander, Hamaia S., Peachey A. R., Slatter D. A., Smethurst P. A., Ouwehand W. H., et al. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. J Biol Chem. 2004;279:47763-72.

[49] G. Tromp, Kuivaniemi H., Stacey A., Shikata H., Baldwin C. T., Jaenisch R., et al. Structure of a full-length cDNA clone for the prepro alpha 1(I) chain of human type I procollagen. Biochem J. 1988;253:919-22.

[50] Y. K. Kim, Que R., Wang S. W., Liu W. F. Modification of biomaterials with a self-protein inhibits the macrophage response. Adv Healthc Mater. 2014;3:989-94.

[51] S. Aota, Nomizu M., Yamada K. M. The short amino acid sequence Pro-His-Ser-Arg-Asn in human fibronectin enhances cell-adhesive function. Journal of Biological Chemistry. 1994;269:24756-61.

[52] T. A. Petrie, Capadona J. R., Reyes C. D., García A. J. Integrin specificity and enhanced cellular activities associated with surfaces presenting a recombinant fibronectin fragment compared to RGD supports. Biomaterials. 2006;27:5459-70.

[53] S.-L. Cheng, Lai C.-F., Blystone S. D., Avioli L. V. Bone Mineralization and Osteoblast Differentiation Are Negatively Modulated by Integrin αvβ3. Journal of Bone and Mineral Research. 2001;16:277-88.

[54] J. D. Humphries, Wang P., Streuli C., Geiger B., Humphries M. J., Ballestrem C. Vinculin controls focal adhesion formation by direct interactions with talin and actin. The Journal of Cell Biology. 2007;179:1043-57.

[55] E. Zamir, Geiger B. Molecular complexity and dynamics of cell-matrix adhesions. Journal of Cell Science. 2001;114:3583-90.

[56] J. Arulmoli, Wright H. J., Phan D. T. T., Sheth U., Que R. A., Botten G. A., et al. Combination scaffolds of salmon fibrin, hyaluronic acid, and laminin for human neural stem cell and vascular tissue engineering. Acta Biomaterialia. 2016;43:122-38.

[57] L. A. Flanagan, Rebaza L. M., Derzic S., Schwartz P. H., Monuki E. S. Regulation of human neural precursor cells by laminin and integrins. Journal of Neuroscience Research. 2006;83:845-56.

[58] R. Milner, Campbell I. The integrin family of cell adhesion molecules has multiple functions within the CNS. Journal of neuroscience research. 2002;69:286-91.

[59] E. A. Wayner, Carter W. G. Identification of multiple cell adhesion receptors for collagen and fibronectin in human fibrosarcoma cells possessing unique alpha and common beta subunits. The Journal of Cell Biology. 1987;105:1873-84.

[60] T. Scholzen, Gerdes J. The Ki-67 protein: From the known and the unknown. Journal of Cellular Physiology. 2000;182:311-22.

[61] M. J. Hendzel, Wei Y., Mancini M. A., Van Hooser A., Ranalli T., Brinkley B. R., et al. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. Chromosoma. 1997;106:348-60.

Chapter 4: Combination Variants: Non-Native Cysteines and GFOGER

4.1 Introduction

In Chapter 3, we demonstrated the ability to introduce the integrin binding motif GFOGER from collagen I into a non-cell adhering collagen III [1]. However, the study was performed only in a two-dimensional context on stiff polystyrene. Most cells natively exist within a three-dimensional environment, and constricting their growth to two-dimensions provides unnatural geometric cues.

Petersen et al. showed that nonmalignant and malignant cells were phenotypically similar in 2D; however, in 3D the nonmalignant cells became growth-arrested organized structures while malignant cells formed disorganized cellular masses [2]. Studies comparing 2D and 3D studies highlight the benefits of 3D models such as providing more physiological distribution of mechanical inputs and cell signaling [3,4]. Endothelial cells when cultured in 2D, are able to migrate, proliferate, and form cordlike structures. However, in 3D cultures, they assemble to form lumen which better demonstrates properties necessary for vasculogenesis *in vivo* [5]. In order to test the effects of GFOGER in a 3D model, we combined the functionality of non-native cysteines and GFOGER into recombinant collagen.

The hexameric peptide sequence GFOGER is implicated as necessary in the formation of vasculature *in vitro* on collagen I gels in 2D [6]. Blocking of α2 integrin, of which GFOGER is a high affinity binding motif, with function-blocking antibodies reduced lumen formation in collagen gels [7]. Further studies utilizing short triple-helical peptides have shown that hydrogels modified with GFOGER peptide leads to larger vascular volumes and number of vessels when compared to RGD functionalized hydrogels without the presence of vascular endothelial growth factor (VEGF) [8]. RGD-functionalized hydrogels containing VEGF yielded vascular volumes similar to those of VEGF-free GFOGER-functionalized hydrogels while VEGF-containing GFOGER-functionalized gels yielded significantly higher vascular volume. However, this study utilized GFOGER in the context of short peptides. Compared against GFOGER in the context of short peptides, significantly less GFOGER in the context of full-length collagen, was necessary to provide full adhesion [1] (Chapter 3). With this

in mind, we were interested to test whether GFOGER in the context of full-length collagen would result in similar or more robust vascularization in 3D.

4.2 Materials and Methods

4.2.1 Cells and reagents

Escherichia coli DH5α was purchased from Strategene. Restriction endonucleases, DNase I, and RNase I_f was acquired from New England Biolabs. Human collagen I, pepsin, and calcein-AM was purchased from BD Biosciences. Dulbecco's modified Eagles's medium (DMEM), and CelLytic M was purchased from Sigma-Aldrich. TOPO vector used to hold PCR generated gene fragments, fetal bovine serum (FBS), and ethidium homodimer-1 were purchased from Life Technologies. KOD Hot Start DNA Polymerase, human collagen III, and bovine collagen I were purchased from EMD Millipore. PfuUltra II DNA polymerase was purchased from Agilent Technologies. 20kDa 4-arm PEG maleimide was acquired from NANOCS. Acrylamide solution and microseal B was obtained from Bio-Rad. HT1080 cells were obtained from ATCC. Thiolated hyaluronic acid(HA), and Extralink® PEGDA acquired from ESI BIO was a generous gift from Lisa Flanagan (University of California, Irvine). NHLF cells, and EGM-2 media from Lonza and ECFC-ECs were a generous gift from Christopher Hughes (University of California, Irvine). All other reagents were purchased from Thermo-Fisher.

4.2.2 Construction of cysteine-GFOGER collagen variants

Full-length cysteine-GFOGER collagen variants were constructed using the previously described modular gene assembly strategy [1,9]. Variants were designed utilizing the rCol-0G, rCol-1G-8, and rCol-4G DNA sequences as backbones while introducing non-native cysteines in the same locations as in rCol-4C (Figure 4-1). Non-native cysteines were introduced into GFOGER variants via site directed mutagenesis utilizing primary fragments (PFs) and secondary fragments (SFs) within TOPO holding vectors as the templates. The oligonucleotides for mutagenesis to introduce non-native

cysteines into PFs 1, 3, 10, and 12 (Table 4-1) were designed to change an amino acid of similar size to cysteine in the X or Y position of the Gly-X-Y tripeptide repeat. 300 nM mutation primer pairs were mixed with 50 ng of their respective DNA template, 2.5 U of PfuUltra II DNA polymerase, 200 µM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 µL. PCR was performed with 5 min at 95°C followed by 12 cycles of 20 s at 95°C, 30 s at 55°C, and 2 min at 72°C, followed by 10 min at 72°C when using PFs as the template. When using SFs as the template, extension time was increased to 2.5 min. Methylated template DNA was digested with 20 U DpnI for 2 h at 37°C, and transformed into *Escherichia coli* DH5α. Colonies with plasmids containing the desired PF sequences were confirmed by DNA sequencing (Genewiz).

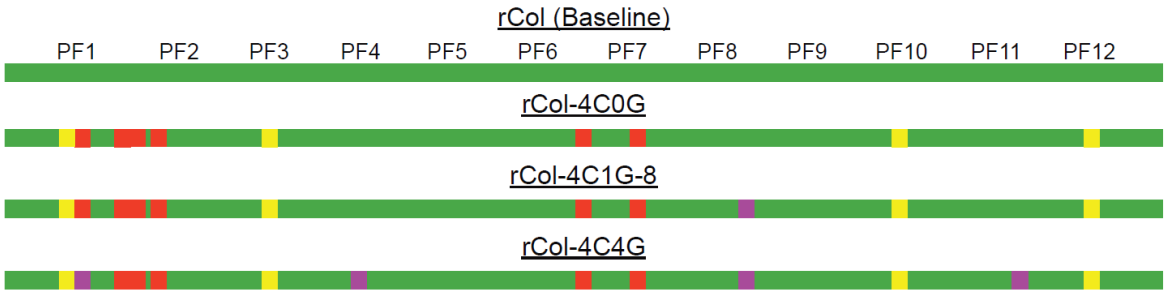


Figure 4-1: Schematic of cysteine-GFOGER variants. Green denotes native hCol III amino acid sequence. Yellow denotes non-native cysteines. Red denotes replacement of integrin binding sequences with non-binding GSPGGK sequences. Purple denotes introduced integrin binding GFOGER sequences. Introduced cysteine locations were chosen to span the length of the triple-helical region of the variants while avoiding integrin binding sites and the MMP cleavage sequence. GFOGER locations were chosen to replace removed binding sites as well as locations with similar sequence homology.

PFs were amplified by PCR by combining 200 nM of each oligo flanking each primary fragment, 50 ng of PF template DNA, 2.5 U of PfuUltra II DNA polymerase, 200 µM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 µL. Reactions were performed using a 10 min denaturing step at 95°C followed by 25 cycles of 20 s at 95°C, 30 s at 62°C, and 15 s at 72°C, followed by 5 min at 72°C. PCR product was then gel purified for use in assembly of secondary fragments.

Table 4-1: List of oligonucleotides used to introduce non-native cysteines for primary fragment.

Primary fragment number	Oligonucleotides sequences (5' → 3'; forward and reverse strands)
PF1	CACCTGGTGAGCCCGGACAATGTGGTCCAAGTGGTCCTCCTGGTC GACCAGGAGGACCACTTGGACCACATTGTCCGGGCTCACCAGGTG
PF3	CGGTCCACCCGGAATTAACGGTTGTCCAGGAGGTAAAGGCGAAATGGGTCCT AGGACCCATTTCGCCTTTACCTCCTGGACAACCGTTAATTCCGGGTGGACCG
PF10	CAGGCATGCCAGGTCCTAGGGGCTGTCCAGGTCCACAAGGCGTTAAAG CTTTAACGCCTTGTGGACCTGGACAGCCCCTAGGACCTGGCATGCCTG
PF12	GACCAAGCGGACCACCTGGTAAATGTGGAACTTCTGGTCATCCAGGTCC GGACCTGGATGACCAGAAGTTCCACATTTACCAGGTGGTCCGCTTGGTC

To assemble secondary fragments, 2 nM of each required primary fragment was mixed together with 200 nM each oligo flanking each secondary fragment, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. Reactions were performed using a 10 min denaturing step at 95°C, followed by 30 cycles of 20 s at 95°C, 30 s at 68°C, and 45 s at 72°C, followed by 5 min at 72°C. Reaction product was agarose gel purified and ligated into pCR-BluntII-TOPO vectors. Colonies with plasmids containing the correct sequence were identified by DNA sequencing.

Secondary fragments were amplified by PCR by combining 200 nM of each oligo flanking each secondary fragment, 50 ng of secondary fragment template DNA, 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, 1 x PfuUltra II reaction buffer in a final volume of 50 μL. PCR was performed using a 10 min denaturing step at 95°C followed by 25 cycles of 20 s at 95°C, 30 s at 62°C, and 45 s at 72°C, followed by 5 min at 72°C. PCR product was then gel purified for use in assembly of full length collagen.

To assemble full length collagen mutants, 2 nM of each secondary fragment was mixed together with 200 nM primers flanking the full length gene and containing a PmeI site upstream of the gene and a NotI site downstream. Full length collagen genes were assembled using 2.5 U of PfuUltra II DNA polymerase, 200 μM dNTPs, and 1x PfuUltra II reaction buffer in a final volume of 50 μL. Reactions were performed using a 10 min denaturing step at 95°C followed by 30 cycles of 20 s at 95°C, 30 s at 68°C, and 1.5 min at 72°C, followed by 5 min at 72°C. The reaction product was

agarose gel purified and ligated into pCR-BluntII-TOPO vectors. Full-length collagen genes (for variants with GFOGER mutations and introduced cysteines) were excised with PmeI and NotI and ligated into the CEN/ARS plasmid. Final sequences for the collagen variants were confirmed by DNA sequencing.

4.2.3 Expression and purification of collagen variants

Full length recombinant biomimetic collagens were expressed in *S. cerevisiae* BYα2β2, which was engineered to contain two copies each of the human prolyl-4-hydroxylase α-subunit and β-subunit genes to achieve hydroxylation of prolines in the collagen biopolymers [10]. The full-length collagen variant genes were cloned into CEN/ARS plasmids and transformed into BYα2β2. Selective SDC(A) and SGC(A) media contain casamino acids (5 g/L), yeast nitrogen base without amino acids (6.7 g/L), adenine sulfate (20 mg/L), and glucose (20 g/L, in SDC(A)) or galactose (20 g/L, in SGC(A)), respectively. Cells were inoculated from -80°C stock cells, grown overnight in 16 × 125 mm culture tubes containing 5 mL SDC(A) medium, and used to inoculate (1% v/v) 2.8 L Fernbach flasks containing 1 L of SGC(A) induction medium. The yeast cells were cultivated at 250 rpm and 30°C (Series 25, New Brunswick), and OD_{600} was monitored. The cells were harvested at an OD_{600} of 2.5 to 3.5 during late exponential phase by centrifugation at 4000 × g, and then frozen at -80°C.

To purify the collagen product from yeast, the yeast cells were thawed and resuspended at 0.1-0.2 g wet cell/mL in cold Tris-buffered saline (TBS, 100 mM Tris, 400 mM NaCl, pH 7.4) with 1 mM PMSF, 2 U/mL DNase, and 1 U/mL RNase. The cells were lysed in a pre-cooled French Press using 25000 psi for two cycles. Ethylenediaminetetraacetic acid (EDTA) and TCEP were added to the lysate to 2 mM and 5 mM final concentration, respectively, and HCl was added to lower the pH to 2. Pepsin (20 mg/mL in 0.1M HCl) was added to a final concentration of 0.2 mg/mL and the solution was incubated at 4°C overnight to digest the background protein and collagen propeptides, leaving the triple-helical component intact. The supernatant was collected by ultracentrifugation (Optima LE80K, Beckman) for 30 min at 100,000 × g and 4°C, and all subsequent steps were performed on ice

or in the cold room. Glacial acetic acid was added to a final concentration of 0.5 M, and NaCl was added to a final concentration of 3 M with stirring. Precipitated protein was collected by ultracentrifugation for 30 min at 70,000 × g and 4°C, and resuspended in cold 0.1 N HCl. The pH of the resuspension was raised to 7.4 with 200 mM Tris pH 8.6, TCEP was added to 5 mM, and 5 M NaCl solution was added to a final concentration of 3 M. Precipitated protein was collected by ultracentrifugation and resuspended in cold 0.1 N HCl. Protein solutions were dialyzed against 50 mM acetic acid (AcOH). Concentration of collagen protein was determined using BCA assay, with bovine collagen type III (BD Biosciences) as a standard. Typical yields ranged from 300 – 500 μg of purified protein per liter of culture.

4.2.4 Characterization of cysteine-GFOGER collagen variants

Triple-helix formation and thermostability were confirmed via circular dichroism (CD). CD scans were performed on a spectropolarimeter (Jasco J-810) equipped with a Peltier temperature controller. The protein samples (collagen variants and native collagen controls) in 50 mM AcOH were scanned between 190–260 nm at 10°C at a scanning speed of 50 nm/min in 0.1 cm path length quartz cells. Three scans were averaged to produce a single reading. For thermal unfolding, we monitored the ellipticity at 221 nm from 25 to 55°C at a heating rate of 1°C/min. The apparent melting temperature (T_m) was defined as the temperature at which 50% of the protein is unfolded. This was determined by fitting the experimental thermal profile to a thermodynamic model via a multi-parameter, non-linear regression using the Levenberg-Marquardt algorithm within MATLAB [11,12]. CD wavelength scans were performed on at least 3 independent batches of protein, and T_m measurements were performed on at least 2 independent batches.

4.2.5 Cellular adhesion assay

HT-1080 cells were grown in DMEM (with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin) and subcultured every 3 days. Cells were trypsinized, washed, and resuspended in DMEM (with 0.1% BSA) before seeding onto the protein-adsorbed surfaces. To prepare these

surfaces, 20 μg/mL native and recombinant collagens (in DPBS) were incubated on 96-well non-tissue culture treated polystyrene plates for 24 h at 4°C.

HT-1080 cells were incubated on the different protein-adsorbed surfaces for 4 h at 37°C. To remove non-specifically bound cells, wells were filled with DMEM containing 0.1% BSA, and the plate was sealed with microseal B adhesive. The plates were inverted, centrifuged at 914 x g for 40 min, and washed with DPBS. Adhered cells were incubated with 4 μM of calcein-AM (in DPBS and 0.1% BSA) at 37°C for 1 h. Solutions were removed, and cells were lysed with CelLytic M. Fluorescence was measured at 494/517 nm (Ex/Em) and normalized to values for hCol I.

4.2.6 Microwell fabrication

Microwells were formed from slabs of PDMS adhered to 1 mm thick glass coverslips to allow for confocal fluorescent imaging. Slygard 184 was mixed at a 10 to 1 ratio and was cast in a dish to form a slab with a thickness of ~3 mm. The solution was degassed and allowed to cure at RT overnight and baked for 1 h at 80°C. A 2 mm biopsy punch was used to punch holes through the slab. The PDMS slab and the glass coverslips were treated in a plasma cleaner for 1.5 min and the coverslips were pressed onto the PDMS to form the bottom of the wells.

4.2.7 Cellular encapsulation

Cysteine-GFOGER variant protein was incubated with 10 times molar excess TCEP for 30 min at RT. 10x DPBS was added to the mixture to bring it to 1x DPBS and pH was adjusted to 7. As a control, bovine collagen I was also treated in the same fashion. NHLFs were suspended in the protein mixture at a concentration of 5 x 10^5 cells per mL and 4 arm PEG-Maleimide 20 kDa crosslinker was added to the solution at a 2 maleimide (Mal) to 1 cysteine ratio. The mixture was incubated at RT for 5 min, deposited into wells and allowed to cure for 1 h at 30°C. Constructs were overlaid with DMEM and incubated at 30°C in a humidified incubator with 5% CO_2 for 5 days. Cells were stained with

calcein-AM and ethidium homodimer-1 to mark living and dead cells respectively and imaged with a fluorescent microscope.

4.2.8 Vasculogenesis assay

Cysteine-GFOGER variant protein was incubated with 10 times molar excess TCEP for 30 min at RT. Concentrated DPBS was added to 1x final concentration and pH was adjusted to 7. The protein was mixed with hyaluronic acid (HA) (for experiments which called for HA) and a co-culture of NHLFs and ECFC-ECs (ratio of cells was varied between 1:1 and 1:2) were suspended in the protein mixture at a concentration of 5×10^4 to 1×10^6 cells per mL. NHLFs were included to remodel the matrix and release the growth factors necessary for the ECFC-ECs to form tubules. To crosslink the hydrogel, 4-arm PEG-maleimide 20 kDa in DPBS was added to the solution at a 2 maleimide to 1 cysteine ratio, or Extralink® PEGDA was added at a 4:1 vol/vol ratio. The mixture was incubated at RT for 5 min, deposited into wells and allowed to cure for 20 min. As a control, fibrin was also mixed with similar concentration of TCEP and cells, and gelation of the mixture was induced by addition of thrombin. Constructs were overlaid with EGM2 media and incubated at 37°C in a humidified incubator with 5% CO_2 for up to 7 days.

4.3 Results and Discussion

4.3.1 Cysteine-GFOGER variants were successfully produced

In Chapters 2 and 3, each variant we produced contained only a single functional sequence introduced into the "baseline" collagen gene or its non-cell adhesion supporting variant. In this study, we were successfully able to assemble variant genes containing a mixture of two functional sequences (non-native cysteine and GFOGER). Cysteines were introduced at locations which avoided reported integrin binding sites and the MMP recognition site. Variants were successfully assembled by mixing-and-matching PFs and SFs containing the desired functionality (Figure 4-1) to produce genes encoding for full length collagen with the specified functionalities at specific locations.

The genes were transferred to our yeast expression system and cysteine-GFOGER variants were successfully expressed and purified (Figure 4-2a) at levels equivalent to rCol. CD spectra for all variants indicated that the triple-helical structure was preserved and equivalent to rCol. Thermostability scans indicated that the triple helixes of cysteine-GFOGER variants were not statistically significantly different from rCol (Figure 4-2b-e) with apparent melting temperatures at ~37°C.

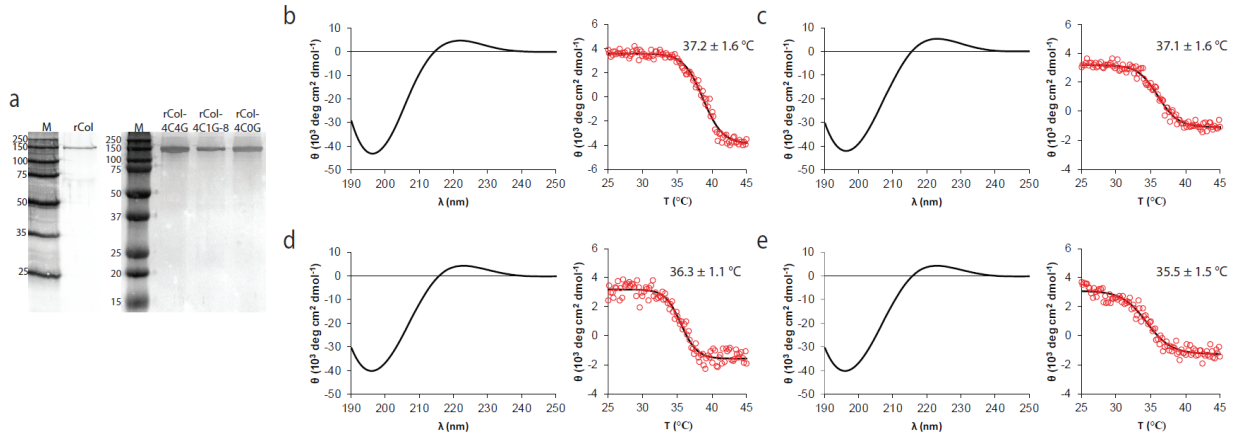


Figure 4-2: Characterization of cysteine-GFOGER variants. (a) SDS-PAGE of purified variants. Molecular weights are given in kDa. All variants are monodisperse in size and similar to that of hCol III running at ~ 150 kDa. Represenative CD wavelength and temperature sweeps of rCol (b), rCol-4C4G (c), rCol-4C1G-8 (d), and rCol-4C0G (e) respectively. All variants display triple-helical CD spectra as well as sharp melting curves characteristic of triple-helical collagen. Variants have apparent melting temperatures similar to that of rCol. Values were an average of at least 2 independent measurements from independent purifications.

4.3.2 Non-native cysteines do not contribute to cellular adhesion

In Chapter 2, we showed that cellular adhesion to cysteine containing variants was not disrupted by the non-native cysteines [13]. However, it does not discount the chance that the non-native cysteines affected cellular adhesion positively, perhaps through disulfide bond formation. The variant rCol-4C0G, which is built off of the non-cell adhesion supporting rCol-0G, would allow testing of this possibility. The 2D adhesion assay showed minimal HT1080 cellular adhesion to rCol-4C0G, similar to that of BSA (Figure 4-3). This result demonstrates that the non-native cysteines did not

increase cellular adhesion and the rCol-4C0G variant would be usable as a non-cell binding control for 3D cell encapsulation assays.

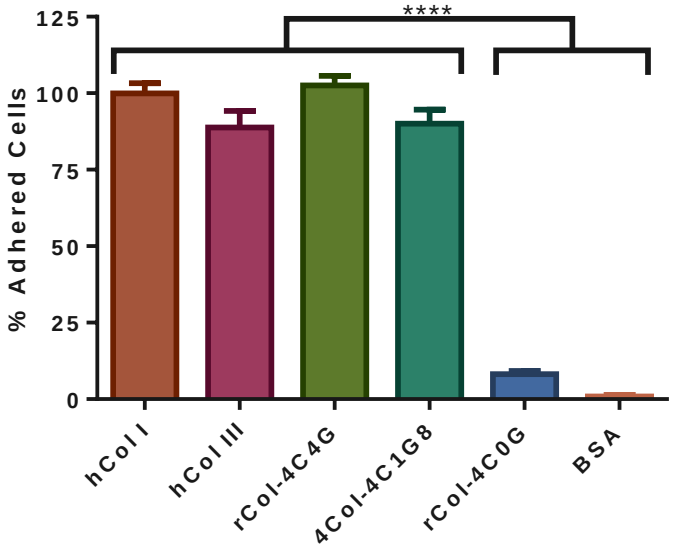


Figure 4-3: HT1080 adhesion assay. Proteins were adsorbed onto non-tissue culture treated polystyrene. HT1080 adhesion was assessed by staining and measurement of calcein-AM. Values were normalized to hCol I. Variant rCol-4C0G displayed levels of cell adhesion similar to BSA while variants with introduced GFOGER displayed levels similar to that of hCol I. No statistically significant differences were detected between cell supporting variants. (Data is from 3 independent experiments of 3 wells in each experiment. **** denotes $p < 0.0001$)

4.3.3 Cell adhesion functionality of introduced GFOGER sequences in 2D are not affected by addition of non-native cysteines

To verify that the introduced non-native cysteines did not disrupt the functionality of the GFOGER sequences, the HT1080 cell adhesion assay was also performed on rCol-4C1G-8 and rCol-4C4G. Both variants were capable of supporting similar numbers of cells at levels comparable to hCol I and hCol III substrates (Figure 4-3). It can be noted that both rCol-4G and rCol-4C4G exhibited higher cell binding numbers than hCol III, rCol-1G-8 and rCol-4C1G-8, though not significantly different (see Chapter 3 for rCol-4G data). These results verified that the non-native cysteines did not significantly affect cell binding to the variants in 2D.

4.3.4 NHLFs spread in hydrogels comprised of rCol-4C1G and rCol-4C4G but not in rCol-4C0G

In Chapter 3, we showed that in 2D, we could modulate cellular adhesion through the introduction of integrin binding sites into a non-cell adhesion supporting variant rCol-0G [1]. However, most cells do not naturally exist in 2D. In Chapter 2, we showed the ability to introduce non-native cysteines into rCol and to form 3D hydrogels, although we did not show the ability to encapsulate cells within them [13]. With the production of the cysteine-GFOGER variants, we combined the functionalities of both the non-native cysteines to produce hydrogels which allows for the testing of the effects of GFOGER within the context of full length collagen on cells in a more physiologically relevant 3D environment.

Within gels of all cysteine-GFOGER variants, the NHLFs were successfully encapsulated. NHLFs were utilized as they would be co-cultured with ECs in subsequent testing for the effects of GFOGER on vasculogenesis. NHLFs were confirmed as viable after 5 days using a live/dead calcein-AM/ethidium homodimer-1 assay (Figure 4-4). Calcein-AM is a cell permeant dye which is non-fluorescent until the acetoxoymethyl ester is hydrolyzed by intracellular esterases. Therefore, calcein-AM only labels cells that are viable [14]. Few cells stained positive with ethidium homodimer-1, which stains the DNA of dead or dying cells but is membrane-impermeable to live cells. Cells within rCol-4C0G, while still viable, retained a spherical morphology. This result was expected as the encapsulated cells would have no adhesion sites with the removal of the integrin binding sites. However, cells within hydrogels formed from rCol-4C1G and rCol-4C4G were able to elongate. These results are consistent with the results of our 2D assay in which cells did not adhere to rCol-4C0G, but were able to adhere to both rCol-4C1G and rCol-4C4G. This also verifies that the crosslinking of the collagen variants through the non-native cysteines does not disrupt the functionality of introduced GFOGER sites. The results demonstrate the ability to combine multiple non-native functionalities within the same collagen variant.

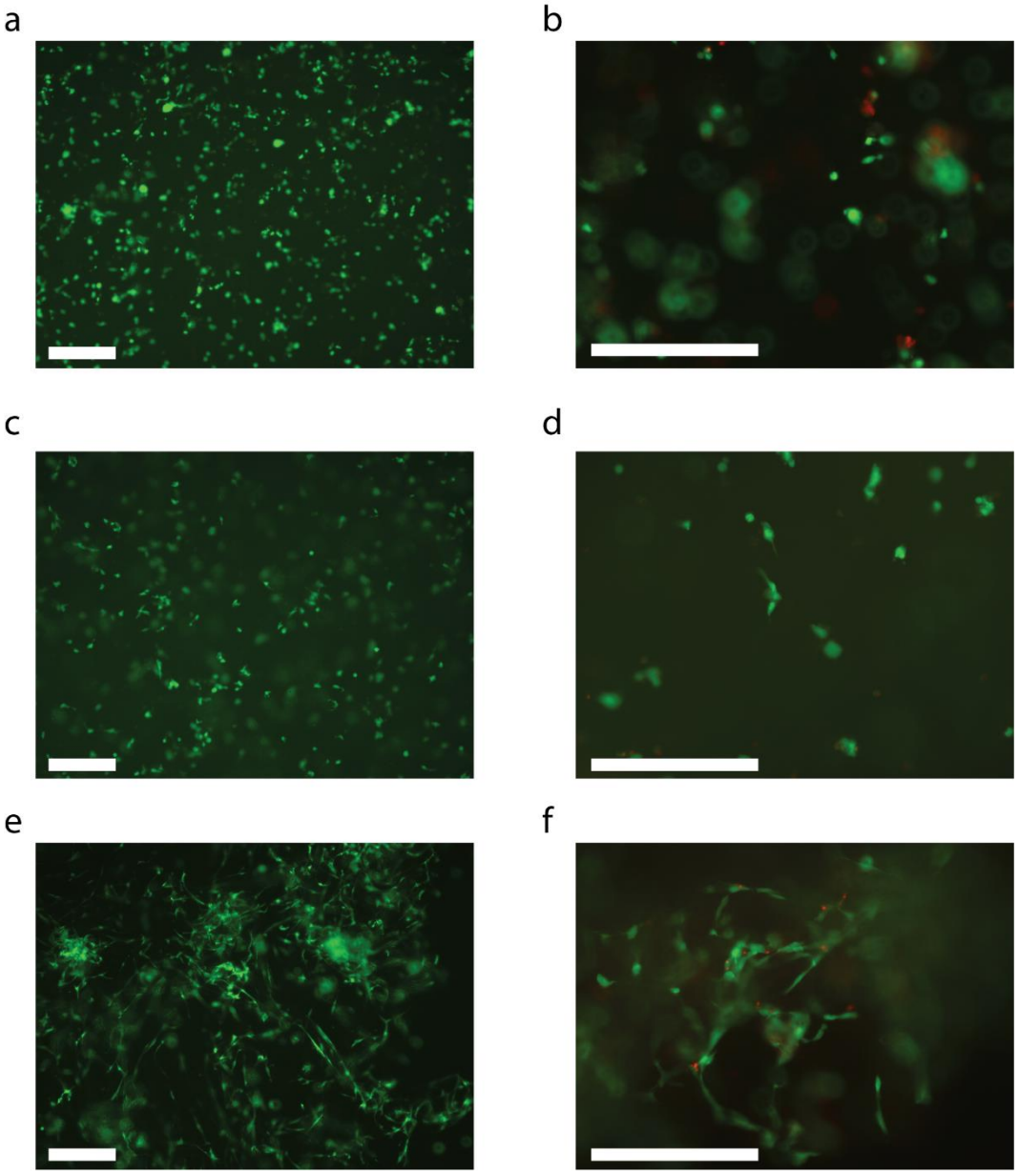


Figure 4-4: Live/dead staining of NHLF encapsulated in collagen variants. NHLFs were incubated with calcein-AM to label live (green) and ethidium homodimer-1 to label dead (red) cells. (a,b) NHLF cells encapsulated in rCol-4C0G. (c,d) NHLF cells encapsulated in rCol-4C1G-8. (e,f) NHLF cells encapsulated in rCol-4C4G. Cells remained rounded in rCol-4C0G while elongating and spreading within the rCol-4C1G-8 and rCol-4C4G gels. Scale bars are 250 μm.

4.3.5 Cysteine-GFOGER variant hydrogels with encapsulated cells do not hold shape at 37°C

For the experiment in the previous section, the NHLFs were incubated at 30°C. However, the human body has a temperature closer to 37°C. When the vasculogenesis experiments were

performed in which NHLFs and ECs were co-cultured, at 37°C in 384 well plate format, the gels did not appear to hold its shape. After the addition of media, the gels would regularly break apart completely or into smaller fragments. The cells, which were not encapsulated within gel, would settle to the bottom of the wells. Switching format from 384-well plates to microwells allowed for the gels to survive the initial addition of media. However, after 24 hours the gels were found to be significantly contracted (Figure 4-5b). These gel clumps had high cell densities, making imaging of the cells within the gels difficult. Therefore we could not conclude on the effect of GFOGER concentration on vasculogenesis. In an attempt to address this contraction, we introduced hyaluronic acid to provide additional structural support.

4.3.6 Addition of hyaluronic acid did not inhibit gel contraction

Hyaluronic acid (HA), also known as hyaluronan, is a glycosaminoglycan found within all tissues of vertebrates. Within the ECM, HA acts as a stabilizer and regulates cell motility [15,16]. We hypothesized that it would provide additional support to prevent the contraction of the gels. HA would also act as a blank slate in terms of cell adhesion [17], allowing for the testing of GFOGER as the only integrin binding domain initially within the gels. The HA was thoilated, allowing it to be crosslinked to the collagen variants using a 4-arm PEG-maleimide crosslinker [17]. The concentration of HA utilized was 6 mg/mL, which has been previously utilized to form hydrogels by the Flanagan Lab (unpublished data). However, the results we observed were similar to gels without HA in the microwell format (Figure 4-5c). The gels with encapsulated ECs and NHLFs would contract within 5 days, sometimes as soon as overnight, which made imaging of formed tubules difficult. It is important to note that, within some replicate hydrogels, the cells remained spherical and the hydrogels did not contract, while the cells would elongate within gels which contracted. Hydrogels made only of HA yielded ECs and NHLF cells which remained spherical; this was expected, as it was reported that fibroblasts do not attach to HA-only gels [17]. These gels did not collapse, though they were only tested at lower cell concentrations (Figure 4-5d). Collagen-HA gels were not tested in the

384 well format nor at temperatures other than 37°C. Due to the concentration of the stock HA, higher concentrations could not be tested.

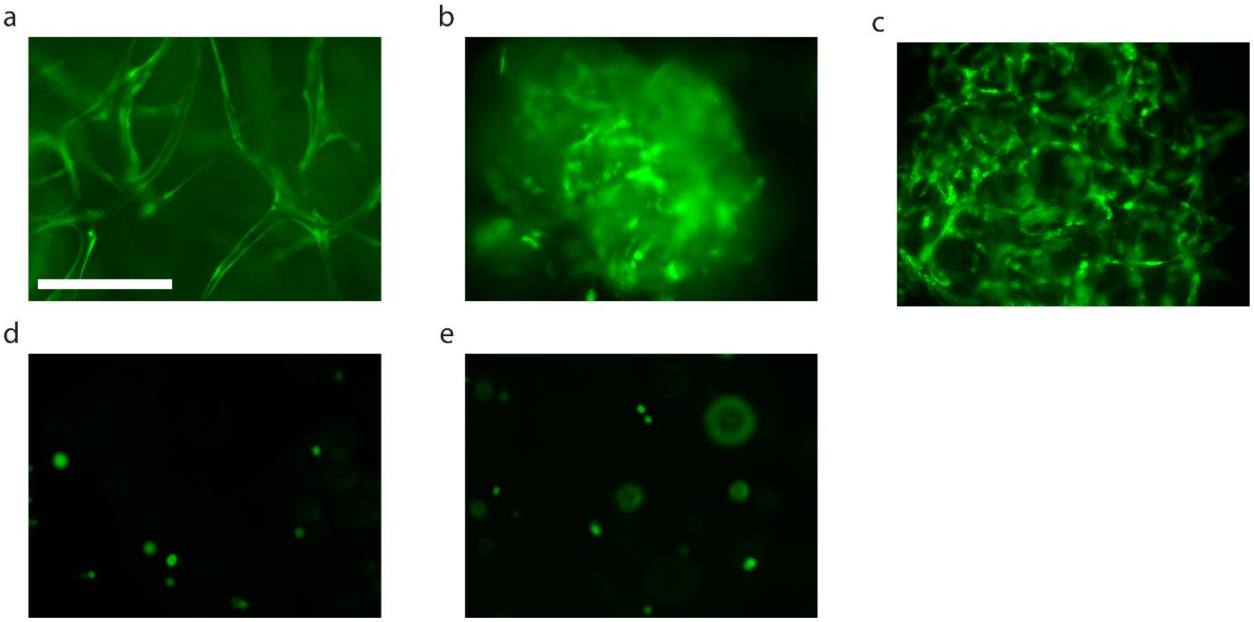


Figure 4-5: Fluorescent image of ECFC-ECs encapsulated in fibrin, collagen, collagen/HA, and HA gels. Endothelial cells transfected to express green fluorescent protein were seeded into gels at a 1:1 ratio with NHLFs (unlabeled) and incubated for 5 days. (a) Endothelial cells migrate and form tubules in fibrin gel. (b) Endothelial cells appear to spread within the rCol-4C4G hydrogel; however, the gel contracts. (c) Endothelial cells still appear to spread within the rCol-4C4G/HA hydrogel; however, the gel continues to contract. (d) In HA-only gels the cells remained spherical. (e) Reduction of cell concentration led to hydrogels in which the cells remained spherical. Scale bar is 250 μm.

4.3.7 Reduction of cell concentration stopped gel contraction, but also inhibited ECFC-EC spreading

Cell seeding density and the ratio of NHLFs to ECs seeded were varied to reduce the contraction of the gels by the encapsulated cells. We hypothesized that by reducing the number of cells, the amount of force exerted on the gels by the cells would also be reduced. Fibroblasts attempt to remodel the ECM through the release of MMPs as well as contraction [18,19]. By reducing the number of cells by a factor of 10 while keeping ratio constant at 1:1, we were able to reduce the frequency of gel contraction. However, the ECs did not appear to spread in these gels (Figure 4-5e). This may be due to the inability of the cells to sense one another at such low densities. At such low

densities, the NHLFs may not be secreting enough growth factors to induce tubule formation. However, higher seeding densities led to gel contraction. The ECs remained viable within the gels as the intensity of the fluorescent tag remained strong, indicating the cells were alive and expressing protein.

Another explanation for the inability of cells to spread could be the possibility of recombinant collagen partially denaturing at 37°C. Previous studies have shown that the triple-helical conformation is necessary for integrin binding to GFOGER. However, if protein denaturation was the cause, we would have expected it to also occur at the higher seeding densities as well. At higher cell concentrations, cells were still able to extend throughout the gels. Gelatin, which is formed by the denaturation of collagen, could hypothetically provide the adhesion sites. This in theory would provide the adhesion sites for the cells to invade the gels. However, we are not aware of previous studies which use gelatin to form gels at concentrations lower than 10% w/v. This is more than 10 times higher than the concentrations used in our study. Therefore, at our current concentrations, collagen, or denatured collagen, may not provide enough integrin binding sites at the concentrations used here.

The fact that the cells did not appear to invade in hydrogels which did not contract could be hinting at the minimum integrin binding site concentrations necessary for cell invasion. It is possible that at the concentrations utilized, there was insufficient integrin binding sites for the cells to bind. However, once the gels contracted, the concentration of integrin binding sites would increase and the cells would display elongated morphologies. Additional experiments would be required to find specific conditions (protein and cell concentrations, crosslinking ratios, or gelation parameters) that would allow for cell spreading while retaining the gel's structural integrity.

4.4 Conclusions

In this study, we further expanded the functionality of the modular collagen system by demonstrating the ability to combine two different functionalities, non-native cysteines and GFOGER sequences, into a single collagen variant. We then utilized both functionalities by using the cysteines for crosslinking and GFOGER sequences for cellular adhesion. This combination allowed the testing of introduced functionality within the context of triple-helical collagen in a 3-dimensional context, more closely mimicking the *in vivo* environment. We attempted to assess the effect of GFOGER on vasculogenesis through the use of ECFC-ECs co-cultured with NHLFs. However, we were unable to produce reliable results, as the gels tended to contract, making imaging of any tubules very difficult. Further optimization of parameters of the vasculogenesis assay is necessary before conclusive results are possible.

4.5 References

[1] R.A. Que, S.W.P. Chan, A.M. Jabaiah, R.H. Lathrop, N.A. Da Silva, S.-W. Wang, Tuning cellular response by modular design of bioactive domains in collagen, Biomaterials. 53 (2015) 309–317. doi:10.1016/j.biomaterials.2015.02.074.

[2] O.W. Petersen, L. Rønnov-Jessen, A.R. Howlett, M.J. Bissell, Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells., Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 9064–9068.

[3] K.M. Yamada, E. Cukierman, Modeling Tissue Morphogenesis and Cancer in 3D, Cell. 130 (2007) 601–610. doi:10.1016/j.cell.2007.08.006.

[4] L.G. Griffith, M.A. Swartz, Capturing complex 3D tissue physiology in vitro, Nat. Rev. Mol. Cell Biol. 7 (2006) 211–224. doi:10.1038/nrm1858.

[5] C.P. Blobel, 3D trumps 2D when studying endothelial cells, Blood. 115 (2010) 5128–5130. doi:10.1182/blood-2010-03-275271.

[6] S.M. Sweeney, G. DiLullo, S.J. Slater, J. Martinez, R.V. Iozzo, J.L. Lauer-Fields, G.B. Fields, J.D.S. Antonio, Angiogenesis in Collagen I Requires α2β1 Ligation of a GFP*GER Sequence and Possibly p38 MAPK Activation and Focal Adhesion Disassembly, J. Biol. Chem. 278 (2003) 30516–30524. doi:10.1074/jbc.M304237200.

[7] G.E. Davis, C.W. Camarillo, An α2β1 Integrin-Dependent Pinocytic Mechanism Involving Intracellular Vacuole Formation and Coalescence Regulates Capillary Lumen and Tube Formation in Three-Dimensional Collagen Matrix, Exp. Cell Res. 224 (1996) 39–51. doi:10.1006/excr.1996.0109.

[8] J.R. García, A.Y. Clark, A.J. García, Integrin-specific hydrogels functionalized with VEGF for vascularization and bone regeneration of critical-size bone defects, J. Biomed. Mater. Res. A. 104 (2016) 889–900. doi:10.1002/jbm.a.35626.

[9] S.W.P. Chan, S.-P. Hung, S.K. Raman, G.W. Hatfield, R.H. Lathrop, N.A. Da Silva, S.-W. Wang, Recombinant Human Collagen and Biomimetic Variants Using a De Novo Gene Optimized for Modular Assembly, Biomacromolecules. 11 (2010) 1460–1469. doi:10.1021/bm100052y.

[10] A. Jabaiah, X. Wang, S.K. Raman, R. Ragan, N.A. Da Silva, S.-W. Wang, Nanoscale architecture and cellular adhesion of biomimetic collagen substrates, J. Biomater. Appl. 28 (2014) 1354–1365. doi:10.1177/0885328213508328.

[11] D. Ren, M. Dalmau, A. Randall, M.M. Shindel, P. Baldi, S.-W. Wang, Biomimetic Design of Protein Nanomaterials for Hydrophobic Molecular Transport, Adv. Funct. Mater. 22 (2012) 3170–3180. doi:10.1002/adfm.201200052.

[12] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nat Protoc. 1 (2007) 2527–2535.

[13] R. Que, A. Mohraz, N.A. Da Silva, S.-W. Wang, Expanding Functionality of Recombinant Human Collagen Through Engineered Non-Native Cysteines, Biomacromolecules. 15 (2014) 3540–3549. doi:10.1021/bm500735d.

[14] X.M. Wang, P.I. Terasaki, G.W. Rankin, D. Chia, H.P. Zhong, S. Hardy, A new microcellular cytotoxicity test based on calcein AM release, Hum. Immunol. 37 (1993) 264–270.

[15] X. Zheng Shu, Y. Liu, F.S. Palumbo, Y. Luo, G.D. Prestwich, In situ crosslinkable hyaluronan hydrogels for tissue engineering, Biomaterials. 25 (2004) 1339–1348. doi:10.1016/j.biomaterials.2003.08.014.

[16] G.P. Dowthwaite, J.C.W. Edwards, A.A. Pitsillides, An Essential Role for the Interaction Between Hyaluronan and Hyaluronan Binding Proteins During Joint Development, J. Histochem. Cytochem. 46 (1998) 641–651. doi:10.1177/002215549804600509.

[17] X.Z. Shu, K. Ghosh, Y. Liu, F.S. Palumbo, Y. Luo, R.A. Clark, G.D. Prestwich, Attachment and spreading of fibroblasts on an RGD peptide–modified injectable hyaluronan hydrogel, J. Biomed. Mater. Res. A. 68A (2004) 365–375. doi:10.1002/jbm.a.20002.

[18] F. Grinnell, Fibroblast biology in three-dimensional collagen matrices, Trends Cell Biol. 13 (2003) 264–269. doi:10.1016/S0962-8924(03)00057-6.

[19] E. Bell, B. Ivarsson, C. Merrill, Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro., Proc. Natl. Acad. Sci. U. S. A. 76 (1979) 1274–1278.

Chapter 5: Conclusions and Recommendations

5.1 Conclusions

The aim of this research was to produce collagen variants which can be utilized to produce a customizable ECM. Utilizing native ECM components makes decoupling of properties, such as stiffness and functional site concentration, difficult. To modulate the stiffness of native ECM protein hydrogels, typically the concentration of the proteins used is modified. However, modifying the protein concentration also changes the concentration of functional sites within the hydrogel, which makes it impossible to separate the effects of stiffness from the concentration of these functional sites. The production of collagen variants containing different numbers of non-native cysteines allowed for the formulation of collagen hydrogels of varying stiffness while holding the concentration of protein constant, decoupling the properties of stiffness and functional site concentrations. The recombinant collagen was tolerant of significant changes throughout the length of the triple-helical region with mutations of up to 44 amino acids in rCol-4C4G, while retaining triple-helical conformation and thermostability near physiological temperature. The non-native cysteines also functioned as sites for specific covalent conjugation of growth factors.

Modulation of cellular adhesion was achieved through the introduction of the integrin binding sequence GFOGER to the non-cell supporting collagen variant rCol-0G, which had the top α2β1 binding sequences removed. The modular collagen platform allows for precise control of the number and location of bioactive sites. The location of the introduced GFOGER within the triple-helical region had an effect on the adhesion capability of the variant, with rCol-1G supporting approximately 30% of the cells that rCol-4G supported. This reduced adhesion showed the importance that the location of the introduced functionality within the polymer can have. Introduction of GFOGER sequences could potentially be replaced by alternative functional peptide sequences to modulate other cellular processes.

A strength of the modular collagen platform is the ability to mix-and-match fragments together when assembling the full-length gene. The previous examples utilized mixing together fragments containing one type of functionality. However, multiple functionalities were also introduced into the same collagen variant at defined locations and quantities to control distinct properties of the resultant hydrogels. By combining non-native cysteines and GFOGER sequences, the functionality of GFOGER within the context of full length collagen was examined in three-dimensional cell culture conditions. As expected, encapsulated normal human lung fibroblasts did not spread in the hydrogels formed with collagen variants that had native α2β1 binding sequences removed, while introduction of even a single GFOGER allowed for the cells to spread within the matrix. Hydrogels contracted when cultured at 37°C, which obfuscated any potential tubule formation within the gels. Overall, this research demonstrated the ability to produce a library of collagen variants with functionalities at varying locations to produce a customizable ECM to guide cell fate.

5.2 Future Directions

We were successfully able to culture cells in 3D within collagen variant hydrogels. However, encapsulation of cells within collagen variants exhibited challenges of gel contraction and reproducibility. This may be due to the low stiffness and protein concentrations of the hydrogels formed. Hydrogels formed at concentrations of 4 mg/mL yielded gels with stiffness of approximately 50 Pa. Other studies utilizing hydrogels at stiffness within this range did not yield significant tubule formation [1]. To increase hydrogel stiffness, variants containing additional cysteines (> 4) could be produced. However, one major limitation is the overall amount of collagen produced and the purification of these variants. Other studies utilizing recombinant elastin have used concentrations upwards of 100 mg/mL [2]. With our current protein expression and purification protocols, these concentrations would require at least 20 L of yeast culture to produce 100 µl of 100 mg/mL protein. Therefore, future directions should include the optimization of the protein expression system and/or development of a more scalable purification procedure.

By switching expression of cultures from shake flasks to larger volume bioreactors, the time requirement for growing larger volumes would be reduced. At the moment we can grow up to 8 L of culture at a time, requiring multiple rounds of expression for greater than 8 L of culture. However, the next bottleneck arises quickly in the purification pipeline. At the current time, the rate limiting steps of purification are cell lysis utilizing the French press and centrifugation of cell lysate. The French press can process approximately 6 g of wet cell pellet per hour. With each L of culture yielding 6 g, processing 20 L would require 20 h. A new method to lyse the cell pellets would need to be investigated. Larger capacity centrifuges would also be necessary to process the increased volume of cell lysate. Both of these solutions may require large capital investments.

An alternative strategy to improve the collagen yields necessary would be to increase the amount of collagen expressed on a per cell basis. For the studies in this thesis, we utilized the *GAL1* promoter for the expression of the recombinant collagen gene as well as the hydroxylase genes. It is possible that further strain engineering or optimizing expression conditions would increase yield. Production of recombinant collagen in *Pichia pastoris* has been reported as high as 1.5 g/L, though this is likely due to growing cultures to much higher densities (>500 g/L wet cell weight) rather than higher protein expression per cell [3]. These yields would still be approximately 30 times higher per wet cell weight than our current system.

Although the stiffness of hydrogels formed in this thesis may be low for vasculogenesis assays, the stiffness range may work well for neural applications. Brain tissue have moduli of ~300-500 Pa [4,5]. Neurons have also been shown to thrive on softer substrates of ~ 50-350 Pa [6]. Furthermore, the concentration of the integrin binding motif RGD has been shown to have an effect on neural differentiation [7]. It may be possible that the effects on differentiation may also occur with modulations of GFOGER concentrations. Initial tests to encapsulate human neural stem/progenitor cells proved elusive with the gels not holding shape when media was applied. However, with higher

yields or reduced hydrogel volume requirements, additional gelation conditions can be tested in future studies. Our group has previously synthesized collagen variants containing the pentapeptide IKVAV from laminin [8]. IKVAV has been shown affect neurite extension, migration, and neural stem cell differentiation [9,10]. With the successful encapsulation of neural stem/progenitor cells, it would be possible to test the effects of GFOGER or other functional peptide sequences such as IKVAV within the context of collagen in 3D. This could perhaps yield substrates that can be further used for neural repair applications such as spinal cord injury.

Within this work, we investigated the use of non-native cysteines to form and control hydrogel stiffness, as well as to immobilize growth factors. We also investigated the ability to modulate integrin binding motifs. Another aspect of the ECM which has been shown to direct cell fate is degradation rate. Vascular invasion has been observed to increase with increasing concentration of MMP sensitive substrates within hydrogels [11]. Our research group has also produced variants in which the number and location of MMP sensitive sequences across the polymer is varied from zero to 3 [12]. One future direction would be to further test the effects of MMP sequences within the context of collagen on vasculogenesis or other cell fate processes. It is possible to introduce other bioactive sequences in our modular collagen system and test their functionality within the context of full length collagen. The fact that the introduction of IKVAV, which disrupts the Gly-X-Y tripeptide repeat, did not destabilize the collagen triple-helix opens the door to test other functional sequences that don't natively appear in collagen. Combining multiple bioactive sequences could result in a highly tunable cellular microenvironment.

Another aspect which was not investigated within this thesis was the ability to form fibrils utilizing recombinant collagen variants. Hydrogels formed from native collagen are formed by the assembly of collagen proteins into fibrils. In fact, it has been shown that neurite outgrowth orients itself along collagen fibers [13]. An initial test to induce fibril formation of rCol yielded the formation

of elongated structures. Testing of additional salt and pH conditions will be necessary to find the proper parameters to form fibrils. Changes to the purification procedure may also be needed. The telopeptides flanking the collagen triple-helical domain are involved in fibril formation [14,15]. Pepsin digestion of yeast background proteins and the collagen pro-peptides may need to be controlled or removed to ensure the integrity of the telopeptides. Previous groups have successfully imaged microfibril formation from recombinant collagen produced in *P. pastoris* which had been pepsin treated [16]. Therefore, it should be possible to achieve this with our system. With the formation of fibrils, PEG crosslinkers to form hydrogels from our collagen variants would no longer be required. Fibril formed hydrogels would better mimic those formed by native collagen and provide additional properties to modulate, such as fibril diameter, to control cell fate.

5.3 References

[1] A.K. Jha, K.M. Tharp, J. Ye, J.L. Santiago-Ortiz, W.M. Jackson, A. Stahl, D.V. Schaffer, Y. Yeghiazarians, K.E. Healy, Enhanced survival and engraftment of transplanted stem cells using growth factor sequestering hydrogels, Biomaterials. 47 (2015) 1–12. doi:10.1016/j.biomaterials.2014.12.043.

[2] Y.-N. Zhang, R.K. Avery, Q. Vallmajo-Martin, A. Assmann, A. Vegh, A. Memic, B.D. Olsen, N. Annabi, A. Khademhosseini, A Highly Elastic and Rapidly Crosslinkable Elastin-Like Polypeptide-Based Hydrogel for Biomedical Applications, Adv. Funct. Mater. 25 (2015) 4814–4826. doi:10.1002/adfm.201501489.

[3] D. Olsen, C. Yang, M. Bodo, R. Chang, S. Leigh, J. Baez, D. Carmichael, M. Perälä, E.-R. Hämäläinen, M. Jarvinen, J. Polarek, Recombinant collagen and gelatin for drug delivery, Adv. Drug Deliv. Rev. 55 (2003) 1547–1567. doi:10.1016/j.addr.2003.08.008.

[4] B.S. Elkin, E.U. Azeloglu, K.D. Costa, B. Morrison III, Mechanical Heterogeneity of the Rat Hippocampus Measured by Atomic Force Microscope Indentation, J. Neurotrauma. 24 (2007) 812–822. doi:10.1089/neu.2006.0169.

[5] Y.-B. Lu, K. Franze, G. Seifert, C. Steinhäuser, F. Kirchhoff, H. Wolburg, J. Guck, P. Janmey, E.-Q. Wei, J. Käs, A. Reichenbach, Viscoelastic properties of individual glial cells and neurons in the CNS, Proc. Natl. Acad. Sci. 103 (2006) 17759–17764. doi:10.1073/pnas.0606150103.

[6] L.A. Flanagan, Y.-E. Ju, B. Marg, M. Osterfield, P.A. Janmey, Neurite branching on deformable substrates, Neuroreport. 13 (2002) 2411–2415. doi:10.1097/01.wnr.0000048003.96487.97.

[7] K. Saha, E.F. Irwin, J. Kozhukh, D.V. Schaffer, K.E. Healy, Biomimetic interfacial interpenetrating polymer networks control neural stem cell behavior, J. Biomed. Mater. Res. A. 81A (2007) 240–249. doi:10.1002/jbm.a.30986.

[8] F. Abdulhadi, Construction of Recombinant Human Collagen III Variants and Genomic Manipulations in Yeast for Collagen Yield Optimization, (2015).

[9] K. Tashiro, G.C. Sephel, B. Weeks, M. Sasaki, G.R. Martin, H.K. Kleinman, Y. Yamada, A synthetic peptide containing the IKVAV sequence from the A chain of laminin mediates cell attachment, migration, and neurite outgrowth., J. Biol. Chem. 264 (1989) 16174–16182.

[10] X. Li, X. Liu, B. Josey, C.J. Chou, Y. Tan, N. Zhang, X. Wen, Short Laminin Peptide for Improved Neural Stem Cell Growth, Stem Cells Transl. Med. 3 (2014) 662–670. doi:10.5966/sctm.2013-0015.

[11] S. Sokic, M.C. Christenson, J.C. Larson, A.A. Appel, E.M. Brey, G. Papavasiliou, Evaluation of MMP substrate concentration and specificity for neovascularization of hydrogel scaffolds, Biomater. Sci. 2 (2014) 1343–1354. doi:10.1039/C4BM00088A.

[12] D. Crakes, Modification of Recombinant Human Collagen with Matrix Metalloproteinase Cleavage Sites, (2015).

[13] T. Liu, J.D. Houle, J. Xu, B.P. Chan, S.Y. Chew, Nanofibrous Collagen Nerve Conduits for Spinal Cord Repair, Tissue Eng. Part A. 18 (2012) 1057–1066. doi:10.1089/ten.tea.2011.0430.

[14] J.P.R.O. Orgel, T.C. Irving, A. Miller, T.J. Wess, Microfibrillar structure of type I collagen in situ, Proc. Natl. Acad. Sci. 103 (2006) 9001–9005. doi:10.1073/pnas.0502718103.

[15] H.J. Chung, A. Steplewski, K.Y. Chung, J. Uitto, A. Fertala, Collagen Fibril Formation, J. Biol. Chem. 283 (2008) 25879–25886. doi:10.1074/jbc.M804272200.

[16] K.E. Williams, D.R. Olsen, Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen, Matrix Biol. 28 (2009) 373–379. doi:10.1016/j.matbio.2009.04.009.

Appendix A: Supplementary Figures

Parts of this chapter are reprinted from Biomaterials, Vol 53, Richard A. Que, Sam Wei Polly Chan, Abeer M. Jabaiah, Richard H. Lathrop, Nancy A. Da Silva, Szu-Wen Wang, Tuning cellular response by modular design of bioactive domains in collagen, 309-317, 2015, with permission from Elsevier

A.1 Chapter 3

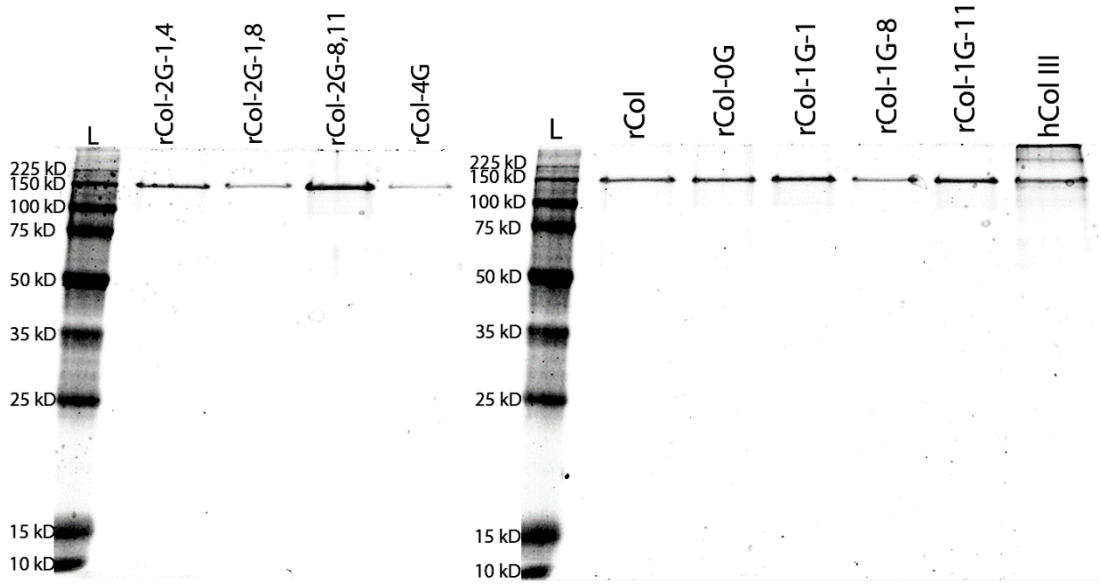


Figure A3-1: SDS-PAGE of purified recombinant collagen mutants and native (commercial) human collagen III. Single distinct bands of the recombinant proteins show purity, with bands running at the same molecular weight as the hCIII controls.

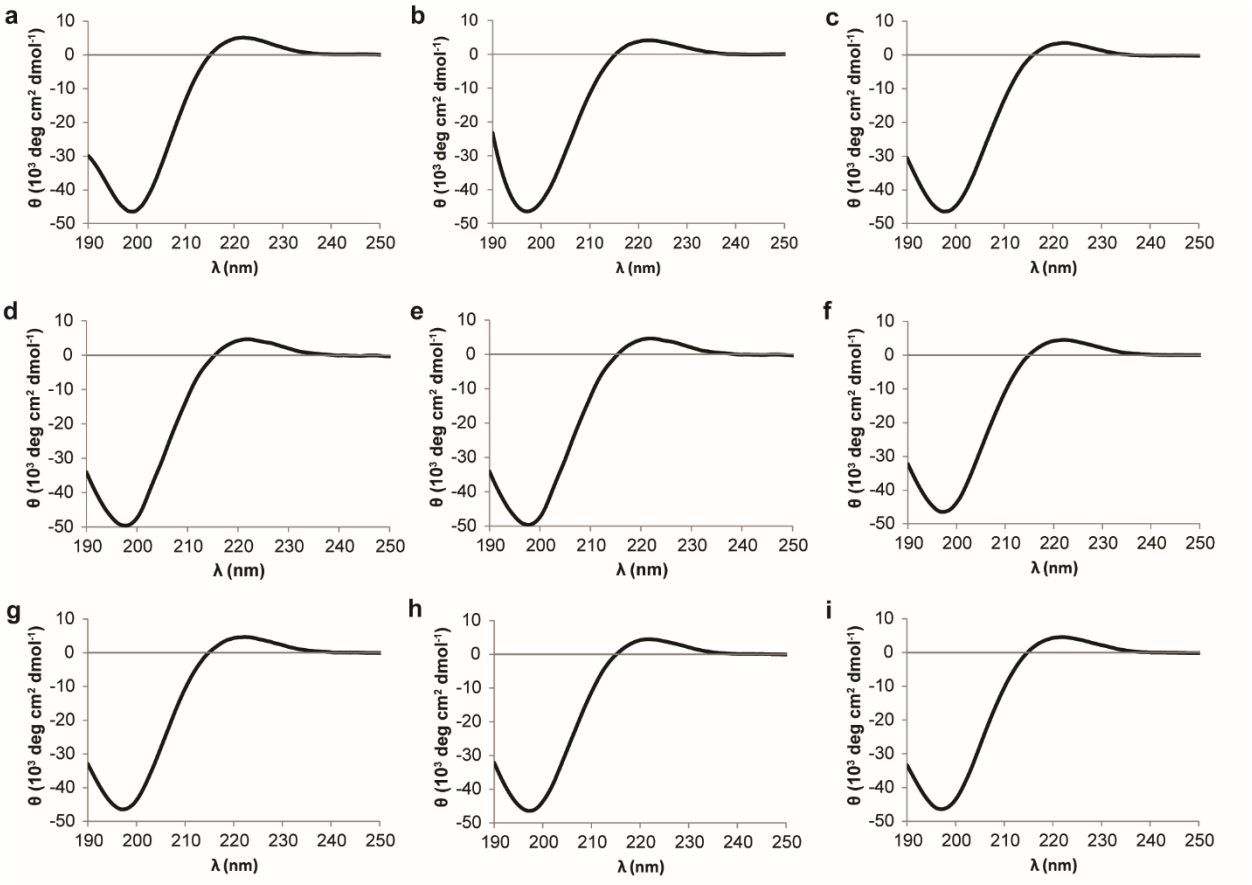


Figure A3-2: CD spectra of recombinant collagen variants confirm triple-helical structure of each mutant. Spectra are of variants (a) rCol; (b) rCol-0G; (c) rCol-1G-1; (d) rCol-1G-8; (e) rCol-1G-11; (f) rCol-2G-1,4; (g) rCol-2G-1,8; (h) rCol-2G-8,11; and (i) rCol-4G.

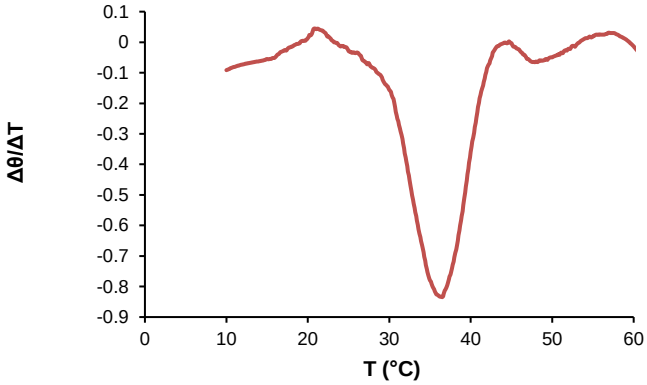


Figure A3-3: Representative first derivative spectrum of recombinant collagen variant rCol-4G. The melting temperatures (T_m) obtained using first derivative analysis also yields 35.6 ± 0.8°C for rCol-4G, which is equivalent to the results from the method described by Greenfield [34].

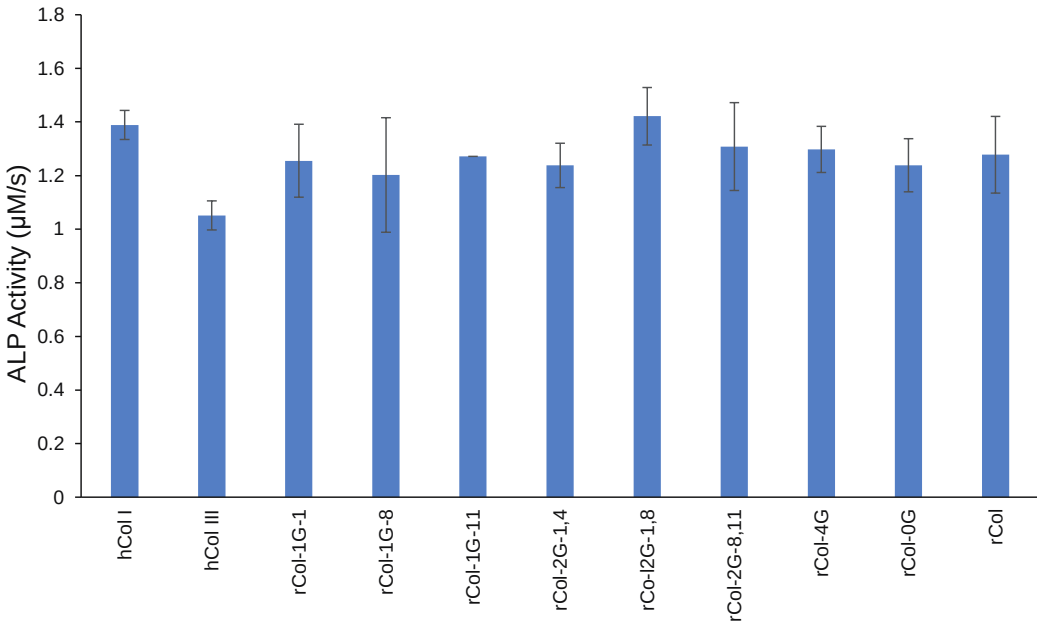


Figure A3-4: An alkaline phosphatase assay was used to assess the amount of surface-adsorbed collagen on the surface of the polystyrene plates [22]. Data confirm that all variants were present at saturation levels and at amounts comparable to one another. Data are reported as average and standard deviation (N=3).

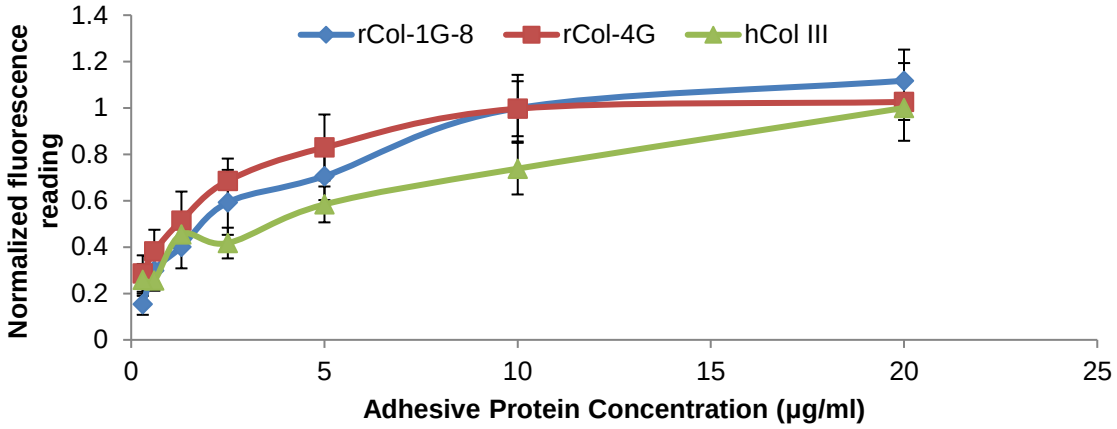


Figure A3-5: Alternate plot of fluorescence measurements quantifying the relative number of cells on surfaces coated with solutions of 0.3 – 20 µg/ml binding collagen (rCol-1G-8, rCol-4G, and hCol I) diluted in non-binding collagen (rCol-0G). Total protein concentrations (binding + non-binding collagens) are 20 µg/ml. Fluorescence values are normalized to average for 20 µg/ml hCol I. Mean ± S.E.M

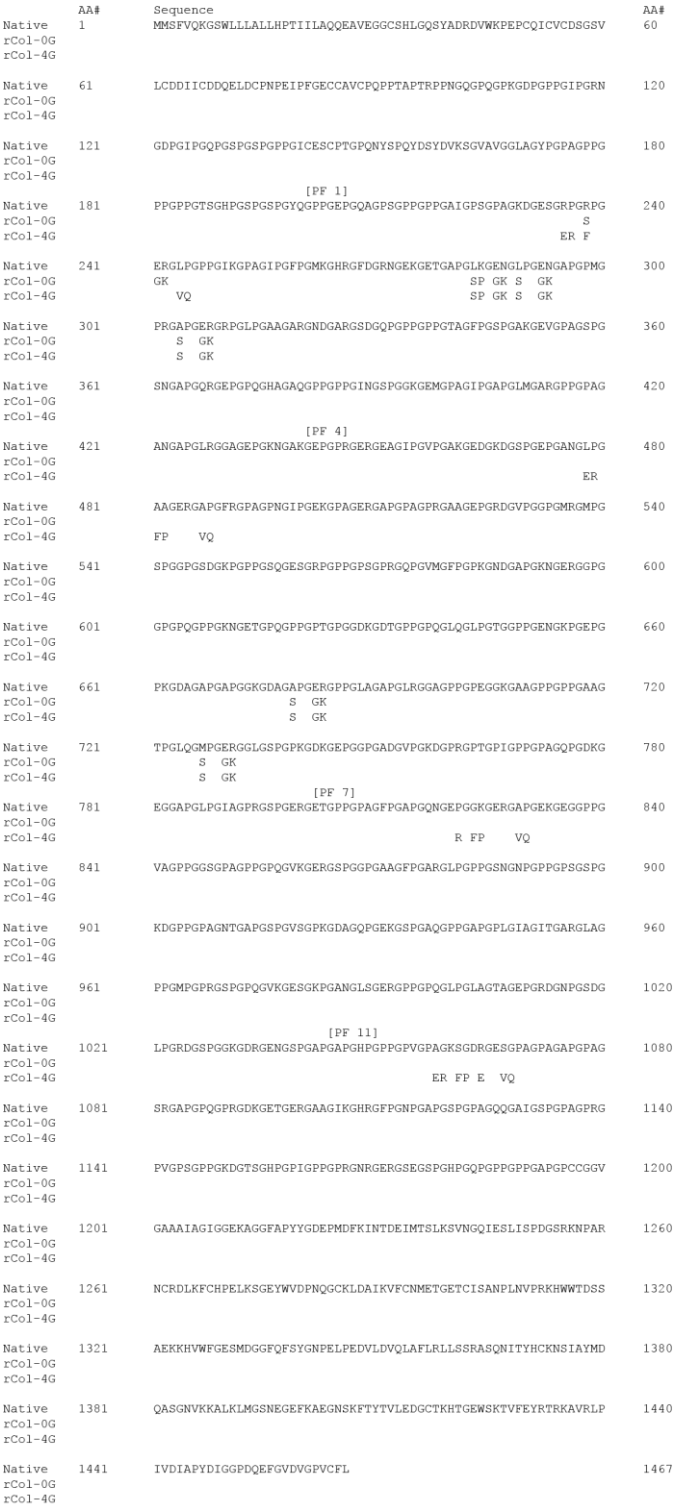


Figure A3-6: Amino acid sequences of native human collagen (top line) and collagen variants rCol-0G and rCol-4G. Changes to the collagen sequence is shown below the native sequence. Removal of native integrin binding sites in rCol-0G are in PF 1, 2, 6, and 7. Introduction of GFOGER sites are in PF 1, 4, 8, and 11.

Table A3-1: Oligonucleotides used in synthesis of genes encoding the triple-helical region of collagen variants rCol-0G and rCol-1G-1. Refer to Chan et al. [12] for sequences encoding rCol and the N- and C- termini.

a. Oligonucleotides used for primary fragment 1 of rCol-0G:

Oligo #	Sequence
1	5'-CCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCTGTTGGT-3'
2	5'-ACCAGCGGGACCTGGATAACCAGCTAGCCCACCAACAGCAACACCAGATTTGACGTCGTA-3'
3	5'-GGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACA-3'
4	5'-TGGTGAACCGGGTGAGCCAGGATGTCCAGATGTACCTGGTGGACCAGGCGGTCCGGGAGG-3'
5	5'-TCTGGACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACA-3'
6	5'-TGGCGGACCAGGAGGACCACTTGGACCAGCTTGTCCGGGCTCACCAGGTGGACCTTGATAACC-3'
7	5'-AGCTGGTCCAAGTGGTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTA-3'
8	5'-GGACTCCCTGGTCTACCAGATTCACCATCCTTACCGGCTGGACCAGATGGACCGATAGCTCC-3'
9	5'-AGGATGGTGAATCTGGTAGACCAGGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGA-3'
10	5'-AAAACCTGGTATTCCCGCAGGGCCTTTTATTCCAGGTGGACCTGGCAAGCCTTTGCCGCCT-3'
11	5'-ATAAAAGGCCCTGCGGGAATACCAGGTTTTCCTGGCATGAAAGGTCACAGAGGTTTTGATGGT-3'
12	5'-AGCACCAGTTTCACCCTTTTCGCCATTTCTACCATCAAAACCTCTGTGACCTTTCATGCCAGG-3

b. Oligonucleotides used for primary fragment 2 of rCol-0G:

Oligo #	Sequence
1	5'-ATAAAAGGCCCTGCGGGAATACCAGGTTTTCCTGGCATGAAAGGTCACAGAGGTTTTGATGGT-3'
2	5'-AGCACCAGTTTCACCCTTTTCGCCATTTCTACCATCAAAACCTCTGTGACCTTTCATGCCAGG-3
3	5'-AGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCGCCAGGTGGTAAGGGTTCGCCAGGA-3'
4	5'-TGGACCCATTGGGCCTGGAGCTCCCTTTCCTCCTGGCGAACCCTTACCACCTGGCGATCCAGG-3'
5	5'-GGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCGGTAAAGGTAGACCTGGT-3'
6	5'-ATTACCACGGGCACCGGCTGCTCCTGGCAAACCAGGTCTACCTTTACCGCCTGGTGAACCCCT-3'
7	5'-TTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGGACAACCTGGC-3'
8	5'-AAAGCCAGCTGTTCCAGGTGGTCCAGGTGGGCCAGGTTGTCCGTCAGATCCTCTAGCACCATC-3'
9	5'-CCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTGG-3'
10	5'-AGCACCATTTGAACCGGGTGAACCTGCTGGACCAACTTCACCTTTGGCGCCGGGACTGCCAGG-3'

c. Oligonucleotides used for primary fragment 3 of rCol-0G:

Oligo #	Sequence
1	5'-CCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTGG-3'
2	5'-AGCACCATTTGAACCGGGTGAACCTGCTGGACCAACTTCACCTTTGGCGCCGGGACTGCCAGG-3'
3	5'-TCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGG-3'
4	5'-GGACCGGGAGGACCTTGAGCACCAGCATGACCCTGAGGTCCGGGTTCTCCCCTTTGTCCAGG-3'
5	5'-GTCATGCTGGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAA-3'
6	5'-CACCAGGTATTCCGGCAGGACCCATTTCGCCTTTACCTCCTGGGCTACCGTTAATTCCGGGT-3'
7	5'-GGCGAAATGGGTCCTGCCGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCT-3'
8	5'-GACCTGGAGCACCGTTAGCTCCAGCTGGACCAGGAGGGCCTCTCGCACCCATTAAGCCTGGAG-3'
9	5'-GGTCCAGCTGGAGCTAACGGTGCTCCAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAA-3'
10	5'-TCCCCTTGGTCCGGGTTCGCCTTTAGCACCATTCTTTCCTGGTTCACCGGCTCCACCTCTTA-3'

d. Oligonucleotides used for primary fragment 4 of rCol-0G:

Oligo #	Sequence
1	5'-GGTCCAGCTGGAGCTAACGGTGCTCCAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAA-3'
2	5'-TCCCCTTGGTCCGGGTTCGCCTTTAGCACCATTCTTTCCTGGTTCACCGGCTCCACCTCTTA-3'
3	5'-TGGTGCTAAAGGCGAACCCGGACCAAGGGGAGAACGTGGAGAAGCCGGCATACCCGGAGTAC-3'
4	5'-CCATCTTTGCCATCCTCACCTTTAGCACCGGGTACTCCGGGTATGCCGGCTTCTCCACGTTC-3'
5	5'-CCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGTGAGCCTGGGGCTAATGGATTG-3'
6	5'-CAGGAGCTCCTCTTTCGCCTGCAGCACCAGGCAATCCATTAGCCCCAGGCTCACCAGGGGAA-3'
7	5'-CCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAGGACCAAATGGAATA-3'
8	5'-CTCTTTCACCTGCTGGGCCTTTTTCACCTGGTATTCCATTTGGTCCTGCGGGTCCTCTGAATC-3'
9	5'-CCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGTGGAGC-3'
10	5'-TGGAACTCCATCTCTGCCGGGTTCTCCTGCTGCTCCACGTGGTCCGGCCGGTCCGGGCGCAC-3'

e. Oligonucleotides used for primary fragment 5 of rCol-0G:

Oligo #	Sequence
1	5'-CCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGTGGAGC-3'
2	5'-TGGAACTCCATCTCTGCCGGGTTCTCCTGCTGCTCCACGTGGTCCGGCCGGTCCGGGCGCAC-3'
3	5'-AGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTT-3'
4	5'-GGTTTACCATCAGATCCAGGACCACCTGGCGAACCTGGCATACCTCTCATTCCTGGACCTCC-3'
5	5'-CGCCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGC-3'
6	5'-TTGGTCCTGAGGGGCCAGGAGGACCGGGTCTGCCTGACTCACCTTGGCTTCCCGGTGGACCG-3'
7	5'-AGACCCGGTCCTCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGC-3'
8	5'-TTTTACCAGGTGCGCCATCGTTGCCTTTTGGGCCAGGAAAGCCCATAACTCCTGGTTGACCCC-3'
9	5'-CCAAAAGGCAACGATGGCGCACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGG-3'
10	5'-TTCACCATTCTTTCCAGGAGGTCCTTGTGGGCCGGGACCACCGGGTCCGCCTCTTTCACCAT-3'

f. Oligonucleotides used for primary fragment 6 of rCol-0G:

Oligo #	Sequence
1	5'-CCAAAAGGCAACGATGGCGCACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGG-3'
2	5'-TTCACCATTCTTTCCAGGAGGTCCTTGTGGGCCGGGACCACCGGGTCCGCCTCTTTCACCAT-3'
3	5'-CCCACAAGGACCTCCTGGAAAGAATGGTGAAACAGGACCACAAGGGCCACCCGGCCCAACAG-3'
4	5'-GGTCCGGTATCACCTTTGTCTCCACCTGGACCTGTTGGGCCGGGTGGCCCTTGTGGTCCTGT-3'
5	5'-GTCCAGGTGGAGACAAAGGTGATACCGGACCACCTGGACCCCAAGGATTGCAAGGACTACCT-3'
6	5'-TACCATTCTCTCCGGGAGGTCCTCCTGTACCAGGTAGTCCTTGCAATCCTTGGGGTCCAGGT-3'
7	5'-GGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACCTGGTCCAAAAGGTGATGCT-3'
8	5'-CACCTTTTCCACCCGGAGCTCCTGGAGCTCCAGCATCACCTTTTGGACCAGGTTCACCTGGTT-3'
9	5'-GGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGAAAAGGTCCACC-3'
10	5'-TCCTCTCAATCCTGGTGCTCCTGCCAAACCTGGTGGACCTTTTCCACCAGGACTACCTGCAT-3'

g. Oligonucleotides used for primary fragment 7 of rCol-0G:

Oligo #	Sequence
1	5'-GGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGAAAAGGTCCACC-3'
2	5'-TCCTCTCAATCCTGGTGCTCCTGCCAAACCTGGTGGACCTTTTCCACCAGGACTACCTGCAT-3'
3	5'-AGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGGTA-3'
4	5'-CCAGGAGGACCAGGTGGTCCTGCAGCTCCTTTACCTCCTTCGGGACCAGGTGGACCAGCTCC-3'
5	5'-AAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCG-3'
6	5'-CGGGTGAACCCAAACCGCCTTTGCCACCAGGCGAACCTTGCAAACCTGGAGTCCCGGCAGCA-3'
7	5'-CCTGGTGGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGT-3'
8	5'-CATCTTTACCGGGAACACCGTCTGCGCCAGGACCACCAGGTTCACCCTTATCACCTTTTGGTC-3'
9	5'-CCTGGCGCAGACGGTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACC-3'
10	5'-TTTATCACCAGGCTGACCTGCGGGTCCGGGAGGTCCAATTGGTCCAGTTGGTCCTCTGGGAC-3'

h. Oligonucleotides used for primary fragment 8 of rCol-0G:

Oligo #	Sequence
1	5'-CCTGGCGCAGACGGTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACC-3'
2	5'-TTTATCACCAGGCTGACCTGCGGGTCCGGGAGGTCCAATTGGTCCAGTTGGTCCTCTGGGAC-3'
3	5'-TCCCGGACCCGCAGGTCAGCCTGGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAA-3'
4	5'-CTTTCACCTGGACTACCCCTTGGGCCAGCAATTCCTGGTAATCCGGGTGCACCACCTTCACC-3'
5	5'-TTGCTGGCCCAAGGGGTAGTCCAGGTGAAAGGGGAGAAACAGGTCCACCAGGCCCAGCTGGA-3'
6	5'-GCTCACCATTTTGTCCAGGGGCACCTGGGAATCCAGCTGGGCCTGGTGGACCTGTTTCTCCC-3'
7	5'-TTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGTAAAGGTGAAAGAGGGGCTCCAGGC-3'
8	5'-CAACTCCTGGTGGTCCACCCTCTCCTTTTTCGCCTGGAGCCCCTCTTTCACCTTTACCACCAG-3'
9	5'-GAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGGTTCCGGACCAGCAGG-3'
10	5'-TTCTCCTTTAACACCCTGGGGACCCGGAGGTCCTGCTGGTCCGGAACCTCCTGGTGGACCTG-3'

i. Oligonucleotides used for primary fragment 9 of rCol-0G:

Oligo #	Sequence
1	5'-GAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGGTTCCGGACCAGCAGG-3'
2	5'-TTCTCCTTTAACACCCTGGGGACCCGGAGGTCCTGCTGGTCCGGAACCTCCTGGTGGACCTG-3'
3	5'-ACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAGCCGCTG-3'
4	5'-GGTCCTGGCAATCCCCTAGCTCCTGGAAAACCAGCGGCTCCTGGACCTCCAGGAGAACCTCT-3'
5	5'-GTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCC-3'
6	5'-GGACCGTCTTTACCAGGAGATCCAGAAGGACCGGGTGGTCCAGGATTACCATTAGATCCGGGC-3'
7	5'-GGTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCG-3'
8	5'-TCTCCCTTTGGTCCAGATACACCGGGTGAACCCGGTGCACCAGTATTTCCCGCCGGTCCAGGT-3'
9	5'-GGTTCACCCGGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCA-3'
10	5'-GCCTGGTGCGCCGGGCGGACCCTGAGCTCCTGGTGAACCTTTTTCACCGGGTTGACCAGCA-3'

j. Oligonucleotides used for primary fragment 10 of rCol-0G:

Oligo #	Sequence
1	5'-GGTTCACCCGGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCA-3'
2	5'-GCCTGGTGCGCCGGGCGGACCCTGAGCTCCTGGTGAACCTTTTTCACCGGGTTGACCAGCA-3'
3	5'-GGAGCTCAGGGTCCGCCCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGG-3'
4	5'-ACCTGGCATGCCTGGCGGTCCAGCCAAGCCCCTAGCACCAGTAATTCCAGCTATTCCCAACGG-3'
5	5'-GGCTTGGCTGGACCGCCAGGCATGCCAGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAA-3'
6	5'-ACCATTAGCACCTGGTTTTCCTGACTCGCCTTTAACGCCTTGTGGACCTGGACTGCCCCTAGG-3'
7	5'-GGCGAGTCAGGAAAACCAGGTGCTAATGGTTTATCGGGAGAAAGAGGTCCGCCTGGGCCTCAG-3'
8	5'-ACCAGCAGTACCAGCCAATCCAGGCAAACCCTGAGGCCCAGGCGGACCTCTTTCTCCCGATAA-3'
9	5'-GGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGAGATGGTAATCCCGGTTCTGAT-3'
10	5'-ACCAGGGGATCCGTCCCTGCCGGGTAATCCATCAGAACCGGGATTACCATCTCTACCTGGCTC-3'

k. Oligonucleotides used for primary fragment 11 of rCol-0G:

Oligo #	Sequence
1	5'-GGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGAGATGGTAATCCCGGTTCTGAT-3'
2	5'-ACCAGGGGATCCGTCCCTGCCGGGTAATCCATCAGAACCGGGATTACCATCTCTACCTGGCTC-3'
3	5'-GGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTGAAAATGGTTCCCCA-3'
4	5'-TCCCGGATGCCCTGGAGCACCGGGAGCACCTGGGGAACCATTTTCACCTCTATCTCCCTTGCC-3'
5	5'-GGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGGCAAATC-3'
6	5'-ACCTGCAGGACCTGACTCTCCTCTATCACCAGATTTGCCGGCTGGTCCTACAGGACCGGGAGG-3'
7	5'-TGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCT-3'
8	5'-ACCTCTAGGGCCTTGGGGACCTGGTGCACCCCTAGATCCGGCGGGGCCAGGAGCTCCTGCGGG-3'
9	5'-AGGGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCT-3'
10	5'-ACCTGGAAAGCCTCTATGACCTTTAATGCCCGCAGCACCTCTCTCACCTGTTTCACCCTTGTC-3'

l. Oligonucleotides used for primary fragment 12 of rCol-0G:

Oligo #	Sequence
1	5'-AGGGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCT-3'
2	5'-ACCTGGAAAGCCTCTATGACCTTTAATGCCCGCAGCACCTCTCTCACCTGTTTCACCCTTGTC-3'
3	5'-GCGGGCATTAAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCA-3'
4	5'-GCCAGGACTACCAATAGCACCTTGTTGACCAGCTGGTCCAGGACTACCAGGTGCACCAGGATT-3'
5	5'-GCTGGTCAACAAGGTGCTATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCA-3'
6	5'-ACCAGAAGTTCCATCTTTACCAGGTGGTCCGCTTGGTCCAACTGGTCCACGAGGACCTGCAGG-3'
7	5'-AGCGGACCACCTGGTAAAGATGGAACTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCA-3'
8	5'-ACCTTCAGAGCCTCTTTCTCCTCTATTGCCCCTTGGTCCAGGAGGACCTATTGGACCTGGATG-3'
9	5'-AGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAGCCCTGGTCATCCAGGACAGCCAGGACCT-3'
10	5'-ACAGGGCCCCGGAGCACCTGGTGGTCCAGGAGGTCCTGGCTGTCCTGGATGACCAGGGCT-3'
11	5'-CCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTGGTGCTGCTGCTATTGCT-3'
12	5'-GAATCCACCAGCTTTTTCACCGCCGATGCCAGCAATAGCAGCAGCACCAACACCACCACA-3'

m. Oligonucleotides used for primary fragment 1 of rCol-1G-1:

Oligo #	Sequence
1	5'-CCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCTGTTGGT-3'
2	5'-ACCAGCGGGACCTGGATAACCAGCTAGCCCACCAACAGCAACACCAGATTTGACGTCGTA-3'
3	5'-GGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACA-3'
4	5'-TGGTGAACCGGGTGAGCCAGGATGTCCAGATGTACCTGGTGGACCAGGCGGTCCGGGAGG-3'
5	5'-TCTGGACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACA-3'
6	5'-TGGCGGACCAGGAGGACCACTTGGACCAGCTTGTCCGGGCTCACCAGGTGGACCTTGATAACC-3'
7	5'-AGCTGGTCCAAGTGGTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTA-3'
8	5'-GGAAAGCCTCTTTCGCCAGATTCACCATCCTTACCGGCTGGACCAGATGGACCGATAGCTCC-3'
9	5'-AGGATGGTGAATCTGGCGAAAGAGGCTTTCCTGGTGAAAGAGGTGTTCAGGGTCCACCTGGA-3'
10	5'-AAAACCTGGTATTCCCGCAGGGCCTTTTATTCCAGGTGGACCCTGAACACCTCTTTCACCA-3'
11	5'-ATAAAAGGCCCTGCGGGAATACCAGGTTTTCCTGGCATGAAAGGTCACAGAGGTTTTGATGGT-3'
12	5'-AGCACCAGTTTCACCCTTTTCGCCATTTCTACCATCAAAACCTCTGTGACCTTTCATGCCAGG-3'

Table A3-2: Primers used in site-directed mutagenesis of the collagen variants.

Mutation Primers		
Primary Fragment	**Sequence**	
PF4G	5'- GTTTTCCAGGAGAAAGAGGCGTTCAGGGATTCAGAGGACCCGCAGGACCAAATGGAATAC	-3'
	5'- CCTCTTTCTCCTGGAAAACCCCTCTCACCATTAGCCCCAGGCTCACCAGGGGAACCATC	-3'
PF8G	5'- GTTTTCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTG	-3'
	5'- CGTTCACCGGGAAAACCTCTTTCTCCATTTTGTCCAGGGGCACCTGGGAATCCAGCTGGG	-3'
PF11G	5'- CGTGGCTTTCCAGGAGAGAGAGGTGTTCAAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGC	-3'
	5'- CTCTCTCCTGGAAAGCCACGTTCGCCTACAGGACCGGGAGGTCCCGGATGCCCTGGAGC	-3'

A.2 Collagen Variant DNA and Protein Sequences

A.2.1 Cysteine Variant DNA Sequences

A.2.1.1 rCol-2C DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGCAGACCTGGCGAAAGGGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGTTTA
AAAGGAGAAAATGGCTTGCCAGGCGAAAATGGAGCTCCAGGCCCAATGGGTCCAAGGGGTGCTCCAGGAG
AGAGGGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAATGTGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGCGCGCCAGGTGAA
CGTGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGG
TAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTATGCCAGGAG
AGAGAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTTGTCCTGGCGCAGACGG
TGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCTG
GTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAAG
GGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGTA
AAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGG
TTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAG
CCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGGT
CCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCGG
TGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCCG

GCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCCA
GGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTAT
CGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGA
GATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTG
AAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGGC
AAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGGG
TGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAAA
GGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTAT
TGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAACTT
CTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAG
CCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTG
GTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCCC
ATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCGT
TGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGAA
TTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGTA
ACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTGAT
AGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGGTA
ATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGTCG
CAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAAGG
CATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACTGT
TCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGAAG
GCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGTTG
ATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.1.2 rCol-4C DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAtgTGGTCCAAGTGG
TCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCAG
GCAGACCTGGCGAAAGGGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCCT
GGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGTTTAA
AAGGAGAAAATGGCTTGCCAGGCGAAAATGGAGCTCCAGGCCCAATGGGTCCAAGGGGTGCTCCAGGAGA
GAGGGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGGA
CAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTGG
TCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCTG
GTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTtGtCCAGGAGGTAAAGGCGAAATGGGTCCTGCCG
GAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTCCA
GGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGAAC

GTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGTGA
GCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAGGAC
CAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGTGG
AGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCGCCA
GGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCCTCC
TGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCGCAC
CTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGGTGA
AACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTGGAC
CCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACCTGG
TCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGCGCGCCAGGTGAACGT
GGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGGTA
AAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTATGCCAGGAGAG
AGAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACGGTG
TTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCTGGT
GATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAAGGG
GAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGTAAA
GGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGGTT
CCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAGCC
GCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGGTCC
TTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCGGTG
TATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCCGGC
GCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCCAGG
TCCTAGGGGCtGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTATCGG
GAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGAGA
TGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTGAA
AATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGGCAA
ATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGGGTG
CACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAAAGG
TCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTATTG
GTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAtgTGGAACTTCTG
GTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAGCCC
TGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTGGTG
CTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCCCATG
GACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCGTTGA
TTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGAATTA
AAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGTAACA
TGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTGATAGT
TCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGGTAATC
CAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGTCGCAG
AATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAAGGCAT
TAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACTGTTCT
GGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGAAGGCT

GTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGTTGATG
TTGGTCCGGTTTGTTTCTTGtaa

A.2.1.3 rCol-6C DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAtgTGGTCCAAGTGG
TCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCAG
GCAGACCTGGCGAAAGGGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCCT
GGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGTTTAA
AAGGAGAAAATGGCTTGCCAGGCGAAAATGGAGCTCCAGGCCCAATGGGTCCAAGGGGTGCTCCAGGAGA
GAGGGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGGA
CAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTGG
TCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCTG
GTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTtGtCCAGGAGGTAAAGGCGAAATGGGTCCTGCCG
GAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTCCA
GGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGAAC
GTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGTGA
GCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAGGAC
CAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGTGG
AGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCGCCA
GGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCCTCC
TGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCGCAC
CTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGGTGA
AACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATtgtGGACCACCTGGACC
CCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACCTGGT
CCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGCGCGCCAGGTGAACGTG
GTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGGTAA
AGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTATGCCAGGAGAGA
GAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACGGTGT
TCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCTGGTG
ATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTtGTCCAGGTGAAAGGGGA
GAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGTAAAGG
TGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGGTTCC
GGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAGCCGC
TGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGGTCCTT
CTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCGGTGTA
TCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCCGGCGC

ACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCCAGGTC
CTAGGGGCtGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTATCGGGA
GAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGAGATG
GTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTGAAAA
TGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGGCAAAT
CTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGGGTGCA
CCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAAAGGTC
ATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTATTGGT
AGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAtgTGGAACTTCTGGT
CATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAGCCCTG
GTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTGGTGCT
GCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCCCATGGA
CTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCGTTGATT
TCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGAATTAAA
ATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGTAACATG
GAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTGATAGTTC
TGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGGTAATCCA
GAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGTCGCAGA
ATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAAGGCATT
AAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACTGTTCTG
GAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGAAGGCTG
TGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGTTGATGT
TGGTCCGGTTTGTTTCTTGtaa

A.2.1.4 rCol-8C DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAtgTGGTCCAAGTGG
TCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCAG
GCAGACCTGGCGAAAGGGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCCT
GGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGTTTAA
AAGGAGAAAATGGCTTGCCAGGCGAAAATGGAGCTCCAGGCCCAATGGGTCCAAGGGGTGCTCCAGGAGA
GAGGGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGGA
CAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTGG
TCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCTG
GTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTtGtCCAGGAGGTAAAGGCGAAATGGGTCCTGCCG
GAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTCCA
GGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGAAC

GTGGAGAAtgtGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGTGAG
CCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAGGACC
AAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGTGGA
GCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCGCCAG
GTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCCTCCT
GGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCGCACC
TGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGGTGAA
ACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATtgtGGACCACCTGGACCC
CAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACCTGGTC
CAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGCGCGCCAGGTGAACGTGG
TCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGGTAAA
GGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTATGCCAGGAGAGAG
AGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACGGTGTTC
CCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCTGGTGAT
AAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTtGTCCAGGTGAAAGGGGAG
AAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGTAAAGGT
GAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGGTTCCG
GACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAGCCGCT
GGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGGTCCTTC
TGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCGGTGTAT
CTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCCGGCGCA
CCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCCAGGTCC
TAGGGGCtGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTATCGGGAG
AAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGAGATGGT
AATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTGAAAATG
GTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTtgtGGACCAGCCGGCAAATCTGG
TGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGGGTGCACCAG
GTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAAAGGTCATAG
AGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTATTGGTAGTC
CTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAtgTGGAACTTCTGGTCATC
CAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAGCCCTGGTCA
TCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTGGTGCTGCTG
CTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCCCATGGACTTT
AAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCGTTGATTTCAC
CCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGAATTAAAATCC
GGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGTAACATGGAAA
CTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTGATAGTTCTGCT
GAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGGTAATCCAGAAT
TGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGTCGCAGAATATA
ACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAAGGCATTAAAGT
TGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACTGTTCTGGAGGA
TGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGAAGGCTGTGAGA

TTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGTTGATGTTGGTCC
GGTTTGTTTCTTGtaa

A.2.2 Cysteine Variant Protein Sequences

A.2.2.1 rCol-2C Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGL
PGENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGECGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGCPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.2.2 rCol-4C Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGL
PGENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDK
GEGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGP
AGPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSG
PKGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGER
GPPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG

PAGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.2.3 rCol-6C Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGL
PGENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDCGPPGPQGL
QGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPPGLAGAPGLRGGAGPPGPEGGKGAA
GPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKGE
GGAPGLPGIAGPRGCPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPAG
PPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGPK
GDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGERGP
PGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDR
GESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.2.4 rCol-8C Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGRPGRPGERGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGLKGENGL
PGENGAPGPMGPRGAPGERGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGECGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDCGPPGPQGL
QGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGAPGERGPPGLAGAPGLRGGAGPPGPEGGKGAA
GPPGPPGAAGTPGLQGMPGERGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKGE
GGAPGLPGIAGPRGCPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPAG
PPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGPK
GDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGERGP

PGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPCGPAGKSGDR
GESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.3 GFOGER Variant DNA Sequences

A.2.3.1 rCol-0G DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT

GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGT
AAAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAG
GTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGA
GCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGG
TCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCG
GTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCC
GGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCC
AGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTA
TCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAG
AGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGT
GAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGG
CAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGG
GTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAA
AGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTA
TTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAACT
TCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTA
GCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTT
GGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCC
CATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCG
TTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGA
ATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGT
AACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTG
ATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGG
TAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGT
CGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAA
GGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACT
GTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGA
AGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGT
TGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.2 rCol-1G-1 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGCGAAAGA
GGCTTTCCTGGTGAAAGAGGTGTTCAGGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG

CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGT
AAAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAG
GTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGA
GCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGG
TCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCG
GTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCC
GGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCC
AGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTA
TCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAG
AGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGT
GAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGG
CAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGG
GTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAA
AGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTA
TTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAACT
TCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTA
GCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTT
GGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCC
CATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCG
TTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGA
ATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGT

AACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTG
ATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGG
TAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGT
CGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAA
GGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACT
GTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGA
AGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGT
TGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.3 rCol-1G-8 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA

GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTT
TCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGA
GGTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGG
AGCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCG
GTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCC
GGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGC
CCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATG
CCAGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTT
TATCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGT
AGAGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAG
GTGAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCC
GGCAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAG
GGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATT
AAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGC
TATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAA
CTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGG
TAGCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTG
TTGGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGA
GCCCATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAA
TCGTTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCC
CGAATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTC
TGTAACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGAC
TGATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTAC
GGTAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGC
GTCGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAG
AAGGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATA
CTGTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAG
GAAGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGG
GTTGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.4 rCol-1G-11 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG

GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGT
AAAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAG
GTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGA
GCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGG
TCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCG
GTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCC
GGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCC
AGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTA
TCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAG
AGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGT
GAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGCGAACGTGG
CTTTCCAGGAGAGAGAGGTGTTCAAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGG
GTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAA
AGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTA
TTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAACT
TCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTA
GCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTT
GGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCC
CATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCG
TTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGA
ATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGT
AACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTG

ATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGG
TAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGT
CGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAA
GGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACT
GTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGA
AGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGT
TGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.5 rCol-2G-1,4 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGGGAAAGA
GGTTTCCCTGGTGAAAGGGGTGTTCAGGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGTGAGAGGGGTTTTCCAGGAGAAAGAGGCGTTCAGGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGT

AAAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAG
GTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGA
GCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGG
TCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCG
GTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCC
GGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCC
AGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTA
TCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAG
AGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGT
GAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGG
CAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGG
GTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAA
AGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTA
TTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAACT
TCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTA
GCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTT
GGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCC
CATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCG
TTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGA
ATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGT
AACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTG
ATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGG
TAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGT
CGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAA
GGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACT
GTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGA
AGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGT
TGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.6 rCol-2G-1,8 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGGGAAAGA
GGTTTCCCTGGTGAAAGGGGTGTTCAGGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG

ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTT
TCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGA
GGTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGG
AGCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCG
GTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCC
GGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGC
CCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATG
CCAGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTT
TATCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGT
AGAGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAG
GTGAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCC
GGCAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAG
GGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATT
AAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGC
TATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAA
CTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGG
TAGCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTG
TTGGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGA
GCCCATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAA
TCGTTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCC
CGAATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTC
TGTAACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGAC
TGATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTAC

GGTAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGC
GTCGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAG
AAGGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATA
CTGTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAG
GAAGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGG
GTTGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.7 rCol-2G-8,11 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTT
TCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGA

GGTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGG
AGCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCG
GTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCC
GGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGC
CCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATG
CCAGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTT
TATCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGT
AGAGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAG
GTGAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGCGAACGT
GGCTTTCCAGGAGAGAGAGGTGTTCAAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAG
GGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATT
AAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGC
TATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAA
CTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGG
TAGCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTG
TTGGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGA
GCCCATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAA
TCGTTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCC
CGAATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTC
TGTAACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGAC
TGATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTAC
GGTAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGC
GTCGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAG
AAGGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATA
CTGTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAG
GAAGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGG
GTTGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.3.8 rCol-4G DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAAGCTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGGGAAAGA
GGTTTCCCTGGTGAAAGGGGTGTTCAGGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG

GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTAGCCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGTGAGAGGGGTTTTCCAGGAGAAAGAGGCGTTCAGGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTT
TCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGA
GGTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGG
AGCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCG
GTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCC
GGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGC
CCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATG
CCAGGTCCTAGGGGCAGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTT
TATCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGT
AGAGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAG
GTGAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGCGAACGT
GGCTTTCCAGGAGAGAGAGGTGTTCAAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAG
GGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATT
AAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGC
TATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAGATGGAA
CTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGG
TAGCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTG
TTGGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGA
GCCCATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAA
TCGTTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCC
CGAATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTC
TGTAACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGAC
TGATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTAC
GGTAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGC

GTCGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAG
AAGGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATA
CTGTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAG
GAAGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGG
GTTGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.4 GFOGER Variant Protein Sequences

A.2.4.1 rCol-0G Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.2 rCol-1G-1 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGERGFPGERGVQGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP

KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.3 rCol-1G-8 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.4 rCol-1G-11 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG

EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGERGFPGER
GVQGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.5 rCol-2G-1,4 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGERGFPGERGVQGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGERGFPGERGVQGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.6 rCol-2G-1,8 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGERGFPGERGVQGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG

LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.7 rCol-2G-8,11 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGERGFPGER
GVQGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.4.8 rCol-4G Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQAGPSGPPGP
PGAIGPSGPAGKDGESGERGFPGERGVQGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGSPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGERGFPGERGVQGFRGPAGPNGIP

GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGERGFPGER
GVQGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKDGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.5 Cysteine-GFOGER Variant DNA Sequences

A.2.5.1 rCol-4C0G DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAATGTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTTGTCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG

TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGTGAGCCTGGTGGT
AAAGGTGAAAGAGGGGCTCCAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAG
GTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGA
GCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGG
TCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCG
GTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCC
GGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCC
AGGTCCTAGGGGCTGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTA
TCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAG
AGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGT
GAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCCGG
CAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGG
GTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAA
AGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTA
TTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAATGTGGAACT
TCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTA
GCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTT
GGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCC
CATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCG
TTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGA
ATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGT
AACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTG
ATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGG
TAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGT
CGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAA
GGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACT
GTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGA
AGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGT
TGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.5.2 rCol-4C1G-8 DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG

TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAATGTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGTAGACCA
GGGAGTCCAGGCGGCAAAGGCTTGCCAGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTTGTCCAGGAGGTAAAGGCGAAATGGGTCCTGC
CGGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTC
CAGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGA
ACGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGT
GAGCCTGGGGCTAATGGATTGCCTGGTGCTGCAGGCGAAAGAGGAGCTCCTGGATTCAGAGGACCCGCAG
GACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACG
TGGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCG
CCAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCC
TCCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCG
CACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGG
TGAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTG
GACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACC
TGGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGA
AAAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAG
GTAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGT
GGCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACG
GTGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCT
GGTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAA
GGGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTT
TCCCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGA
GGTTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGG
AGCCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCG
GTCCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCC
GGTGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGC
CCGGCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATG
CCAGGTCCTAGGGGCTGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTT
TATCGGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGT
AGAGATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAG
GTGAAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGACCAGCC
GGCAAATCTGGTGATAGAGGAGAGTCAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAG
GGGTGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATT
AAAGGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGC

TATTGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAATGTGGAA
CTTCTGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGG
TAGCCCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTG
TTGGTGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGA
GCCCATGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAA
TCGTTGATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCC
CGAATTAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTC
TGTAACATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGAC
TGATAGTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTAC
GGTAATCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGC
GTCGCAGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAG
AAGGCATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATA
CTGTTCTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAG
GAAGGCTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGG
GTTGATGTTGGTCCGGTTTGTTTCTTGtaa

A.2.5.3 rCol-4C4G DNA Sequence

ATGATGTCTTTCGTTCAAAAGGGGAGCTGGTTGTTGTTGGCTCTACTCCACCCGACCATTATTTTGGCCCA
GCAGGAAGCTGTTGAAGGTGGATGTTCTCATTTGGGTCAAAGCTACGCGGATAGAGATGTTTGGAAGCCA
GAACCATGTCAAATCTGTGTCTGTGATAGCGGTTCGGTCCTGTGTGATGACATTATTTGTGACGATCAGG
AGCTCGACTGTCCTAACCCTGAGATCCCATTCGGTGAGTGTTGTGCTGTTTGTCCGCAACCGCCCACTGCT
CCAACTCGTCCTCCCAATGGTCAAGGTCCCCAAGGTCCAAAAGGTGATCCGGGGCCACCAGGCATACCAGG
TAGAAATGGTGACCCAGGAATACCCGGACAGCCAGGTTCCCCAGGGAGTCCAGGACCTCCCGGAATTTGTG
AGTCTTGTCCAACTGGACCCCAGAACTACAGCCCGCAATACGATTCTTACGACGTCAAATCTGGTGTTGCT
GTTGGTGGGCTAGCTGGTTATCCAGGTCCCGCTGGTCCTCCCGGACCGCCTGGTCCACCAGGTACATCTGG
ACATCCTGGCTCACCCGGTTCACCAGGTTATCAAGGTCCACCTGGTGAGCCCGGACAATGTGGTCCAAGTG
GTCCTCCTGGTCCGCCAGGAGCTATCGGTCCATCTGGTCCAGCCGGTAAGGATGGTGAATCTGGCGAAAGA
GGCTTCCCTGGTGAAAGAGGTGTTCAGGGTCCACCTGGAATAAAAGGCCCTGCGGGAATACCAGGTTTTCC
TGGCATGAAAGGTCACAGAGGTTTTGATGGTAGAAATGGCGAAAAGGGTGAAACTGGTGCTCCTGGATCG
CCAGGTGGTAAGGGTTCGCCAGGAGGAAAGGGAGCTCCAGGCCCAATGGGTCCAAGGGGTTCACCAGGCG
GTAAAGGTAGACCTGGTTTGCCAGGAGCAGCCGGTGCCCGTGGTAATGATGGTGCTAGAGGATCTGACGG
ACAACCTGGCCCACCTGGACCACCTGGAACAGCTGGCTTTCCTGGCAGTCCCGGCGCCAAAGGTGAAGTTG
GTCCAGCAGGTTCACCCGGTTCAAATGGTGCTCCTGGACAAAGGGGAGAACCCGGACCTCAGGGTCATGCT
GGTGCTCAAGGTCCTCCCGGTCCACCCGGAATTAACGGTtGtCCAGGAGGTAAAGGCGAAATGGGTCCTGCC
GGAATACCTGGTGCTCCAGGCTTAATGGGTGCGAGAGGCCCTCCTGGTCCAGCTGGAGCTAACGGTGCTCC
AGGTCTAAGAGGTGGAGCCGGTGAACCAGGAAAGAATGGTGCTAAAGGCGAACCCGGACCAAGGGGAGAA
CGTGGAGAAGCCGGCATACCCGGAGTACCCGGTGCTAAAGGTGAGGATGGCAAAGATGGTTCCCCTGGTG
AGCCTGGGGCTAATGGTGAGAGGGGTTTTCCAGGAGAAAGAGGCGTTCAGGGATTCAGAGGACCCGCAGG
ACCAAATGGAATACCAGGTGAAAAAGGCCCAGCAGGTGAAAGAGGTGCGCCCGGACCGGCCGGACCACGT
GGAGCAGCAGGAGAACCCGGCAGAGATGGAGTTCCAGGAGGTCCAGGAATGAGAGGTATGCCAGGTTCGC
CAGGTGGTCCTGGATCTGATGGTAAACCCGGTCCACCGGGAAGCCAAGGTGAGTCAGGCAGACCCGGTCCT
CCTGGCCCCTCAGGACCAAGGGGTCAACCAGGAGTTATGGGCTTTCCTGGCCCAAAAGGCAACGATGGCGC
ACCTGGTAAAAATGGTGAAAGAGGCGGACCCGGTGGTCCCGGCCCACAAGGACCTCCTGGAAAGAATGGT
GAAACAGGACCACAAGGGCCACCCGGCCCAACAGGTCCAGGTGGAGACAAAGGTGATACCGGACCACCTGG

ACCCCAAGGATTGCAAGGACTACCTGGTACAGGAGGACCTCCCGGAGAGAATGGTAAACCAGGTGAACCT
GGTCCAAAAGGTGATGCTGGAGCTCCAGGAGCTCCGGGTGGAAAAGGTGATGCAGGTAGTCCTGGTGGAA
AAGGTCCACCAGGTTTGGCAGGAGCACCAGGATTGAGAGGAGGAGCTGGTCCACCTGGTCCCGAAGGAGG
TAAAGGAGCTGCAGGACCACCTGGTCCTCCTGGTGCTGCCGGGACTCCAGGTTTGCAAGGTTCGCCTGGTG
GCAAAGGCGGTTTGGGTTCACCCGGACCAAAAGGTGATAAGGGTGAACCTGGTGGTCCTGGCGCAGACGG
TGTTCCCGGTAAAGATGGTCCCAGAGGACCAACTGGACCAATTGGACCTCCCGGACCCGCAGGTCAGCCTG
GTGATAAAGGTGAAGGTGGTGCACCCGGATTACCAGGAATTGCTGGCCCAAGGGGTAGTCCAGGTGAAAG
GGGAGAAACAGGTCCACCAGGCCCAGCTGGATTCCCAGGTGCCCCTGGACAAAATGGAGAAAGAGGTTTTC
CCGGTGAACGTGGGGTACAAGGCGAAAAAGGAGAGGGTGGACCACCAGGAGTTGCAGGTCCACCAGGAGG
TTCCGGACCAGCAGGACCTCCGGGTCCCCAGGGTGTTAAAGGAGAAAGAGGTTCTCCTGGAGGTCCAGGAG
CCGCTGGTTTTCCAGGAGCTAGGGGATTGCCAGGACCGCCCGGATCTAATGGTAATCCTGGACCACCCGGT
CCTTCTGGATCTCCTGGTAAAGACGGTCCACCTGGACCGGCGGGAAATACTGGTGCACCGGGTTCACCCGG
TGTATCTGGACCAAAGGGAGATGCTGGTCAACCCGGTGAAAAAGGTTCACCAGGAGCTCAGGGTCCGCCCG
GCGCACCAGGCCCGTTGGGAATAGCTGGAATTACTGGTGCTAGGGGCTTGGCTGGACCGCCAGGCATGCCA
GGTCCTAGGGGCtGTCCAGGTCCACAAGGCGTTAAAGGCGAGTCAGGAAAACCAGGTGCTAATGGTTTATC
GGGAGAAAGAGGTCCGCCTGGGCCTCAGGGTTTGCCTGGATTGGCTGGTACTGCTGGTGAGCCAGGTAGA
GATGGTAATCCCGGTTCTGATGGATTACCCGGCAGGGACGGATCCCCTGGTGGCAAGGGAGATAGAGGTG
AAAATGGTTCCCCAGGTGCTCCCGGTGCTCCAGGGCATCCGGGACCTCCCGGTCCTGTAGGCGAACGTGGC
TTTCCAGGAGAGAGAGGTGTTCAAGGTCCTGCAGGTCCCGCAGGAGCTCCTGGCCCCGCCGGATCTAGGGG
TGCACCAGGTCCCCAAGGCCCTAGAGGTGACAAGGGTGAAACAGGTGAGAGAGGTGCTGCGGGCATTAAA
GGTCATAGAGGCTTTCCAGGTAATCCTGGTGCACCTGGTAGTCCTGGACCAGCTGGTCAACAAGGTGCTAT
TGGTAGTCCTGGCCCTGCAGGTCCTCGTGGACCAGTTGGACCAAGCGGACCACCTGGTAAAtgTGGAACTTC
TGGTCATCCAGGTCCAATAGGTCCTCCTGGACCAAGGGGCAATAGAGGAGAAAGAGGCTCTGAAGGTAGC
CCTGGTCATCCAGGACAGCCAGGACCTCCTGGACCACCAGGTGCTCCGGGGCCCTGTTGTGGTGGTGTTGG
TGCTGCTGCTATTGCTGGCATCGGCGGTGAAAAAGCTGGTGGATTCGCTCCATATTATGGTGATGAGCCCA
TGGACTTTAAGATTAATACGGATGAGATCATGACGTCCCTGAAATCGGTCAATGGGCAAATTGAATCGTT
GATTTCACCCGATGGTTCCAGAAAGAATCCAGCTAGGAACTGTCGTGACCTAAAATTCTGTCACCCCGAAT
TAAAATCCGGTGAATATTGGGTCGATCCAAATCAAGGCTGTAAGTTGGACGCTATTAAGGTGTTCTGTAA
CATGGAAACTGGCGAAACGTGTATTTCTGCGAATCCCCTCAATGTTCCTAGAAAACACTGGTGGACTGATA
GTTCTGCTGAGAAGAAGCATGTTTGGTTCGGTGAATCCATGGACGGTGGTTTTCAATTTAGCTACGGTAA
TCCAGAATTGCCAGAGGATGTTTTGGACGTTCAATTGGCATTCTTGCGTTTGTTGTCATCAAGGGCGTCGC
AGAATATAACTTATCATTGTAAGAACTCCATCGCCTATATGGATCAAGCATCCGGCAACGTGAAGAAGGC
ATTAAAGTTGATGGGTTCTAATGAGGGCGAATTTAAGGCTGAAGGTAACAGCAAGTTCACCTATACTGTT
CTGGAGGATGGTTGTACTAAACACACGGGTGAATGGTCTAAGACGGTGTTCGAATACAGAACGAGGAAGG
CTGTGAGATTGCCCATTGTTGACATAGCTCCATACGATATAGGTGGTCCAGATCAAGAATTTGGGGTTGA
TGTTGGTCCGGTTTGTTTCTTGtaa

A.2.6 Cysteine-GFOGER Variant Protein Sequences

A.2.6.1 rCol-4C0G Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS

PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.6.2 rCol-4C1G-8 Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP
TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGRPGSPGGKGLPGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLPGAAGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGD
RGESGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPG
PAGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIA
GIGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYW
VDPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPE
DVLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCT
KHTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

A.2.6.3 rCol-4C4G Protein Sequence

MMSFVQKGSWLLLALLHPTIILAQQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQEL
DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCP

TGPQNYSPQYDSYDVKSGVAVGGLAGYPGPAGPPGPPGPPGTSGHPGSPGSPGYQGPPGEPGQCGPSGPPGP
PGAIGPSGPAGKDGESGERGFPGERGVQGPPGIKGPAGIPGFPGMKGHRGFDGRNGEKGETGAPGSPGGKGS
PGGKGAPGPMGPRGSPGGKGRPGLPGAAGARGNDGARGSDGQPGPPGPPGTAGFPGSPGAKGEVGPAGSP
GSNGAPGQRGEPGPQGHAGAQGPPGPPGINGCPGGKGEMGPAGIPGAPGLMGARGPPGPAGANGAPGLRG
GAGEPGKNGAKGEPGPRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGERGFPGERGVQGFRGPAGPNGIP
GEKGPAGERGAPGPAGPRGAAGEPGRDGVPGGPGMRGMPGSPGGPGSDGKPGPPGSQGESGRPGPPGPSG
PRGQPGVMGFPGPKGNDGAPGKNGERGGPGGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQG
LQGLPGTGGPPGENGKPGEPGPKGDAGAPGAPGGKGDAGSPGGKGPPGLAGAPGLRGGAGPPGPEGGKGA
AGPPGPPGAAGTPGLQGSPGGKGGLGSPGPKGDKGEPGGPGADGVPGKDGPRGPTGPIGPPGPAGQPGDKG
EGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGERGFPGERGVQGEKGEGGPPGVAGPPGGSGPA
GPPGPQGVKGERGSPGGPGAAGFPGARGLPGPPGSNGNPGPPGPSGSPGKDGPPGPAGNTGAPGSPGVSGP
KGDAGQPGEKGSPGAQGPPGAPGPLGIAGITGARGLAGPPGMPGPRGCPGPQGVKGESGKPGANGLSGERG
PPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRDGSPGGKGDRGENGSPGAPGAPGHPGPPGPVGERGFPGER
GVQGPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHRGFPGNPGAPGSPGPAGQQGAIGSPGP
AGPRGPVGPSGPPGKCGTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAPGPCCGGVGAAAIAG
IGGEKAGGFAPYYGDEPMDFKINTDEIMTSLKSVNGQIESLISPDGSRKNPARNCRDLKFCHPELKSGEYWV
DPNQGCKLDAIKVFCNMETGETCISANPLNVPRKHWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPED
VLDVQLAFLRLLSSRASQNITYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTK
HTGEWSKTVFEYRTRKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL

Appendix B: Protocols

Protocol: Collagen Expression

A. Prepare SDC(A):
   - A.1 Suspend in ddH_2O: 2% w/v Dextrose 0.5% w/v Casamino Acids, 0.5% w/v Ammonium Sulfate, 0.17% w/v Yeast Nitrogen Base w/o Amino Acids & Ammonium Sulfate, 0.01% w/v Adenine Hemisulfate
   - A.2 Sterile Filter (0.22 µm)

B. Prepare SGC(A):
   - B.1 Suspend in ddH_2O: 2% w/v Dextrose 0.5% w/v Casamino Acids, 0.5% w/v Ammonium Sulfate, 0.17% w/v Yeast Nitrogen Base w/o Amino Acids & Ammonium Sulfate, 0.01% w/v Adenine Hemisulfate, 0.02% Ascorbic Acid
   - B.2 Sterile Filter (0.22 µm)

1. Innoculate overnight culture; 5 mL of SDC(A) with a single colony.
2. Incubate overnight culture at 30 °C, 250 RPM overnight (~16 h)
3. Innoculate expansion culture; 11 mL of SDC(A) with overnight culture to bring to final OD_{600} of 0.25
4. Incubate expansion culture at 30 °C, 250 RPM until OD_{600} of 1 (~4 h)
5. Innoculate expression culture; 1 L of SGC(A) with 10 mL of expansion culture
6. Incubate expression culture at 30 °C, 250 RPM for 24 hours
7. Add 1 mL of sterile 20% Ascorbic acid to expression culture
8. Incubate expression culture at 30 °C, 250 RPM until OD_{600} of 3 (~16 h)
9. Transfer expression culture to 500 mL bottles
10. Centrifuge in Beckman Alegra-25R at 5500 RPM, 4 °C, Acceleration 9, Deceleration 1 for 10 minutes
11. Decant supernatant
12. Resuspend pellet in 40 ddH_2O, transfer to pre-weighed 50 mL conical tube
13. Centrifuge in Beckman Alegra-25R at 5500 RPM, 4 °C, Acceleration 9, Deceleration 9 for 5 minutes
14. Decant supernatant, aspirate all liquid
15. Weight tube with pellet to determine pellet weight
16. Freeze pellet in -80 °C freezer for at least 2 hours

Comments or other information:

- Prewarm all media to reduce incubation time required to reach proper OD
- If expansion culture passes OD_{600} of 2 restart expression protocol
- Decant solution immediately after centrifugation.

Protocol: Collagen Purification

A. Buffers / Solutions
   a. Breaking Buffer (100 mM Tris, 400 mM NaCl)
      i. Tris base (F.W. 121.14)
      ii. Sodium Chloride (F.W. 58.44)
      iii. 6 N HCl to pH 7.
      iv. ddH_2O
   b. 0.2 M PMSF
      i. Phenylmethanesulfonylfluoride (F.W. 174.19)
      ii. 100% Isopropano
      - NOTE: Store at -20°C
   c. 0.5 M EDTA pH 8.0
      i. Ethylenediamine Tetraacetic Acid, Disodium Salt, Dihydrate (F.W. 372.24)
      ii. Sodium Hydroxide (F.W. 40)
      iii. ddH_2O
   d. 1 M $MgCl_2$
      i. Magnesium Chloride Hexahydrate (F.W. 203.30)
      ii. ddH_2O
   e. 0.1 N HCl Solution
      i. ddH_2O
      ii. 12.1N HCl
   f. Pepsin (20 mg/mL)
      i. Pepsin
      ii. 0.1 N HCl
      - NOTE: Prepare fresh, right before use
   g. 5M NaCl Solution
      i. Sodium Chloride (F.W. 58.44)
      ii. ddH_2O
      iii. Slightly heat to dissolve if necessary

B. Break cells
   - NOTE: Volumes are given for a pellet size of 3 g
   a. Resuspend cell pellet to a concentration of 0.1-0.2 g/mL with Breaking Buffer
      i. Add DNase to 2 U/mL
      ii. Add RNase to 1 U/mL
      iii. Add $MgCl_2$ to 1 mM
      iv. Add PMSF to 1 mM (right before breaking)
   b. Freeze French Press in -20°C freezer for 30 minutes
   c. French Press at High Ratio setting at 1580 psi gauge reading twice (25000 Cell Pressure)
   d. Wait 30 min, add EDTA to 2 mM final concentration to the lysed cells.
   e. For variants containing non-native cysteines, add TCEP to 5 mM
   f. Repeat for each tube of cells
      - Note: French Press Cell should be chilled for 30 min in -20°C Freezer between uses

C. Pepsin Digest
    a. Lower pH from 7.4 to 2 using 6 N HCl
    b. Add Pepsin to a final concentration of 0.2 mg/mL
    c. Digest at 4 °C for 12 h with gentle agitation
    d. Centrifuge for 30 min at 30,000 RPM at 4°C using the Beckman Ti 45 rotor
    e. Transfer supernatant to a container for NaCl precipitation.
D. NaCl Precipitation
    a. Add 17 M Glacial acetic acid to a final concentration of 0.5 M Acetic Acid
    b. Add Solid NaCl to bring solution to a final concentration of 3 M NaCl
        i. Mix for a minimum of 30 min after finish adding the salt (solution not completely opaque)
    c. Centrifuge for 30 min at 30,000 RPM at 4°C using the Beckman Ti 45 rotor
    d. Remove supernatant and pipet out any residue solution after inverting the centrifuge tube
    e. Dissolve pellet in 10 mL 0.1 N HCl
    f. Add 10 mL of 200 mM Tris pH 8.6
    g. Titrate to pH 7.4 with NaOH or HCl
    h. Add 5M NaCl solution to bring solution to a final concentration of 3 M NaCl
        i. Mix for a minimum of 30 min after adding salt solution
    i. Centrifuge for 30 min at 30,000 RPM at 4°C using the Beckman Ti 45 rotor
    j. Remove supernatant and pipet out any residue solution after inverting the centrifuge tube
    k. Resuspend pellet in 0.5 mL 0.1N HCl
E. Dialysis
    a. Dialyze sample into 50 mM AcOH using 300kDa MWCO dialysis membrane overnight at 4 °C.
    (1 mL sample vs 1 L AcOH, replace AcOH after 1 h and 2 h)
    b. Transfer pellet to 1.5 mL microcentrifuge tube and centrifuge at 15k x g at 4 °C for 1 minute
    c. Transfer supernatant to new microcentrifuge tube
        i. Repeat steps b and c until no pellet is visible
    d. Store sample at -20°C

Comments or other information:

- Read manual for French Press cell assembly and operation
- Decant solutions immediately after centrifugation.
- Remove all drops of supernatant during NaCl precipitation steps to increase purity of product
- After low pH precipitation, pellets can be dissolved in the same 10 mL to reduce reagent use
    - Do not combine the pellet of more than 180 mL of low pH precipitation solution in this manner
- After neutral pH prepetition, pellets can be dissolved in the same 0.5 mL to increase concentration

Protocol: HT1080 Cell Adhesion Assay (MTT version)

1) Culture HT1080 human fibrosarcoma cells in DMEM + 10% FBS
2) Prepare 96-well non-tissue culture treated plates
    a. Deposit 40 µl of 20 µg/ml protein in DPBS into each well. (make sure surface fully coated)
    b. Incubate at 4 °C for at least 12 hours, 24 hours preferred
    c. Remove protein solution
    d. Deposit 100 µl of DMEM + 0.5% BSA
    e. Incubate at RT for 1 hour
    f. Remove DMEM + 0.5% BSA
3) Harvest HT1080s, Suspend to concentration of 5×10^5 cells/ml
4) Seed 100 µl of cell mixture into each well
5) Incubate at 30 °C for 4 hours
6) Remove supernatant slowly (pipet from top of liquid)
7) Wash each well with 125 µl of DPBS thrice
8) Add 160 µL of DMEM + 10% FBS
9) Add 25 µL of a 5 mg/ml MTT solution (in PBS)
10) Incubate for 2 hours at 37 °C
11) Add 70 µL of Lysis Buffer
12) Incubate overnight at 37 °C
13) Read A_{570} for each well

Solutions

- MTT solution (5 mg/mL)
    - 10 mL MTT solution (in PBS): 50 mg MTT + 10 mL PBS. Filter Sterilize.

- Lysing Buffer (20% w/v)
    - 80% acetic acid (10mL): 8mL glacial acetic acid + 2mL ddH2O
    - 1 N HCI (12mL): 1mL 12N HCl + 11mL ddH2O
    - 10 mL of 50% DMF solution: 5mL DMF+5mL demineralised water
    - 10 mL lysing buffer: 2g SDS + 10 mL DMF solution.
    - pH=4.7 (Adjust by adding 0.25 mL of an 80% acetic acid and 0.25mL 1N HCl)
    - Mix by magnetic stirring.

Protocol: Integrin Inhibition Cellular Adhesion Assay

1) Culture HT1080 human fibrosarcoma cells in DMEM + 10% FBS
2) Prepare 96-well non-tissue culture treated plates
    a. Deposit 40 µl of 20 µg/ml protein in DPBS into each well. (make sure surface fully coated)
    b. Incubate at 4 °C for at least 12 hours, 24 hours preferred
    c. Remove protein solution
    d. Deposit 100 µl of DMEM + 0.5% BSA
    e. Incubate at RT for 1 hour
    f. Remove DMEM + 0.5% BSA
3) Harvest HT1080s, Suspend to concentration of 1 x 10^6 cells/ml
4) Incubate HT1080s (1 x 10^5 cells per mL) with required antibody (3600 µl final volume for triplicate of 12 surfaces)
    a. Dilute antibody in DMEM to 10 µg/ml to 10% final volume (360 µl)
    b. Add 1% final volume of µl 10% BSA (36 µl)
    c. Add 10% final volume of HT1080s at a concentration of 1 x 10^6 Cells/ml (360 µl)
    d. Add DMEM to final volume (2844 µl)
    e. Incubate at 37 °C for 15 minutes
5) Seed 100 µl of Antibody/Cell mixture into each well
6) Incubate at 30 °C for 4 hours
7) Add 300 µl of DMEM + 0.1% BSA to each well. Seal Plate with PCR sealing tape
8) Spin plate at 2000 RPM (using GS-3 Rotor) ~= 900 x g for 40 minutes
9) Remove supernatant slowly (pipet from top of liquid)
10) Wash each well with 100 µl of PBS

- **Important: All subsequent steps should be performed away from light. Calcein-AM is light sensitive.**

11) Prepare 4 µM calcein-AM solution
    a. 36 µl of 1 mM Calcein-AM, 90 µl of 10% BSA, 9 ml PBS
12) Add 50 µl of 4 µM calcein-AM to each well
13) Incubate at 37 °C for 1 hour
14) Remove calcein-AM
15) Add 100 µl of CelLytic M to each well
16) Incubate at RT for 15 minutes
17) Transfer 90 µl from each well to a separate 384-well black fluorescent plate well
18) Read fluorescence reading using SpectraMax M2 plate reader
    a. 494/517 nm (Em/Ex), 515nm cutoff

Protocol: Lyophilization of Collagen

1) Filter collagen using a sterile 0.22 μm syringe filter if sterility is necessary.
    a. For collagen variants containing cysteine, add 10 x molar excess TCEP
    b. Pass an additional 100 μL of buffer through filter afterward. Combine with filtered protein
2) Cover tube with sterile filter paper. Use rubber band to hold filter paper in place.
3) Freeze protein using dry ice. Be sure sample is completely frozen
4) Place in lyophilzer for 18 hours
5) Remove, cap, and store at -20°C until ready to resuspend/use.