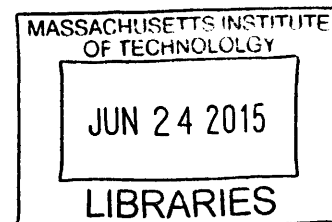
Understanding Collagen-I Folding and Misfolding

ARCHIVES



by

Rebecca J. Taylor

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Chemistry and Biology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

©Massachusetts Institute of Technology, 2015
All rights reserved

Author: Signature redacted

Department of Chemistry
May 26, 2015

Certified by: Signature redacted

Matthew D. Shoulders
Assistant Professor of Chemistry
Chemistry Thesis Supervisor

Accepted by: Signature redacted

Rick L. Danheiser
A. C. Cope Professor of Chemistry
Undergraduate Officer, Department of Chemistry

Understanding Collagen-I Folding and Misfolding

by

Rebecca J. Taylor

Submitted to the Department of Chemistry on May 26th, 2015

Abstract

Chapter One: Introduction to Type I Collagen and Osteogenesis Imperfecta

Collagen-I is the primary proteinaceous component of skin, bone, and tendon. Disruptions in collagen-I homeostasis, typically due to non-synonymous mutations in collagen-I-encoding genes, cause a variety of severe incurable diseases, including Osteogenesis Imperfecta (OI). OI phenotypes include brittle, deformed bones, frequent fractures, and growth deficiency. In order to fill the need for treatments that target the underlying causes of collagen-I-related diseases like OI, a better understanding of the collagen-I proteostasis network and how it differentially engages mutant and wild type collagen-I, is required.

Chapter Two: Creation and Characterization of a Cell-Based Platform for Delineating the Wild Type and Mutant Collagen-I Proteostasis Network

Previous studies of the collagen biosynthetic pathway have been limited by the lack of a biochemically tractable system to allow manipulation of the collagen-I genes (and other genes of interest) and especially by the lack of immunoprecipitation-grade antibodies for collagen-I which has prevented the broad study of the complete set of collagen-I interacting proteins. We have overcome the challenges of working with the collagen-I genes and have created stable cell lines that inducibly express epitope-tagged versions of both wild type and mutant collagen-I. This platform is greatly facilitating studies of the collagen-I proteostasis network.

Chapter Three: Mechanistic Exploration of Novel Collagen-I Interacting Proteins Identified by SILAC Mass Spectrometry

Using the model cell platform described in Chapter 2, we have performed an unbiased and quantitative investigation into the network of collagen-I interacting proteins using SILAC-assisted, quantitative mass spectrometry. The method allowed us to identify more than 25 novel collagen-I interactors. We are currently investigating the mechanistic roles of these proteins in collagen-I processing using shRNA knockdown of proteins of interest.

Chapter Four: Creation and Validation of Constructs for the Independent Expression of the Collagen-I C-Propeptide Domains

The collagen-I C-propeptide domains are responsible for collagen-I chain selectivity and triple helix nucleation. Many unanswered questions remain relating to the mechanistic details of C-propeptide function both in collagen-I folding and also in diverse biological processes. In order to address these questions, we created and validated constructs that allow the independent expression of the C-propeptide domains.

Thesis Supervisor: Matthew D. Shoulders
Title: Assistant Professor of Chemistry

Acknowledgements

I especially want to thank Professor Matthew Shoulders for all of the opportunities and mentorship he has provided me, from working in his lab doing really interesting and exciting research to serving as a teaching assistant for 5.111. I greatly appreciate everything he has done for me and especially all of his encouragement. Working in the Shoulders lab has been the most inspirational and influential part of my time at MIT and I have enjoyed my work immensely. I will truly miss working in the Shoulders lab next fall.

I also want to thank Andrew DiChiara, who has been my day-to-day supervisor and research partner in the lab. Working with Andrew has been a great experience and I'm extremely grateful for all the help he has given me. All of the work presented here has been done in collaboration with Andrew and in particular, the mass spectrometry data presented in Chapter Three is Andrew's work.

It has also been wonderful to work with the other members of the collagen project, Madeline Wong and Dr. N. Duc Doan, and I would like to thank them for the help they have given me. I would also like to thank all of the other members of the Shoulders lab, especially Dr. Mahender Dewal for letting me include one of his figures and Dr. Emmanuel Nekongo for help he has given me during my time in the lab.

Finally, I would like to thank my parents, my sisters, and my friends for all of their support.

Table of Contents

Abstract 2

Acknowledgements 3

Table of Contents 4

List of Figures and Tables 5

Chapter One: Introduction to Type I Collagen and Osteogenesis Imperfecta 6

1.1 Introduction to Type I Collagen 6

1.2 Introduction to Osteogenesis Imperfecta 9

1.3 Thesis Outline 11

Chapter Two: Creation and Characterization of a Cell-Based Platform for Delineating the Wild-Type and Mutant Collagen-I Proteostasis Network 13

2.1 Introduction 13

2.2 Results 16

2.3 Discussion 24

2.4 Conclusions and Future Work 28

2.5 Experimental Procedures 29

Chapter Three: Mechanistic Exploration of Novel Collagen-I Interacting Proteins Identified by SILAC Mass Spectrometry 35

3.1 Introduction 35

3.2 Results 39

3.3 Discussion 48

3.4 Conclusions and Future Work 52

3.5 Experimental Procedures 53

Chapter Four: Creation and Validation of Constructs for the Independent Expression of the Collagen-I C-Propeptide Domains 59

4.1 Introduction 59

4.2 Results 64

4.3 Discussion and Future Directions 67

4.4 Experimental Procedures 69

Chapter Five: Conclusions and Future Work 72

Bibliography 74

List of Figures and Tables

Figure 1.1.	A section of the assembled collagen-I triple helix	6
Figure 1.2.	The collagen-I assembly pathway	8
Figure 2.1.	Collagen-I Vector Design	16
Table 2.1.	Collagen-1 mutants chosen for OI-model cell lines	17
Figure 2.2.	Schematic of Transcriptional Regulation in HT1080 Tet- Off cells	18
Figure 2.3.	Inducible expression of either Colα1(I), Colα2(I) or both in HT1080 Tet-Off cells	20
Figure 2.4.	Inducible expression of OI-mutant Colα1(I) and wild type Colα2(I) in HT1080 cells.	21
Figure 2.5.	Co-Immunoprecipitation of Colα1(I) with Colα2(I) and vice versa	21
Figure 2.6.	Disulfide-dependent assembly of secreted collagen-I triple helices	22
Figure 2.7.	N-glycosylation of collagen-I	22
Figure 2.8.	Hydroxylation of the Yaa proline residue	23
Figure 2.9.	Trypsin treatment of collagen-I from primary fibroblasts and HT1080 cells	24
Table 2.2.	Primers used Collagen-I qPCR	32
Figure 3.1.	Workflow for stable isotope labeling by amino acids in cell culture (SILAC)	38
Figure 3.2.	Immunoprecipitation of stable collagen-I interactors	40
Figure 3.3.	Optimization of the crosslinking HA (Colα1(I)) immunoprecipitation	41
Table 3.1.	Results of SILAC mass spectrometry experiments	43
Figure 3.4.	Crosslinking) co-IP of the newly identified collagen-I interacting proteins	44
Figure 3.5.	shRNA knockdown of several collagen-I interactors of interest	46
Figure 3.6.	Optimized shRNA knockdown of each collagen-I interactor of interest	47
Table 3.2.	Descriptions and functions for key new proteins identified by SILAC-assisted mass spectrometry	51
Table 3.3.	Primers used for Quantitative PCR	57
Figure 4.1.	Schematic of the disulfide linkages in the Type I and Type III Collagen C-propeptides	62
Figure 4.2.	Collagen-I C-propeptide vector design	64
Figure 4.3.	N-glycosylation of Collagen-I. Colα1(I) C-propeptide	65
Figure 4.4.	Variable DTT treatment of HEK293 cells expressing the Colα1(I) and Colα2(I) C-propeptides	66
Figure 4.5.	Activation of XBP1S shifts the relative populations of Colα1(I) C-propeptide N-glycoforms	69

Chapter One:

Introduction to Type I Collagen and Osteogenesis Imperfecta

1.1 Introduction to Type I Collagen

The collagens are a family of trimeric extracellular matrix proteins distinguished by the presence of one or more triple helical domains.[1] The 29 collagen types serve as the molecular scaffold for life, constituting the main component of connective tissues ranging from cartilage and bone to basement membranes.[2] Together, they account for approximately one-third of the total human proteome by mass.[3]

The most abundant form of collagen by far is type I collagen, which provides the three-dimensional structure of skin, bone, and tendon.[4]

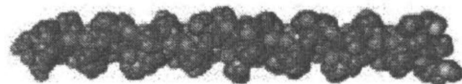


Figure 1.1. A section of the assembled collagen-I triple helix.

(1) The biosynthetic precursor of extracellular collagen-I, referred to as procollagen-I, consists of two strands of proα1 collagen-I (Colα1(I)) and one strand of proα2 collagen-I (Colα2(I)) assembled into a heterotrimeric triple helical structure with globular domains at both the amino and carboxyl termini called the N- and C-propeptides.[5] (3,4) Within the triple helical region, the primary structure of each polypeptide is characterized by a repeating triplet pattern (Gly-Xaa-

[1] Van der Rest, M. & Garrone, R. Collagen family of proteins. *FASEB J.* **5,** 2814–2823 (1991).
[2](a) Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* **3,** 1–19 (2011).
(b) Veit, G. *et al.* Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *J. Biol. Chem.* **281,** 3494–3504 (2006).
[3] Ricard-Blum, S. (2011).
[4] Ricard-Blum, S. (2011).
[5] (a) Canty, E. G. & Kadler, K. E. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* 118, 1341–53 (2005).
(b) Engel, J. & Bächinger, H. P. Structure, stability and folding of the collagen triple helix. *Top. Curr. Chem.* 247, 7–33 (2005).

Yaa), where a glycine residue is required at every third position due to the structural constraints of the helix, and the Xaa and Yaa amino acids are most commonly (2*S*)-proline and (2*S*,4*R*)-4-hydroxyproline, respectively.[6]

Nascent procollagen-I alpha strands are co-translationally imported into the endoplasmic reticulum (ER), the organelle responsible for the synthesis, folding, and processing of all secreted and membrane proteins (Figure 1.2). Triple helix chain selection and nucleation is controlled by the C-propeptide domains, which means that the entirety of each procollagen polypeptide must enter the ER and the C-propeptide domains must fold before triple helix assembly can occur.[7] While the Colα1(I) and Colα2(I) strands are in a monomeric state, they undergo extensive post-translational modifications including prolyl 4-hydroxylation, prolyl 3-hydroxylation, and lysyl hydroxylation followed by hydroxylysl glycosylation.[8] Triple helix assembly is initiated by non-covalent interactions between the three C-propeptide domains, followed by the formation of inter-strand disulfide bonds which covalently link the polypeptides that constitute the collagen-I trimer.[9] Once the three collagen-I strands are aligned with a one

---

[6] (a) Shoulders, M. D. & Raines, R. T. Collagen structure and stability. Annu. Rev. Biochem. 78, 929–958 (2009).
(b) Jenkins, C. L., Vasbinder, M. M., Miller, S. J. & Raines, R. T. Peptide bond isosteres: Ester or (E)-alkene in the backbone of the collagen triple helix. *Org. Lett.* **7,** 2619–2622 (2005).
(c) Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266,** 75–81 (1994).
[7] Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* 44, 21–32 (2012).
[8] (a) Pihlajaniemi, T., Myllylä, R. & Kivirikko, K. I. Prolyl 4-hydroxylase and its role in collagen synthesis. *J. Hepatol.* 13 Suppl 3, S2–S7 (1991).
(b) Vranka, J. a., Sakai, L. Y. & Bächinger, H. P. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J. Biol. Chem.* 279, 23615–23621 (2004).
(c) Myllyharju, J. & Kivirikko, K. I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* 20, 33–43 (2004).
[9] Boudko, S. P., et al. (2012).

residue stagger, triple helix folding proceeds in a zipper-like manner with a folding rate that is limited by proline *cis-trans* isomerization.[10]

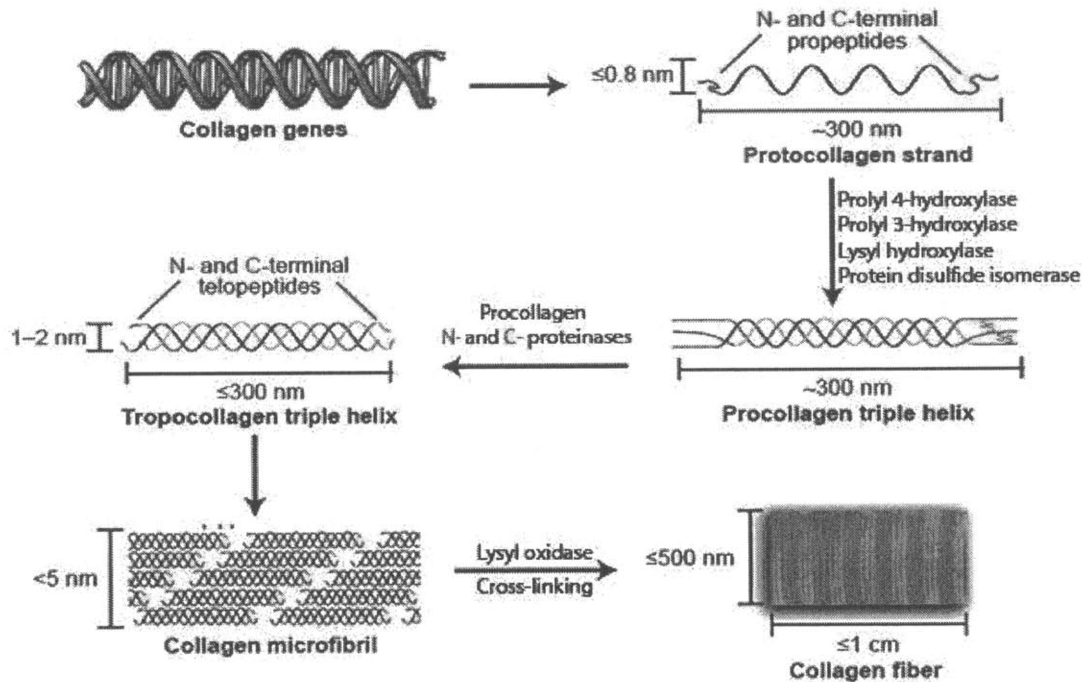


Figure 1.2. The collagen-I assembly pathway. Collagen-I molecules are initially synthesized as procollagen polypeptides termed Colα1(I) and Colα2(I). The Colα1(I) and Colα2(I) strands are extensively post-translationally modified by prolyl 4-hydroxylase, lysyl hydroxylase, and prolyl 3-hydroxylase. After collagen-1 is exported to the extracellular matrix, the N-terminal and C-terminal propeptides are cleaved by the N- and C-proteinases (ADAMTS2 and BMP1 respectively). This cleavage even triggers the self-assembly of collagen-I triple helices into microfibrils which subsequently are cross-linked together to form the collagen-I fibers which constitute the three dimensional basis of skin, bone, and tendon.

After the triple helix is assembled, procollagen-I traffics through the secretory pathway to the extracellular matrix where the N- and C-propeptides are cleaved by specific proteinases,

[10] (a) Engel, J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* 20, 137–152 (1991).
(b) Bächinger, H. P., Bruckner, P., Timpl, R. & Engel, J. The role of cis-trans isomerization of peptide bonds in the coil leads to and comes from triple helix conversion of collagen. *Eur. J. Biochem.* 90, 605–613 (1978).

ADAMTS2 and BMP-1, respectively.[11] This cleavage initiates the self-assembly of collagen-I triple helices into microfibrils, which are subsequently cross-linked together to form the collagen fibers that constitute the three-dimensional basis of skin, bone, and tendon (Figure 1.2).[12]

1.2 Introduction to Osteogenesis Imperfecta

Disruptions to either the structure of collagen-I itself or to components in its complex synthesis and assembly pathway can lead to disorders in collagen-I homeostasis. The prototypical example of a collagen homeostasis disorder is Osteogenesis Imperfecta (OI) which is an often-severe connective tissue disorder with clinical features that include low bone mass, bone deformity, and growth deficiency.[13] OI is most commonly caused (approximately 90% of OI cases) by heterozygous mutations in either COL1A1 or COL1A2, the genes that encode Colα1(I) and Colα2(I) respectively, with over 1,500 different mutations have been identified.[14] These mutations are most often located within the triple helical domain and replace the required glycine residue with a different amino acid. All other known OI cases are caused by

[11] (a) Tuderman, L., Kivirikko, K. I. & Prockop, D. J. Partial Purification and Characterization of a Neutral Protease Which Cleaves the N-Terminal Propeptides from Procollagen. *Biochemistry* 78, 2948–2954 (1978).
(b) Hojima, Y., van der Rest, M. & Prockop, D. J. Type I Procollagen Carboxyl-terminal Proteinase from Chick Embryo Tendons. *J. Biol. Chem.* 260, 15996–16003 (1985).

[12] Jenkins, C. L., Vasbinder, M. M., Miller, S. J. & Raines, R. T. Peptide bond isosteres: Ester or (E)-alkene in the backbone of the collagen triple helix. *Org. Lett.* 7, 2619–2622 (2005).

[13] Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J. C. New Perspectives on Osteogenesis Imperfecta. *Nat. Rev. Endocrinol.* 7, 540–557 (2011).

[14] Marini, J. C. *et al.* Consortium for Osteogenesis Imperfecta Mutations in the Helical Domain of Type I Collagen : Regions Rich in Lethal Mutations Align With Collagen Binding Sites for Integrins and Proteoglycans. 28, 209–221 (2007).

homozygous mutations in proteins responsible for post-translational modification of collagen-I or chaperoning its folding.[15]

OI phenotypes can range from very mild (minimal bone deformity or growth deficiency, occasional fractures) to perinatally lethal.[16] The mildest cases of OI are caused by the presence of one null *COL1A1* allele, leading to synthesis of half the normal amount of collagen-I, but the collagen-I produced is wild type. The more severe forms of OI are caused by either the glycine replacement mutations that are the most common OI mutations or by mutations in collagen-I chaperone or modification proteins.[17] In both of these cases, procollagen folding and triple helix assembly are delayed, resulting in the production of highly over post-translationally modified collagen-I that is much less stable than the wild type version.[18]

The current standard of care for OI is palliative, such as the treatment of symptoms (e.g., broken bones) or the administration of bisphosphonates to increase bone matrix density by reducing bone resorption.[19] Treatments that target the underlying cause of the disease do not exist largely because, while the genetic basis of OI has long been known, the molecular mechanisms of the disease pathogenesis are still poorly understood. One possible hypothesis is

---

[15] (a) Marini, J. C., Cabral, W. a. & Barnes, A. M. Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. *Cell Tissue Res.* 339, 59–70 (2010).
(b) Pyott, S. M. *et al.* Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum. Mol. Genet.* 20, 1595–1609 (2011).
(c) Christiansen, H. E. *et al.* Homozygosity for a Missense Mutation in SERPINH1, which Encodes the Collagen Chaperone Protein HSP47, Results in Severe Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86, 389–398 (2010).
(d) Alanay, Y. *et al.* Mutations in the Gene Encoding the RER Protein FKBP65 Cause Autosomal-Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86, 551–559 (2010).

[16] Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J. C. New Perspectives on Osteogenesis Imperfecta. Nat. Rev. Endocrinol. 7, 540–557 (2011).

[17] Bodian, D. L. *et al.* Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype-phenotype relationships. *Hum. Mol. Genet.* **18,** 463–71 (2009).

[18] Forlino A. *et al* (2011)

[19] Gloreux, F. H. *et al.* Cyclic Administration of Pamidronate in Children with Severe. *N. Engl. J. Med.* 339, 947–952 (1998).

that the OI phenotype results entirely from the structural defects in collagen-I, either due to mutations that disrupt the local triple-helical structure or excessive post-translational modifications due to slow helix assembly in the ER. Alternatively, it has recently been suggested that osteoblast apoptosis due to the prolonged ER stress caused by the presence of mutant collagen-I may also play a role in OI pathogenesis.[20]

In addition to incomplete knowledge about the molecular basis for OI pathogenesis, the development of treatments that target the underlying causes of the disease is further hindered because many aspects of the collagen-I biosynthesis and folding pathway remain obscure, especially with regard to how OI-mutant collagen-I is processed by the cellular machinery.[21] There are numerous open questions, for example: How does a cell assess whether collagen-I has achieved its proper folded state? How does a cell decide whether mutant or misfolding collagens will be secreted or targeted for degradation?[22] What factors influence how cellular machinery processes normal and OI-mutant collagen-I, and with what outcomes?

1.3 Thesis Outline

The research presented in this thesis takes critical steps toward the systematic exploration of the collagen-I proteostasis network, with specific contributions that include:

---

[20] Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of Osteogenesis imperfecta. *PLoS Genet.* 4, (2008).

[21] Ishikawa, Y. & Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833, 2479–2491 (2013).

[22] (a) Fitzgerald, J., Lamandé, S. R. & Bateman, J. F. Proteasomal degradation of unassembled mutant type I collagen pro-alpha1(I) chains. *J. Biol. Chem.* 274, 27392–27398 (1999).
(b) Bottomley, M. J., Batten, M. R., Lumb, R. a. & Bulleid, N. J. Quality control in the endoplasmic reticulum: PDI mediates the ER retention of unassembled procollagen C-propeptides. *Curr. Biol.* 11, 1114–1118 (2001).
(c) Ishida, Y. *et al.* Autophagic Elimination of Misfolded Procollagen Aggregates in the Endoplasmic Reticulum as a Means of Cell Protection. 20, 2744–2754 (2009).

1. Development of stable isogenic cell lines that allow the inducible expression of epitope-tagged wild type and OI-mutant collagen-I, thereby creating a biochemically tractable system in which to study collagen-I proteostasis and how the cellular proteostasis machinery differentially engages mutant and wild type collagens. This work is described in Chapter Two.
2. Application of the stable incorporation of labeled amino acids in cell culture (SILAC) method to systematically identify proteins that interact with collagen-I during its biosynthesis and folding, resulting in the identification of more than 25 novel collagen-I interactors. This work is described in Chapter Three.
3. Experiments to optimize shRNA knockdown of collagen-I interactors. Future work will focus on screening for collagen-I secretion phenotypes to determine the roles of collagen-I interactors in its biosynthetic pathway and their contributions to collagen-I secretion defects. This work is also described in Chapter Three.
4. Development of a system that allows the collagen-I C-propeptide trimers to be independently expressed in the absence of the remainder of the collagen-I molecule. This system can be used to study the details of the roles of the C-propeptide domains in collagen-I folding. This work is described in Chapter Four.

In the future, these cell lines and methods will be used to fully delineate the collagen-I proteostasis network and to investigate how the cell differentially engages mutant and wild-type collagens with the goal of understanding how cells deal with mutant collagens so that treatments can be developed, for example, to increase the quality of the collagen secreted or help the cell more efficiently target mutant collagen-I for degradation, thus ameliorating OI.

Chapter Two:

Creation and Characterization of a Cell-Based Platform for Delineating the Wild-Type and Mutant Collagen-I Proteostasis Network

2.1 Introduction

Type I collagen is one of the most abundant proteins in the body and it has a critical structural function, making up the three-dimensional proteinaceous scaffold for skin, bone, and tendon.[23] Collagen-I biosynthesis occurs in the endoplasmic reticulum (ER) and therefore the proteostasis machinery in the endoplasmic reticulum is responsible for ensuring the proper folding and assembly of collagen-I, as well as enabling various quality control mechanisms to deal with the presence of mutant or misfolding variants of collagen-I, and providing specialized transport machinery for the secretion of mature collagen-I triple helices.[24] Defects in collagen-I biosynthesis and secretion, caused by either primary mutations in the collagen-I molecule itself or by mutations in proteins responsible for assisting its maturation process result in the severe and incurable disease, Osteogenesis Imperfecta (OI).[25] Multiple decades of effort have been put into studying the molecular mechanisms of OI and developing treatments, but these efforts have been hindered by the major gaps in our knowledge of collagen proteostasis.[26]

Most of our knowledge of the collagen-I proteostasis machinery derives from a few types of studies, and the proteins that interact with collagen-I have been identified either due

[23] Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* 3, 1–19 (2011).
[24] Canty, E. G. & Kadler, K. E. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* 118, 1341–53 (2005).
[25] Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J. C. New Perspectives on Osteogenesis Imperfecta. *Nat. Rev. Endocrinol.* 7, 540–557 (2011).
[26] Ishikawa, Y. & Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833, 2479–2491 (2013).

to their high abundance, by a rational prediction, or because mutations in these proteins cause human diseases. The collagen specific chaperone, HSP47, was discovered as an abundant protein produced by heat shock of chick embryo fibroblasts, and was subsequently found to bind to collagen-I both *in vitro* and *in vivo.*[27] The importance of HSP47's role in collagen-I folding was further clarified through study of HSP47 null mice and the discovery that HSP47 is one of the proteins that causes recessive Osteogenesis Imperfecta.[28] The role of peptidyl-prolyl isomerases in collagen folding was predicted based on the high proline content of collagen-I, leading to *in vitro* and *in vivo* studies showing that peptidyl-prolyl isomerases did indeed interact with collagen-I.[29] Cartilage-associated protein (CRTAP), a member of the collagen-I prolyl 3-hydroxylation complex, was first identified as being involved in collagen-I biosynthesis when a recessive mutation in the CRTAP gene was discovered to be the causative mutation in one form of OI.[30]

These types of traditional genetic, primary cell-based, and mouse model methods have led to the discovery of these and a limited set of other components of the collagen-I proteostasis machinery, but a broad characterization of the complete collagen-I proteostasis machinery has never been done before. This type of investigation has been hindered by the

---

[27] (a) Nagata, K., Saga, S. & Yamada, K. M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. J. Cell Biol. 103, 223–229 (1986).
(b) Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. J. Cell Biol. 117, 903–914 (1992).
[28] (a) Nagai, N. et al. Embryonic lethality of molecular chaperone Hsp47 knockout mice is associated with defects in collagen biosynthesis. J. Cell Biol. 150, 1499–1505 (2000).
(b) Christiansen, H. E. et al. Homozygosity for a Missense Mutation in SERPINH1, which Encodes the Collagen Chaperone Protein HSP47, Results in Severe Recessive Osteogenesis Imperfecta. Am. J. Hum. Genet. 86, 389–398 (2010).
[29] Bächinger, H. P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262, 17144–17148 (1987).
[30] Morello, R. *et al.* CRTAP Is Required for Prolyl 3- Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* 127, 291–304 (2006).

lack of a system where collagen-I can easily be manipulated and isolated for study, and especially by the lack of immunoprecipitation-grade antibodies for collagen-I. Development of such a system has been impeded by the difficulty of working with the collagen-I genes and the complexity of the machinery that is required for expression of a properly modified and folded collagen-I molecule. Collagen genes are difficult to work with because they are extremely large, their GC content is high, which complicates amplification by standard PCR methods, and they have a repetitive sequence that results in frequent recombination events during molecular biology manipulations. Additionally, expression of properly folded and assembled collagen-I triple helices requires a higher-eukaryotic expression system due to collagen-I's extensive post-translational modifications.[31]

We set out to overcome these obstacles by developing a cell-based platform that would allow the systematic elucidation and study of the collagen-I proteostasis network. We created cell lines that singly express either wild type Colα1(I) or Colα2(I), that express both wild type Colα1(I) and Colα2(I), or that express wild type Colα2(I) and OI-causing mutant variants of Colα1(I). We characterized the collagen-I produced by the wild type Colα1(I)/Colα2(I) (the cell line that expresses both Colα1(I) and Colα2(I))cell line and found that the collagen-I secreted by this system has properties that match those of endogenously-produced collagen-I, indicating that the our cell-based platform produces collagen-I that is properly folded and post-translationally modified. These cell lines allowed us to perform unbiased proteomics experiments to broadly probe the full set of proteins that interact with collagen-I and compare how the wild type and mutant collagen-I variants are differentially engaged by the cellular

[31] Myllyharju, J. in *Methods Mol. Biol. Extracell. Matrix Protoc.* 552, 51–62 (2009).

proteostasis machinery (work described in later chapters). The mechanistic knowledge gained through these experiments has already led us to new ideas for therapeutic approaches to collagen-I-related diseases like OI.

2.2 Results

In order to create a system that would allow a proteomics-based investigation into the collagen-I proteostasis machinery, we designed collagen-I expression vectors that contain a pre-protrypsin signal sequence at the 5' end of the gene construct, followed by an antibody epitope tag and the procollagen cDNA sequence (Figure 2.1). The pre-protrypsin signal sequence was included to ensure ER localization, while the antibody epitope tags were included to allow immunoprecipitation and efficient Western blot detection of the collagen-I protein. In order to allow selective identification or immunoprecipitation of Colα1(I) versus Colα2(I), we used a HA tag for all Colα1(I) constructs and a FLAG tag for all Colα2(I) constructs. These epitope tags were appended to the N-terminus of the protein in order to minimize the disruption to collagen-I folding, which initiates at the C-terminus.[32]

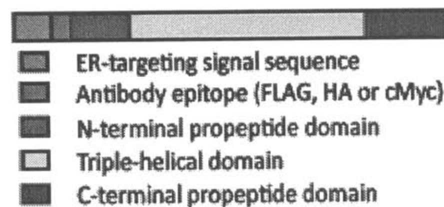


Figure 2.1. Collagen-I Vector Design.

In addition to vectors for expression of wild type collagen-I we also designed vectors that allow expression of selected OI-mutants. We selected total of eight different OI-causing collagen-I mutants that are located in various places along the Colα1(I) and Colα2(I)

[32] Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* 44, 21–32 (2012).

polypeptides, with three of the selected mutants located in the triple helical domain and a fourth located in the C-propeptide domain (see Table 2.1). These particular mutations were chosen because primary OI patient fibroblasts are available that express each of these collagen-I mutations, thus providing a way to verify the physiological relevance of all experimental results.

Table 2.1. Collagen-1 mutants chosen for OI-model cell lines.

Protein	Epitope Tag	Mutation
Colα1(I)	HA	Wild Type
	cMyc	Gly247Ser
		Gly352Ser
		Gly883Ser
		Cys1299Trp
Colα2(I)	FLAG	Wild Type
	cMyc	Gly502Ser
		Gly610Cys
		Gly907Asp
		Cys1163Arg

To create the collagen-I expression vectors, we had to overcome the challenges associated with manipulating the collagen-I genes, including extremely large size, high GC content, and a repetitive sequence that leads to frequent recombinations. We used a recombination-incompetent *E. coli* cell line for all collagen-I cloning and the PCR protocols used to amplify the collagen-I gene were optimized by extensive experimentation, as were protocols for site-directed mutagenesis and other manipulations of the collagen-I genes. These approaches enabled us to generate collagen-I expression vectors for wild type Colα1(I) and Colα2(I) and their four OI-mutant variants of each using the pTRE-Tight vector backbone (ClonTech), which has a tetracycline-responsive promoter that allows inducible expression in the HT1080 Tet-Off system.

HT1080 Tet- Off® cells are a human fibrosarcoma cell line that stably expresses the tetracycline-regulated transactivator, Tet-Off. The Tet-Off system allows the inducible expression (in the absence of doxycycline) of genes that are under the control of a tetracycline-

responsive promoter (Figure 2.2).[33] The HT1080 cell line was selected because it has been shown to express all of the known components of the collagen-I biosynthetic machinery (HT1080 cells endogenously produce type IV collagen) but it does not produce any endogenous collagen-I itself, thus simplifying interpretation of experimental results.[34] We chose to use an inducible expression system because constitutive expression of collagen-I could potentially stress cells, leading to a selective disadvantage for cells that express high levels of collagen-I, thus making it difficult to generate a cell line that robustly expresses high levels of collagen-I. The inducible expression system allows us to circumvent the selective disadvantage of constitutively expressing high levels of collagen-I in order to enable us to generate stable isogenic cell lines.

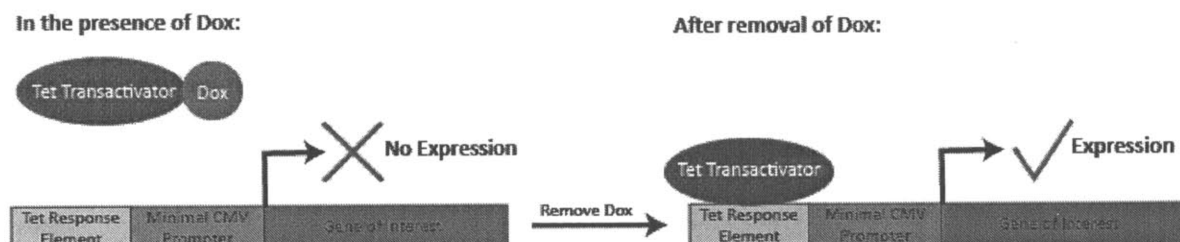


Figure 2.2. Schematic of Transcriptional Regulation in HT1080 Tet- Off cells. If doxycycline (dox) is present, it binds to the tetracycline-regulated transactivator preventing its binding to the tet-responsive element in the gene promoter and thus inhibiting gene transcription. When dox is removed, the tet-transactivator binds to the tet-response element and gene transcription occurs.

---

[33] Urlinger, S. *et al.* Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7963–7968 (2000).

[34] (a) Geddis, A. E. & Prockop, D. J. Expression of Human *COL1A1* Gene in Stably Transfected HT1080 Cells: The Production of a Thermostable Homotrimer of Type I Collagen in a Recombinant System. *Matrix* 13, 399–405 (1993).
(b) Fertala, A. *et al.* Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). *Biochem. J.* 298 (Pt 1, 31–37 (1994).
(c) Chung, H. J., Jensen, D. a., Gawron, K., Steplewski, A. & Fertala, A. R992C (p.R1192C) Substitution in Collagen II Alters the Structure of Mutant Molecules and Induces the Unfolded Protein Response. *J. Mol. Biol.* 390, 306–318 (2009).
(d) Ito, H. *et al.* Guilty by association: some collagen ii mutants alter the formation of ecm as a result of atypical interaction with fibronectin. *J. Mol. Biol.* 352, 382–395 (2005).

To generate cell lines that stably express collagen-I, we co-transfected either the Colα1(I) or Colα2(I) expression vector and a linear drug resistance marker into the HT1080 Tet-off cells. The Colα1(I) vector was co-transfected with the puromycin resistance marker and the Colα2(I) vector was co-transfected with the hygromycin resistance marker. Achieving a successful transfection of the large collagen-I genes was difficult, and required extensive optimization of the transfection protocols that were used. After transfection and selection, we obtained heterogenous populations that stably express either Colα1(I) or Colα2(I). We generated many genetically homogeneous single colony populations for each Colα1(I) and Colα2(I). We used both qPCR and Western blotting to test each single colony for collagen-I expression levels and the Colα1(I) and Colα2(I) colonies with the highest levels of collagen-I expression were retained. Successful identification of a single colony with high collagen-I expression, required us to generate and evaluate large numbers of candidate single colonies (approximately 25 - 30 for each Colα1(I) and Colα2(I). To create an isogenic cell line expressing both Colα1(I) and Colα2(I), we co-transfected the Colα1(I) expression construct and the puromycin linear resistance marker into the Colα2(I)-expressing cell line using our optimized transfection protocol. Again, we generated many (approximately 25 – 30) genetically homogenous single colonies, and the colony with the highest expression levels of both Colα1(I) and Colα2(I) was retained. Figure 2.3 shows the successful inducible protein-level expression of the Colα1(I), Colα2(I), and Colα1(I)/Colα2(I) cell lines in lysate and media samples from the chosen single colonies.

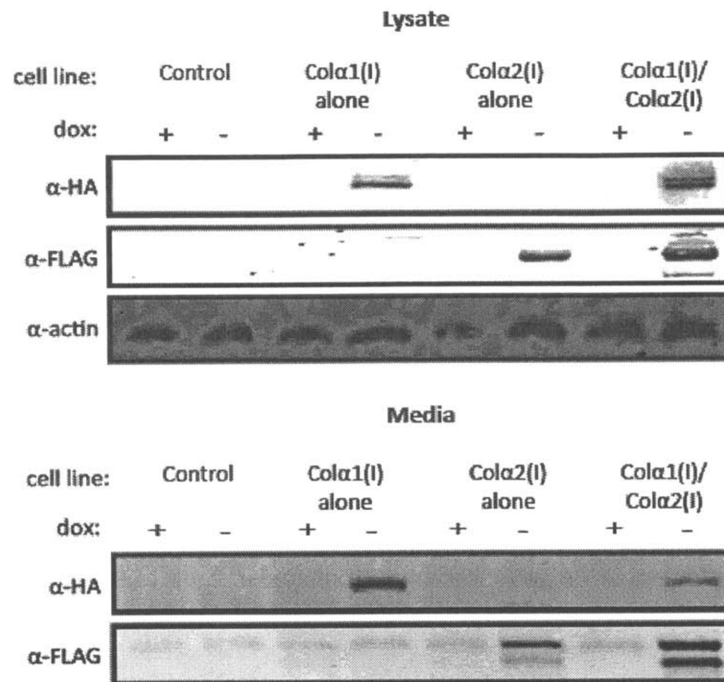


Figure 2.3. Inducible expression of either Colα1(I), Colα2(I) or both in HT1080 Tet-Off cells. Expression of Colα1(I) and Colα2(I) is activated by removal of Dox for 72 h.

In addition to the cell lines that express wild type Colα1(I) and/or Colα2(I), we also generated cell lines that express versions of Colα1(I) with the OI-causing G352S and C1229W mutations. We generated these OI-model cell lines by co-transfecting each of the mutant Colα1(I) expression constructs along with the puromycin linear resistance marker into the Colα2(I)-expressing cell line. For both the G352S and C1229W mutant versions of Colα1(I), we generated many genetically homogenous single colonies, and the colonies with the highest expression levels of both mutant Colα1(I) and wild type Colα2(I) were retained. Figure 2.4 shows the inducible expression of G352S Colα1(I) or C1229W Colα1(I) together with wild type Colα2(I) in HT1080 cells.

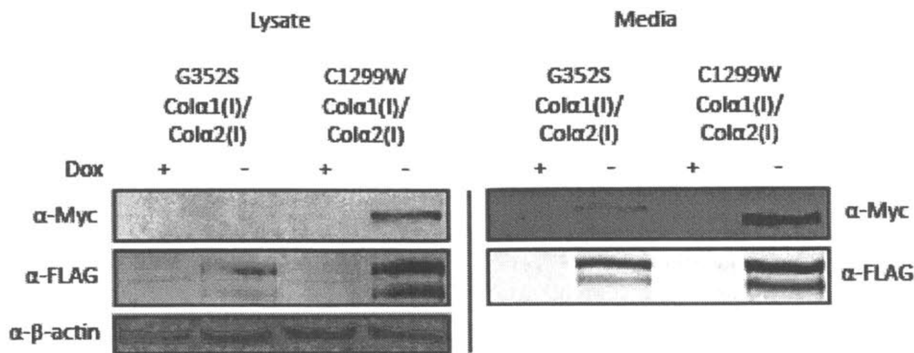


Figure 2.4. Inducible expression of OI-mutant Colα1(I) and wild type Colα2(I) in HT1080 cells.

In order to validate that the HT1080-based collagen-I expression system was capable of producing stable, properly post-translationally modified, and correctly folded collagen-I, we characterized the collagen-I produced by the cell line expression both wild type Colα1(I) and Colα2(I). As a first step, we probed for intracellular interaction between the Colα1(I) and Colα2(I) polypeptides. Co-immunoprecipitation of Colα2(I) was observed when Colα1(I) was targeted for immunoprecipitation via the HA epitope tag and the reverse (co-immunoprecipitation of Colα1(I) with Colα2(I) immunoprecitipation) was also observed, indicating that the two collagen-I constituents are indeed associating intracellularly (Figure 2.5). We also tested whether the secreted collagen-I was assembled into trimers. As shown in Figure 2.6, we found that the secreted collagen-I is

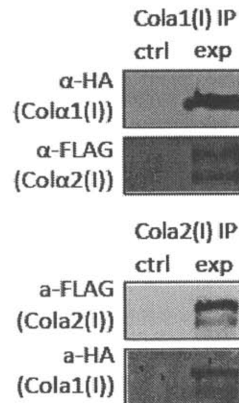


Figure 2.5. Co-Immunoprecipitation of Colα1(I) with Colα2(I) and vice versa. Colα2(I) was pulled down by HA (Colα1(I)) immunoprecipitation, and Colα1(I) was pulled down by FLAG (Colα2(I)) Immunoprecipitation.

indeed assembled into trimers in a disulfide-dependent matter, as evidenced by comparison of DTT-treated and untreated samples.

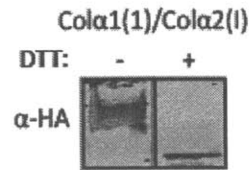


Figure 2.6. Disulfide-dependent assembly of secreted collagen-I triple helices.

Another important check on the validity of the Colα1(I)/Colα2(I) collagen-I expression system is whether the collagen-I produced is properly post-translationally modified. The collagen-I C-propeptide domain contains an N-glycosylation consensus sequence that is modified with an N-glycan in endogenously-produced collagen-I.[35] In order to probe for the presence of this N-glycan modification in the collagen-I produced by the HT1080 system, cell lysate from the HT1080 cells was enriched for collagen-I by HA immunprecipitation and treated or untreated with PNGase F, an endoglycosidase that cleaves N-linked glycans. As shown in Figure 2.7, the Western blot results show a decrease in the molecular weight of the Colα1(I) band after treatment with PNGase F. Figure 2.7 also shows a loss of Concanovalin A (ConA is a lectin that binds specifically to N-linked glycans) reactivity after treatment with PNGase

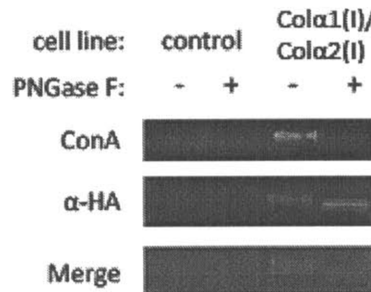


Figure 2.7. N-glycosylation of collagen-I. A decrease in the molecular weight of the Colα1(I) band is observed on treatment with PNGase F, an enzyme that cleaves N-linked glycans. Additionally, treatment with PNGase F causes a loss of ConA reactivity. Together these results indicate that the collagen produced in the HT1080 cells is N-glycosylated.

[35] Lamandé, S. R. & Bateman, J. F. The Type I Collagen proalpha(I) COOH-terminal Propeptide N-Linked Oligosaccharide. *J. Biol. Chem.* 270, 17858–17865 (1995).

F, further confirmation that the collagen-I produced by the HT1080 system is properly N-glycosylated.[36]

Another critical post-translational modification present in endogenously-produced collagen-I is 4-hydroxylation of proline residues located at the Yaa position in the repeating Gly-Xaa-Yaa triplet motif within the triple helical domain. This modification confers stability on the collagen-I triple helix and without it, secretion of collagen-I is reduced or delayed.[37] Secretion of stable collagen-I from primary patient fibroblasts requires treatment with ascorbate, a cofactor required for prolyl 4-hydroxylase function.[38] The hydroxylation level of the collagen-I produced by the HT1080 cells was tested by mass spectrometry in the presence or absence of ascorbate. As shown in Figure 2.8, the number of hydroxylated peptides in the ascorbate-treated sample increased by greater than ten-fold compared to the untreated sample. Furthermore, a wide range of hydroxylated collagen-I peptides and hydroxylation states were observed in the mass spectrometry data. These results strongly suggest that the collagen-I produced by the HT1080 cells is being properly hydroxylated.

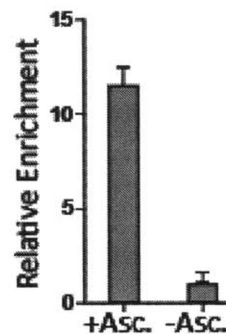


Figure 2.8. Hydroxylation of the Yaa proline residue. Treatment with sodium ascorbate increases the number of peptides with hydroxylated Yaa-residues observed by mass spectrometry

[36] Gotkin, M. G., Ripley, C. R., Lamande, S. R., Bateman, J. F. & Bienkowski, R. S. Intracellular trafficking and degradation of unassociated proα2 chains of collagen type I. *Exp. Cell Res.* 296, 307–316 (2004).
[37] Pihlajaniemi, T., Myllylä, R. & Kivirikko, K. I. Prolyl 4-hydroxylase and its role in collagen synthesis. *J. Hepatol.* 13, S2–S7 (1991).
[38] Murad, S. *et al.* Regulation of collagen synthesis by ascorbic acid. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2879–2882 (1981).

An additional property of endogenously-produced collagen-I that is useful in assaying the folding of the collagen-I produced in the HT1080 system is that properly folded collagen-I triple helices are resistant to treatment with proteases, while the C-terminal and N-terminal propeptides remain susceptible to protease digestion. Western blot analysis of collagen-I produced by primary fibroblasts produces a characteristic banding pattern when digested with the proteases trypsin and chymotrypsin. In order to assay whether the collagen-I produced by the HT1080 cells is stable and properly folded, precipitated media samples from the HT1080 system were digested with trypsin and chymotrypsin and analyzed by Western blotting. As shown in Figure 2.9, the pattern produced by these digestions was compared to a control sample using a precipitated media sample isolated from primary fibroblasts. The banding pattern observed after protease treatment is the same for both samples, indicating that the collagen-I secreted by the HT1080 system is properly folded and stable.

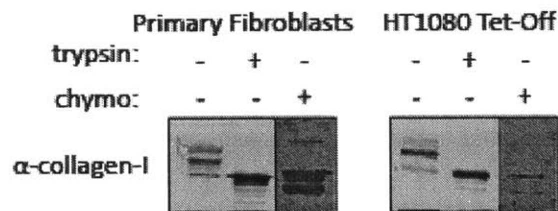


Figure 2.9. Trypsin treatment of collagen-I from primary fibroblasts and HT1080 cells. The collagen-I produced by the HT1080 cells is resistant to protease treatment and the banding pattern observed after protease treatment matches that obtained using collagen-I secreted by primary fibroblasts.

2.3 Discussion

Despite collagen-I's abundance, its essential function as the structural basis of skin, bone, and tendon, and the known complexity of its biosynthetic pathway, very little is known about the cellular mechanisms that assist in collagen-I folding and regulate its proteostasis.[39]

---

[39] Ishikawa, Y. & Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833, 2479–2491 (2013).

The little that is known about the collagen-I proteostasis machinery has either been discovered through genetic investigations or studies done in mouse model systems or primary patient fibroblasts.[40] A thorough and systematic investigation into the collagen-I folding and homeostasis machinery has not yet been accomplished. This type of study has been precluded by the lack of a biochemically tractable system that allows facile manipulation of collagen-I genes and immunoprecipitation and detection of the collagen-I protein.

This type of system has been difficult to develop because production of stable collagen-I molecules requires extensive post-translational modifications, therefore requiring a higher-eukaryotic expression system, and the collagen-I genes are challenging to work with due to their large size, high GC content, and repetitive sequence that results in frequent recombination events.[41] We have overcome these significant challenges using optimized molecular biology methods to generate collagen-I expression vectors that allow the inducible expression of epitope-tagged wild type Colα1(I) and Colα2(I) polypeptides as well as four OI-

---

[40] (a) Nagata, K., Saga, S. & Yamada, K. M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J. Cell Biol.* 103, 223–229 (1986).
(b) Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol.* 117, 903–914 (1992).
(c) Bächinger, H. P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262, 17144–17148 (1987).
(d) Morello, R. *et al.* CRTAP Is Required for Prolyl 3- Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* 127, 291–304 (2006).
(e) Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of osteogenesis imperfecta. *PLoS Genet.* 4, e7 (2008).
(f) Bodian, D. L. *et al.* Mutation and polymorphism spectrum in osteogenesis imperfecta type II: Implications for genotype-phenotype relationships. *Hum. Mol. Genet.* 18, 463–471 (2009).
(g) Gotkin, M. G., Ripley, C. R., Lamande, S. R., Bateman, J. F. & Bienkowski, R. S. Intracellular trafficking and degradation of unassociated proα2 chains of collagen type I. *Exp. Cell Res.* 296, 307–316 (2004).
(h) Fitzgerald, J., Lamandé, S. R. & Bateman, J. F. Proteasomal degradation of unassembled mutant type I collagen pro-alpha1(I) chains. *J. Biol. Chem.* 274, 27392–27398 (1999).
[41] Myllyharju, J. in *Methods Mol. Biol. Extracell. Matrix Protoc.* 552, 51–62 (2009).

causing mutant versions of each. We then used these vectors to create a cell-based platform that recombinantly produces epitope-tagged collagen-I.

As the first step toward creating cell lines that express the physiologically-relevant heterotrimeric collagen-I, we first generated stable single colony cell lines that express only either the *COLA1* or *COL1A2* genes, and therefore produce collagen-I homotrimers. We found that both Colα1(I) and Colα2(I) were detectable in both lysate and media fractions (Figure 2.3). Given that it is highly unlikely that a significant fraction of non-triple helical collagen would be able to pass through the secretory pathway, this suggests that both Colα1(I) and Colα2(I) are forming homotrimers.

Secretion of Colα1(I) homotrimers has been previously reported both natively (in cases where both collagen-I genes are intact), for example in fetal, fibrotic, or cancerous tissues, and also in cases where the *COL1A2* gene is absent due to mutations, such as in the *oim* mouse model which completely lacks *COL1A2* mRNA.[42] In all of these cases, the tissues containing the Colα1(I) homotrimers exhibit OI-like phenotypes, with overmodified Colα1(I) strands and altered fibril formation. Intriguingly, secretion of Colα2(I) homotrimers has never before been observed until our work here. In fact, the mutation in the Mov-13 mouse model system which causes expression of only Colα2(I) results in embryonic lethality, and the Colα2(I) protein produced by those cells has been found to be immediately degraded by a autophagy-

[42] (a) Bailey, A. J., Sims, T. J. & Knott, L. Phenotypic expression of osteoblast collagen in osteoarthritic bone: Production of type I homotrimer. *Int. J. Biochem. Cell Biol.* 34, 176–182 (2002).
(b) Han, S. *et al.* Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *J. Biol. Chem.* 285, 22276–22281 (2010).
(c) Chipman, S. D. *et al.* Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1701–1705 (1993).
(d) McBride, D. J., Choe, V., Shapiro, J. R. & Brodsky, B. Altered collagen structure in mouse tail tendon lacking the alpha 2(I) chain. *J. Mol. Biol.* 270, 275–284 (1997).

independent lysosomal pathway, although the mechanism by which these strands are targeted to degradation remains unclear.[43] Characterization of the Colα2(I) (and Colα1(I)) produced in our system will therefore be an interesting future research topic.

After creating isogenic cell lines singly expressing either Colα1(I) or Colα2(I), we next created three cell lines expressing both proteins (either wild type, G352S, or C1299W Colα1(I) with wild type Colα2(I)). Thus far, our characterization of the collagen-I produced by these cell lines has focused on the wild-type Colα1(I) and Colα2(I) cell line in order to confirm that the HT1080 cell type is capable of producing collagen-I that is properly folded and post-translationally modified. We have shown that the collagen-I secreted by the Colα1(I)/Colα2(I) cell line is properly post-translationaly modified, as shown by evidence of N-glycosylation detected by treatment with PNGase F and Concanovalin A (Figure 2.7) and the ascorbate-dependent increase in hydroxylation of proline and lysine residues detected by LC-MS.MS analysis.[44] We have further shown that the Colα1(I)/Colα2(I) cell line secretes stable, disulfide-dependent, heterotrimeric collagen-I as evidenced by protease resistance treatments (Figure 2.9) and treatment with DTT (Figure 2.6). Altogether this evidence shows that the collagen-I secreted by this system has properties that match those of endogenously-produced collagen-I, indicating that the HT1080 cells are producing collagen-I that is properly folded and post-

---

[43] (a) Harbers, K., Kuehn, M., Delius, H. & Jaenisch, R. Insertion of retrovirus into the first intron of alpha 1(I) collagen gene to embryonic lethal mutation in mice. *Proc. Natl. Acad. Sci. U. S. A.* 81, 1504–1508 (1984).
(b) Ishida, Y. et al. Autophagic Elimination of Misfolded Procollagen Aggregates in the Endoplasmic Reticulum as a Means of Cell Protection. 20, 2744–2754 (2009).
[44] (a) Lamandé, S. R. & Bateman, J. F. The Type I Collagen proalpha(I) COOH-terminal Propeptide N-Linked Oligosaccharide. *J. Biol. Chem.* 270, 17858–17865 (1995).
(b) Pihlajaniemi, T., Myllylä, R. & Kivirikko, K. I. Prolyl 4-hydroxylase and its role in collagen synthesis. *J. Hepatol.* 13, S2–S7 (1991).

translationally modified. This system will therefore provide an ideal platform for future systematic study of the collagen-I proteostasis network.

2.4 Conclusions and Future Work

We have created vectors for expression of both wild type and four different OI-mutants for each Colα1(I) and Colα2(I) and associated cell-based platforms that express either Colα1(I) alone, Colα2(I) alone or wild type Colα1(I) with wild type Colα2(I), as well as two OI model cell lines that express G352S Colα1(I) with wild type Colα2(I) or and C1299W Colα1(I) with wild type Colα2(I). We have shown that the wild type Colα1(I)/ Colα2(I) cell line produces collagen that is properly folded and post-translationally modified, which indicates that the HT1080 system is physiologically relevant. This cell line has already been used to conduct a systematic proteomics-based investigation into the wild type collagen-I proteostasis network (described in Chapter 3).

Important future work will include the creation of additional OI model cell lines expressing each of the mutant collagen-I expression vectors we have created and the characterization of the collagen-I produced by each of the OI model cell lines we generate. The cell lines described here were created using standard transfection methods. Even with highly optimized transfection protocols, the efficiency of collagen-I expression plasmid transfection was low and we had to screen a large number of single colonies in order to obtain a colony with high levels of collagen-I expression. For future creation of additional OI model cell lines, we plan to adopt a lentiviral gene transduction strategy using the Tet-Off lentiviral vectors pLenti-X

(ClonTech) that should serve as a much more robust strategy for introducing collagen-I genes into the HT1080 cells.

Development of these OI model cell lines will allow the future study of the properties of the OI-mutant collagen-I and the investigation of how the cellular proteostasis machinery differentially engages misfolding mutant collagen-I versus wild type collagen-I with the ultimately goal of identifying targets that will allow the development of new treatments for Osteogenesis Imperfecta that target the underlying cause of the disease. An additional intriguing direction for future work is the characterization of the collagen-I produced by the cell lines that singly express the wild type version of Colα1(I) or Colα2(I).

2.5 Experimental Procedures

Plasmid DNA: An oligonucleotide containing the pre-protrypsin signal sequence attached to either a FLAG or HA epitope tag was ordered from IDT and inserted into a pTRE-Tight vector (ClonTech). The *COL1A1* and *COL1A2* genes were obtained from the Origene True Clone Repository (accession numbers SC112997 and SC126717 respectively). *COL1A1* was amplified using primers that introduced a Not1 site at the 5′ end of the gene and an Xba1 site at the 3′ end and was inserted into the pre-protrypsin.HA.pTRE-Tight vector using the Not1 and Xba1 restriction sites. *COL1A2* was amplified using primers that introduced a Not1 site at the 5′ end of the gene and an EcoRV site at the 3′ end and was inserted into the pre-protrypsin.FLAG.pTRE-Tight vector using the Not1 and EcoRV restriction sites. SURE 2 supercompetent cells (Agilent Technologies), an *E. coli* cell line that is engineered to lack components of the cellular machinery that catalyze rearrangement or deletion of DNA

segments and is therefore recombination deficient, were used for all steps of collagen-I expression vector creation. To generate mutant *COL1A1* and *COL1A2* vectors with OI mutations of interest, the pre-protrypsin.HA.COL1A1.pTRE-Tight and pre-protrypsin.FLAG.COL1A2.pTRE-Tight vectors were mutagenized using a QuickChange XL kit (Agilent Technologies) and the HA or FLAG epitope tags were replaced by a cMyc epitope tag. A GFP.pTRE-Tight vector was used as a control. All plasmids were purified using the Omega BioTech Midi Prep Kit according to manufacturer's instructions.

Antibodies: Mouse monoclonal anti-HA (Santa Cruz; sc7392), anti-cMyc (Sigma Aldrich; M4439), anti-β-actin (Sigma Aldrich; A5441), anti-FLAG-M1 (Sigma Aldrich; F3040), anti-KDEL (Enzo Life Sciences; ADI-SPA-827), rat monoclonal anti-FLAG-M2 (Agilent Technologies; 200472), and rabbit polyclonal anti-COL1A1 internal (MyBioSource; MBS502153), anti-COL1A2 internal (Sigma Aldrich; SAB4500363) antibodies were used.

Transfection and Selection of Cells Stably Expressing COL1A1 and/or COL1A2: The HT1080 Tet- Off® cells (ClonTech) were co-transfected with either the COL1A1.pTRE-Tight plasmid and a puromycin linear drug resistance marker (ClonTech) or the COL1A2.pTRE-Tight plasmid and a hygromycin linear drug resistance marker (ClonTech) using the Xfect transfection reagent (ClonTech) according to manufacturer's instructions. Cells containing the COL1A1.pTRE-Tight or COL1A2.pTRE-Tight plasmids were selected for over 14 days by culturing the cells in the presence of puromycin (0.25 μg/mL; Corning) or hygromycin (150 μg/mL; Enzo Life Sciences). Heterogeneous populations of HT1080 Tet- Off® cells stably expressing either COL1A1.pTRE-Tight or COL1A2.pTRE-Tight were plated at dilute concentrations (1:5000, 1:10,000, and 1:20,000 dilutions) and single colonies were selected and analyzed for expression

of either Colα1(I) or Colα2(I) by qPCR and Western blotting. The single colony stably expressing COL1A2.pTRE-Tight under hygromycin selection was co-transfected with the COL1A1.pTRE-Tight plasmid (or one of the mutant versions, G352S or C1299W) and a puromycin linear drug resistance marker (ClonTech) and cells containing the COL1A1.pTRE-Tight plasmid (or mutants) were selected for over 14 days by culturing the cells in the presence of puromycin (0.25 μg/mL; Corning). The COL1A2.pTRE-Tight single colony cell line heterogeneously expressing COL1A1.pTRE-Tight (or mutants) was plated at dilute concentrations (1:5000, 1:10,000, and 1:20,000 dilutions) and single colonies were selected and analyzed for expression of both Colα1(I) and Colα2(I) by qPCR and Western blotting.

Mammalian Cell Culture: HT1080 Tet- Off® cells (ClonTech; a human fibrosarcoma cell line stably expressing the tetracycline-regulated transactivator Tet-Off) stably expressing Colα1(I) and Colα2(I) were cultured at 37 °C in a 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS supplemented with penicillin, streptomycin, glutamine (Mediatech). The cell lines were continuously cultured in G418 (100 μg/mL; Enzo Life Sciences), puromycin (0.25 μg/mL; Corning) and hygromycin (150 μg/mL; Enzo Life Sciences) to maintain the presence of the transgenes, and doxycycline (1 μg/mL; EMD) was added to suppress expression of Colα1(I) and Colα2(I).

Induction of Collagen-I Expression: To induce expression of Colα1(I) and Colα2(I), the HT1080 Tet- Off® cell lines stably expressing either COL1A1.pTRE-Tight, COL1A2.pTRE-Tight, both, or mutant COL1A1.pTRE-Tight with COL1A2.pTRE-Tight were cultured in tetracycline-free media (DMEM supplemented with penicillin, streptomycin, glutamine, and Tet Approved FBS

(ClonTech; tetracycline-free FBS) for 48 –72 h, supplementing with fresh sodium ascorbate (50 μM) every 24 h.

Immunoblotting: Cells with collagen-I expression induced were harvested using 0.25% Trypsin + EDTA (Cellgro), washed twice with 1X phosphate-buffered saline (PBS), and lysed using lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM $MgCl_2$, 1% Triton X-100, protease inhibitor tablets (Thermo Fisher), 1.5 mM phenylmethylsulfonyl fluoride). Cell lysate or media samples were boiled for 10 minutes with 1X gel loading buffer (50 mM Tris-HCl pH 6.8, 3% glycerol, 1% SDS, 20 μg/mL bromophenol blue, 16.6 mM DTT) analyzed by SDS-PAGE (8% polyacrylamide), and transferred to nitrocellulose membranes (Bio-Rad Laboratories). The membranes were blocked with 5% milk in 1X Tris Buffered Saline (TBS) for 30 minutes and incubated with the appropriate primary antibodies for 2 h at room temperature or overnight at 4 °C. The immunoblots were washed three times with TBS + 1% Tween (TBS-T) and were incubated with 680 nm or 800 nm fluorophore-labeled secondary antibodies (LI-COR Biosciences) for 1 h at room temperature. The blots were washed twice with TBS-T, once with TBS, and were imaged with the Odyssey Infrared Imaging System (LI-COR Biosciences).

Table 2.2. Primers used Collagen-I qPCR.

Primer Name	Sequence
RPLP2 Fwd	5'-CCATTCAGCTCACTGATAACCTT-3'
RPLP2 Rev	5'-CGTCGCCTCCTACCTGCT-3'
COL1A1 Fwd	5'-TGGTAGCCGTGGTTTCCCTG-3'
COL1A1 Rev	5'-TCCAGTCAGACCCTTGGCAC-3'
COL1A2 Fwd	5'-TGGCTCGAGAGGTGAACGTG-3'
COL1A2 Rev	5'-AGCACCGTTGACTCCAGGAC-3'

Quantitative RT-PCR: mRNA samples were isolated using the Omega mRNA Isolation Kit and cDNA was synthesized using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems Technology) according to manufacturer's instructions. Primers for qPCR analysis were designed

using Primer3 version 4.0.0 (Human mispriming library, product ranges 130 – 180 bp, Tm 57 – 62 – 65, one CG clamp). Quantitative PCR was done using primers specific for either *COL1A1*, *COL1A2*, or an *RPLP2* control and Sybr Green (Applied Biosystems Technology) according to the manufacturer's instructions (see Table 2.2).

PNGase F Digestions: Collagen-I expression was induced in the HT1080 Tet- Off® cell line stably expressing Colα1(I) and Colα2(I) by removal of doxycycline for 72 h and the resulting cell lysate was digested with PNGase F (New England BioLabs) according to manufacturer's instructions. The samples were analyzed by SDS-PAGE (8% polyacrylamide) and Western blotting.

Protease Digestion of Triple Helical Collagen-I: Collagen-I expression was induced in the HT1080 Tet- Off® cell line stably expressing Colα1(I) and Colα2(I) by removal of doxycycline for 72 h. A sample of the resulting media (30 mL) was harvested and precipitated following a previously reported protocol. The precipitated samples were centrifuged at 3000 rpm for 30 min and the pellets were resuspended in buffer (400 mM sodium chloride, 150 mM Tris, pH 7.5). Samples were aliquoted, treated with trypsin (0.1 mg/mL final concentration; Sigma Aldrich) or chymotrypsin (0.1 mg/mL final concentration; Sigma Aldrich), and analyzed by SDS-PAGE and Western blotting.

Immunoprecipitations: HT1080 cells with collagen-I expression induced by removal of doxycycline for 72 h were harvested using 0.25% Trypsin + EDTA (Cellgro), washed twice with 1X phosphate-buffered saline (PBS), and lysed using lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM $MgCl_2$, 1% Triton X-100, protease inhibitor tablets (Thermo Fisher), 1.5 mM phenylmethylsulfonyl fluoride). The lysed samples were centrifuged

at 21,100 x g for 15 minutes at 4 °C. The protein content of the cell lysate supernatant samples was normalized and the lysate samples were incubated for 16 h with antibody-conjugated beads (either HA, FLAG-M2, FLAG-M1, or cMyc). The antibody-conjugated beads were washed with lysis buffer, resuspended in a Tris-SDS solution (300 mM Tris, pH 7.5, 6% SDS), and boiled for 10 minutes to elute the immunoprecipitated proteins. The samples were analyzed by SDS-PAGE and Western blotting.

Hydroxyproline Content: The hydroxyproline content of the collagen-I produced by the HT1080 cells stably expressing Colα1(I) and Colα2(I) was analyzed by mass spectrometry (see Chapter 3).

Chapter Three:
Mechanistic Exploration of Novel Collagen-I Interacting Proteins Identified by SILAC Mass Spectrometry

3.1 Introduction

Osteogenesis Imperfecta (OI) is an often-severe connective tissue disorder with clinical features that include low bone mass, bone deformity, and growth deficiency.[45] OI is most-commonly caused (approximately 90% of all cases) by autosomal dominant mutations in the genes that encode type I collagen.[46] These mutations result in delayed procollagen-I folding, slow triple helix assembly, and ultimately, cause production of mature collagen-I proteins with reduced stability. The presence of these malformed triple helices in bone leads to structural defects and increased bone remodeling, ultimately resulting in the brittle bone phenotype that is characteristic of OI.[47] Alternatively, it has also been suggested that the OI phenotype may also result from apoptosis of collagen-I producing cells due to the prolonged stress caused by the presence of the misfolding mutant collagen-I molecules inside the ER.[48] Despite decades of effort, there is currently no cure for OI, and treatments are limited to palliative care. Development of treatments that target the underlying cause of OI will require a greater knowledge of the cellular protein homeostasis machinery that assists in collagen-I biosynthesis

---

[45] Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J. C. New Perspectives on Osteogenesis Imperfecta. *Nat. Rev. Endocrinol.* 7, 540–557 (2011).
[46] Marini, J. C. *et al.* Consortium for Osteogenesis Imperfecta Mutations in the Helical Domain of Type I Collagen : Regions Rich in Lethal Mutations Align With Collagen Binding Sites for Integrins and Proteoglycans. 28, 209–221 (2007).
[47] Forlino, A., *et al.* (2011).
[48] Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of Osteogenesis imperfecta. *PLoS Genet.* 4, (2008].

and folding, and a better understanding of how this proteostasis network differentially engages mutant and wild type collagen-I.

Despite collagen-I's abundance, essential function as the structural basis of skin, bone, and tendon, and the known complexity of its biosynthetic pathway, very little is known about the cellular mechanisms that assist in collagen-I folding and regulate its proteostasis. The known components of the collagen-I proteostasis machinery were identified either due to their abundance, by rational prediction, or by low throughput methods such as genetic investigations or studies done in either mouse model systems or primary patient fibroblasts.[49] For example, the role of CRTAP, a member of the collagen-I prolyl 3-hydroxylation complex was discovered because CRTAP null mutations cause recessive Osteogenesis Imperfecta.[50] The collagen-specific chaperone, HSP47, was discovered due to its abundance and tight collagen-I binding interaction.[51] The role of peptidyl-prolyl isomerases in collagen-I folding was predicted due to the high proline content of collagen-I, leading to *in vitro* and *in vivo* studies showing that

---

[49] (a) Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of Osteogenesis imperfecta. *PLoS Genet.* 4, (2008).
(b) Bodian, D. L. *et al.* Mutation and polymorphism spectrum in osteogenesis imperfecta type II: Implications for genotype-phenotype relationships. *Hum. Mol. Genet.* 18, 463–471 (2009).
(c) Gotkin, M. G., Ripley, C. R., Lamande, S. R., Bateman, J. F. & Bienkowski, R. S. Intracellular trafficking and degradation of unassociated proα2 chains of collagen type I. *Exp. Cell Res.* 296, 307–316 (2004).
(d) Fitzgerald, J., Lamandé, S. R. & Bateman, J. F. Proteasomal degradation of unassembled mutant type I collagen pro-alpha1(I) chains. *J. Biol. Chem.* 274, 27392–27398 (1999).
(e) Nagata, K., Saga, S. & Yamada, K. M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J. Cell Biol.* 103, 223–229 (1986).
(f) Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol.* 117, 903–914 (1992).
(g) Bächinger, H. P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262, 17144–17148 (1987).
(h) Morello, R. *et al.* CRTAP Is Required for Prolyl 3- Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* 127, 291–304 (2006).
[50] Morello, R. *et al.* (2006).
[51] (a) Nagata, K., *et al.* (1986).
(b) Nakai, A., *et al.* (1992).

peptidyl-prolyl isomerases served to increase collagen-I folding speed.[52] While these traditional genetic and primary cell-based methods have led to the discovery of these and a limited set of other components of the collagen-I proteostasis machinery, an alternative, unbiased, comprehensive, and higher-throughput method for studying the system would greatly help to facilitate the study of the collagen-I proteostasis network.

A thorough and systematic investigation into the collagen-I folding and homeostasis machinery has been precluded until recently by the lack of a biochemically tractable system that allows facile manipulation of collagen-I genes and the immunoprecipitation and detection of collagen-I protein. This type of system has been difficult to develop since the production of stable collagen-I molecules requires extensive post-translational modifications, therefore requiring a higher-eukaryotic expression system, and the collagen-I genes are challenging to work with due to their large size, high GC content, and repetitive sequence that results in frequent recombination events.[53]

We have overcome these obstacles and have developed a cell-based platform for the systematic elucidation of the collagen-I proteostasis network. The Colα1(I)/Colα2(I) collagen-I model cell line we created (see Chapter 2 description) stably and inducibly expresses antibody epitope-tagged collagen-I and provides an ideal system for studying the collagen-I proteostasis network by proteomic methods because the epitope tags appended to both Colα1(I) and Colα2(I) allow immunoprecipitation of these proteins. This system will allow us to perform mass spectrometry proteomics studies using stable isotope labeling of amino acids in cell culture

[52] Bächinger, H. P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262, 17144–17148 (1987).
[53] Myllyharju, J. in *Methods Mol. Biol. Extracell. Matrix Protoc.* 552, 51–62 (2009).

(SILAC) to assess the relative quantities of each protein that interact with collagen-I under a particular treatment condition, thereby establishing a mass spectrometry-based proteomics as an unbiased and comprehensive technique to characterize the collagen-I proteostasis network.

Stable isotope labeling of amino acids in cell culture (SILAC) is a simple and unbiased method for studying cellular protein interaction networks that has been successfully applied to a wide variety of biological problems.[54] In the SILAC method, the proteins in samples that undergo different treatments are labeled with either light (no isotopes), medium (^{13}C-labeled lysine and arginine) or heavy (^{13}C- and ^{15}N-labeled lysine and arginine) amino acids (Figure 3.1). This labeling allows the relative abundance of individual proteins to be quantitatively compared across samples via comparison of the relative ratios of the intensities of the peaks observed for a given peptide identified by mass spectrometry.[55]

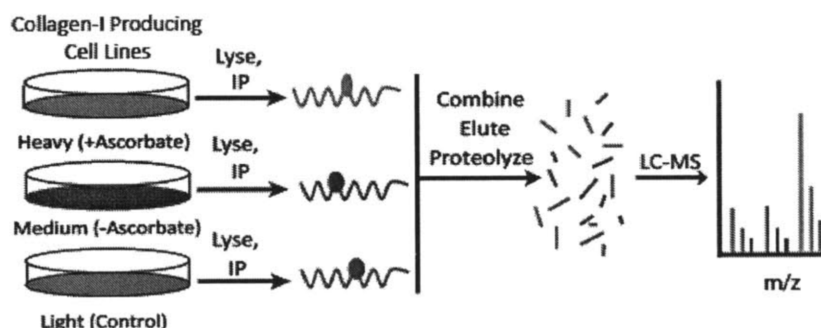


Figure 3.1. Workflow for stable isotope labeling by amino acids in cell culture (SILAC). In the SILAC experimental workflow, separate cell populations are grown in the presence of light, medium, or heavy media (distinguished by the presence of amino acids labeled with stable isotopes) for 6 passages to ensure complete incorporation of the isotopically-labeled amino acids. The labeled cell populations are subjected to different treatments and then samples are lysed, mixed, and analyzed by mass spectrometry. The relative ratios of the peak intensities for a given peptide indicate the relative abundance of that protein in the corresponding samples.

---

[54] Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958 (2006).

[55] Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386 (2002).

We used SILAC-assisted mass spectrometry to investigate the protein interaction network of wild type collagen-I. The Colα1(I)/Colα2(I) collagen-I cell line was cultured in the presence or absence of ascorbate, a required cofactor for prolyl 4-hydroxylase. The interactomes of collagen-I under these two experimental condition were compared, using cells that do not express epitope-tagged collagen-I as a control. Impressively, in addition to identifying all of the proteins previously known to interact with collagen-I, this method also identified more than 25 previously unknown collagen-I interactors. We are currently optimizing procedures for the shRNA-based knockdown of these novel collagen-I interactors, most of which appear to be valid and reproducible. Future work will establish their mechanistic roles in the collagen-I proteostasis machinery.

3.2 Results

In order to delineate the proteostasis machinery that is involved in wild type collagen-I biosynthesis and folding, we used SILAC-assisted mass spectrometry to characterize the wild type collagen-I interactome. The Colα1(I)/Colα2(I) collagen-I model cell line (see Chapter 2 description) stably and inducibly expresses antibody epitope-tagged collagen-I. This cell line provides an ideal system for studying the collagen-I proteostasis network by proteomic methods because the HA epitope tag appended to Colα1(I) and the FLAG epitope tag appended to Colα2(I) allow orthogonal immunoprecipitation of these two proteins.

As a first step in our study of the network of proteins that interact with wild type collagen-I, we immunoprecipitated the Colα1(I) produced in the Colα1(I)/Colα2(I) cell line using the attached HA epitope tag. The resulting samples were analyzed by Western blot, and probed for co-immunoprecipitation of known collagen-I interacting proteins. Using the

traditional (stable) immunoprecipitation protocol (without crosslinker), only a few of the known collagen-I interacting proteins were pulled down at detectable levels (see Figure 3.2). These proteins were Colα2(I), Hsp47 (a collagen-specific chaperone protein), and PDIA1 (a protein disulfide isomerase and the beta subunit of prolyl-4-hydroxylase).[56] These are proteins that are well-known to have stable interactions with collagen-I. Since many of the interactions between collagen-I and the proteins involved in regulating its folding and homeostasis are likely to be transient, the limited number of proteins detected after stable co-immunoprecipitation was not surprising.

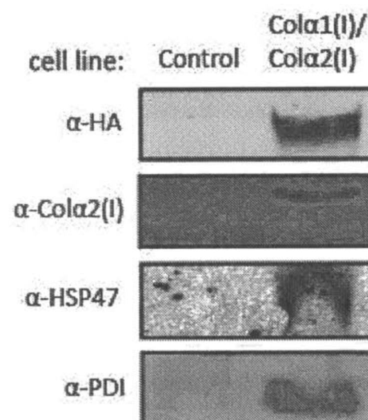


Figure 3.2. Immunoprecipitation of stable collagen-I interactors. Using a traditional IP protocol (without a covalent crosslinker) only the most stable collagen-I interacting proteins were successfully pulled down. This result is not surprising given the transient nature of most associations between chaperones and their clients.

In order to stabilize the transient associations between Colα1(I) and other interacting proteins that were not detectable by stable immunoprecipitation, we treated the Colα1(I)/Colα2(I) cells with a bifunctional crosslinking reagent known as Lomant's Reagent (dithiobis(succinimidylpropionate), DSP), in order to immortalize the contacts between associating proteins.[57] Lomant's Reagent contains two succinimide moieties that react with the

---

[56] (a) Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol.* 117, 903–914 (1992).
(b) Wilson, R., Lees, J. F. & Bulleid, N. J. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. J. Biol. Chem. 273, 9637–9643 (1998).
[57] (a) Adir, N. & Ohad, I. Probing for the interaction of the 32 kDa-Qb protein with its environment by use of bifunctional cross-linking reagents. *Biochim. Biophys. Acta* 850, 264–274 (1986).

lysine residues in proteins resulting in a covalent linkage that can later be broken by treating with DTT to reduce the disulfide bond.[58] Both Colα1(I) and Colα2(I) contain over 50 lysine residues that are evenly distributed throughout the protein, meaning that treatment with Lomant's Reagent likely immortalizes the large majority of transient interactions with collagen-I. To optimize the crosslinking HA immunoprecipitation, we incubated the cells with various concentrations of Lomant's reagent (ranging from 0.1 mM to 2.0 mM) prior to lysis (Figure 3.3). This range of crosslinking reagent treatments allowed us to identify the treatment conditions that enabled the immunoprecipitation of significant levels of known transient collagen-I interacting proteins (FKBP10 and calreticulin (CRT)).[59] To avoid over-crosslinking and the immunoprecipitation of proteins that do not interact with collagen-I, we chose the mildest crosslinking

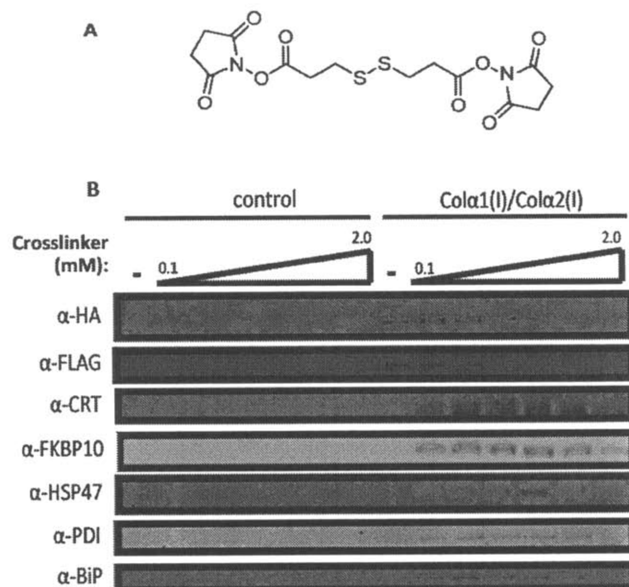


Figure 3.3. Optimization of the crosslinking HA (Colα1(I)) immunoprecipitation. A) Stucture of Lomant's Reagent
B) In order to identify the mildest covalent crosslinking treatments that allowed the detection of transient interactions with collagen-I, cells were treated for 15 min with various concentrations of Lomant's reagent.

(b) Agre, P., Gardner, K. & Bennett, V. Association between human erythrocyte calmodulin and the cytoplasmic surface of human erythrocyte membranes. *J. Biol. Chem.* 258, 6258–6265 (1983).

[58] Mattson, G. *et al.* A practical approach to crosslinking. *Mol. Biol. Rep.* **17,** 167–183 (1993).

[59] (a) Kelley, B. P. *et al.* Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. *J. Bone Miner. Res.* 26, 666–72 (2011).
(b) Van Duyn Graham, L., Sweetwyne, M. T., Pallero, M. a. & Murphy-Ullrich, J. E. Intracellular calreticulin regulates multiple steps in fibrillar collagen expression, trafficking, and processing into the extracellular matrix. *J. Biol. Chem.* 285, 7067–7078 (2010).

treatment that resulted in the reliable immunoprecipitation of the known transient collagen-I interatorsto. Treatment with 0.2 mM Lomant's Reagent for 15 min was identified as the optimal treatment condition to allow reliable co-immunoprecipitation of transient collagen-I interactors.

In order to probe for previously unknown collagen-I interacting proteins, we used our optimized crosslinking immunoprecipitation protocol in conjunction with the stable isotope labeling of amino acids in cell culture approach (introduced in Section 3.1). These techniques allowed us to compare the relative interactomes of collagen-I isolated from cells cultured in the presence or absence of ascorbate, a cofactor for prolyl 4-hydroxylation that is required for the formation of stable collagen-I molecules. Both of these treatment conditions were compared to a control sample lacking HA-tagged collagen-I in order to normalize for any proteins non-specifically pulled down with the HA-linked beads. The optimized crosslinking immunoprecipitation protocol was used to isolate the full complement of collagen-I interacting proteins and the samples were analyzed by mass spectrometry. The optimized crosslinking immunoprecipitation protocol was used to isolate the full complement of collagen-I interacting proteins and the samples were analyzed by mass spectrometry. As shown in Table 3.1, in addition to identifying all of the proteins previously known to interact with collagen-I (Table 3.1-A), this method also identified more than 25 previously unknown collagen-I interactors (Table 3.1-B). Interestingly, all of the proteins detected were present in both the ascorbate-treated and -untreated samples. There were no proteins detected exclusively in one treatment condition. In some cases, however, the amount of a particular protein detected varied between the two samples.

A Protein	Protein Abbreviation	With Ascorbate Fold Enrichment	Without Ascorbate Fold Enrichment	Peptide Count***
Peptidyl-prolyl cis-trans isomerase B	PPIB	41.2	22.0	9
Serpin H1	HSP47	30.5	37.6	4
Protein disulfide-isomerase	PDI	19.4	29.3	7
Procollagen galactosyltransferase 1	ColGalT1	14.0	8.5	3
78 kDa glucose-regulated protein	BiP	13.6	22.9	20
Cartilage-associated protein	CRTAP	13.2	23.5	4
Secreted protein acidic, rich in cysteine	SPARC	9.8	5.3	2
Collagen alpha-1(II) chain	Col 1(II)	9.7	4.4	3
Prolyl 4-hydroxylase subunit alpha-1	P4H 1	9.0	14.9	5
Prolyl 3-hydroxylase 1	P3H1	7.9	14.6	2
Collagen alpha-2(I) chain	Col 2(I)	7.8	9.2	55
Fibronectin	FN	5.2	4.5	9
Prolyl 3-hydroxylase 3	P3H3	4.8	9.4	3
Prolyl 4-hydroxylase subunit alpha-2	P4H 2	4.5	7.0	2
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	4.3	8.8	2
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	4.1	6.9	2
Peptidyl-prolyl cis-trans isomerase FKBP10	FKBP10	4.0	1.4	2
Calreticulin	Calr	3.4	5.5	4
Peptidyl-prolyl cis-trans isomerase FKBP9	FKBP9	2.4	1.7	2
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	2.3	2.4	2
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	PLOD1	2.0	5.3	3

B Protein	Protein Abbreviation	With Ascorbate Fold Enrichment	Without Ascorbate Fold Enrichment	Peptide Count***
Calumenin	Calu	20.4	25.3	5
Golgi integral membrane protein 4**	Golim4	14.3	19.0	2
Protein canopy homolog 2	CNPY2	12.6	17.9	3
Cytoskeleton-associated protein 4	CKAP4	11.5	11.1	5
Thioredoxin domain-containing protein 5	TXNDC5	8.4	13.5	3
Protein disulfide-isomerase A4	PDIA4	8.2	15.4	3
Reticulocalbin-1	RCN1	7.5	8.1	2
Protein disulfide-isomerase A3	PDIA3	7.4	14.3	10
Aspartyl-asparaginyl beta hydroxylase*	AspH	7.4	9.4	2
Glucosidase 2 subunit beta	PKCSH	6.9	7.2	4
Annexin A2	AnxA2	6.7	0.9	2
Endoplasmic reticulum resident protein 29	Erp29	4.8	8.3	2
ATP synthase subunit alpha, mitochondrial	ATP5A1	4.6	0.5	3
Endoplasmin	Grp94	4.4	6.4	13
ERO1-like protein alpha**	Ero1L	4.3	6.6	2
Protein disulfide-isomerase A6	PDIA6	4.1	8.3	5
Neutral alpha-glucosidase AB	GANAB	2.8	2.9	3
Major vault protein	MVP	2.6	1.9	19
Heat shock protein HSP 90-beta	HSP90	2.5	0.4	4
Heat shock protein HSP 90-alpha	HSP90	2.3	0.4	4
Poly [ADP-ribose] polymerase 4	PARP4	2.3	1.1	4
Hypoxia up-regulated protein 1*	HYOU1	2.2	6.1	3
Protein-glutamine gamma-glutamyltransferase 2	TGM2	2.0	1.6	10
Tubulin alpha-1A chain	TUBA1A	2.0	0.9	7
DnaJ homolog subfamily B member 11	Erdj3	2.0	2.6	3
Peroxiredoxin-1	PDX1	0.8	2.9	3
Endoplasmic reticulum resident protein 44*	Erp44	0	2.0	3

Table 3.1. Results of SILAC mass spectrometry experiments. An all-inclusive list of the proteins that appeared in at least two replicates, with a fold enrichment of greater than 2, and with values that were calculated from more than one peptide. A) Previously-known collagen-I interacting proteins. B) Newly-identified collagen-I interactors. Key: * Present in all three replicates, even if there were not ratios for all; **Present in two replicates, including those without ratios; ***Average peptide counts across biological triplicate/duplicate

Using this list of newly identified collagen-I interacting proteins, we have now begun to investigate the mechanistic roles of these proteins in the collagen-I proteostasis network. For our initial experiments, we are focusing on a subset of eight novel collagen interactors chosen either because their known functions could potentially enable them to play a role in collagen-I biosynthesis (ERO1L, RCN1, AnxA2, Calu, Erdj3, PDIA3, ERp44, and Golim4). Three proteins that are previously-known collagen-I interactors with known roles in the collagen-I biosynthesis pathway (FKBP10, HSP47, and P4HB (PDI)) are serving as positive controls.[60]

In order to confirm that the newly-identified collagen-I interacting proteins actually associate with the Colα1(I) collagen-I chain, we repeated the crosslinking HA immunoprecipitation, as was done prior to the mass spectrometry experiments, and probed for pull down of the novel collagen-I interacting proteins. As shown in Figure 3.4, we were able to successfully

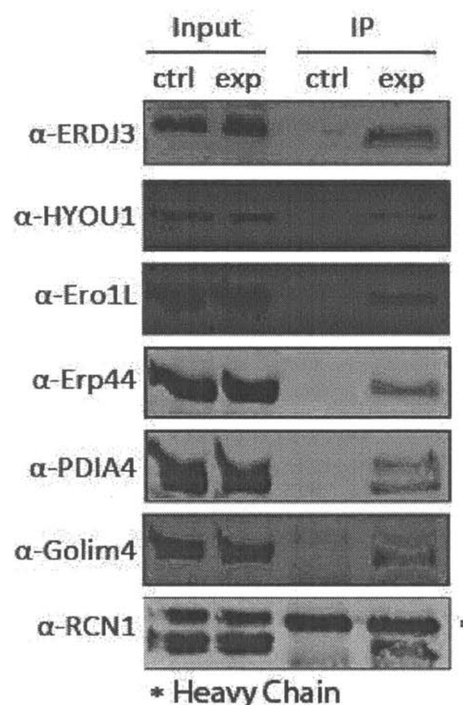


Figure 3.4. Crosslinking) co-IP of the newly identified collagen-I interacting proteins. Many of the novel collagen-I interactors identified by mass spectrometry were successfully co-immunoprecipitated by HA pull down.

---

[60] (a) Kelley, B. P. et al. Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. J. Bone Miner. Res. 26, 666–72 (2011).
(b) Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. J. Cell Biol. 117, 903–914 (1992).
(c) Wilson, R., Lees, J. F. & Bulleid, N. J. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. J. Biol. Chem. 273, 9637–9643 (1998).

immunoprecipitate many of the proteins of interest, and plan to continue to confirm that all of the proteins identified by the mass spectrometry experiment can be co-immunoprecipitated by crosslinking HA immunoprecipitation as we obtain antibodies that allow the Western blot detection of each of these proteins.

In order to investigate the mechanistic roles of the collagen-I interacting proteins, we plan to knockdown each target proteins of interest using shRNA and screen for defects in collagen-I secretion or triple helix stability. We are currently optimizing protocols to knockdown proteins of interest using lentiviral shRNA constructs in order to determine the appropriate time course that allows for the most efficient shRNA knockdown of each protein of interest. Figure 3.5 shows the qPCR data for a few proteins of interest from the experiments done to optimize their knockdown by shRNA. The samples were harvested 7 days (R1), 9 days (R2), or 11 days (R3) post-transduction with shRNA-producing lentivirus and each sample is normalized to a sample transduced with the CFP lentiviral control. The data for these proteins is representative of what was seen in many cases: for some of the proteins, shRNA knockdown was maintained at similar levels across all three time point, while for other proteins, the expression of the mRNA was rescued at the later time points, likely due to the lack of a selective agent to maintain the shRNA knockdown and the competitive advantage that the cells without the protein knockdown have. At the mRNA level, each of the proteins exhibited a slightly different time course of shRNA knockdown and we have identified the optimal time course for shRNA knockdown for most of our proteins of interest (see Figure 3.6). The next step will be to confirm that each protein of interest is also knocked down at the protein level. Once knockdown at both the mRNA and protein levels is confirmed and optimal time course for

each protein of interest is determines, these experimental protocols will be used to test for collagen-I secretion or triple helix stability defects.

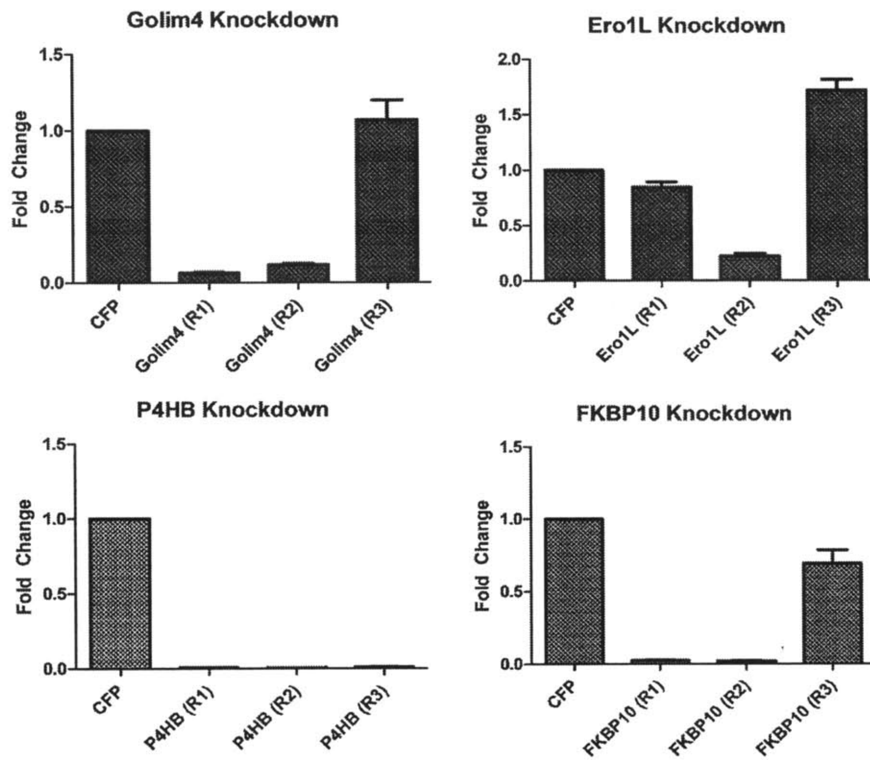


Figure 3.5. shRNA knockdown of several collagen-I interactors of interest. The mRNA levels of several collagen-I interactors (Golim4, Ero1L, P4HB, and FKBP10) were tested in samples harvested 7 (R1), 9 (R2), or 11 (R3) days after the Colα1(I)/Colα2(I) cell line was transduced with the corresponding lentiviral shRNA. All samples were normalized to a CFP lentivirus control.

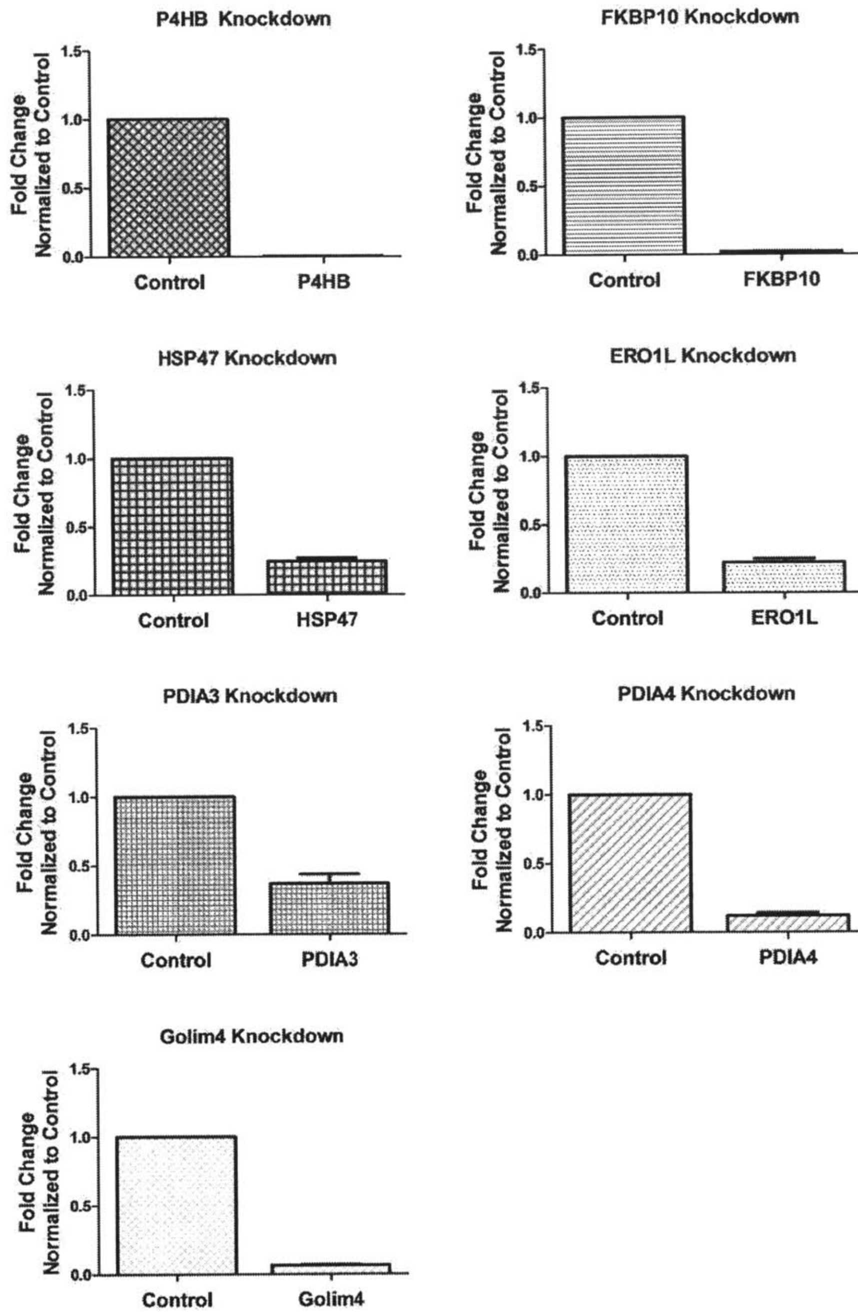


Figure 3.6. Optimized shRNA knockdown of each collagen-I interactor of interest. The knockdown of Golim4 was maximal 5 days after lentiviral transduction. The knockdown of P4HB, FKBP10, HSP47, ERO1L PDIA3, and PDIA4 were maximal 7 days after lentiviral transduction.

3.3 Discussion

The cellular proteostasis machinery that is responsible for assisting with the folding and secretion of type I collagen remains poorly characterized, in spite of the high abundance of collagen-I and its important role providing the structural basis of skin, bone, and tendon. This lack of knowledge has largely been due to the absence of an appropriate system that would allow the systematic study of the full complement of collagen-I interacting proteins by immunoprecipitation, mass spectrometry, or other techniques. To address this issue, we created an HT1080-based cell line that stably and inducibly expresses epitope-tagged wild type collagen-I that is properly post-translationally modified, assembled, and secreted (see Chapter 2). We have now used this Colα1(I)/Colα2(I) cell line to investigate the full network of proteins that interact with Colα1(I), generating a much more complete identification of the Colα1(I) interacome, and demonstrating that SILAC-assisted mass spectrometry will be a useful tool for the study of the collagen-I proteostasis networks and the study of other fibrous proteins.

Using SILAC-assisted mass spectrometry, we identified, not only all of the known collagen-I interacting proteins (Table 3.1-A), but also more than 25 previously unidentified collagen-I interactors (Table 3.1-B). Almost all of the proteins identified are ER-localized (except for tubulin alpha-1A chain (TUBA1A), major vault protein (MVP), ATP synthase subunit alpha (ATP5A1), and Hsp90) and they have a diverse variety of known functions. For example, reticulocalbin (RCN1) and calumenin (Calu) are members of a family of calcium-binding, multiple EF-hand motif proteins, while ERO1-like protein alpha (ERO1Lα), ER-resident protein 44 (ERp44) and thioredoxin domain-containing protein 5 (TXNDC5) are all involved in oxidative

protein folding and maintaining the oxidative balance in the ER.[61] Thus the SILAC-assisted mass spectrometry technique was able to identify a large variety of novel collagen-I interactors that have various known functions in the ER and likely have additional previously unknown roles in collagen-I proteostasis.

In addition to identifying many previously-unknown collagen-I interacting proteins, we also compared the network of Colα1(I) interacting proteins in cells that were either treated or untreated with ascorbate. Ascorbate is a cofactor that is required for the proper function of proyly 4-hydroxylase, the protein responsible for the hydroxylation of the Yaa residues in the repeating Gly-Xaa-Yaa motif that characterizes the collagen-I triple helix.[62] In the absence of 4-hydroxylation of the Yaa proline residues, the stability of the triple helix is decreased, and production of collagen-I triple helices is typically reduced or delayed.[63] We hypothesized that the protein interaction networks of Colα1(I) would differ between cell populations grown in the presence or absence of ascorbate due to delayed triple helix production in the sample lacking ascorbate. Interestingly, there were no proteins that were detected exclusively in one

---

[61] (a) Tachikui, H., Navet, a F. & Ozawa, M. Identification of the Ca(2+)-binding domains in reticulocalbin, an endoplasmic reticulum resident Ca(2+)-binding protein with multiple EF-hand motifs. *J. Biochem.* **121,** 145–149 (1997).
(b) Yabe, D., Nakamura, T., Kanazawa, N., Tashiro, K. & Honjo, T. Calumenin, a Ca2+ -binding protein retained in the endoplasmic reticulum with a novel carboxyl-terminal sequence, HDEF. *J. Biol. Chem.* **272,** 18232–18239 (1997).
(c) Zhang, L. *et al.* Different Interaction Modes for Protein-disulfide Isomerase (PDI) as an Efficient Regulator and a Specific Substrate of Endoplasmic Reticulum Oxidoreductin-1 (Ero1). *J. Biol. Chem.* **289,** 31188–31199 (2014).
(d) Anelli, T. *et al.* ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *EMBO J.* **21,** 835–844 (2002).
(e) Cabibbo, a *et al.* ERO1-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum [In Process Citation]. *J. Biol. Chem.* **275,** 4827–4833 (2000).
(f) Horna-Terrón, E., Pradilla-Dieste, A., Sánchez-de-Diego, C. & Osada, J. TXNDC5, a Newly Discovered Disulfide Isomerase with a Key Role in Cell Physiology and Pathology. *Int. J. Mol. Sci.* **15,** 23501–23518 (2014).
[62] Pihlajaniemi, T., Myllylä, R. & Kivirikko, K. I. Prolyl 4-hydroxylase and its role in collagen synthesis. J. Hepatol. 13, S2–S7 (1991).
[63] Murad, S. et al. Regulation of collagen synthesis by ascorbic acid. Proc. Natl. Acad. Sci. U. S. A. 78, 2879–2882 (1981).

treatment condition, indicating that these two populations of Colα1(I) largely engage the same protein interaction partners. In some cases, however, the amount of a particular protein detected varied between the two samples. Based purely on the enrichment ratios detected by mass spectrometry, it is not possible to make any definite statements about the degree to which collagen-I engages different protein interaction partners in the presence or absence of ascorbate. Future experiments can be done to biochemically characterize these interactions and determine whether there are differences in the ascorbate-treated and –untreated cases and what these differences are.

Using the list of proteins identified by SILAC-assisted mass spectrometry (Table 3.1), we have identified a subset of these proteins as the most promising targets (ERO1L, RCN1, AnxA2, Calu, Erdj3, ERp44, PDIA3, and Golim4) based on their known roles in diverse processes within the ER or the secretory pathway that may relate to collagen-I folding and proteostasis (see Table 3.2 for a summary of known functions). We are currently optimizing our procedures for shRNA knockdown of each of these proteins of interest. We have already determined the time course of mRNA-level knockdown for several proteins of interest (Figure 3.4). In the future, we will first finish determining the time course of shRNA knockdown for all target proteins of interest at the both the mRNA and protein levels and then use these optimized shRNA knockdown protocols in order to investigate the mechanistic roles for the collagen-I interacting proteins in the collagen-I biosynthesis and folding pathway.

Table 3.2. Descriptions and functions for key new proteins identified by SILAC-assisted mass spectrometry.

Protein	Brief Description	Selected known functions
Annexin A2 (AnxA2)	A Ca^{2+}-dependent phospholipid binding protein	• Functions in plasma membrane docking of vesicles and vesicle fusion during collagen-VI secretion[64] • Expressed by osteoblast cells and may play a role in matrix mineralization[65] • Appears to play a role in post-transcriptionally downregulating collagen-I synthesis in human fibroblasts[66]
Calumenin (Calu)	A member of the CREC family of Ca^{2+} binding proteins containing multiple (7) EF hand motifs	• Binds Ca^{2+} at low affinity ($K_D \approx 0.6$ mM); may play a role in ER Ca^{2+} storage[67] • Interacts with the ER Ca^{2+} ATPase Pump, SERCA2[68] • Overexpression of Calu has been shown to alleviate ER stress by an unknown mechanism[69]
DnaJ homolog subfamily B member 11 (Erdj3)	A dimeric type I DnaJ cochaperone of the HSP70 BiP	• Involved in delivery of clients to BiP and also stimulates BiP ATPase activity[70] • Is a UPR-induced secreted chaperone protein that binds to misfolded proteins in the extracellular space[71]
ERO1-like protein alpha (ERO1Lα)	A flavoprotein that catalyses re-oxidation of reduced PDI	• Functions in oxidative protein folding by re-oxidizing PDI[72] • Accumulated at ER-mitochondrial contacts and plays a role in regulation of Ca^{2+} release[73] • Plays a role in ER stress-induced apoptosis[74]
Endoplasmic reticulum resident protein 44 (ERp44)	A thioredoxin domain-containing multifunctional chaperone protein	• Forms a covalently-linked complex with ERO1Lα and ERO1Lβ to maintain their ER-localization[75] • Inhibits ER Ca^{2+} release by the IP3R1 receptor[76]
Golgi integral membrane protein 4 (Golim4)	A type II early Golgi intergral membrane protein	• Normally localized to the cis-Golgi, but undergoes pH-dependent localization to endosomes. Retrieved via the bypass transport pathway that goes from the endosomes directly to the cis-Golgi; required for pathway function.[77]
Protein disulfide-isomerase A3 (PDIA3)	A member of the PDI family	• Involved in PDI functions: disulfide oxidation, reduction, and isomerization[78] • Interacts with calnexin and calreticulin and plays a role in chaperoning the folding of glycosylated proteins[79]
RCN1	A member of the CREC family of Ca^{2+} binding proteins containing multiple (6) EF hand motifs	• Binds Ca^{2+} at low affinity ($K_D \approx 0.6$ mM); may play a role in ER Ca^{2+} storage. However only a subset of the EF hand domains in RCN1 are capable of binding Ca^{2+}[80] • Overexpression of RCN1 may play a role in tumorigenesis and tumor invasion[81]

[64] Dassah, M. *et al.* Annexin A2 mediates secretion of collagen VI, pulmonary elasticity and apoptosis of bronchial epithelial cells. *J. Cell Sci.* **127,** 828–44 (2014).
[65] Genetos, D. C., Wong, A., Watari, S. & Yellowley, C. E. Hypoxia increases Annexin A2 expression in osteoblastic cells via VEGF and ERK. *Bone* **47,** 1013–1019 (2010).
[66] Schäfer, G., Hitchcock, J. K., Shaw, T. M., Katz, A. a. & Parker, M. I. A Novel Role of Annexin A2 in Human Type I Collagen Gene Expression. *J. Cell. Biochem.* **116,** 408–417 (2015).

3.4 Conclusions and Future Work

We have used the SILAC-assisted method of mass spectrometry to perform a systematic and comprehensive quantitative identification of the proteins that interact with Colα1(I). This method has allowed us to identify many novel Colα1(I) interactors, most of which are ER-localized and have diverse known functions within the ER. This result shows that mass spectrometry will be a valuable tool for further investigations of the collagen-I proteostasis network.

The most promising future experiments center on the comparison between the proteostasis networks that engage wild type collagen-I versus OI-mutant variants. These

---

[67] Yabe, D., Nakamura, T., Kanazawa, N., Tashiro, K. & Honjo, T. Calumenin, a Ca2+ -binding protein retained in the endoplasmic reticulum with a novel carboxyl-terminal sequence, HDEF. J. Biol. Chem. 272, 18232–18239 (1997).
[68] Vandecaetsbeek, I., Vangheluwe, P., Raeymaekers, L., Wuytack, F. & Vanoevelen, J. The Ca2+ Pumps of the Endoplasmic Reticulum and Golgi Apparatus. *Cold Spring Harb. Perspect. Biol.* **3,** a004184–a004184 (2011).
[69] Lee, J. H., Kwon, E. J. & Kim, D. H. Calumenin has a role in the alleviation of ER stress in neonatal rat cardiomyocytes. *Biochem. Biophys. Res. Commun.* **439,** 327–332 (2013).
[70] Guo, F. & Snapp, E. L. ERdj3 regulates BiP occupancy in living cells. *J. Cell Sci.* **126,** 1429–39 (2013).
[71] Genereux, J. C. *et al.* Unfolded protein response-induced ERdj 3 secretion links ER stress to extracellular proteostasis. **34,** 4–19 (2015).
[72] Zhang, L. et al. Different Interaction Modes for Protein-disulfide Isomerase (PDI) as an Efficient Regulator and a Specific Substrate of Endoplasmic Reticulum Oxidoreductin-1 (Ero1). J. Biol. Chem. 289, 31188–31199 (2014)
[73] Anelli, T. *et al.* Fluxes at the Endoplasmic Reticulum–Mitochondria Interface (MAM). *Antioxid. Redox Signal.* **16,** 1077–1087 (2012).
[74] Li, G. *et al.* Role of ERO1-α–mediated stimulation of inositol 1, 4, 5-triphosphate receptor activity in endoplasmic reticulum stress–induced apoptosis. *J. Cell Biol.* **186,** 783–792 (2009).
[75] Otsu, M. *et al.* Dynamic Retention of ERO1-alpha and ERO1-beta in the Endoplasmic Reticulum by Interactions with PDI and ERp44. **8,** 274–282 (2006).
[76] Vandecaetsbeek, I. *et al.* (2013).
[77] Natarajan, R. & Linstedt, A. D. A cycling cis-Golgi Protein Mediates Endosome-to-Golgi Traffic. *Mol. Biol. Cell* **15,** 4798–4806 (2004).
[78] Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* **6,** 28–32 (2005).
[79] Ellgaard, L. & Ruddock, L. W. (2005).
[80] Tachikui, H., Navet, a F. & Ozawa, M. Identification of the Ca(2+)-binding domains in reticulocalbin, an endoplasmic reticulum resident Ca(2+)-binding protein with multiple EF-hand motifs. J. Biochem. 121, 145–149 (1997).
[81] Fukuda, T. *et al.* Distribution and variable expression of secretory pathway protein reticulocalbin in normal human organs and non-neoplastic pathological conditions. *J. Histochem. Cytochem.* **55,** 335–345 (2007).

experiments are especially significant because they will likely result in the identification of new proteins that could serve as promising drug target to treat the underlying cause of OI. Another important future experiment is the determination of the wild type Colα2(I) interactome. While the wild type Colα1(I) and Colα2(I) interacomes are likely very similar, it will be interesting to determine whether the Colα2(I) is involved in any unique interactions while it is still monomeric inside the ER.

3.5 Experimental Procedures

Antibodies: Mouse monoclonal anti-HA (Santa Cruz; sc7392), anti-cMyc (Sigma Aldrich; M4439), anti-β-actin (Sigma Aldrich; A5441), anti-FLAG-M1 (Sigma Aldrich; F3040), anti-KDEL (Enzo Life Sciences; ADI-SPA-827), anti-HSP47 (Enzo Life Sciences; ADI-SPA-470), rat monoclonal anti-FLAG-M2 (Agilent Technologies; 200472), goat polyclonal anti-CRTAP (Santa Cruz; sc-99367), and rabbit polyclonal anti-COL1A1 internal (MyBioSource; MBS502153), anti-COL1A2 internal (Sigma Aldrich; SAB4500363), anti-PDI (Santa Cruz; sc-20132) antibodies were used.

Mammalian Cell Culture: HT1080 Tet- Off® cells (ClonTech; a human fibrosarcoma cell line stably expressing the tetracycline-regulated transactivator Tet-Off) stably expressing Colα1(I) and Colα2(I) were cultured at 37 °C in a 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS supplemented with penicillin, streptomycin, glutamine (Mediatech). The cell lines were continuously cultured in G418 (100 μg/mL; Enzo Life Sciences), puromycin (0.25 μg/mL; Corning) and hygromycin (150 μg/mL; Enzo Life Sciences) to maintain the presence of the transgenes, and doxycycline (1 μg/mL; EMD) was added to suppress expression of Colα1(I) and Colα2(I).

Induction of Collagen-I Expression: To induce expression of Colα1(I) and Colα2(I), the HT1080 Tet- Off® cell line stably expressing COL1A1.pTRE-Tight and COL1A2.pTRE-Tight was cultured in tetracycline-free media (DMEM supplemented with penicillin, streptomycin, glutamine, and Tet Approved FBS (ClonTech; tetracycline-free FBS) for 48 – 72 h, supplementing with fresh sodium ascorbate (50 μM; Amresco every 24 h.

Immunoprecipitations: Samples from the Colα1(I)/Colα2(I) cell line with collagen-I expression induced were harvested using 0.25% Trypsin + EDTA (Cellgro), washed twice with 1X phosphate-buffered saline (PBS). For a stable (no covalent crosslinking) IP, the samples were lysed using lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM $MgCl_2$, 1% Triton X-100, protease inhibitor tablets (Thermo Fisher), 1.5 mM phenylmethylsulfonyl fluoride (PMSF)). For a cross-linking IP, the cells were resuspended in 1 mL of 1x PBS with dithiobis(succinimidyl propionate) (0.2 mM) and were allowed to rotate at room temperature for 30 min. Tris (1 M, pH 8.0) was added to the cross-linking cellular suspension (100 μL per 1 mL suspension) to quench the cross-linking reaction. The cross-linked samples were lysed in RIPA buffer (150 mM sodium chloride, 50 mM Tris-HCl, pH 7.5, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium docecyl sulfate (SDS), protease inhibitor tablets (Thermo Fisher) and 1.5 mM PMSF). All lysed samples were centrifuged at 21,100 x g for 15 min at 4 °C. The protein content of the cell lysate supernatant samples was normalized and the lysate samples were incubated for 16 h with antibody-conjugated beads (either HA, FLAG-M2, FLAG-M1, or cMyc). The antibody-conjugated beads were washed with lysis buffer, resuspended in a Tris-SDS solution (300 mM Tris, pH 7.5, 6% SDS), and boiled for 10 min to

elute the immunoprecipitated proteins. The samples were analyzed by SDS-PAGE and Western blotting or prepared for mass spectrometry analysis.[82]

Mass Spectrometry Sample Preparation and Analysis: The stable isotope labeling of amino acids in cell culture (SILAC) method was used to quantitatively investigate the collagen-I proteostasis machinery. All SILAC reagents were purchased from Cambridge Isotope Labs. The Colα1(I)/Colα2(I) cells were labeled in either heavy, medium, or light media for 6 passages (approximately 21 – 24 days), and the resulting samples were analyzed to ensure greater than 95% incorporation of isotropically labeled amino acids into the cellular proteome.

After crosslinking immunoprecipitation (as described above), the IP elution samples were treated with dithiothreitol (DTT, 100 µM) to reduce the disulfide bonds and release the collagen-I interacting proteins from the Colα1(I) bait protein. Each sample was centrifuged over a 100 kilodalton molecular weight cut off filter (100 MWCO, Millipore) in order to reduce the amount of collagen-I present in the sample. The samples were precipitated by vortexing methanol (450 µL) with the IP elution sample (volume of 150 µL or less). Chloroform (150 µL) and then water (450 µL) were added to the mixture, vortexing between addition of each reagent and the samples were centrifuged at 10,000 x g for three min. The top aqueous phase was removed, preserving the precipitate at the solvent interface. The precipitate samples were washed three time with methanol (1 mL), centrifuging between washes. The resultant pellet was dried, and resuspended in a solution of urea (8 M) and ammonium bicarbonate (50 mM). DTT (10 mM) was added, and the samples were mixed and incubated in a 56 °C water bath for

---

[82] Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of Osteogenesis imperfecta. *PLoS Genet.* **4**, (2008).

45 min. The samples were cooled and rotated with iodoacetamide (IAM, 55 mM) for 1 h in the dark prior to incubation with sequencing-grade trypsin (1 μg, Promega) overnight at room temperature while rotating. The proteolyzed samples were acidified to a final concentration of 5% formic acid and subjected to C18 Stage Tips for desalting. Prior to sample addition, the C18 Stage Tips were washed with TFA (0.1%), TFA (0.1%) with acetonitrile (90%), and then TFA (0.1%). Samples were then loaded onto the C18 Stage Tips, ensuring all the volume passed through the column. Each column was washed with formic acid (0.1%) and eluted with formic acid with acetonitrile (0.1% and 80% respectively). The elutions were dried by speedvac, resuspended with up to 20 μL of formic acid (0.1%), and injected onto a high performance nanoflow HPLC, with MS/MS data acquired on a Thermo QExactive mass spectrometer (LC-MS/MS). Protein identification was carried out using the Mascot database (Matrix Science). Database search results were assembled using Proteome Discoverer (Thermo ScientificTM).

shRNA Lentiviral Production and Transduction: Based on the MS interactomic data, lentiviral plasmids encoding shRNA constructs targeting proteins of interest were obtained from the TRC shRNA Clone Library (Sigma Aldrich). HEK293FT cells (a human embryonic kidney cell line optimized for the production of high-titer lentivirus using the ViraPower™ Lentiviral Expression System; Life Technologies) were transfected according to the Third Generation Lentiviral production protocols to prepare viral shRNA stocks for each protein of interest. For transduction of the viral shRNA constructs, the HT1080 cells stably expressing Colα1(I) and Colα2(I) were plated at low cell density (approximately 50% confluence) and were treated with virus for 12 – 24 h. The cells were induced for collagen-I expression by removal of doxycycline

either 5 days, 7 day, or 9 days post-transduction, harvested 48 h later, and the samples were analyzed for knockdown of the genes of interest by both qPCR.

Quantitative RT-PCR: mRNA samples were isolated using the Omega mRNA Isolation Kit and cDNA was synthesized using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems Technology) according to the manufacturer's instructions. Primers for qPCR analysis were designed using Primer3 version 4.0.0 (Human mispriming library, product ranges 130 – 180 bp, Tm 57 – 62 – 65, one CG clamp). DNA sequences for each gene were obtained from the CCDS database and where more than one isoform of the gene existed, primers were chosen that could target each isoform. Quantitative PCR was done using primers specific to each gene of interest or an RPLP2 control and Sybr Green (Applied Biosystems Technology) according to the manufacturer's instructions (see Table 3.3).

Table 3.3. Primers used for Quantitative PCR.

Primer Name	Sequence
RPLP2 Fwd	5'-CCATTCAGCTCACTGATAACCTT-3'
RPLP2 Rev	5'-CGTCGCCTCCTACCTGCT-3'
ERO1L Fwd	5'-AAGAGGCCGTGTCCTTTCTG-3'
ERO1L Rev	5'-TCCACTGCTCCAAGTCGTTC-3'
FKBP10 Fwd	5'-GCATGTGTGTGGGAGAGAGG-3'
FKBP10 Rev	5'-TGCCACACAAACAGGTAGCC-3'
Golim4 Fwd	5'-AACGAGAAGCAGCCAACCTC-3'
Golim4 Rev	5'-GCAAAGCTTCCTGGTGTTCC-3'
HSP47 Fwd	5'-CAGCCTCATCATCCTCATGC-3'
HSP47 Rev	5'-TTCTGCAGGTCATGGGTCAC-3'
PDIA3 Fwd	5'-AAATCAAGCCCCACCTGATG-3'
PDIA3 Rev	5'-TATCCCAAATGGGAGCCAAC-3'
PDIA4 Fwd	5'-TCACGGACGACAACTTCGAG-3'
PDIA4 Rev	5'-GCAGTGCAATCAACCTTTGC-3'
P4HB Fwd	5'-AAATCAAGCCCCACCTGATG-3'
P4HB Rev	5'-TATCCCAAATGGGAGCCAAC-3'

Immunoblotting: Samples were boiled for 10 min with 1X gel loading buffer (50 mM Tris-HCl pH 6.8, 3% glycerol, 1% SDS, 20 μg/mL bromophenol blue, 16.6 mM DTT), analyzed by SDS-

PAGE (8% polyacrylamide), and transferred to nitrocellulose membranes (Bio-Rad Laboratories). The membranes were blocked with 5% milk in 1X Tris Buffered Saline (TBS) for 30 min and incubated with the appropriate primary antibodies for 2 h at room temperature or overnight at 4 °C. The immunoblots were washed three times with TBS + 1% Tween (TBS-T) and were incubated with 680 nm or 800 nm fluorophore-labeled secondary antibodies (LI-COR Biosciences) for 1 h at room temperature. The blots were washed twice with TBS-T, once with TBS, and were imaged with the Odyssey Infrared Imaging System (LI-COR Biosciences).

Chapter Four:

Creation and Validation of Constructs for the Independent Expression of the Collagen-I C-Propeptide Domains

4.1 Introduction

The C-propeptide domains are globular domains located at the extreme C-terminus of a collagen-I molecule that are responsible for chain selection and triple helix nucleation, thus ensuring that the collagen-I molecule assembles with its proper heterotrimeric composition of two Colα1(I) strands and one Colα2(I) strand.[83] There are many unanswered questions about the details of the role of C-propeptides in collagen-I folding, and also to their potential roles in other biological processes. One long-standing topic of interest is understanding the mechanisms behind how the C-propeptides ensure collagen-I chain selectivity.[84] During collagen-I biosynthesis, the C-propeptide domains fold and associate to initiate triple helix formation.[85] This association is presumably driven by noncovalent interactions prior to the formation of the disulfide bonds that covalently link the C-propeptide domains.[86] The initial noncovalent interactions appear to mediate the chain selectivity of collagen-I (and other collagens) via a discontinuous sequence of 15 amino acids termed the chain recognition

[83] (a) Khoshnoodi, J., Cartailler, J. P., Alvares, K., Veis, A. & Hudson, B. G. Molecular recognition in the assembly of collagens: Terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. J. Biol. Chem. 281, 38117–38121 (2006).
(b) Canty, E. G. & Kadler, K. E. Procollagen trafficking, processing and fibrillogenesis. J. Cell Sci. 118, 1341–53 (2005).
(c) Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* 44, 21–32 (2012).
[84] (a) Lees, J. F., Tasab, M. & Bulleid, N. J. Identifcation of the molecular recognition sequence which determines the type-specific assembly of procollagen. 16, 908–916 (1997).
(b) Bulleid, N. J. Solving the mystery of procollagen chain selectivity. 19, 977–979 (2012).
(c) Malone, J. P., Alvares, K. & Veis, A. Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of type I collagen: Significance of the α2(I) chain. *Biochemistry* 44, 15269–15279 (2005).
[85] Boudko S. P., *et al.* (2012).
[86] Lees, J. F., *et al.* (1997).

sequence (CRS), although it remains unclear whether other features in the C-propeptide sequence or structure also play a role in the chain selectivity.[87] In a recent paper on the crystal structure of the type III collagen (Colα1(III)) C-propeptide trimer, for example, the authors observed that one of the three constituent C-propeptide domains had a slightly different conformation and suggested that a similar variation in collagen-I conformational requirements might contribute to explaining collagen-I heterotrimer formation.[88]

In addition to understanding the mechanisms of collagen-I chain selectivity, many other questions relating to C-propeptide roles in collagen-I folding remain unanswered. One important set of questions relates to chain association mechanisms between C-propeptides.[89] Logically, there are three alternatives for the chain association mechanism: (1) a Colα1(I)-Colα2(1) dimer forms first, followed by association with the second Colα1(I); (2) a Colα1(I)-Colα1(I) dimer forms first, followed by association with Colα2(1); or (3) all three chains associate simultaneously. It is not yet known which of these alternatives is correct. Similarly, it is again not known which cellular mechanisms (if any) promote the formation of the correct, heterotrimeric C-propeptides associations and discourage formation of incorrect homotrimeric associations.

Another interesting topic relating to the collagen-I C-propeptides relates to the cysteine residues involved in the disulfide bonds that covalently link them. The Colα1(I) C-propeptide

---

[87] (a) Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* 44, 21–32 (2012).
(b) Lees, J. F., Tasab, M. & Bulleid, N. J. Identifcation of the molecular recognition sequence which determines the type-specific assembly of procollagen. 16, 908–916 (1997).
[88] Bourhis, J.-M. *et al.* Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* 19, 1031–6 (2012).
[89] Boudko, S. P., *et al.* (2012).

contains eight cysteine residues (Cys41, Cys47, Cys64, Cys73, Cys81, Cys152, Cys197, and Cys244; residue numbers given with reference to the start of the C-propeptide domain), while the Colα2(I) C-propeptide contains only seven cysteine residues (Cys61, Cys84, Cys93, Cys101, Cys170, Cys215, Cys262) and is missing the cysteine equivalent to the second cysteine in the Colα1(I) C-propeptide (Cys47).[90] Until recently, the was some potential controversy over whether the first and fourth cysteines (Cys41 and Cys73 for Colα1(I) and Cys 61 and Cys 93 for Colα2(I)) were involved in the formation of inter- or intra-strand disulfide bonding.[91] However, the crystal structure of the type III collagen C-propeptide revealed that, at least for collagen-III, the C-propeptides are linked by only one set of inter-strand disulfide bonds which occurs between the second cysteine (Cys47) on one C-propeptide and the third cysteine (Cys 64) on an adjacent C-propeptide. The crystal structure shows that the first (Cys41) and fourth (Cys73) cysteines are involved in an intra-strand disulfide bond (see Figure 4.1.A).[92)] Due to the high degree of sequence homology between fibrillar collagen C-propeptide domains (46%), it is likely that the collagen-I Colα1(I) C-propeptides also have an intra-strand linkage between Cys41–Cys73 and the Colα2(I) C-propeptide has the equivalent intra-strand linkage between Cys61–Cys84.[93] The connectivity of the inter-strand linkages in the collagen-I C-propeptide remains unclear, although one hypothesis is that Cys47 and Cys64 on the Colα1(I) C-propeptides are

---

[90] Dion, A. S. & Myers, J. C. COOH-terminal propeptides of the major human procollagens. Structural, functional and genetic comparisons. *J. Mol. Biol.* 193, 127–143 (1987).

[91] (a) Mclaughlin, S. H. & Bulleid, N. J. Molecular recognition in procollagen chain assembly. *Matrix Biol.* 16, 369–377 (1998).

(b) Koivu, J. Disulfide bonding as a determinant of the molecular composition of types I, II and III procollagen. *FEBS Lett.* 217, 216–220 (1987).

(c) Lees, J. F. & Bulleid, N. J. The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J. Biol. Chem.* 269, 24354–24360 (1994).

[92] Bourhis, J.-M. *et al.* Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* 19, 1031–6 (2012).

[93] Exposito, J. Y., Valcourt, U., Cluzel, C. & Lethias, C. The fibrillar collagen family. *Int. J. Mol. Sci.* 11, 407–426 (2010).

linked to the equivalent residues on the adjacent Colα1(I) strand (See Figure 4.1.B). Since the Colα2(I) C-propeptide has one fewer cysteine than the Colα1(I) C-propeptides, there must be at least one cysteine in the collagen-I C-propeptide trimer that is not involved in disulfide bonding. The function of this reduced cysteine, if any, is also unknown.

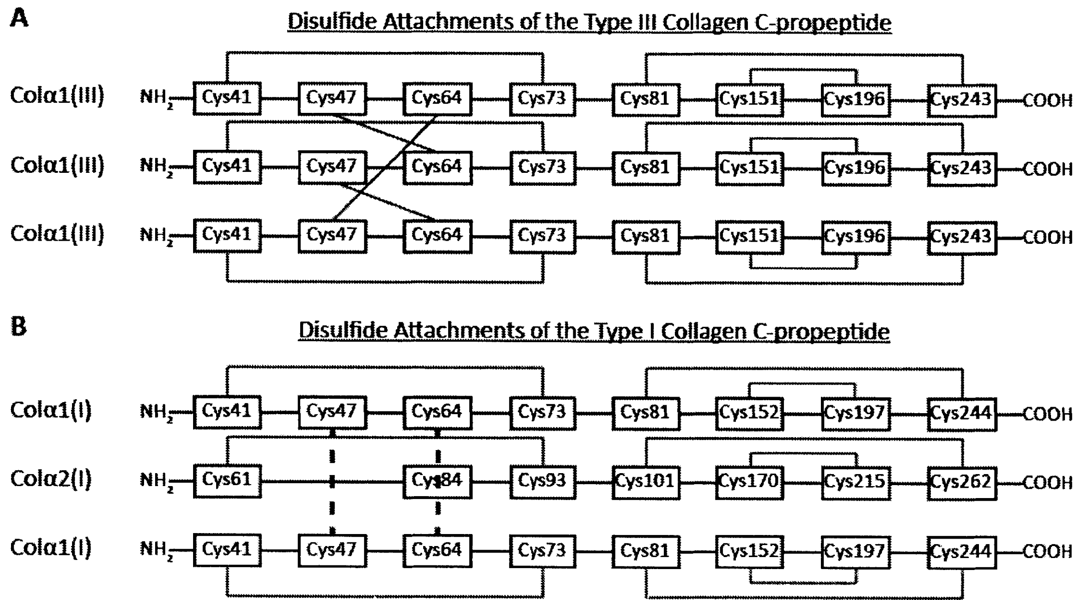


Figure 4.1. Schematic of the disulfide linkages in the Type I and Type III Collagen C-propeptides. A) The disulfide linkages in the Type III collagen C-propeptide B) The disulfide linkages in the Type I collagen C-propeptide. The intrastrand linkages between Cys81 – Cys244 and Cys152 – Cys197 (and their equivalents in the Colα2(I) C-propeptide) are well established. The Cys41 – Cys73 linkage is predicted given the high degree of homology between the C-propeptide domains (46% homology across Types I-III). The connectivity of the inter-strand linkages in the collagen-I C-propeptide remain unclear, although it is postulated that Cys47 and Cys64 on the Colα1(I) C-propeptides are linked to the corresponding residues on the adjacent Colα1(I) strand. The role of the unbonded Cys84 in the Colα2(I), if any, is also unknown.

In addition to their function in collagen-I folding and assembly, the C-propeptide domains also confer solubility to the collagen-I molecules until they are cleaved by the matrix metalloproteinase, BMP1, following collagen-I secretion, thereby initiating collagen-I fibril

formation.[94] The fate of the C-propeptide trimers after cleavage is also unknown, although a number of possible roles have been suggested in diverse processes such as osteoblast differentiation, feedback inhibition of collagen-I synthesis mediated by interaction with integrin receptors, and angiogenesis leading to tumor progression.[95] In each of these cases, the molecular mechanisms by which the collagen-I C-propeptides affect these biological processes are not well-understood.

In order to facilitate study of the folding, assembly, and possible downstream roles of the collagen-I C-propeptide domains we created a system for expression of the C-propeptide domains in the absence of the remainder of the collagen-I molecule. We designed expression vectors that allow the independent expression of the C-propeptide domains and expressed these vectors in HEK293T cells by transient transfection. We found that the recombinant C-

---

[94] (a) Prockop, D. J., Sieron, A. L. & Li, S. W. Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important: Roles in development and cell signaling. *Matrix Biol.* 16, 399–408 (1998).
(b) Muir, A. & Greenspan, D. S. Metalloproteinases in Drosophila to humans that are central players in developmental processes. *J. Biol. Chem.* 286, 41905–41911 (2011).

[95] (a) Mizuno, M., Fujisawa, R. & Kuboki, Y. The effect of carboxyl-terminal propeptide of type I collagen (c-propeptide) on collagen synthesis of preosteoblasts and osteoblasts. *Calcif. Tissue Int.* 67, 391–399 (2000).
(b) Mizuno, M., Fujisawa, R. & Kuboki, Y. Carboxyl-terminal propeptide of type I collagen (c-propeptide) modulates the action of TGF-β on MC3T3-E1 osteoblastic cells. *FEBS Lett.* 479, 123–126 (2000).
(c) Mizuno, M., Kitafima, T., Tomita, M. & Kuboki, Y. The osteoblastic MC3T3-E1 cells synthesized C-terminal propeptide of type I collagen, which promoted cell-attachment of osteoblasts. *Biochim. Biophys. Acta* 1310, 97–102 (1996).
(d) Wu, C. H., Walton, C. M. & Wu, G. Y. Propeptide-mediated regulation of procollagen synthesis in IMR-90 human lung fibroblast cell cultures: Evidence for transcriptional control. *J. Biol. Chem.* 266, 2983–2987 (1991).
(e) Bhattacharyya-Pakrasi, M., Dickeson, S. K. & Santoro, S. A. Alpha2beta1 integrin recognition of the carboxyl-terminal propeptide of type I procollagen: Integrin recognition and feed-back regulation of matrix biosynthesis are mediated by distinct sequences. *Matrix Biol.* 17, 223–232 (1998).
(f) Weston, S. a, Hulmes, D. J., Mould, a P., Watson, R. B. & Humphries, M. J. Identification of integrin alpha 2 beta 1 as cell surface receptor for the carboxyl-terminal propeptide of type I procollagen. *J. Biol. Chem.* 269, 20982–20986 (1994).
(g) Palmieri, D. *et al.* Procollagen I COOH-terminal fragment induces VEGF-A and CXCR4 expression in breast carcinoma cells. *Exp. Cell Res.* 314, 2289–2298 (2008).
(h) Vincourt, J. B. *et al.* C-propeptides of procollagens Iα1 and II that differentially accumulate in enchondromas versus chondrosarcomas regulate tumor cell survival and migration. *Cancer Res.* 70, 4739–4748 (2010).

propeptides are efficiently secreted by the HEK293T cells as disulfide-mediated trimers and are correctly post-translationally modified. This system, therefore, will be an ideal one to use for future investigation of the many unanswered questions relating to the collagen-I C-propeptides.

4.2 Results

In order to create a system that allows the study of the collagen-I C-propeptide domains, we designed and created expression vectors to allow the independent expression of both the Colα1(I) and Colα2(I) C-propeptide domains in the absence of the remainder of the collagen-I molecule. Expressing the C-propeptide domains alone eliminates the many of the challenges of working with the full-length collagen-I genes since the shorter gene can be transfected into cells more readily and the smaller protein is much more efficiently expressed. The expression vectors we designed, similar to the full-length collagen-I expression vectors (see Chapter 2), contain a pre-protrypsin signal sequence at the 5' end of the gene construct, followed by an antibody epitope tag (Figure 4.2). The pre-protrypsin signal sequence was included to ensure ER localization, while the antibody epitope tags were included to allow efficient Western blot detection of the collagen-I C-propeptide proteins. To allow selective identification of the Colα1(I) versus Colα2(I) C-propeptides, we used a HA tag for the Colα1(I) C-propeptide construct and a FLAG tag for the Colα2(I) C-propeptide construct. We created the C-propeptide expression plasmids using a pcDNA3.1(+) vector backbone that allows constitutive

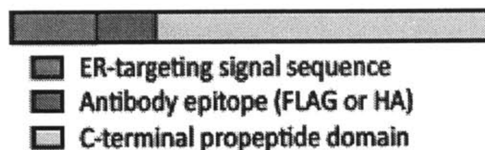


Figure 4.2. Collagen-I C-propeptide vector design.

expression. We used a constitutive expression vectors since we planned to express the C-propeptide domains by transient transfection rather than creating stable cell lines.

The collagen-I C-propeptide vectors were expressed in HEK293T cells by transient co-transfection of both the Colα1(I) and Colα2(I) C-propeptides. We tested for the intracellular presence of the Colα1(I) C-propeptide domain by Western blot analysis, treating one sample with MG132, a proteasome inhibitor, in order to assess whether the C-propeptide domains would be targeted for degradation by the ubiquitin-proteasome system when expressed without the remainder of the collagen-I molecule. This sample, along with a portion of the sample that was not treated with MG132 was also digested with the endoglycosidase, PNGase F, which cleaves N-linked glycans, in order to test whether the recombinant Colα1(I) C-propeptide protein produced by the HEK293T cells would be properly N-glycosylated. As shown in Figure 4.3, the Colα1(I) C-propeptide is present in the cell lysate. The intensity of the observed band does not change appreciably with MG132 treatment, suggesting that the intracellular expression levels of the C-propeptide constructs are not affected by proteasome inhibition. Furthermore, the bands observed in the two samples that were treated with PNGase F are present at a lower molecular weight than the untreated sample, suggesting that the Colα1(I) C-propeptide is properly N-glycosylated.

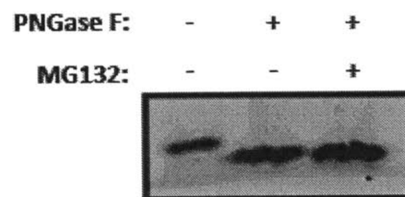


Figure 4.3. N-glycosylation of Collagen-I. Colα1(I) C-propeptide. A decrease in the molecular weight of the Colα1(I) C-propeptide is observed on treatment with PNGase F, indicating that the Colα1(I) C-propeptide is N-glycosylated.

In order to test whether the recombinant collagen-I C- propeptide domain proteins produced in this system would assemble into disulfide bond-mediated trimers and be secreted by the HEK293T cells, we treated samples of cell lysate and media with varying concentrations of DTT (untreated, 2.5 mM and 25 mM). Figure 4.4 shows the DTT-dependent drop in the molecular weight of the Colα1(I) band from approximately 100 kDa (the trimeric molecular weight) to approximately 35 kDa (the monomeric molecular weight). This pattern is observed for both the cell lysate and media samples. For Colα2(I), a similar DTT-dependent drop in molecular weight is observed for the media samples, but in the lysate samples, band detected for the Colα2(I) C-propeptide at the higher trimeric molecular weight is extremely faint and the protein also appears at the monomeric size, even when untreated with DTT.

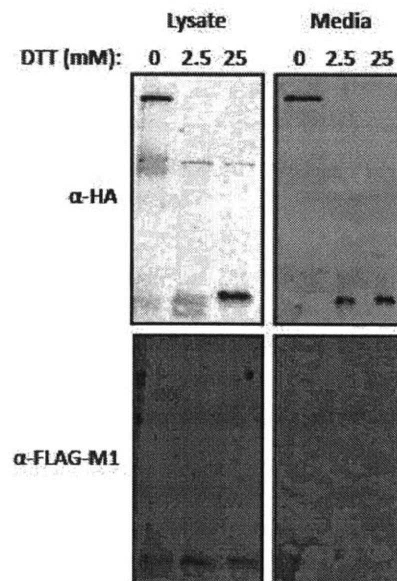


Figure 4.4. Variable DTT treatment of HEK293 cells expressing the Colα1(I) and Colα2(I) C-propeptides. Cell lysate and media samples isolated from HEK293 cells expressing the Colα1(I) and Colα2(I) C-propeptides were boiled with loading buffer containing variable concentrations of DTT. The Colα1(I) band in sample lacking DTT is observed at approximately 100 kDa (trimeric MW) while the Colα1(I) bands in the DTT-treated samples are observed at approximately 33 kDa (monomeric MW).

The presence of the Colα2(I) C-propeptide in the trimeric molecular weight in the media sample indicates that the secreted C-propeptide trimer contains both Colα1(I) and Colα2(I) C-propeptides. Altogether, this data indicates that the C-propeptide domains are indeed able to be efficiently secreted as properly modified, disulfide-mediated trimers when expressed independently of the remainder of the collagen-I molecule.

4.3 Discussion and Future Directions

In order to create a system that facilitates the study of the collagen-I C-propeptide domains and their roles in collagen-I biosynthesis and folding and as well as in other biological processes, we designed vectors that allow the expression of the C-propeptide domains without the remainder of the collagen-I molecule. We have shown that these vectors can be introduced into HEK293T cells by transient transfection, the C-propeptide domains can be detected by Western blot in the cell lysate, and they are properly N-glycosylated. We have also shown that heterotrimeric C-propeptide molecules linked by disulfide bonds are efficiently secreted. Intracellularly, the Colα1(I) C-propeptide show a similar pattern DTT-dependent drop in molecular weight as is observed in the media, but the Colα2(I) C-propeptide was only faintly detected at the higher, trimeric molecular weight, and is also present at the monomeric size, even when untreated with DTT. This could potentially be due to non-native relative levels of the Colα1(I) and Colα2(I) C-propeptide domains intracellularly, although this hypothesis has not yet been verified.

We have shown that the collagen-I C-propeptide domains can be properly post-translationally modified and secreted as disulfide-mediated trimers in the absence of the rest of

the collagen-I molecule. This system will therefore be an excellent platform for studying the properties of the collagen-I C-propeptide. In fact, the Colα1(I) C-propeptide construct has already been used by another lab member, Dr. Mahender Dewal, in his study of the effect of the unfolded protein response (UPR; the cellular pathway that regulates protein homeostasis in the ER) on the mature molecular architecture of N-glycans[96]. He showed that activation of the XBP1S branch on the UPR increases the population of the higher molecular weight Colα1(I) C-propeptide N-glycoforms relative to the population of the second N-glycoform without altering the extent of secretion of the protein. He also showed that activation of the ATF6 branch of the UPR decreased secretion of the Colα1(I) C-propeptide, while the combined activation of the both ATF6 and XBP1S had an additive effect. This investigation is one example of how the C-propeptide expression vectors we created will prove useful for investigation of the properties of the collagen-I C-propeptides. In the future, variations of these constructs should prove useful for many other investigations in to the mechanisms of collagen-I C-propeptide function.

In addition to their roles in collagen-I biosynthesis and folding as well as other biological processes, the collagen-I C-propeptide domains are also interesting due to their potential for use as tools in protein engineering and biomaterial design.[97] The ability of the collagen-I C-propeptides to assemble into disulfide-mediated heterotrimers that are efficiently secreted could potentially be harnessed in order to bring proteins together in a 2:1 ratio or to build biomaterials. Trimerization domains from other collagen types have already been used for these purposes. For example, the collagen XVIII trimerization domain has been used to develop

[96] Dewal, M. B. *et al.* XBP1s Links the Unfolded Protein Response to the Molecular Architecture of Mature N-Glycans, *in review*.

[97] Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. Int. J. Biochem. Cell Biol. 44, 21–32 (2012).

a multivalent antibody system called "trimerbodies" where three antibody fragments are linked the trimerization domain.[98] Exploring similar protein engineering adaptations of the collagen-I trimerization domain is another highly interesting direction for future research.

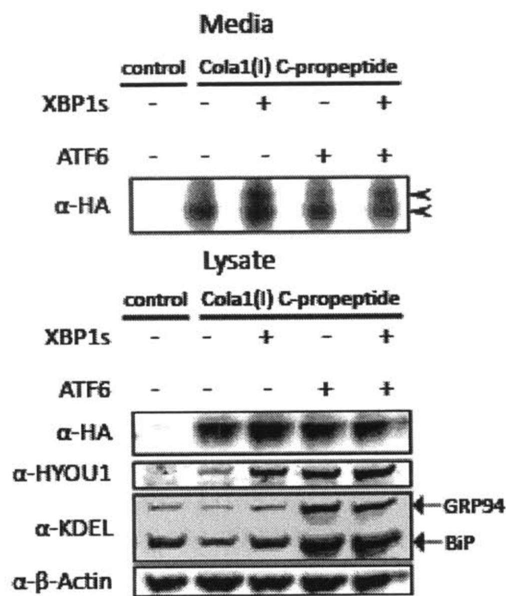


Figure 4.5. Activation of XBP1S shifts the relative populations of Colα1(I) C-propeptide N-glycoforms. Comparison of the relative abundance of the two Colα1(I) C-propeptide N-glycoforms in the media sample with and without XBP1S activation shows the selective increase in the population of the higher molecular weight N-glycoform upon activation of XBP1S. Comparison of the lysate samples show that the intracellular levels of Colα1(I) C-propeptide are unchanges with XBP1S and/or ATF6 activation, and demonstrate that this activation is occurring (HYOU1, GRP94, and BiP are markers of UPR activation).

4.4 Experimental Procedures

Plasmid DNA: The C-propeptide domain of *COL1A1* (residues 1218 – 1464) was amplified from the *COL1A1* gene obtained from the Origene True Clone Repository (accession number SC112997) and inserted between BamH1 and Xba1 restriction sites in the pcDNA3.1(+) plasmid. The C-propeptide domain of *COL1A2* (residues 1103 – 1366) was amplified from the *COL1A2*

---

[98] Sánchez-Arévalo Lobo, V. J. *et al.* Enhanced antiangiogenic therapy with antibody-collagen XVIII NC1 domain fusion proteins engineered to exploit matrix remodeling events. *Int. J. Cancer* **119,** 455–462 (2006).

gene obtained from the Origene True Clone Repository (accession number SC126717) and inserted between BamH1 and EcoRV restriction sites in the pcDNA3.1(+) plasmid. The control plasmid was either GFP-N3 or mtGFP. All plasmids were purified using the Omega BioTech Midi Prep Kit according to manufacturer's instructions.

Antibodies: Mouse monoclonal anti-HA (Santa Cruz; sc7392), anti-β-actin (Sigma Aldrich; A5441), anti-FLAG-M1 (Sigma Aldrich; F3040), and rabbit polyclonal antibodies anti-COL1A1 C-propeptide (Sigma Aldrich; HPA008405), anti-COL1A2 C-propeptide (Abcam; 96723) were used.

Mammalian Cell Culture: HEK293T cells (a human embryonic kidney cell line stably expressing the tet-repressor) were cultured at 37 °C in a 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS supplemented with penicillin, streptomycin, glutamine (Mediatech), and G418 (Enzo Life Sciences).

Transfection: HEK293T cells were transfected either with pcDNA3.1(+) vectors containing the Colα1(I) or Colα2(I) C-propeptides or a GFP control using the Xtremegene9 transfection reagent (Roche) according to manufacturer's instructions. Cells were harvested for lysis between 72 – 96 h post-transfection.

Immunoblotting: HEK29T cells transiently transfected with the Colα1(I) or Colα2(I) C-propeptides or a GFP control were harvested 72 – 96 h post-transfection using 0.25% Trypsin + EDTA (Cellgro), washed twice with 1X phosphate-buffered saline (PBS), and lysed using lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM $MgCl_2$, 1% Triton X-100, protease inhibitor tablets (Thermo Fisher), 1.5 mM phenylmethylsulfonyl fluoride). Cell lysate or media samples were boiled for 10 min with 1X gel loading buffer (50 mM Tris-HCl pH 6.8, 3% glycerol, 1% SDS, 20 μg/mL bromophenol blue, 16.6 mM DTT) and analyzed by SDS-

PAGE (8% or 12% polyacrylamide), and transferred to nitrocellulose membranes (Bio-Rad Laboratories). The membranes were blocked with 5% milk in 1X Tris Buffered Saline (TBS) for 30 min and were incubated with the appropriate primary antibodies for 2 h at room temperature or overnight at 4 °C. The immunoblots were washed three times with TBS + 1% Tween (TBS-T) and were incubated with 680 nm or 800 nm fluorophore-labeled secondary antibodies (LI-COR Biosciences) for 1 h at room temperature. The blots were washed twice with TBS-T, once with TBS, and were imaged with the Odyssey Infrared Imaging System (LI-COR Biosciences).

PNGase F Digestions: Cell lysate samples (40 μg total protein) isolated from Hek293T cells transiently transfected with either the Colα1(I) or Colα2(I) C-propeptide plasmids, both, or a GFP control (transfection described above) were digested with PNGase F (New England BioLabs) according to manufacturer's instructions. The samples were analyzed by SDS-PAGE (8% polyacrylamide) and Western blotting.

Variable Concentrations of DTT Treatments: Cell lysate (100 μg total protein) and media samples isolated from Hek293T cells transiently transfected with the Colα1(I) C-propeptide and Col1α2 C-propeptide plasmids were boiled for 10 min with 1X gel loading buffer with variable concentrations of DTT (0 mM, 0.417 mM, and 4.17 mM). The samples were analyzed by SDS-PAGE (12% polyacrylamide) and Western blotting.

Chapter Five:
Conclusions and Future Work

Type I collagen is one of the most abundant proteins in the body and it plays a critical role in providing the molecular scaffold for skin, bone, and tendon. Despite this fundamental importance, there are still many unanswered questions relating to the details of the collagen-I biosynthetic pathway and the cellular mechanisms that assist in collagen-I folding. Fully delineating these pathways is essential, not only in order to gain a complete understanding of collagen-I biosynthesis, but also because understanding how the cell deals with collagen-I biosynthesis both under normal conditions and in cases where there are misfolding mutant collagen-I molecules will allow development of novel treatment strategies for collagen-I homeostasis diseases, such as Osteogenesis Imperfecta.

The research presented in this thesis represents critical steps toward the goal of gaining a complete understanding of the collagen-I proteostasis network. We have developed a cell-based platform that allows the inducible expression of epitope-tagged heterotrimeric wild type collagen-I and we have used this system to generate a much more complete picture of the full complement of proteins that interact with wild type collagen-I using SILAC-assisted mass spectrometry. We have optimized protocols for the mRNA-level knockdown of the newly-identified collagen-I interacting proteins of interest and we are currently investigating their mechanistic roles in collagen-I proteostasis.

Several future research directions will serve to further expand our knowledge of the collagen-I proteostasis network. We have already developed OI-model cell lines that express heterotrimeric collagen-I containing a mutant version of the Colα1(I) strand. Using these and

other OI-model cell lines we plan to create, we will fully characterize the properties of the mutant collagen-I produced by these cells. We will also use SILAC-assisted mass spectrometry to quantitatively compare the proteins that interact with mutant versus wild type collagen-I.

In addition to the cell lines expressing heterotrimeric collagen-I, we have also generated cell lines that express either Colα1(I) or Colα2(I) homotrimers. This represents the first time Colα2(I) homotrimer formation has been observed. Characterization of the collagen-I produced by the cell lines that express Colα1(I) or Colα2(I) homotrimers will be another exciting direction for future work.

In order to facilitate study of the collagen-I C-propeptide domains, we have created vectors that allow the independent expression of the collagen-I C-terminal propeptide domains and have shown that the C-propeptide domains are properly modified and secreted as disulfide-mediated trimers in this system. These vectors will be used to investigate both the mechanisms of collagen-I C-propeptide function, and also the use of the collagen-I C-propeptide domains in protein engineering or biomaterial design.

Bibliography

1. Adir, N. & Ohad, I. Probing for the interaction of the 32 kDa-Qb protein with its environment by use of bifunctional cross-linking reagents. *Biochim. Biophys. Acta* 850, 264–274 (1986).
2. Agre, P., Gardner, K. & Bennett, V. Association between human erythrocyte calmodulin and the cytoplasmic surface of human erythrocyte membranes. *J. Biol. Chem.* 258, 6258–6265 (1983).
3. Alanay, Y. *et al.* Mutations in the Gene Encoding the RER Protein FKBP65 Cause Autosomal-Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86, 551–559 (2010).
4. Anelli, T. et al. ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. EMBO J. 21, 835–844 (2002).
5. Anelli, T. et al. Fluxes at the Endoplasmic Reticulum–Mitochondria Interface (MAM). Antioxid. Redox Signal. 16, 1077–1087 (2012).
6. Bächinger, H. P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262, 17144–17148 (1987).
7. Bächinger, H. P., Bruckner, P., Timpl, R. & Engel, J. The role of cis-trans isomerization of peptide bonds in the coil leads to and comes from triple helix conversion of collagen. *Eur. J. Biochem.* 90, 605–613 (1978).
8. Bailey, A. J., Sims, T. J. & Knott, L. Phenotypic expression of osteoblast collagen in osteoarthritic bone: Production of type I homotrimer. *Int. J. Biochem. Cell Biol.* 34, 176–182 (2002).
9. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* 266, 75–81 (1994).
10. Bharadwaj, A., Bydoun, M., Holloway, R. & Waisman, D. *Annexin A2 heterotetramer: Structure and function. Int. J. Mol. Sci.* 14, (2013).
11. Bhattacharyya-Pakrasi, M., Dickeson, S. K. & Santoro, S. A. Alpha2beta1 integrin recognition of the carboxyl-terminal propeptide of type I procollagen: Integrin recognition and feed-back regulation of matrix biosynthesis are mediated by distinct sequences. *Matrix Biol.* 17, 223–232 (1998).
12. Bodian, D. L. *et al.* Mutation and polymorphism spectrum in osteogenesis imperfecta type II: Implications for genotype-phenotype relationships. *Hum. Mol. Genet.* 18, 463–471 (2009).
13. Bottomley, M. J., Batten, M. R., Lumb, R. a. & Bulleid, N. J. Quality control in the endoplasmic reticulum: PDI mediates the ER retention of unassembled procollagen C-propeptides. *Curr. Biol.* 11, 1114–1118 (2001).
14. Boudko, S. P., Engel, J. & Bächinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* 44, 21–32 (2012).
15. Bourhis, J.-M. *et al.* Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* 19, 1031–6 (2012).
16. Bulleid, N. J. Solving the mystery of procollagen chain selectivity. 19, 977–979 (2012).
17. Cabibbo, a *et al.* ERO1-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum [In Process Citation]. *J. Biol. Chem.* 275, 4827–4833 (2000).

18. Canty, E. G. & Kadler, K. E. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* 118, 1341–53 (2005).
19. Chipman, S. D. *et al.* Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1701–1705 (1993).
20. Cho, J. H. *et al.* Pleiotropic roles of calumenin (calu-1), a calcium-binding ER luminal protein, in Caenorhabditis elegans. *FEBS Lett.* 583, 3050–3056 (2009).
21. Christiansen, H. E. *et al.* Homozygosity for a Missense Mutation in SERPINH1, which Encodes the Collagen Chaperone Protein HSP47, Results in Severe Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86, 389–398 (2010).
22. Chung, H. J., Jensen, D. a., Gawron, K., Steplewski, A. & Fertala, A. R992C (p.R1192C) Substitution in Collagen II Alters the Structure of Mutant Molecules and Induces the Unfolded Protein Response. *J. Mol. Biol.* 390, 306–318 (2009).
23. Dassah, M. et al. Annexin A2 mediates secretion of collagen VI, pulmonary elasticity and apoptosis of bronchial epithelial cells. J. Cell Sci. 127, 828–44 (2014).
24. Dion, A. S. & Myers, J. C. COOH-terminal propeptides of the major human procollagens. Structural, functional and genetic comparisons. *J. Mol. Biol.* 193, 127–143 (1987).
25. Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. EMBO Rep. 6, 28–32 (2005).
26. Engel, J. & Bächinger, H. P. Structure, stability and folding of the collagen triple helix. *Top. Curr. Chem.* 247, 7–33 (2005).
27. Engel, J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* 20, 137–152 (1991).
28. Exposito, J. Y., Valcourt, U., Cluzel, C. & Lethias, C. The fibrillar collagen family. *Int. J. Mol. Sci.* 11, 407–426 (2010).
29. Fertala, A. *et al.* Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). *Biochem. J.* 298 (Pt 1, 31–37 (1994).
30. Fitzgerald, J., Lamandé, S. R. & Bateman, J. F. Proteasomal degradation of unassembled mutant type I collagen pro-alpha1(I) chains. *J. Biol. Chem.* 274, 27392–27398 (1999).
31. Forlino, A., Cabral, W.A., Barnes, A.M., Marini, J. C. New Perspectives on Osteogenesis Imperfecta. *Nat. Rev. Endocrinol.* 7, 540–557 (2011).
32. Fukuda, T. et al. Distribution and variable expression of secretory pathway protein reticulocalbin in normal human organs and non-neoplastic pathological conditions. J. Histochem. Cytochem. 55, 335–345 (2007).
33. Geddis, A. E. & Prockop, D. J. Expression of Human *COL1A1* Gene in Stably Transfected HT1080 Cells: The Production of a Thermostable Homotrimer of Type I Collagen in a Recombinant System. *Matrix* 13, 399–405 (1993).
34. Genereux, J. C. et al. Unfolded protein response-induced ERdj 3 secretion links ER stress to extracellular proteostasis. 34, 4–19 (2015).
35. Genetos, D. C., Wong, A., Watari, S. & Yellowley, C. E. Hypoxia increases Annexin A2 expression in osteoblastic cells via VEGF and ERK. Bone 47, 1013–1019 (2010).
36. Gloreux, F. H. *et al.* Cyclic Administration of Pamidronate in Children with Severe. *N. Engl. J. Med.* 339, 947–952 (1998).

37. Gotkin, M. G., Ripley, C. R., Lamande, S. R., Bateman, J. F. & Bienkowski, R. S. Intracellular trafficking and degradation of unassociated proα2 chains of collagen type I. *Exp. Cell Res.* 296, 307–316 (2004).
38. Guo, F. & Snapp, E. L. ERdj3 regulates BiP occupancy in living cells. J. Cell Sci. 126, 1429–39 (2013).
39. Han, S. *et al.* Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *J. Biol. Chem.* 285, 22276–22281 (2010).
40. Harbers, K., Kuehn, M., Delius, H. & Jaenisch, R. Insertion of retrovirus into the first intron of alpha 1(I) collagen gene to embryonic lethal mutation in mice. *Proc. Natl. Acad. Sci. U. S. A.* 81, 1504–1508 (1984).
41. Hojima, Y., van der Rest, M. & Prockop, D. J. Type I Procollagen Carboxyl-terminal Proteinase from Chick Embryo Tendons. *J. Biol. Chem.* 260, 15996–16003 (1985).
42. Horna-Terrón, E., Pradilla-Dieste, A., Sánchez-de-Diego, C. & Osada, J. TXNDC5, a Newly Discovered Disulfide Isomerase with a Key Role in Cell Physiology and Pathology. *Int. J. Mol. Sci.* 15, 23501–23518 (2014).
43. Ishida, Y. et al. Autophagic Elimination of Misfolded Procollagen Aggregates in the Endoplasmic Reticulum as a Means of Cell Protection. 20, 2744–2754 (2009).
44. Ishikawa, Y. & Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833, 2479–2491 (2013).
45. Ito, H. *et al.* Guilty by association: some collagen ii mutants alter the formation of ecm as a result of atypical interaction with fibronectin. *J. Mol. Biol.* 352, 382–395 (2005).
46. Jenkins, C. L., Vasbinder, M. M., Miller, S. J. & Raines, R. T. Peptide bond isosteres: Ester or (E)-alkene in the backbone of the collagen triple helix. *Org. Lett.* 7, 2619–2622 (2005).
47. Kelley, B. P. *et al.* Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. *J. Bone Miner. Res.* 26, 666–72 (2011).
48. Khoshnoodi, J., Cartailler, J. P., Alvares, K., Veis, A. & Hudson, B. G. Molecular recognition in the assembly of collagens: Terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. J. Biol. Chem. 281, 38117–38121 (2006).
49. Koivu, J. Disulfide bonding as a determinant of the molecular composition of types I, II and III procollagen. *FEBS Lett.* 217, 216–220 (1987).
50. Lamandé, S. R. & Bateman, J. F. The Type I Collagen proalpha(I) COOH-terminal Propeptide N-Linked Oligosaccharide. *J. Biol. Chem.* 270, 17858–17865 (1995).
51. Lee, J. H., Kwon, E. J. & Kim, D. H. Calumenin has a role in the alleviation of ER stress in neonatal rat cardiomyocytes. Biochem. Biophys. Res. Commun. 439, 327–332 (2013).
52. Lees, J. F. & Bulleid, N. J. The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J. Biol. Chem.* 269, 24354–24360 (1994).
53. Lees, J. F., Tasab, M. & Bulleid, N. J. Identifcation of the molecular recognition sequence which determines the type-specific assembly of procollagen. 16, 908–916 (1997).
54. Li, G. *et al.* Role of ERO1-α–mediated stimulation of inositol 1, 4, 5-triphosphate receptor activity in endoplasmic reticulum stress–induced apoptosis. *J. Cell Biol.* 186, 783–792 (2009).
55. Lisse, T. S. *et al.* ER stress-mediated apoptosis in a new mouse model of Osteogenesis imperfecta. *PLoS Genet.* 4, (2008).

56. Malone, J. P., Alvares, K. & Veis, A. Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of type I collagen: Significance of the α2(I) chain. *Biochemistry* 44, 15269–15279 (2005).
57. Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958 (2006).
58. Marini, J. C. *et al.* Consortium for Osteogenesis Imperfecta Mutations in the Helical Domain of Type I Collagen : Regions Rich in Lethal Mutations Align With Collagen Binding Sites for Integrins and Proteoglycans. 28, 209–221 (2007).
59. Marini, J. C., Cabral, W. a. & Barnes, A. M. Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. *Cell Tissue Res.* 339, 59–70 (2010).
60. Mattson, G. et al. A practical approach to crosslinking. Mol. Biol. Rep. 17, 167–183 (1993).
61. McBride, D. J., Choe, V., Shapiro, J. R. & Brodsky, B. Altered collagen structure in mouse tail tendon lacking the alpha 2(I) chain. *J. Mol. Biol.* 270, 275–284 (1997).
62. Mclaughlin, S. H. & Bulleid, N. J. Molecular recognition in procollagen chain assembly. *Matrix Biol.* 16, 369–377 (1998).
63. Mizuno, M., Fujisawa, R. & Kuboki, Y. Carboxyl-terminal propeptide of type I collagen (c-propeptide) modulates the action of TGF-β on MC3T3-E1 osteoblastic cells. *FEBS Lett.* 479, 123–126 (2000).
64. Mizuno, M., Fujisawa, R. & Kuboki, Y. The effect of carboxyl-terminal propeptide of type I collagen (c-propeptide) on collagen synthesis of preosteoblasts and osteoblasts. *Calcif. Tissue Int.* 67, 391–399 (2000).
65. Mizuno, M., Kitafima, T., Tomita, M. & Kuboki, Y. The osteoblastic MC3T3-E1 cells synthesized C-terminal propeptide of type I collagen, which promoted cell-attachment of osteoblasts. *Biochim. Biophys. Acta* 1310, 97–102 (1996).
66. Morello, R. *et al.* CRTAP Is Required for Prolyl 3- Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* 127, 291–304 (2006).
67. Muir, A. & Greenspan, D. S. Metalloproteinases in Drosophila to humans that are central players in developmental processes. *J. Biol. Chem.* 286, 41905–41911 (2011).
68. Murad, S. *et al.* Regulation of collagen synthesis by ascorbic acid. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2879–2882 (1981).
69. Myllyharju, J. & Kivirikko, K. I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* 20, 33–43 (2004).
70. Myllyharju, J. in *Methods Mol. Biol. Extracell. Matrix Protoc.* 552, 51–62 (2009).
71. Nagai, N. et al. Embryonic lethality of molecular chaperone Hsp47 knockout mice is associated with defects in collagen biosynthesis. J. Cell Biol. 150, 1499–1505 (2000).
72. Nagata, K., Saga, S. & Yamada, K. M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J. Cell Biol.* 103, 223–229 (1986).
73. Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol.* 117, 903–914 (1992).
74. Natarajan, R. & Linstedt, A. D. A cycling cis-Golgi Protein Mediates Endosome-to-Golgi Traffic. *Mol. Biol. Cell* 15, 4798–4806 (2004).
75. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386 (2002).

76. Otsu, M. et al. Dynamic Retention of ERO1-alpha and ERO1-beta in the Endoplasmic Reticulum by Interactions with PDI and ERp44. 8, 274–282 (2006).
77. Palmieri, D. *et al.* Procollagen I COOH-terminal fragment induces VEGF-A and CXCR4 expression in breast carcinoma cells. *Exp. Cell Res.* 314, 2289–2298 (2008).
78. Pihlajaniemi, T., Myllylä, R. & Kivirikko, K. I. Prolyl 4-hydroxylase and its role in collagen synthesis. *J. Hepatol.* 13 Suppl 3, S2–S7 (1991).
79. Prockop, D. J., Sieron, A. L. & Li, S. W. Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important: Roles in development and cell signaling. *Matrix Biol.* 16, 399–408 (1998).
80. Pyott, S. M. *et al.* Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum. Mol. Genet.* 20, 1595–1609 (2011).
81. Ricard-Blum, S. The Collagen Family. *Cold Spring Harb. Perspect. Biol.* 3, 1–19 (2011).
82. Sánchez-Arévalo Lobo, V. J. et al. Enhanced antiangiogenic therapy with antibody-collagen XVIII NC1 domain fusion proteins engineered to exploit matrix remodeling events. Int. J. Cancer 119, 455–462 (2006).
83. Schäfer, G., Hitchcock, J. K., Shaw, T. M., Katz, A. a. & Parker, M. I. A Novel Role of Annexin A2 in Human Type I Collagen Gene Expression. J. Cell. Biochem. 116, 408–417 (2015).
84. Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* 78, 929–958 (2009).
85. Tachikui, H., Navet, a F. & Ozawa, M. Identification of the Ca(2+)-binding domains in reticulocalbin, an endoplasmic reticulum resident Ca(2+)-binding protein with multiple EF-hand motifs. *J. Biochem.* 121, 145–149 (1997).
86. Tuderman, L., Kivirikko, K. I. & Prockop, D. J. Partial Purification and Characterization of a Neutral Protease Which Cleaves the N-Terminal Propeptides from Procollagen. *Biochemistry* 78, 2948–2954 (1978).
87. Urlinger, S. *et al.* Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7963–7968 (2000).
88. Vandecaetsbeek, I., Vangheluwe, P., Raeymaekers, L., Wuytack, F. & Vanoevelen, J. The Ca2+ Pumps of the Endoplasmic Reticulum and Golgi Apparatus. Cold Spring Harb. Perspect. Biol. 3, a004184–a004184 (2011).
89. Van der Rest, M. & Garrone, R. Collagen family of proteins. FASEB J. 5, 2814–2823 (1991).
90. Van Duyn Graham, L., Sweetwyne, M. T., Pallero, M. a. & Murphy-Ullrich, J. E. Intracellular calreticulin regulates multiple steps in fibrillar collagen expression, trafficking, and processing into the extracellular matrix. *J. Biol. Chem.* 285, 7067–7078 (2010).
91. Veit, G. et al. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J. Biol. Chem. 281, 3494–3504 (2006).
92. Vincourt, J. B. *et al.* C-propeptides of procollagens Iα1 and II that differentially accumulate in enchondromas versus chondrosarcomas regulate tumor cell survival and migration. *Cancer Res.* 70, 4739–4748 (2010).
93. Vranka, J. a., Sakai, L. Y. & Bächinger, H. P. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J. Biol. Chem.* 279, 23615–23621 (2004).

94. Weston, S. a, Hulmes, D. J., Mould, a P., Watson, R. B. & Humphries, M. J. Identification of integrin alpha 2 beta 1 as cell surface receptor for the carboxyl-terminal propeptide of type I procollagen. *J. Biol. Chem.* 269, 20982–20986 (1994).
95. Wilson, R., Lees, J. F. & Bulleid, N. J. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *J. Biol. Chem.* 273, 9637–9643 (1998).
96. Wu, C. H., Walton, C. M. & Wu, G. Y. Propeptide-mediated regulation of procollagen synthesis in IMR-90 human lung fibroblast cell cultures: Evidence for transcriptional control. *J. Biol. Chem.* 266, 2983–2987 (1991).
97. Yabe, D., Nakamura, T., Kanazawa, N., Tashiro, K. & Honjo, T. Calumenin, a Ca2+ -binding protein retained in the endoplasmic reticulum with a novel carboxyl-terminal sequence, HDEF. *J. Biol. Chem.* 272, 18232–18239 (1997).
98. Zhang, L. et al. Different Interaction Modes for Protein-disulfide Isomerase (PDI) as an Efficient Regulator and a Specific Substrate of Endoplasmic Reticulum Oxidoreductin-1 (Ero1). J. Biol. Chem. 289, 31188–31199 (2014).