Type I Collagen Proteostasis

by

Andrew Stephen DiChiara
B.S. in Chemistry (2012)
Boston College, Chestnut Hill, MA

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

at the

Massachusetts Institute of Technology

February 2018

Signature redacted

Signature of Author: _____
Department of Chemistry
December 19, 2017

Signature redacted

Certified by: _____
Matthew D. Shoulders
Whitehead Career Development Associate Professor
Thesis supervisor

Signature redacted

Accepted by: _____
Robert W. Field
Chairman, Department Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Signature redacted

Alexander M. Klibanov
Novartis Professor, Chemistry and Bioengineering
Thesis committee chair

Signature redacted

Matthew D. Shoulders
Whitehead Career Development Associate Professor
Thesis supervisor

Signature redacted

Collin M. Stultz
Professor of Electrical Engineering and Computer Science
Institute for Medical Engineering and Science
Thesis committee member

Type I Collagen Proteostasis

by

Andrew Stephen DiChiara

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

ABSTRACT

The folding, quality control, and secretion of collagen presents a significant challenge to collagen-producing cells. Each monomeric polypeptide must undergo extensive post-translational modifications, folding and assembly of the C-terminal propeptide globular domain, and isomerization of hundreds of prolyl bonds into the trans conformation before the mature triple helix can form. The triple-helical domain lacks a traditional hydrophobic core that often drives the assembly and folding of globular proteins. Even once folded, the triple helix is at best marginally stable at body temperature and is prone to local regions of unwinding. The process must be highly orchestrated by the endoplasmic reticulum's chaperone network, including maintaining newly synthesized collagen polypeptides in an unfolded, non-aggregated, and unassembled form until after the extreme C-terminus adopts its folded structure. Despite decades of work, however, the mechanisms of collagen folding remain poorly understood, and collagen quality control is largely unexplored. Misfolding collagen variants cause disease, including osteogenesis imperfecta in the case of collagen-I variants. The origins of pathology in osteogenesis imperfecta and the other collagenopathies are still debated, but what is clear is that collagen proteostasis is strongly disrupted.

In this thesis, I report a series of studies targeted at elucidating mechanisms of proteostasis for collagen type-I, the most abundant collagen in the human body. First, I explore the folding and assembly of the critically important C-terminal, globular propeptide domains of collagen-I. My data demonstrate the importance of a single amino acid in the globular C-terminal domain that regulates the oligomerization propensity of 30 kDa C-propeptide monomers. Ultimately, I show that this single amino acid guides the assembly of the C-propeptides of all the fibrillar collagens. Second, I use a quantitative mass spectrometry-based interactomics approach and a novel human cell-based model system to map the components of the wild-type collagen-I proteostasis network. This approach resulted in the discovery of > 25 new putative proteostasis mechanisms that engage collagen-I, as well as led to the discovery of a previously unidentified post-translational modification in collagen-I. Finally, I leveraged this improved understanding of the collagen-I proteostasis network to discover a strategy to selectively increase the secretion of collagen-I from primary cells expressing a collagen-I variant that causes osteogenesis imperfecta. Resolving the disease-associated collagen-I secretion defect may prove valuable in disease.

Thesis Supervisor: Matthew D. Shoulders
Title: Whitehead Career Development Associate Professor

Acknowledgements

For me, graduate school has been a team effort, in more ways than I can count. I was lucky enough to have an unbelievably strong support system both at home and in the lab, that made the journey that much easier and more enjoyable. Without the support of my mentors, colleagues, friends, and family, graduate school would have been a much longer, drawn out task. Instead, surrounded by the support of everyone around me, I can honestly say that graduate school has been some of the best years of my life. And for that, I am eternally grateful.

I must first thank my advisor, Dr. Matthew Shoulders. I will never forget the excitement and enthusiasm for science that emanated from Matt the first time we met. We planned to meet for an hour, and he planned to outline three potential projects for me to consider if I were to join the lab. We went 15 minutes over our hour time slot, and he hadn't even finished explaining the possibilities of the first project. Of course, most know that was the collagen project, and I was hooked immediately. We met three more times to discuss the potential of joining his lab, and when I received my letter confirming that I would join the Shoulders Lab, I was elated.

Over the years, my relationship with Matt has continued to grow. He is an exceedingly talented, understanding, and guiding mentor who thrives on seeing his students succeed. No matter the size of the lab or how busy his calendar got, Matt always makes time for his students. I have always felt comfortable going directly to his office with exciting or confusing data, and even if he couldn't meet to discuss it immediately, I knew that I was on his priority list and would talk to him as soon as he had a minute. In preparing this thesis and thinking about all of the writing and experiments that I have done over the past five years, it is eminently clear now that Matt has had a huge impact on me. He has forced me to think critically, solve problems in unique ways, and encouraged me all along to do the best science that I can. Matt has made me a better scientist, a better writer, and a better person over the past five years, and for that I am very thankful.

I'd also like to thank my thesis committee, Prof. Alexander Klibanov and Collin Stultz. Thank you both for your advice and helpful conversations over the years. Alex it was a pleasure taking your class, TAing for you, and being able to stop in down the hall any time I needed to discuss something with you. Collin, your input during my oral exams was always extremely insightful and helpful, and I sincerely appreciate you taking the time to be on my committee.

Right from the start of the lab, we have always had an excellent lab culture that fosters productivity and collaboration, while also making it enjoyable to come to work everyday. I cannot thank the Shoulders Lab members enough for always being there for a laugh, a break from the work, or a sounding board to vent with. To the Shoulders Lab members, both past and present, Becca Taylor, Chris M, Emmanuel, Mahender, Pyae, Patreece, Madeline, Chet, Angela, Duc, Louis, Chris R, Azade, Anna, Kenny, Sam, Chichi, Rebecca, and Michelle – thank you for everything. Throughout the past five years I had the pleasure of mentoring some of the brightest, most talented undergraduate students that I know. Becca, Patreece, and Michelle you all pushed me to be a better mentor, and very quickly I grew to consider us colleagues rather than mentor and mentee. To the whole Shoulders lab: you all truly make coming to work more pleasant every day.

Special shout out to Chris Richardson and Tsehai Grell for proofreading some of this thesis for me. I owe you guys one!

I also must thank my friends for being some of the most supportive and understanding people I know. Brett, Danny, Nick, and Dan, thank you for everything over the years. You've always been there to take my mind off work and really emphasize and remind me that there is a lot more to life than work, school, and research.

Last but certainly not least, there literally are no words to properly thank my family. Each and every one of you has made me into the person I am today, and I am so grateful for

everything that you do for me. The amount of love and support that I feel from each and every one of you is overwhelming, but I don't know what I'd do without you all.

Mom and Dad, you two have been my rock for my whole life, and graduate school was no exception. You are always encouraging me and supporting my every action, from the time I wanted to be a painter, all the way to this moment. You've always taught me to put my best foot forward and the rest will fall into place. I thank you both for being the kind, understanding and compassionate people that you are and pushing me to be the best that I can be. Love you guys.

Sam, thank you for being my best friend for my entire life. You are such a strong and supportive sister who always keeps me on my toes! Not to mention that you always keep me laughing, and always know exactly what to say to cheer me up. You're my best friend, and I love you for all the support you've given me.

Finally, last but certainly not least, Glenn and Einstein. My boys! Glenn you have put up with more than anyone probably knows. From working weekends, to late nights, or all-nighters, you're the one that never complains; never give me a hard time. That level of understanding and compassion is hard to come by. I am so lucky to have someone so patient, caring and understanding in my life. I love you both so much, and I can't wait to finally plan the wedding!

This thesis is dedicated to my family,
Mom, Dad, Sam, Glenn and Einstein

TABLE OF CONTENTS

List of Figures

Chapter 1: Collagen structure, folding, and function in health and disease

Chapter 2: A Molecular Code for Collagen C-Propeptide Assembly

Chapter 3: Mapping and Exploring the Collagen-I Proteostasis Network

Chapter 4: The XBP1s arm of the unfolded protein response can selectively resolve collagen-I secretion defects in osteogenesis imperfecta primary cells

List of Tables

Abbreviations

$\bar{v}$	partial specific volume of the particle
A	absorbance
Ad	adenovirus
Asc	ascorbate
ATF6	cyclic AMP-dependent transcription factor 6
BiP	binding immunoglobulin protein
BSA	bovine serum albumin
C	Celsius
C-Pro	C-propeptide
C1, C2...etc	cysteine in position 1, cysteine in position 2, etc.
C1S	cysteine in position 1 mutated to serine
C2S	cysteine in position 2 mutated to serine
C3S	cysteine in position 3 mutated to serine
C4S	cysteine in position 4 mutated to serine
cDNA	complimentary DNA
cm	centimeters
co-IP	co-immunoprecipitate
CRS	collagen recognition sequence
CRTAP	cartilage-assocated protein
CyPB	cyclophilin B
Cys/Met	cysteine/methionine
d	days
DHFR	dihydrofolate reductase
DMEM	Dulbecco's modified Eagle medium
DNAJB11	DnaJ homolog subfamily B member 11
dox	doxycycline
DSP	dithobis(succinimidyl propionate)
DTT	dithothreitol
e. coli	Escherichia coli
*ec*DHFR	Escherichia coli dihydrofolate reductase
EDTA	ethylenediaminetetraacetic acid
EMEM	essential modified Eagle medium
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated degradation
Ero1L	ERO1-like protein alpha
Erp29	endoplasmic reticulum protein 29
Erp57	endoplasmic reticulum protein 57
Erp72	endoplasmic reticulum protein 72
FBS	fetal bovine serum
FKBP10	FK506 binding protein 10
g	gravity
GFP	green fluorescent protein
GM130	golgin subfamily A member 2
Golim4	golgi-integral membrane protein 4
GRP94	glucose regulated protein 94
h	hours
HEK293	human embryonic kidney cells 293
HPLC	high pressure liquid chromatography

HRV-3C	human rhinovirus 3C protease
HSP47	heat shock protein 47
HYOU1	hypoxia upregulated protein 1
IMAC	immobilized metal affinity chromatography
IP	immunoprecipitation
IRE1	inosito-requiring protein 1
kDa	kilodalton
LAMP-1	lysosome-associated membrane glycoprotein 1
LC-MS/MS	liquid chromatography tandem mass spectrometry
MALDI-TOF	matrix assisted laser desorption ionization – time of flight
mCi	millicurie
min	minute
mM	millimolar
MOI	multiplicity of infection
Mp	weighted average molecular weight
Mr	reduced molecular weight
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS1	mass spectrum 1
MS2	mass spectrum 2
MWCO	molecular weight cut off
m/z	mass to charge ratio
NaCl	sodium chloride
nm	nanometers
o/n	over night
OI	osteogenesis imperfecta
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PERK	PRKR-like endoplasmic reticulum kinase
PMSF	phenyl methyl sulfonyl fluoride
PPIase	peptidyl prolyl isomerase
Proteostasis	protein homeostasis
qPCR	quantitative polymerase chain reaction
R	ideal gas constant
r	radius
RIPA	radioimmunoprecipitation assay buffer
rpm	revolutions per minute
rt	room temperature
S2C	serine in position 2 mutated to a cysteine
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEM	standard error of the mean
shRNA	short-hairpin RNA
SILAC	stable isotope labeling of amino acids in cell culture
T	temperature
TBS	Tris buffered saline
Tet	tetracycline
TFA	trifluoroacetic acid
Tg	thapsigargin
Tm	tunicamycin

TMP	trimethoprim
UPR	unfolded protein response
Xaa	any amino acid, in the X position
XBP1s	X-box binding protein 1, spliced form
XBP1u	X-box binding protein 1, unspliced form
Yaa	any amino acid, in the Y position
mM	micromolar
ρ	solvent density
ω	frequency

Chapter 1:
Collagen structure, folding, and function in health and disease

1.1 The collagen family of proteins is an indispensable component of the extracellular matrix

The extracellular matrix (ECM) is an essential component of life that extends far beyond solely structural support of tissues and organs. Composed of over 300 proteins,[1-3] the ECM is a dynamic scaffold that simultaneously supports organ structure, transduces signals between cells, and plays an important role in regulating immune system responses and transduction of cytokines.[4] The major components of the ECM can be categorized as glycoproteins, proteoglycans, and collagens,[5-8] with some molecules spanning multiple categories.

Glycoproteins, such as the laminin family, are ubiquitous throughout the ECM, and are often intermediate molecules receiving signals from the ECM and propagating them intracellularly.[9] They also play essential roles in cell adhesion and migration,[10] ensuring proper assembly of tissues and organs.[11,12] Proteoglycans are proteins that have one or more di-sugar group called glycosaminoglycans (GAGs) attached to their structure.[13] Proteoglycans are also important molecules involved in signal transduction and immune responses.[4,14,15]

The third ECM protein category, collagens, comprises the most fundamental component of the ECM and the most abundant protein in the human body. There are 28 types of human collagen that can be further classified into one of five main categories based on their structures and functions: fibrillar, network, anchoring, fibril-associated collagen with interrupted triple helices (FACIT) and membrane-associated collagen with interrupted triple helices (MACIT).[16] Each type of collagen plays important roles in maintaining the structural integrity of the ECM. For example, collagen type-I, the most prevalent collagen, is the main proteinaceous component of bone and skin.[17]

While the details of collagen type-specific structure and function vary, the unifying feature of the collagen family is the triple-helical domain, which is the primary structural feature observed in all mature collagens deposited into the ECM.[18] Proper folding and secretion of triple-helical molecules are essential for maintaining a functioning ECM, and therefore for human health. The protein homeostasis (proteostasis) network of the endoplasmic reticulum is

responsible for producing and secreting properly folded triple helices, and selectively degrading misfolded collagen molecules.

The collagenopathies are a class of currently incurable diseases that affect almost every organ system in the human body. The collagenopathies are most commonly caused by autosomal dominant missense mutations in a collagen gene (**Table 1.1**).[19-22] In rare cases, autosomal recessive mutations to essential collagen chaperones prevent collagen from folding properly, and also result in disease.[23-27] Regardless of the genetic cause, a unifying feature of many collagenopathies is a decrease in collagen secretion, reducing the structural integrity of the ECM. Treatment for these diseases is primarily palliative, and knowing the genotypic cause of each disease is insufficient information for the design of adequate treatment, as genotype-phenotype relationships are inconsistent.[28,29] The collagenopathies manifest as a result of (1) misfolding variants causing a blockage to secretion, decreasing the total amount of collagen that reaches the ECM, and/or (2) allowing misfolding variants to be secreted and contribute to malformation of the ECM. One way to identify alternative treatment strategies from what is currently available is to probe the mechanisms of collagen assembly and quality control regulating these underlying manifestations of the collagenopathies.

Table 1.1 List of collagenopathies

Type of collagen	Collagenopathy
Collagen-I	Osteogenesis imperfecta; [20] Ehlers-Danlose syndrome[137]
Collagen-II	Achondrogenesis;[138] sponyloepiphyseal dysplasia[139]
Collagen-III	Ehlers-Danlos syndrome[140-142]
Collagen-IV	Alport syndrome[143]
Collagen-V	Ehlers-Danlos syndrome[144]
Collagen-VI	Bethlem myopathy[19,145]
Collagen-VII	Epidermolysis bullosa dystrophica[114]
Collagen-IX	Multiple epiphyseal dysplasia[146]
Collagen-X	Metaphyseal chondrodysplasia[58,119]
Collagen-XI	Stickler syndrome[147]
Collagen-XVII	Epidermolysis bullosa dystrophica
Collagen-XVIII	Knobloch syndrome[148]

1.2 Domain architecture and structure of the collagens

Although the different types of collagen have unique distributions throughout human tissue and assemble into distinctive supramolecular structures extracellularly, the general types of domains present in any given collagen can be divided into two unifying categories: non-collagenous (NC) and triple-helical domains. In the case of fibrillar collagens, both the N- and C-terminal NC domains are commonly referred to as N- and C-propeptides, respectively. The propeptide domains of fibrillar collagens are cleaved extracellularly by specific proteinases in order to free the mature, triple-helical molecule to assemble into fibrils. Other types of collagen can also undergo proteolytic cleavage of their NC domains, but references to propeptides hereafter will refer to the N- or C-terminal propeptides of the fibrillar collagens, which are always cleaved after secretion, unless specifically noted otherwise.

The C-terminal NC domain is of particular importance for every collagen molecule. Collagen folds from the C- to N-direction with the C-terminal NC domain guiding assembly, nucleation, and composition of the triple helix.[30-32] Depending on the gene composition of a given type of collagen, the C-terminal NC domain will assemble trimers into one of three possible patterns: homotrimers, 2:1 heterotrimers, or 1:1:1 heterotrimers (the last only in the case where a type of collagen includes three or more genes). The C-propeptide domains of the fibrillar collagens form interstrand disulfide bonds between monomers after or during trimeric assembly.[33] In contrast, network collagen NC domains rely primarily on hydrophobic interfaces to drive trimer assembly.[34] While the stoichiometry of different types of collagen varies, the code for assembly into a homotrimer or a heterotrimer resides within the C-terminal NC domain (see Chapter 2 for work towards understanding the requirements for fibrillar collagen assembly).

Once the C-terminal NC domain nucleates assembly of the three monomeric strands, the triple-helical domain can begin to fold. The triple helix is the defining structural feature of all types of collagen. Collagen triple helices are composed of hundreds of Gly-Xaa-Yaa repeats, with Xaa and Yaa most commonly being either (2S)-proline (Pro) or (2*S*,4*R*)-4-hydroxyproline

(Hyp), respectively. Not every Xaa or Yaa position is, however, occupied by Pro or Hyp. There is considerable sequence diversity in the triple helical region of collagen providing a diverse array of binding sites for cell surface receptors and other ECM proteins, as well as impacting the stability of local triple-helix regions.[35-39]

Although we still lack a high-resolution structure of the intact, mature collagen molecule, structural biologists have studied collagen structure for > 70 years. The first X-ray diffraction pattern of collagen was obtained from rat tail tendon in 1955.[40,41] Analysis of the data led to the first structural model of the collagen triple-helical domain, suggesting the fundamental architecture and the presence of a ladder of interstrand hydrogen bonds between backbone amides that stitch the triple helix together. In 1994, the first high resolution crystal structure of a short, triple-helical collagen-like peptide with the sequence $(Pro\text{-}Hyp\text{-}Gly)_{10}$ containing a glycine to alanine mutation was solved. The structure showed that three left-handed polyproline II helices coil together to form a parallel, right-handed triple helix largely consistent with the structure proposed in 1955 (**Figure 1.1**).[42-44] Since that time, > 20 high-resolution structures of Pro-rich triple-helical peptides have been solved.[45-50] Despite the extensive structural information we now have for these short triple helices, we still lack detailed insight into the structure of the full-length ~1000 amino acid-long triple helix of native collagens, and into the effects of sequence in Pro-poor regions on the structure of the protein.

Full-length collagen does not crystallize, instead forming liquid crystals,[51] so an x-ray structure is currently not possible to obtain. While attaining an atomic level structure of full-length collagen is problematic, tunneling and scanning electron microscopy have provided significant advancement in our understanding of how collagen molecules assemble.[52-55] Several different types of structures are detected by electron microscopy, depending in part on collagen type, as illustrated in **Figure 1.2**.

Beyond the triple-helical domain of mature collagen, the structures of the N- and C-terminal NC domains of collagen are also important to understand. In fibrillar collagens, each

monomeric N-propeptide is ~15 kDa while monomeric C-propeptides are ~30 kDa. While the function of the N-propeptide in fibrillar collagens remains unknown, the C-propeptide structure is of particular importance. The C-propeptide is the domain that initiates assembly of collagen trimers.[30] In cases where the C-Pro domain is absent due to mutations that insert early stop codons into collagen genes, collagen monomers cannot be incorporated into triple helices.[56-58] Given its indispensable role in collagen assembly, a considerable amount of effort has been applied to obtaining high-resolution structures of assembled C-propeptide trimers. Three efforts have succeeded to date: two structures of homotrimeric fibrillar collagen C-Pro domains,[33,59] and one structure of the distinctive C-Pro domain of collagen-IV[34] (**Figure 1.3A and 1.3B**).

Importantly, many of the most abundant fibrillar collagens, including collagen-I, are heterotrimers. Despite numerous attempts, the structure of collagen C-propeptide heterotrimers has still not been obtained. As a consequence, the molecular mechanisms that distinguish collagen homotrimerization from heterotrimerization remain unclear, a question that is addressed in detail in Chapter 2 of this thesis using a combined biophysical and biochemical approach.

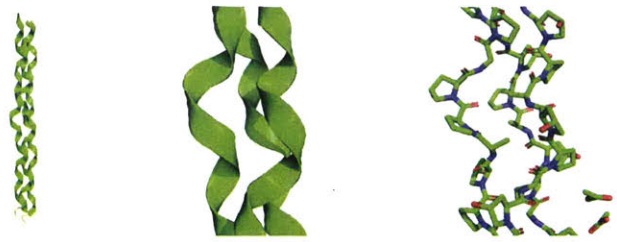


Figure 1.1 | Structural studies of the triple helical domain of collagens

Structure of a collagen-like peptide containing a glycine to alanine mutation, at 1.9 angstroms. Three views are shown, from left to right: the full triple helical peptide structure in cartoon form, a zoomed in version of the cartoon form to show the kink caused by the alanine mutation, and the stick model showing the orientation of the alanine side chain causing the kink.[44]

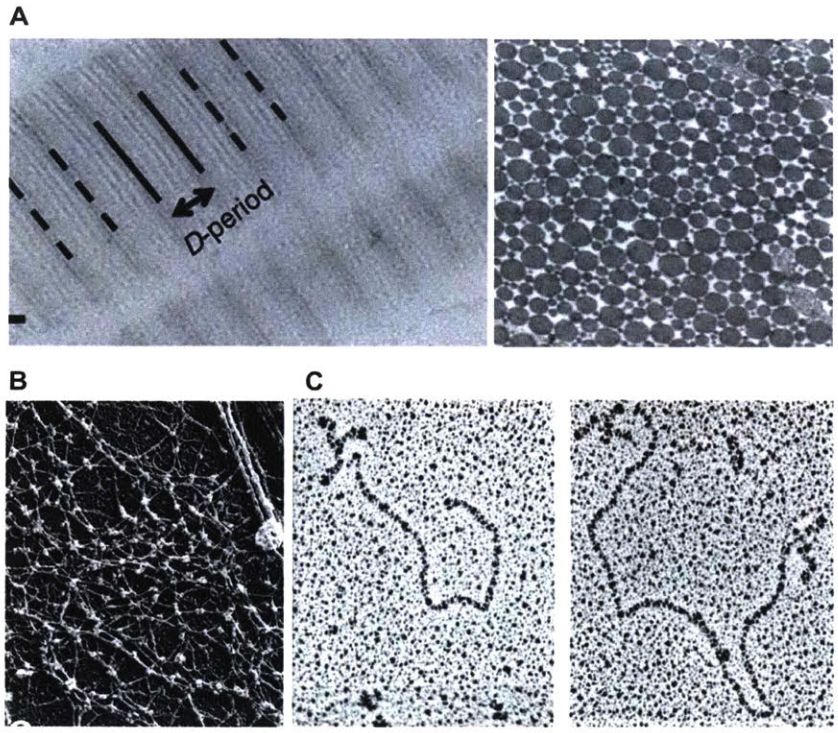


Figure 1.2 | Electron microscopy of different types of collagen structures

(A) Fibril formation showing consistent D-period spacing and cross-section of collagen-I fibrils, illustrating the average diameter of the fibrils.[134] **(B)** Collagen VI electron microscopy data demonstrating the bead on a string structure of the network.[135] **(C)** Collagen VII electron microscopy shows the free amino-terminal NC domain three individual extensions at one end of the molecule (left image) The right image shows the head to head assembly of two collagen-VII monomers at their carboxy-terminal ends. [136] **(A-C)** images demonstrate the variety of structures that the collagen family can assume in the ECM.

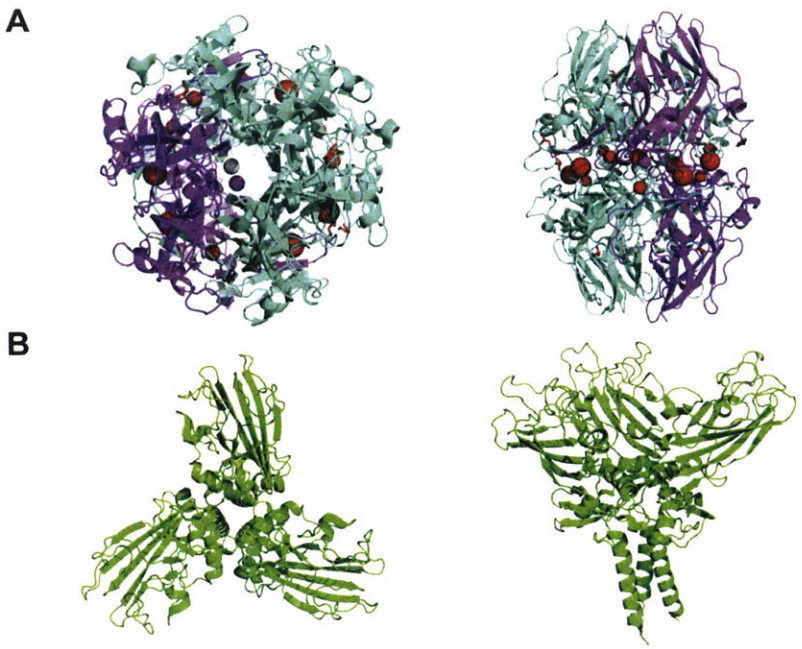


Figure 1.3 | Crystal structures of globular C-terminal NC domains

(A) Solved crystal structures the head-to-head dimer of trimers of collagen IV NC1 domain. Strands are differentially colored to identify different amino acid chains, forming a dimer of 2:1 heterotrimers. Left view is oriented down the central axis of the structure, as if the triple helical domain were in the z-plane (where the two dimensional page is the xy-plane); right view is a side view of the head to head dimer where the triple helices would be oriented from top to bottom of the page. **(B)** Structures of the Colα1(I) homotrimer C-propeptide domain. Left is a top down view of the NC domain. Right is a side view, where the helix would orient towards the bottom of the page.

1.3 Post-translational modifications

Regardless of the type of collagen in question, extracellular assembly and stability of all collagen molecules is dictated largely by events and modifications that occur intracellularly. The most abundant modification of collagen is 4*R*-hydroxylation of proline residues in the Yaa position of triple-helical domains by the enzyme prolyl-4-hydroxylase to form Hyp.[60] The 4*R*-hydroxyl group stabilizes the C^{γ}-exo pucker of proline's pyrrolidine ring. This pucker is necessary to preorganize the peptide backbone of monomeric collagen strands for triple helix formation, and is required for formation of thermally stable, folded collagen triple helices.[61,62]

Hydroxylation of lysine is another essential modification, as hydroxylated lysines are the dominant site for O-linked glycosylation, and for formation of extracellular crosslinks to other collagen molecules in assembling fibrils. Collagen galactosyltransferase enzymes are responsible for addition of a galactose sugar to hydroxylysine residues of collagen.[63] Lysyl hydroxylase 3 then adds a second sugar, glucose, resulting in a dicarbohydrate moiety on the monomeric collagen strand.[64,65] These sugar modifications are thought to play a role in regulating fibril size, mediating interactions extracellularly, and are necessary for endocytosis of damaged collagen targeted for degradation by the lysosome.[66]

Hydroxylation of each collagen strand is performed by prolyl and lysyl hydroxylases.[67] The hydroxylase enzymes are non-heme iron containing proteins, localized to the endoplasmic reticulum. The enzymes use a P450-like mechanism to hydroxylate substrates, with molecular oxygen as the source of the oxygen molecule.[67] Each enzyme has a pocket in the active site to bind ascorbate, an essential co-factor that acts as a built-in rescue mechanism to prevent the iron atom from falling into a high oxidation state that renders the protein inactive.[68] It is unclear whether the hydroxylation events occur co- or post-translationally, but the modifying enzymes only show activity toward monomeric collagen in vitro and therefore must act prior to triple-helix formation.[67,69]

Delays in triple helix formation increase the amount of time each monomer is exposed to the hydroxylase enzymes, resulting in over hydroxylation in the triple helical domain. Excessive hydroxylation alters the structural and functional characteristics of the triple helix, especially causing excessive hydroxylysine formation. An increase in hydroxylysine modification decreases N-propeptide cleavage efficiency, effectively reducing the amount of collagen available to incorporate into fibrils.[70] Therefore the kinetics of folding are essential to ensure that collagen can function properly.

Collagen molecules are also N-glycosylated, specifically in their NC domains. A single N-glycan sequon is highly conserved across all types of fibrillar collagen, and each collagen-IV gene also has a putative N-glycosylation site, although the actual presence of the modification has not been confirmed. Given that the N-glycan is located in the NC domain for fibrillar collagens, a domain that is cleaved after secretion, an extracellular function of the N-glycan seems unlikely. N-Linked glycosylation allows nascent proteins to engage the extensive lectin-based chaperone and quality control machineries in the ER.[71] Thus, a more likely role for this conserved N-glycan is enhancing collagen folding. Surprisingly, in the case of otherwise wild-type collagen-I, removal of the N-glycosylation site by substituting the N-glycosylated Asn with Gln resulted in no measurable effect on trafficking, expression, or secretion.[72] Thus, unlike other common collagen post-translational modifications, the function of the conserved NC N-glycan on collagen remains unknown.

1.4 Collagen chaperones and ER function

Folding collagen into a functional triple helix presents a formidable task to the cell. The protein is large, with each collagen monomer for type-I being > 1400 amino acids in length and the assembled protein being > 400 kDa. Moreover, unlike most or perhaps all other large proteins, collagen folds from the C- to N- direction.[30] Thus, each strand must remain in a monomeric, non-aggregated form until the extreme C-terminus of the protein enters the ER and nucleation

can begin. The C-propeptide domain must then fold into its globular structure and nucleate the three monomeric strands in the correct stoichiometry. Proper kinetics of folding are essential because delays in assembly expose the triple-helical domains to the hydroxylases for more time than is necessary, causing over-modified helices to be formed.[73] Furthermore, the properly modified triple helix is unstable at body temperature, likely to provide dynamics necessary for formation of supramolecular structures, but nonetheless meaning that collagen folding faces a challenging thermodynamics problem.[74] Finally, collagen cannot be secreted by normal methods due to its > 300 nm length, when traditional vesicles for secretion average ~70 nm in diameter.[75]

Only limited efforts have been made to map the folding and quality control networks required for collagen biogenesis owing to the following difficulties: (1) collagen is a massive, rigid protein making biochemical purification and analysis difficult; (2) collagen lacks any type of enzymatic activity or significant small-molecule binding affinity that could be used to track its subcellular localization or assess its foldedness; (3) genetically malleable systems to study collagen are not available, or are unwieldy to produce;[76,77] (4) a lack of high quality, specific antibodies precludes immunoprecipitation of collagen to identify protein-protein interactions involved in its folding pathway; and (5) the collagen genes are over 4000 base pairs in length, and contain highly repetitive GC-rich regions that are readily recombined in bacterial cells. As a result, until recently (see Chapter 3), only a handful of ER chaperones were known to play a role in collagen biogenesis. A schematic of the ER nodes that collagen engages is shown in **Figure 1.4**.

Most of the known collagen chaperones were identified due their abundance in collagen-producing cells (e.g., HSP47),[78,79] the existence of collagenopathies caused by genetic mutations in a chaperone (e.g., FKBP10),[80-82] or by logical prediction (e.g., peptidyl prolyl isomerases or protein disulfide isomerases likely to be critical for a cysteine-containing, ER protein).[83] Although seminal to our understanding of collagen folding, the three aforementioned

methods of chaperone discovery are inherently limited in their scope, and an unbiased discovery approach had not been applied to the problem prior to the work described in Chapter 3. Fewer than 4 chaperone families were proposed to have a role in collagen biogenesis prior to our work, and virtually nothing was known about collagen quality control. The previously known chaperone families are described in greater detail below.

Most striking is the existence of a collagen-specific chaperone, known as heat shock protein 47 (HSP47). HSP47 was discovered based on its sheer abundance in avian collagen-producing cells.[78] HSP47 is thought to act as a clamp at certain locations on the triple helix during folding in the ER, ensuring that newly folded regions do not locally unfold. A co-crystal structure of HSP47 and a collagen-like peptide highlights the GPRG binding motif in the collagen triple helix that is bound by HSP47.[84] While this sequence is present at nine and ten locations throughout the triple helical domains of Colα1(I) and Colα2(I), respectively (collagen-I was the model used for the collagen-like peptides), the register of the heterotrimeric collagen-I triple helix remains unknown. Therefore, the precise locations for HSP47 binding have not yet been elucidated.

HSP47 escorts collagen from the ER to the Golgi apparatus, where the lower pH protonates histidine residues in the collagen-binding domain of HSP47, causing a conformational change that releases the folded collagen molecule. HSP47 is then recycled back to the ER.[85] HSP47 knockout causes a lethal phenotype in mice. In mouse embryonic fibroblasts knocked out for HSP47, collagen-I is retained in the ER and eventually degraded (possibly by autophagy although the mechanism remains unclear), significantly decreasing the amount of collagen secreted from the cell.[86]

Peptidyl prolyl isomerases (PPIases) were predicted to play a role in collagen folding and maturation owing to the high abundance of proline residues in the collagen triple-helical domain. These enzymes assist the isomerization of all collagen's proline residues into the *trans* configuration, which is required for triple-helix formation, as in the case of the hydroxylase

enzymes. The PPIase family of enzymes act only on monomeric collagen in vitro.[87] The ER contains a considerable number of distinctive PPIases, each with different specificity for different types of collagen in vitro.[88] The molecular factors and interactions that define which collagen proteins interact with each PPIase domain are not yet clear, and possible rules for the PPIases in collagen quality control have not been elucidated.

Protein disulfide isomerases (PDI) are also implicated in collagen folding and maturation. PDIA1 is the beta subunit of the prolyl 4-hydroxylase enzyme complex responsible for the large majority of proline hydroxylation in collagen.[89,90] PDIA1 has two catalytically active sites [91] that could carry out disulfide bond isomerization or assist disulfide bond formation. While each active site has been probed for activity towards different substrates,[92] such analysis has not been performed for collagen. In fact, it is likely that PDIA1 is not the sole PDI family member that interacts with collagen, and may only be known as a collagen-interacting protein because of its function as a member of the prolyl 4-hydroxylase complex.

Finally, collagen engages the HSP70/90 system in the ER by interacting with BiP and GRP94.[93-97] BiP is responsible for many roles in the ER, including binding nascent protein chains as they enter the ER, preventing each chain from misfolding or aggregating.[98] However, the specifics of collagen binding to these chaperones, or the requirements for each in the biogenesis of collagen are not known. Furthermore, this ATP-dependent chaperoning family contains seven HSP40 proteins[99] in the ER that shuttle unfolded or nascent protein chains to BiP, to then be delivered to a variety of fates. Our recent work identifies Erdj3 as the specific HSP40 that interacts with collagen (see Chapter 3), but the molecular details of this interaction are not yet known.

In Chapter 3 I describe an unbiased approach to more comprehensively define the collagen proteostasis network. The work resulted in the identification of an additional 25–30 proteostasis factors that putatively assist collagen-I biogenesis. Mechanistic follow-up confirmed roles of a number of these new interactors, including the discovery of a new collagen post-

translational modification and the first collagen quality control pathway identified to date. Building on these discoveries and the platform reported in Chapter 2, ongoing work to define how cells engage misfolding, disease-causing collagens is providing further important insights into collagen proteostasis, and suggesting potential new targets for treatment of the collagenopathies.

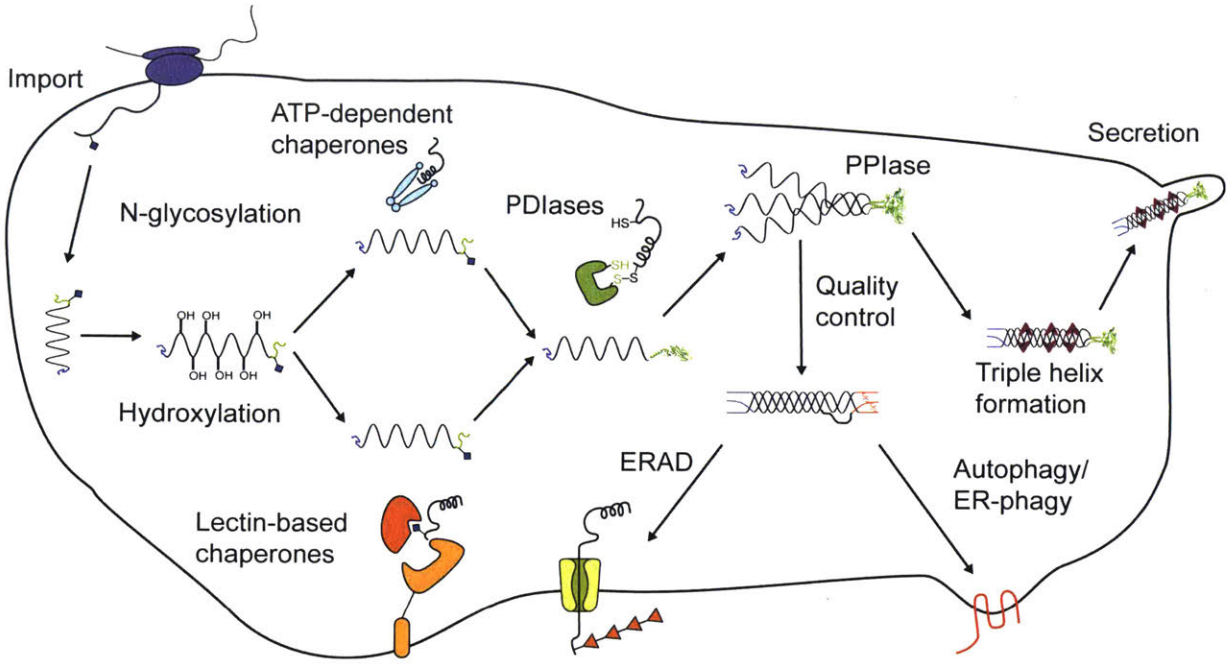


Figure 1.4 | Schematic of collagen folding and maturation in the ER

Collagen traverses the ER and engages each of the nodes highlighted here. The solved Colα1(I) crystal structure was used to demonstrate the point at which the C-terminal NC domain would be completely folded (prior to triple helix folding). The schematic is not intended to indicate the order in which collagen engages each node, but instead to indicate that collagen is known to interact with proteins within each functionally distinct group of enzymes and chaperones in the ER.

1.5 Extracellular assembly of secreted collagens

Each sub-classification of the collagen family has a different extracellular assembly pattern for establishing the ECM. For the fibrillar collagens, such as types I, II, and III, proteinases cleave the globular NC domains anytime starting in the late stages of the secretory pathway through to post-secretion,[100-102] freeing the mature molecule composed of a lengthy triple helix and short ~15 amino acid telopeptides at its N- and C-termini. Cleavage of the NC domains allows the triple helical domain to begin self-assembly into fibrils (**Figure 1.5**). The non-triple helical telopeptides are essential for proper collagen crosslinking in fibrils, contributing to the strength of the overall fibrils.[103]

In the case of most other types of collagen, the NC domains are maintained extracellularly, and play an essential role in the function of the collagen molecule.[104-107] For example, the N- and C-terminal NC domains of type-IV collagen differentially oligomerize, initiating the creation of a network of collagen molecules that form the proteinaceous foundation of basement membranes. The amino-terminal NC domain tetramerizes in a non-covalent manner, driven by hydrophobic/hydrophilic interactions creating a node of four individual collagen-IV molecules projecting in four different directions of the xy-plane, maximizing spatial coverage. The carboxy-terminal NC domain forms a head-to-head dimer with a proximal collagen-IV molecule, initiating the formation of a sulfilimine crosslink between the two NC domains (**Figure 1.5**).[108-112] In the case of collagen-VI, the NC domains are also maintained, and monomers tetramerize intracellularly in an antiparallel structure, forming the precursor for networks that resemble beads on a string.[104]

Anchoring fibrils of collagen-VII only have their carboxy-terminal NC domain proteolyzed, keeping the amino-terminal NC domain intact. Once cleaved, the carboxy-terminal ends form a head-to-head dimer with one another while the amino-terminal ends bind to cell surface receptors, forming interstitial loops throughout the ECM.[113,114] FACIT and MACIT collagen molecules are more complex, with triple-helical domains interrupted by multiple NC domains

(**Figure 1.5**).[106] These types of collagen maintain their NC domains, as they are essential for binding cell surface receptors or other ECM proteins, maintaining connections throughout the ECM.

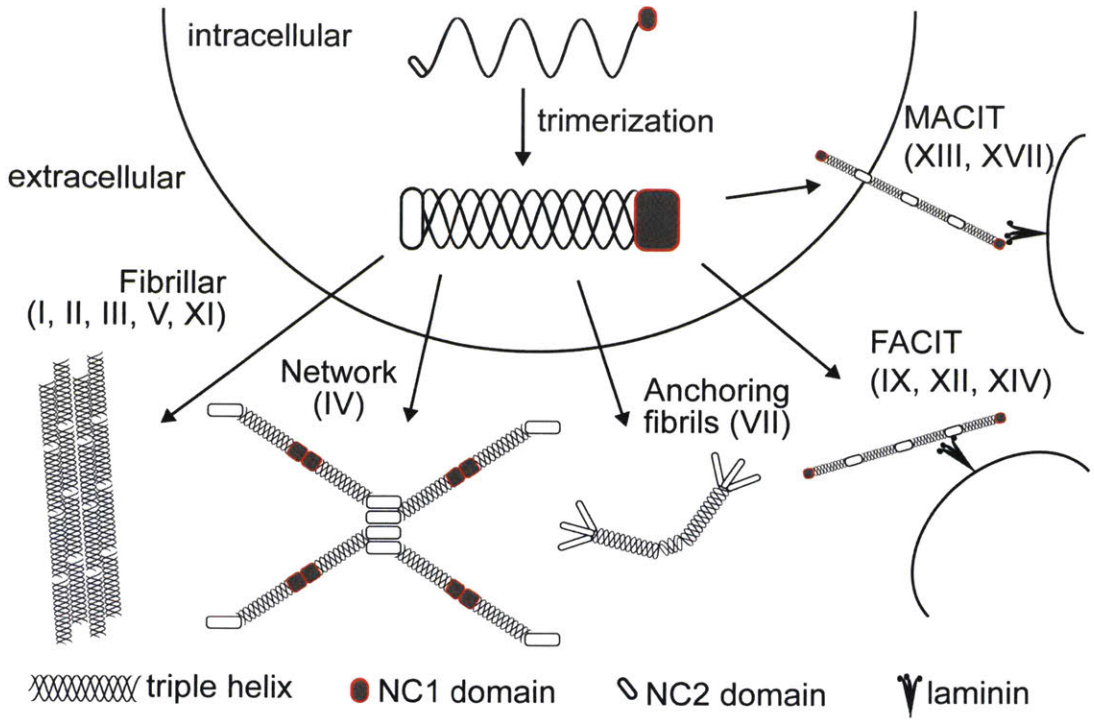


Figure 1.5 | Schematic representation of different types of collagen assembly

Collagen assembles intracellularly in a similar way from monomer to trimer for all types of collagen. At that point, each type of collagen diverges to form different extracellular structures, depicted as fibrils, networks, anchoring fibrils, or cell-cell contact molecules. Examples of the types of collagen that fall into each category are given in parenthesis (not an exhaustive list). NC1 and NC2 indicate functionally distinct NC domains.

1.6 Non-canonical secretion of large, rigid collagen molecules

Collagen triple helices are typically ~300 nm in length, and can be larger in collagen types with more complex domain architectures.[62] Although collagen secretion is COPII-dependent, standard COPII vesicles that traffic proteins from the ER to the Golgi cannot accommodate such a long, relatively rigid protein, as their average diameter is only about 70 nm.[75] Collagen must, therefore, be packaged and trafficked by a non-canonical COPII pathway, involving expansion of vesicles to accommodate collagen. A mechanism requiring monoubiquitination of COPII vesicles signals for vesicle expansion, accommodating and secreting collagen molecules. A localized increase in calcium concentration recruits the necessary E3 ubiquitin ligase, KLHL12, to the site of COPII vesicle formation. KLHL12 then monoubiquitinates a coat protein, Sec31, allowing for enlargement of the vesicle.[115,116] This expansion accommodates the large, 300–400 nm rod structure of collagen, and allows for trafficking to the Golgi for subsequent secretion (**Figure 1.6**).

While many essential details have been elucidated, the molecular-level details of signaling for this expanded Sec31-dependent expansion are not yet clear. How does folded collagen in the ER lumen transduce a signal in the cytosol to modify the vesicular machinery? What is the source of the calcium ions that induce activation of the ubiquitin ligase complex on the cytosolic side of the ER membrane? Further, how is collagen loaded into the expanded vesicles? Recent work suggests that TANGO, an integral membrane protein required for collagen secretion,[117] may use HSP47 as a handle to load triple-helical, folded collagen molecules into these expanded vesicles.[118] Although compelling, the work was performed in vitro and biological validation is required. In summary, collagen secretion is an area of ECM biology that remains incompletely understood, although important given that many of the collagenopathies involve collagen trafficking defects.

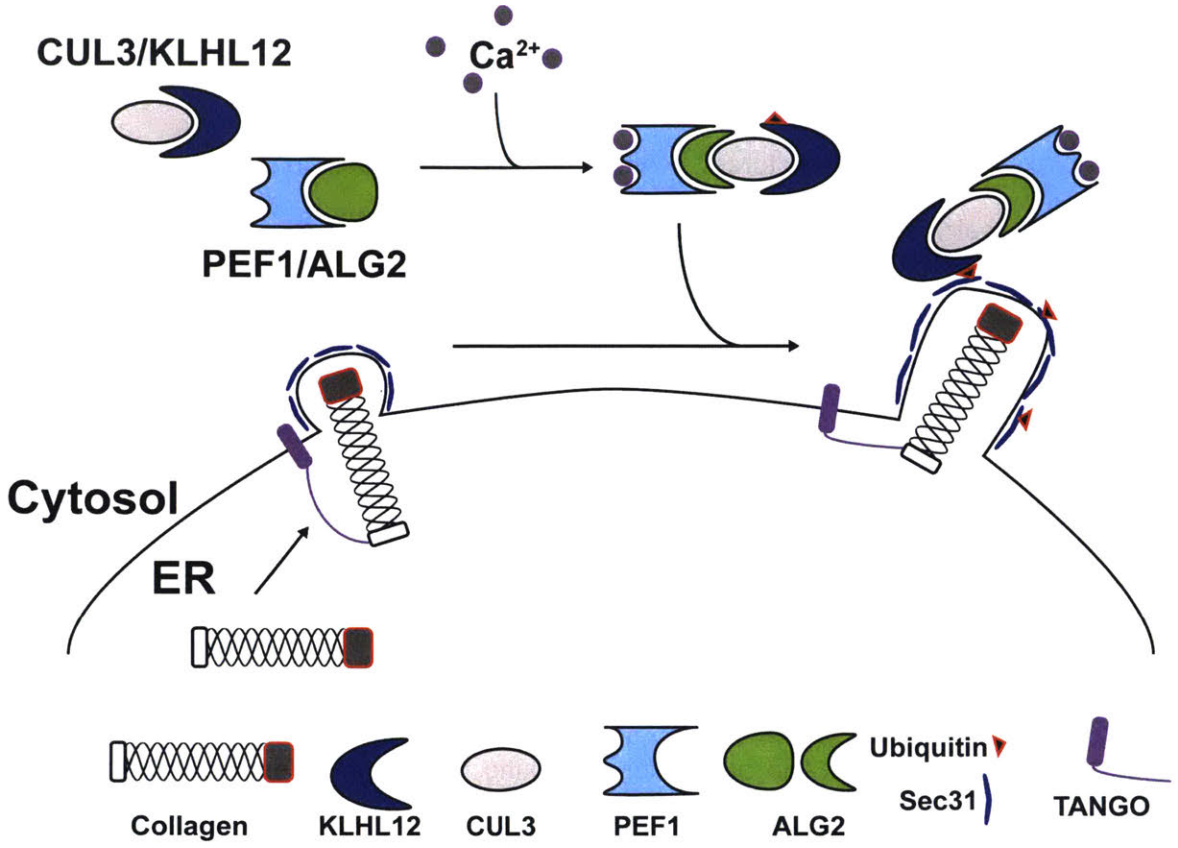


Figure 1.6 | Expanded COPII vesicles are required for collagen secretion and trafficking

Collagen molecules require expanded COPII vesicles owing to their long, rigid structures. Signaling in the cytosol initates vesicle expansion in a calcium dependent manner. Calcium triggers a conformational change in the PEF1/ALG2 complex, enabling binding to the CUL3/KLHL12 complex. PEF1/ALG2 then acts as a scaffold for the interaction between CUL3/KLHL12 and Sec31 to facilitate monoubiquitination and vesicle expansion.

1.7 Collagen quality control and disease

Due to the inherent complexity of collagen molecules, it is clear that they present a formidable folding task to cells. Inevitably, misfolding events will occur even for wild-type collagen, and misfolding is an even bigger problem in the case of disease-related misfolding collagen variants or when collagen chaperones are genetically disrupted. Failed quality control of misfolded collagen results in either toxic intracellular accumulation of collagen aggregates or secretion of non-functional triple helices.[119,120] Thus, one of the most important gaps in our understanding of collagen proteostasis is a lack of knowledge regarding collagen quality control mechanisms.

Very little is known about how misfolded collagens are recognized or how they are degraded. Autophagy and endoplasmic reticulum associated degradation (ERAD)[121] have both been proposed to be involved, but the data are inconclusive.[93,122] In order for collagen to be degraded by ERAD, the misfolded molecules must be selectively recognized by the cell, disulfides must be cleaved to facilitate retrotranslocation, residual structure must be denatured, and any trimerization must be reversed. In order for collagen to be degraded by autophagy, there must be some mechanism for selective autophagy of a misfolded ER client protein. Despite the recent provocative discovery of ER-phagy,[123,124] no such mechanism is currently known.

The prototypical collagenopathy is osteogenesis imperfecta (OI), an incurable brittle bone disease.[20] OI is most commonly caused by missense mutations in one of the collagen-I genes, causing a range of phenotypes from mild reductions in bone strength to perinatal lethality. Mutations occur throughout the entire triple helical region, as well as the NC domains of *Co1A1* and *Col1A2*, with no obvious genotype-phenotype link based on location in the gene.[125-127] In general, however, mutations closer to the 3'- end of either gene are more severe, possibly because mutations in or near the C-terminus of the protein cause more severe delays in trimerization and folding, resulting in excessive post-translational modifications in the triple

helical region. In the presence of OI-causing variants of collagen, collagen-I secretion is reduced and misfolded proteins are retained and accumulate in the ER.

In general, accumulation of a misfolded protein activates the unfolded protein response (UPR), a transcriptional response to an unfolded protein stress sensed in the ER.[128-131] The UPR has three signaling pathways, referred to as arms, that sense misfolded proteins in the ER lumen and signal for the transcriptional upregulation of ER protein chaperones to help ameliorate the folding stress. ATF6 and IRE1 are the two UPR arms that predominantly upregulate ER chaperones to buffer the folding capacity and clear the misfolded protein.[99] When a protein misfolding stress is chronic and cannot be addressed by upregulation of chaperones, PERK, the third arm, induces cellular apoptosis[132] (**Figure 1.7**).

While select collagen variants may activate the UPR or a non-canonical stress pathway,[120,128] the vast majority of OI-variants that cause collagen-I misfolding do not activate the UPR.[133] Cells do, however, retain collagen intracellularly and prevent it from being secreted. Therefore, the cells likely can recognize misfolded collagen, but may not have the capacity to respond to collagen misfolding by activating the UPR to address the defect. One possibility is that collagen's non-traditional structure escapes notice by the UPR's sensors of protein misfolding, which tend to recognize exposed hydrophobic domains that are absent from an unfolded or misfolded triple helix.

Given that most collagenopathies are autosomal dominant like OI, each collagen-producing cell has the capacity to synthesize and selectively secrete wild type collagen, while degrading disease-variant collagen. Therefore, identification of differential mechanisms between wild type and disease-variant collagen would illuminate a therapeutic window for treatment of the underlying cause of the collagenopathies. Before treatment can be considered, however, we must have a better understanding of wild type collagen proteostasis to then understand how disease-variant collagen proteostasis diverges from the wild type mechanisms.

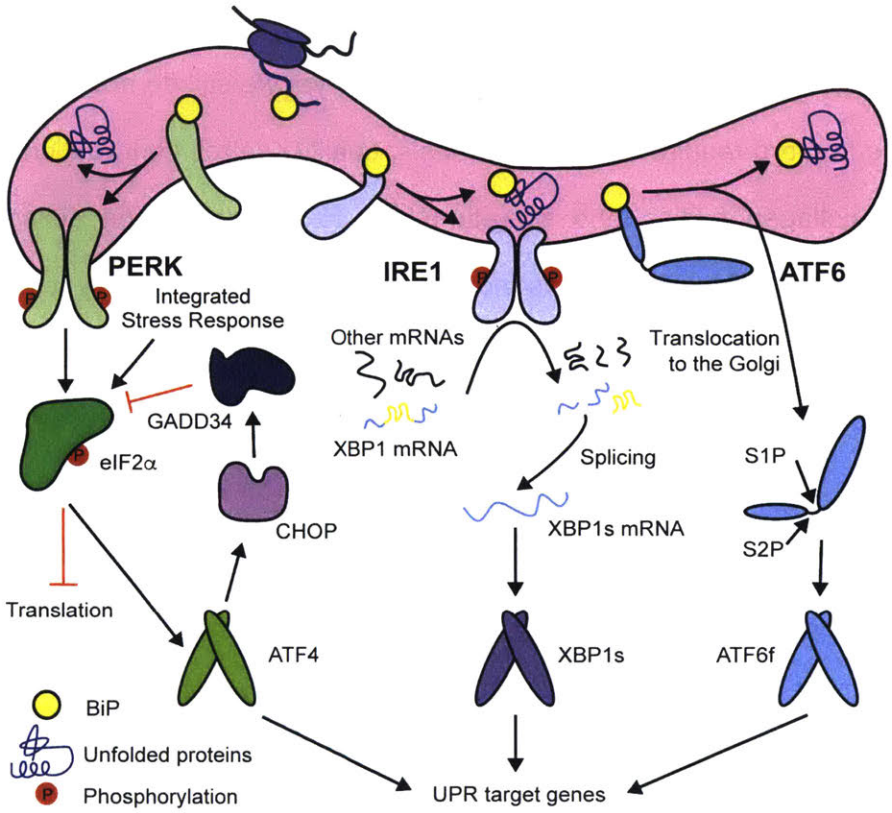


Figure 1.7 | Schematic representation of the unfolded protein response

The ER's UPR has three main arms: IRE1, ATF6, and PERK. Each arm has a subset of transcriptional targets that assist in ameliorating protein folding stress, and induce apoptosis if the stress cannot be remedied. XBP1s and ATF6 predominantly upregulate ER chaperones, while PERK induces a signal cascade to induce apoptosis in the event of chronic unfolded protein stress.

1.8 Summary

Collagen is the most abundant protein the human body, yet we still do not understand many of the mechanisms that are required for its biogenesis, quality control, degradation, and secretion. Malfunctions in collagen folding and secretion lead to the debilitating collagenopathies, a category of diseases that currently has no cure, and treatment is only palliative. Without molecular-level detail of interactions that govern folding, assembly, secretion and degradation, the design of viable therapeutic strategies for the collagenopathies will likely remain challenging.

In Chapter 2, I describe our analysis of the C-terminal NC domains of collagen-I. Collagen-I is a heterotrimeric triple helix, whose stoichiometry and register is determined by C-propeptide-initiated assembly, prior to triple helix folding. Our biochemical and biophysical data reveal a simple, cysteine-based molecular code that controls whether fibrillar collagen strands homo- versus hetero-trimerize.

In Chapter 3, I describe our efforts to better define the collagen proteostasis network. Prior to our work, only a few factors were known to play a role in collagen folding and secretion. Development of the first platform for quantitative collagen-I interactomic studies using mass spectrometry-based proteomics provided an expanded, higher resolution map of the collagen proteostasis network and led to a number of interesting discoveries. These include the identification of a previously unknown collagen post-translational modification and identification of the first collagen quality control-related ER proteostasis network component.

In Chapter 4, I outline our efforts using stress-independent UPR activation to improve collagen folding and secretion from OI patient primary cells. A subset of the interactors identified in Chapter 3 is transcriptionally regulated by the UPR. We hypothesized that activation of the UPR may both ameliorate collagen misfolding stress in OI-variant lines by enhancing folding, and improve collagen quality control. Indeed, we find that forced XBP1s expression resolves the OI-associated collagen-I secretion defect, at least for selected OI genotypes. Our data suggest

that XBP1s enhances autophagic clearance of misfolded collagen-I, while simultaneously improving the capacity of the ER to fold and secrete OI collagen-I variants.

In the appendices, I outline our work aimed at elucidating the role in regulating assembly of each cysteine residue in the collagen-I C-Pro domain, and understanding how select cysteine variants affect full-length collagen assembly.

1.9 References

(1) Naba, A., Clauser, K.R., Hoersch, S., Liu, H., Carr, S.A., Hynes, R.O. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell Proteomics* **2012**, *11*, M111.014647.

(2) Wilson, R.; Diseberg, A. F.; Gordon, L.; Zivkovic, S.; Tatarczuch, L.; Mackie, E. J.; Gorman, J. J.; Bateman, J. F. Comprehensive profiling of cartilage extracellular matrix formation and maturation using sequential extraction and label-free quantitative proteomics. *Mol Cell Proteomics* **2010**, *9*, 1296-1313.

(3) Brachvogel, B.; Zaucke, F.; Dave, K.; Norris, E. L.; Stermann, J.; Dayakli, M.; Koch, M.; Gorman, J. J.; Bateman, J. F.; Wilson, R. Comparative proteomic analysis of normal and collagen IX null mouse cartilage reveals altered extracellular matrix composition and novel components of the collagen IX interactome. *J. Biol. Chem.* **2013**, *288*, 13481-13492.

(4) Frey, H.; Schroeder, N.; Manon-Jensen, T.; Iozzo, R. V.; Schaefer, L. Biological interplay between proteoglycans and their innate immune receptors in inflammation. *FEBS J.* **2013**, *280*, 2165-2179.

(5) Mouw, J. K.; Ou, G.; Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 771-785.

(6) Jarvelainen, H.; Sainio, A.; Koulu, M.; Wight, T. N.; Penttinen, R. Extracellular matrix molecules: potential targets in pharmacotherapy. *Pharmacol. Rev.* **2009**, *61*, 198-223.

(7) Iozzo, R. V.; Schaefer, L. Proteoglycans in health and disease: novel regulatory signaling mechanisms evoked by the small leucine-rich proteoglycans. *FEBS J.* **2010**, *277*, 3864-3875.

(8) Schaefer, L.; Schaefer, R. M. Proteoglycans: from structural compounds to signaling molecules. *Cell Tissue Res.* **2010**, *339*, 237-246.

(9) Harburger, D. S.; Calderwood, D. A. Integrin signalling at a glance. *J. Cell Sci.* **2009**, *122*, 159-163.

(10) Xu, H.; Raynal, N.; Stathopoulos, S.; Myllyharju, J.; Farndale, R. W.; Leitinger, B. Collagen binding specificity of the discoidin domain receptors: binding sites on collagens II and III and molecular determinants for collagen IV recognition by DDR1. *Matrix Biol.* **2011**, *30*, 16-26.

(11) Xian, X.; Gopal, S.; Couchman, J. R. Syndecans as receptors and organizers of the extracellular matrix. *Cell Tissue Res.* **2010**, *339*, 31-46.

(12) Schmidt, S.; Friedl, P. Interstitial cell migration: integrin-dependent and alternative adhesion mechanisms. *Cell Tissue Res.* **2010**, *339*, 83-92.

(13) Iozzo, R. V.; Murdoch, A. D. Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB J.* **1996**, *10*, 598-614.

(14) Singh, B.; Fleury, C.; Jalalvand, F.; Riesbeck, K. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. *FEMS Microbiol. Rev.* **2012**, *36*, 1122-1180.

(15) Piccinini, A. M.; Midwood, K. S. DAMPening inflammation by modulating TLR signalling. *Mediators Inflamm.* **2010**, *2010*.

(16) Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a004978.

(17) Di Lullo, G. A.; Sweeney, S. M.; Korkko, J.; Ala-Kokko, L.; San Antonio, J. D. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **2002**, *277*, 4223-4231.

(18) Brodsky, B.; Ramshaw, J. A. The collagen triple-helix structure. *Matrix Biol.* **1997**, *15*, 545-554.

(19) Myllyharju, J.; Kivirikko, K. I. Collagens and collagen-related diseases. *Ann. Med.* **2001**, *33*, 7-21.

(20) Marini, J. C.; Forlino, A.; Bachinger, H. P.; Bishop, N. J.; Byers, P. H.; Paepe, A.; Fassier, F.; Fratzl-Zelman, N.; Kozloff, K. M.; Krakow, D.; Montpetit, K.; Semler, O. Osteogenesis imperfecta. *Nat. Rev. Dis. Primers* **2017**, *3*, 17052.

(21) Clements, K. A.; Acevedo-Jake, A. M.; Walker, D. R.; Hartgerink, J. D. Glycine Substitutions in Collagen Heterotrimers Alter Triple Helical Assembly. *Biomacromolecules* **2017**, *18*, 617-624.

(22) Tooley, L. D.; Zamurs, L. K.; Beecher, N.; Baker, N. L.; Peat, R. A.; Adams, N. E.; Bateman, J. F.; North, K. N.; Baldock, C.; Lamande, S. R. Collagen VI microfibril formation is abolished by an {alpha}2(VI) von Willebrand factor type A domain mutation in a patient with Ullrich congenital muscular dystrophy. *J. Biol. Chem.* **2010**, *285*, 33567-33576.

(23) Marini, J. C.; Reich, A.; Smith, S. M. Osteogenesis imperfecta due to mutations in non-collagenous genes: lessons in the biology of bone formation. *Curr. Opin. Pediatr.* **2014**, *26*, 500-507.

(24) Valli, M.; Barnes, A. M.; Gallanti, A.; Cabral, W. A.; Viglio, S.; Weis, M. A.; Makareeva, E.; Eyre, D.; Leikin, S.; Antoniazzi, F.; Marini, J. C.; Mottes, M. Deficiency of CRTAP in non-lethal recessive osteogenesis imperfecta reduces collagen deposition into matrix. *Clin. Genet.* **2012**, *82*, 453-459.

(25) Takagi, M.; Ishii, T.; Barnes, A. M.; Weis, M.; Amano, N.; Tanaka, M.; Fukuzawa, R.; Nishimura, G.; Eyre, D. R.; Marini, J. C.; Hasegawa, T. A novel mutation in LEPRE1 that eliminates only the KDEL ER- retrieval sequence causes non-lethal osteogenesis imperfecta. *PLoS One* **2012**, *7*, e36809.

(26) Pyott, S. M.; Schwarze, U.; Christiansen, H. E.; Pepin, M. G.; Leistritz, D. F.; Dineen, R.; Harris, C.; Burton, B. K.; Angle, B.; Kim, K.; Sussman, M. D.; Weis, M.; Eyre, D. R.; Russell, D. W.; McCarthy, K. J.; Steiner, R. D.; Byers, P. H. Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum. Mol. Genet.* **2011**, *20*, 1595-1609.

(27) Christiansen, H. E.; Schwarze, U.; Pyott, S. M.; AlSwaid, A.; Al Balwi, M.; Alrasheed, S.; Pepin, M. G.; Weis, M. A.; Eyre, D. R.; Byers, P. H. Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* **2010**, *86*, 389-398.

(28) Ben Amor, I. M.; Glorieux, F. H.; Rauch, F. Genotype-phenotype correlations in autosomal dominant osteogenesis imperfecta. *J. Osteoporos.* **2011**, *2011*, 540178.

(29) Bodian, D. L.; Chan, T. F.; Poon, A.; Schwarze, U.; Yang, K.; Byers, P. H.; Kwok, P. Y.; Klein, T. E. Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype-phenotype relationships. *Hum. Mol. Genet.* **2009**, *18*, 463-471.

(30) Boudko, S. P.; Engel, J.; Bachinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 21-32.

(31) Wirz, J. A.; Boudko, S. P.; Lerch, T. F.; Chapman, M. S.; Bachinger, H. P. Crystal structure of the human collagen XV trimerization domain: a potent trimerizing unit common to multiplexin collagens. *Matrix Biol.* **2011**, *30*, 9-15.

(32) Boudko, S. P.; Sasaki, T.; Engel, J.; Lerch, T. F.; Nix, J.; Chapman, M. S.; Bachinger, H. P. Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. *J. Mol. Biol.* **2009**, *392*, 787-802.

(33) Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R. N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671.

(34) Sundaramoorthy, M.; Meiyappan, M.; Todd, P.; Hudson, B. G. Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. *J. Biol. Chem.* **2002**, *277*, 31142-31153.

(35) An, B.; Abbonante, V.; Xu, H.; Gavriilidou, D.; Yoshizumi, A.; Bihan, D.; Farndale, R. W.; Kaplan, D. L.; Balduini, A.; Leitinger, B.; Brodsky, B. Recombinant Collagen Engineered to Bind to Discoidin Domain Receptor Functions as a Receptor Inhibitor. *J. Biol. Chem.* **2016**, *291*, 4343-4355.

(36) An, B.; Abbonante, V.; Yigit, S.; Balduini, A.; Kaplan, D. L.; Brodsky, B. Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. *J. Biol. Chem.* **2014**, *289*, 4941-4951.

(37) Boisvert, M.; Chetoui, N.; Gendron, S.; Aoudjit, F. Alpha2beta1 integrin is the major collagen-binding integrin expressed on human Th17 cells. *Eur. J. Immunol.* **2010**, *40*, 2710-2719.

(38) Chhum, P.; Yu, H.; An, B.; Doyon, B. R.; Lin, Y. S.; Brodsky, B. Consequences of Glycine Mutations in the Fibronectin-binding Sequence of Collagen. *J. Biol. Chem.* **2016**, *291*, 27073-27086.

(39) Yigit, S.; Yu, H.; An, B.; Hamaia, S.; Farndale, R. W.; Kaplan, D. L.; Lin, Y. S.; Brodsky, B. Mapping the Effect of Gly Mutations in Collagen on alpha2beta1 Integrin Binding. *J. Biol. Chem.* **2016**, *291*, 19196-19207.

(40) Rich, A., Crick, F.H.C The molecular structure of collagen. *J. Mol. Biol.* **1961**, *3*, 483-506.

(41) Rich, A., Crick, F.H.C The structure of collagen. *Nature* **1955**, *176*, 915-916.

(42) Bella, J.; Brodsky, B.; Berman, H. M. Disrupted collagen architecture in the crystal structure of a triple-helical peptide with a Gly-->Ala substitution. *Connect Tissue Res.* **1996**, *35*, 401-406.

(43) Bella, J.; Brodsky, B.; Berman, H. M. Hydration structure of a collagen peptide. *Structure* **1995**, *3*, 893-906.

(44) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **1994**, *266*, 75-81.

(45) Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. Collagen model peptides: Sequence dependence of triple-helix stability. *Biopolymers* **2000**, *55*, 436-450.

(46) Kramer, R. Z.; Venugopal, M. G.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J. Mol. Biol.* **2000**, *301*, 1191-1205.

(47) Buevich, A. V.; Dai, Q. H.; Liu, X.; Brodsky, B.; Baum, J. Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. *Biochemistry* **2000**, *39*, 4299-4308.

(48) Shah, N. K.; Brodsky, B.; Kirkpatrick, A.; Ramshaw, J. A. Structural consequences of D-amino acids in collagen triple-helical peptides. *Biopolymers* **1999**, *49*, 297-302.

(49) Liu, X.; Kim, S.; Dai, Q. H.; Brodsky, B.; Baum, J. Nuclear magnetic resonance shows asymmetric loss of triple helix in peptides modeling a collagen mutation in brittle bone disease. *Biochemistry* **1998**, *37*, 15528-15533.

(50) Kramer, R. Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M. X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **1998**, *280*, 623-638.

(51) Giraud Guille, M. M.; Mosser, G.; Helary, C.; Eglin, D. Bone matrix like assemblies of collagen: from liquid crystals to gels and biomimetic materials. *Micron* **2005**, *36*, 602-608.

(52) Ward, N. P.; Hulmes, D. J.; Chapman, J. A. Collagen self-assembly in vitro: electron microscopy of initial aggregates formed during the lag phase. *J. Mol. Biol.* **1986**, *190*, 107-112.

(53) Brodsky, B.; Eikenberry, E. Supramolecular collagen assemblies. *Ann. N. Y. Acad. Sci.* **1985**, *460*, 73-84.

(54) Hulmes, D. J.; Holmes, D. F.; Cummings, C. Crystalline regions in collagen fibrils. *J. Mol. Biol.* **1985**, *184*, 473-477.

(55) Hulmes, D. J.; Jesior, J. C.; Miller, A.; Berthet-Colominas, C.; Wolff, C. Electron microscopy shows periodic structure in collagen fibril cross sections. *Proc. Natl. Acad. Sci. U S A* **1981**, *78*, 3567-3571.

(56) Fang, Y.; Bateman, J. F.; Mercer, J. F.; Lamande, S. R. Nonsense-mediated mRNA decay of collagen -emerging complexity in RNA surveillance mechanisms. *J. Cell. Sci.* **2013**, *126*, 2551-2560.

(57) Lamande, S. R.; Bateman, J. F.; Hutchison, W.; McKinlay Gardner, R. J.; Bower, S. P.; Byrne, E.; Dahl, H. H. Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum. Mol. Genet.* **1998**, *7*, 981-989.

(58) Chan, D.; Weng, Y. M.; Graham, H. K.; Sillence, D. O.; Bateman, J. F. A nonsense mutation in the carboxyl-terminal domain of type X collagen causes haploinsufficiency in schmid metaphyseal chondrodysplasia. *J. Clin. Invest.* **1998**, *101*, 1490-1499.

(59) Bourhis, J. M.; Mariano, N.; Zhao, Y.; Harlos, K.; Exposito, J. Y.; Jones, E. Y.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1031-1036.

(60) Gorres, K. L.; Raines, R. T. Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45*, 106-124.

(61) Shoulders, M. D.; Kamer, K. J.; Raines, R. T. Origin of the stability conferred upon collagen by fluorination. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3859-3862.

(62) Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

(63) Schegg, B.; Hulsmeier, A. J.; Rutschmann, C.; Maag, C.; Hennet, T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. *Mol. Cell Biol.* **2009**, *29*, 943-952.

(64) Sricholpech, M.; Perdivara, I.; Nagaoka, H.; Yokoyama, M.; Tomer, K. B.; Yamauchi, M. Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. *J. Biol. Chem.* **2011**, *286*, 8846-8856.

(65) Sricholpech, M.; Perdivara, I.; Yokoyama, M.; Nagaoka, H.; Terajima, M.; Tomer, K. B.; Yamauchi, M. Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. *J. Biol. Chem.* **2012**, *287*, 22998-23009.

(66) Jurgensen, H. J.; Madsen, D. H.; Ingvarsen, S.; Melander, M. C.; Gardsvoll, H.; Patthy, L.; Engelholm, L. H.; Behrendt, N. A novel functional role of collagen glycosylation: interaction with the endocytic collagen receptor uparap/ENDO180. *J. Biol. Chem.* **2011**, *286*, 32736-32748.

(67) Myllyharju, J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **2003**, *22*, 15-24.

(68) Zito, E.; Hansen, H. G.; Yeo, G. S.; Fujii, J.; Ron, D. Endoplasmic reticulum thiol oxidase deficiency leads to ascorbic acid depletion and noncanonical scurvy in mice. *Mol. Cell* **2012**, *48*, 39-51.

(69) Pekkala, M.; Hieta, R.; Kursula, P.; Kivirikko, K. I.; Wierenga, R. K.; Myllyharju, J. Crystallization of the proline-rich-peptide binding domain of human type I collagen prolyl 4-hydroxylase. *Acta. Crystallogr. D. Biol. Crystallogr.* **2003**, *59*, 940-942.

(70) Torre-Blanco, A.; Adachi, E.; Hojima, Y.; Wootton, J. A.; Minor, R. R.; Prockop, D. J. Temperature-induced post-translational over-modification of type I procollagen. Effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self-assembly of collagen into fibrils. *J. Biol. Chem.* **1992**, *267*, 2650-2655.

(71) Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **2004**, *73*, 1019-1049.

(72) Lamande, S. R.; Bateman, J. F. The type I collagen pro alpha 1(I) COOH-terminal propeptide N-linked oligosaccharide. Functional analysis by site-directed mutagenesis. *J. Biol. Chem.* **1995**, *270*, 17858-17865.

(73) Pace, J. M., Kuslich, C.D., Willing, M.C. Byers, P.H. Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the pro-alpha1(I) chain of type I collagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta. *J. Med. Genet.* **2001**, *38*, 443-449.

(74) Leikina, E.; Mertts, M. V.; Kuznetsova, N.; Leikin, S. Type I collagen is thermally unstable at body temperature. *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 1314-1318.

(75) Stagg, S. M.; LaPointe, P.; Razvi, A.; Gurkan, C.; Potter, C. S.; Carragher, B.; Balch, W. E. Structural basis for cargo regulation of COPII coat assembly. *Cell* **2008**, *134*, 474-484.

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

(77) An, B.; Kaplan, D. L.; Brodsky, B. Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Fron. Chem.* **2014**, *2*, 40.

(78) 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.* **1986**, *103*, 223-229.

(79) Nagata, K. HSP47 as a collagen-specific molecular chaperone: function and expression in normal mouse development. *Semin. Cell. Dev. Biol.* **2003**, *14*, 275-282.

(80) Lietman, C. D.; Rajagopal, A.; Homan, E. P.; Munivez, E.; Jiang, M. M.; Bertin, T. K.; Chen, Y.; Hicks, J.; Weis, M.; Eyre, D.; Lee, B.; Krakow, D. Connective tissue alterations in Fkbp10-/- mice. *Hum. Mol. Genet.* **2014**, *23*, 4822-4831.

(81) Schwarze, U.; Cundy, T.; Pyott, S. M.; Christiansen, H. E.; Hegde, M. R.; Bank, R. A.; Pals, G.; Ankala, A.; Conneely, K.; Seaver, L.; Yandow, S. M.; Raney, E.; Babovic-Vuksanovic, D.; Stoler, J.; Ben-Neriah, Z.; Segel, R.; Lieberman, S.; Siderius, L.; Al-Aqeel, A.; Hannibal, M.; Hudgins, L.; McPherson, E.; Clemens, M.; Sussman, M. D.; Steiner, R. D.; Mahan, J.; Smith, R.; Anyane-Yeboa, K.; Wynn, J.; Chong, K.; Uster, T.; Aftimos, S.; Sutton, V. R.; Davis, E. C.; Kim, L. S.; Weis, M. A.; Eyre, D.; Byers, P. H. Mutations in FKBP10, which result in Bruck syndrome and recessive forms of osteogenesis imperfecta, inhibit the hydroxylation of telopeptide lysines in bone collagen. *Hum. Mol. Genet.* **2013**, *22*, 1-17.

(82) Barnes, A. M.; Cabral, W. A.; Weis, M.; Makareeva, E.; Mertz, E. L.; Leikin, S.; Eyre, D.; Trujillo, C.; Marini, J. C. Absence of FKBP10 in recessive type XI osteogenesis imperfecta leads to diminished collagen cross-linking and reduced collagen deposition in extracellular matrix. *Hum. Mutat.* **2012**, *33*, 1589-1598.

(83) Ishikawa, Y.; Vranka, J.; Wirz, J.; Nagata, K.; Bachinger, H. P. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 is a molecular chaperone that interacts with collagens. *J. Biol. Chem.* **2008**, *283*, 31584-31590.

(84) Widmer, C.; Gebauer, J. M.; Brunstein, E.; Rosenbaum, S.; Zaucke, F.; Drogemuller, C.; Leeb, T.; Baumann, U. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U S A* **2012**, *109*, 13243-13247.

(85) Oecal, S.; Socher, E.; Uthoff, M.; Ernst, C.; Zaucke, F.; Sticht, H.; Baumann, U.; Gebauer, J. M. The pH-dependent Client Release from the Collagen-specific Chaperone HSP47 Is Triggered by a Tandem Histidine Pair. *J. Biol. Chem.* **2016**, *291*, 12612-12626.

(86) Ishida, Y.; Yamamoto, A.; Kitamura, A.; Lamande, S. R.; Yoshimori, T.; Bateman, J. F.; Kubota, H.; Nagata, K. Autophagic elimination of misfolded procollagen aggregates in the endoplasmic reticulum as a means of cell protection. *Mol. Biol. Cell* **2009**, *20*, 2744-2754.

(87) Ishikawa, Y.; Wirz, J.; Vranka, J. A.; Nagata, K.; Bachinger, H. P. Biochemical characterization of the prolyl 3-hydroxylase 1.cartilage-associated protein.cyclophilin B complex. *J. Biol. Chem.* **2009**, *284*, 17641-17647.

(88) Ishikawa, Y.; Bachinger, H. P. A substrate preference for the rough endoplasmic reticulum resident protein FKBP22 during collagen biosynthesis. *J. Biol. Chem.* **2014**, *289*, 18189-18201.

(89) Kivirikko, K. I.; Pihlajaniemi, T. Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, *72*, 325-398.

(90) Kivirikko, K. I.; Myllyharju, J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol.* **1998**, *16*, 357-368.

(91) Zhang, L.; Niu, Y.; Zhu, L.; Fang, J.; Wang, X.; Wang, L.; Wang, C. C. Different interaction modes for protein-disulfide isomerase (PDI) as an efficient regulator and a specific substrate of endoplasmic reticulum oxidoreductin-1alpha (Ero1alpha). *J. Biol. Chem.* **2014**, *289*, 31188-31199.

(92) Ellgaard, L.; Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.* **2005**, *6*, 28-32.

(93) Lamande, S. R.; Chessler, S. D.; Golub, S. B.; Byers, P. H.; Chan, D.; Cole, W. G.; Sillence, D. O.; Bateman, J. F. Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro alpha 1 (I) chain carboxyl-terminal propeptide which impair subunit assembly. *J. Biol. Chem.* **1995**, *270*, 8642-8649.

(94) Chessler, S. D.; Byers, P. H. BiP binds type I procollagen pro alpha chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. *J. Biol. Chem.* **1993**, *268*, 18226-18233.

(95) Ferreira, L. R.; Norris, K.; Smith, T.; Hebert, C.; Sauk, J. J. Association of Hsp47, Grp78, and Grp94 with procollagen supports the successive or coupled action of molecular chaperones. *J. Cell Biochem.* **1994**, *56*, 518-526.

(96) Ferreira, L. R.; Norris, K.; Smith, T.; Hebert, C.; Sauk, J. J. Hsp47 and other ER-resident molecular chaperones form heterocomplexes with each other and with collagen type IV chains. *Connect. Tissue Res.* **1996**, *33*, 265-273.

(97) Malone, J. P.; Alvares, K.; Veis, A. Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of type I collagen: significance of the alpha2(I) chain. *Biochemistry* **2005**, *44*, 15269-15279.

(98) Gething, M. J. Role and regulation of the ER chaperone BiP. *Semin. Cell Dev. Biol.* **1999**, *10*, 465-472.

(99) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R., 3rd; Su, A. I.; Kelly, J. W.; Wiseman, R. L. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013**, *3*, 1279-1292.

(100) Canty-Laird, E. G.; Lu, Y.; Kadler, K. E. Stepwise proteolytic activation of type I procollagen to collagen within the secretory pathway of tendon fibroblasts in situ. *Biochem. J.* **2012**, *441*, 707-717.

(101) Humphries, S. M.; Lu, Y.; Canty, E. G.; Kadler, K. E. Active negative control of collagen fibrillogenesis in vivo. Intracellular cleavage of the type I procollagen propeptides in tendon fibroblasts without intracellular fibrils. *J. Biol. Chem.* **2008**, *283*, 12129-12135.

(102) Nishimura, M.; Nishie, W.; Shirafuji, Y.; Shinkuma, S.; Natsuga, K.; Nakamura, H.; Sawamura, D.; Iwatsuki, K.; Shimizu, H. Extracellular cleavage of collagen XVII is essential for correct cutaneous basement membrane formation. *Hum. Mol. Genet.* **2016**, *25*, 328-339.

(103) Terajima, M.; Taga, Y.; Chen, Y.; Cabral, W. A.; Hou-Fu, G.; Srisawasdi, S.; Nagasawa, M.; Sumida, N.; Hattori, S.; Kurie, J. M.; Marini, J. C.; Yamauchi, M. Cyclophilin-B

Modulates Collagen Cross-linking by Differentially Affecting Lysine Hydroxylation in the Helical and Telopeptidyl Domains of Tendon Type I Collagen. *J. Biol. Chem.* **2016**, *291*, 9501-9512.

(104) Cescon, M.; Gattazzo, F.; Chen, P.; Bonaldo, P. Collagen VI at a glance. *J. Cell Sci.* **2015**, *128*, 3525-3531.

(105) Tanaka, T.; Wakabayashi, T.; Oizumi, H.; Nishio, S.; Sato, T.; Harada, A.; Fujii, D.; Matsuo, Y.; Hashimoto, T.; Iwatsubo, T. CLAC-P/collagen type XXV is required for the intramuscular innervation of motoneurons during neuromuscular development. *J. Neurosci.* **2014**, *34*, 1370-1379.

(106) Tu, H.; Huhtala, P.; Lee, H. M.; Adams, J. C.; Pihlajaniemi, T. Membrane-associated collagens with interrupted triple-helices (MACITs): evolution from a bilaterian common ancestor and functional conservation in C. elegans. *BMC Evol. Biol.* **2015**, *15*, 281.

(107) Fidler, A. L.; Darris, C. E.; Chetyrkin, S. V.; Pedchenko, V. K.; Boudko, S. P.; Brown, K. L.; Gray Jerome, W.; Hudson, J. K.; Rokas, A.; Hudson, B. G. Collagen IV and basement membrane at the evolutionary dawn of metazoan tissues. *Elife* **2017**, *6*.

(108) Cummings, C. F.; Pedchenko, V.; Brown, K. L.; Colon, S.; Rafi, M.; Jones-Paris, C.; Pokydeshava, E.; Liu, M.; Pastor-Pareja, J. C.; Stothers, C.; Ero-Tolliver, I. A.; McCall, A. S.; Vanacore, R.; Bhave, G.; Santoro, S.; Blackwell, T. S.; Zent, R.; Pozzi, A.; Hudson, B. G. Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J. Cell Biol.* **2016**, *213*, 479-494.

(109) McCall, A. S.; Cummings, C. F.; Bhave, G.; Vanacore, R.; Page-McCaw, A.; Hudson, B. G. Bromine is an essential trace element for assembly of collagen IV scaffolds in tissue development and architecture. *Cell* **2014**, *157*, 1380-1392.

(110) Fidler, A. L.; Vanacore, R. M.; Chetyrkin, S. V.; Pedchenko, V. K.; Bhave, G.; Yin, V. P.; Stothers, C. L.; Rose, K. L.; McDonald, W. H.; Clark, T. A.; Borza, D. B.; Steele, R. E.; Ivy, M. T.; Aspirnauts; Hudson, J. K.; Hudson, B. G. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. *Proc. Natl. Acad. Sci. U S A* **2014**, *111*, 331-336.

(111) Glentis, A.; Gurchenkov, V.; Matic Vignjevic, D. Assembly, heterogeneity, and breaching of the basement membranes. *Cell Adh. Migr.* **2014**, *8*, 236-245.

(112) Kalluri, R. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat. Rev. Cancer* **2003**, *3*, 422-433.

(113) Bachinger, H. P.; Morris, N. P.; Lunstrum, G. P.; Keene, D. R.; Rosenbaum, L. M.; Compton, L. A.; Burgeson, R. E. The relationship of the biophysical and biochemical characteristics of type VII collagen to the function of anchoring fibrils. *J. Biol. Chem.* **1990**, *265*, 10095-10101.

(114) Chung, H. J.; Uitto, J. Type VII collagen: the anchoring fibril protein at fault in dystrophic epidermolysis bullosa. *Dermatol. Clin.* **2010**, *28*, 93-105.

(115) Jin, L.; Pahuja, K. B.; Wickliffe, K. E.; Gorur, A.; Baumgartel, C.; Schekman, R.; Rape, M. Ubiquitin-dependent regulation of COPII coat size and function. *Nature* **2012**, *482*, 495-500.

(116) McGourty, C. A.; Akopian, D.; Walsh, C.; Gorur, A.; Werner, A.; Schekman, R.; Bautista, D.; Rape, M. Regulation of the CUL3 Ubiquitin Ligase by a Calcium-Dependent Co-adaptor. *Cell* **2016**, *167*, 525-538 e514.

(117) Wilson, D. G.; Phamluong, K.; Li, L.; Sun, M.; Cao, T. C.; Liu, P. S.; Modrusan, Z.; Sandoval, W. N.; Rangell, L.; Carano, R. A.; Peterson, A. S.; Solloway, M. J. Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J. Cell Biol.* **2011**, *193*, 935-951.

(118) Ishikawa, Y.; Ito, S.; Nagata, K.; Sakai, L. Y.; Bachinger, H. P. Intracellular mechanisms of molecular recognition and sorting for transport of large extracellular matrix molecules. *Proc. Natl. Acad. Sci. U S A* **2016**, *113*, E6036-E6044.

(119) Gawron, K. Endoplasmic reticulum stress in chondrodysplasias caused by mutations in collagen types II and X. *Cell Stress Chaperon.* **2016**, *21*, 943-958.

(120) Mirigian, L. S.; Makareeva, E.; Mertz, E. L.; Omari, S.; Roberts-Pilgrim, A. M.; Oestreich, A. K.; Phillips, C. L.; Leikin, S. Osteoblast Malfunction Caused by Cell Stress Response to Procollagen Misfolding in alpha2(I)-G610C Mouse Model of Osteogenesis Imperfecta. *J. Bone Miner. Res.* **2016**, *31*, 1608-1616.

(121) Wong, M. Y.; DiChiara, A. S.; Suen, P. H.; Chen, K.; Doan, N. D.; Shoulders, M. D. Adapting Secretory Proteostasis and Function Through the Unfolded Protein Response. *Curr. Top. Microbiol. Immunol.* **2017**.

(122) Ishida, Y.; Kubota, H.; Yamamoto, A.; Kitamura, A.; Bachinger, H. P.; Nagata, K. Type I collagen in Hsp47-null cells is aggregated in endoplasmic reticulum and deficient in N-propeptide processing and fibrillogenesis. *Mol. Biol. Cell* **2006**, *17*, 2346-2355.

(123) Fumagalli, F.; Noack, J.; Bergmann, T. J.; Cebollero, E.; Pisoni, G. B.; Fasana, E.; Fregno, I.; Galli, C.; Loi, M.; Solda, T.; D'Antuono, R.; Raimondi, A.; Jung, M.; Melnyk, A.; Schorr, S.; Schreiber, A.; Simonelli, L.; Varani, L.; Wilson-Zbinden, C.; Zerbe, O.; Hofmann, K.; Peter, M.; Quadroni, M.; Zimmermann, R.; Molinari, M. Translocon component Sec62 acts in endoplasmic reticulum turnover during stress recovery. *Nat. Cell Biol.* **2016**, *18*, 1173-1184.

(124) Khaminets, A.; Heinrich, T.; Mari, M.; Grumati, P.; Huebner, A. K.; Akutsu, M.; Liebmann, L.; Stolz, A.; Nietzsche, S.; Koch, N.; Mauthe, M.; Katona, I.; Qualmann, B.; Weis, J.; Reggiori, F.; Kurth, I.; Hubner, C. A.; Dikic, I. Regulation of endoplasmic reticulum turnover by selective autophagy. *Nature* **2015**, *522*, 354-358.

(125) Marini, J. C.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; San Antonio, J. D.; Milgrom, S.; Hyland, J. C.; Korkko, J.; Prockop, D. J.; De Paepe, A.; Coucke, P.; Symoens, S.; Glorieux, F. H.; Roughley, P. J.; Lund, A. M.; Kuurila-Svahn, K.; Hartikka, H.; Cohn, D. H.; Krakow, D.; Mottes, M.; Schwarze, U.; Chen, D.; Yang, K.; Kuslich, C.; Troendle, J.; Dalgleish, R.; Byers, P. H. 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. *Hum. Mutat.* **2007**, *28*, 209-221.

(126) Dalgleish, R. The human type I collagen mutation database. *Nucleic Acids Res.* **1998**, *25*, 181-187.

(127) Dalgleish, R. The human collagen mutation database 1998. *Nucleic Acids Res.* **1998**, *26*, 253-255.

(128) Jones, F. E.; Bailey, M. A.; Murray, L. S.; Lu, Y.; McNeilly, S.; Schlotzer-Schrehardt, U.; Lennon, R.; Sado, Y.; Brownstein, D. G.; Mullins, J. J.; Kadler, K. E.; Van Agtmael, T. ER stress and basement membrane defects combine to cause glomerular and tubular renal disease resulting from Col4a1 mutations in mice. *Dis. Model Mech.* **2016**, *9*, 165-176.

(129) Jaronen, M.; Goldsteins, G.; Koistinaho, J. ER stress and unfolded protein response in amyotrophic lateral sclerosis-a controversial role of protein disulphide isomerase. *Front. Cell Neurosci.* **2014**, *8*, 402.

(130) Matus, S.; Valenzuela, V.; Medinas, D. B.; Hetz, C. ER Dysfunction and Protein Folding Stress in ALS. *Int. J. Cell Biol.* **2013**, *2013*, 674751.

(131) Prell, T.; Lautenschlager, J.; Witte, O. W.; Carri, M. T.; Grosskreutz, J. The unfolded protein response in models of human mutant G93A amyotrophic lateral sclerosis. *Eur. J. Neurosci.* **2012**, *35*, 652-660.

(132) Woo, C. W.; Cui, D.; Arellano, J.; Dorweiler, B.; Harding, H.; Fitzgerald, K. A.; Ron, D.; Tabas, I. Adaptive suppression of the ATF4-CHOP branch of the unfolded protein response by toll-like receptor signalling. *Nat. Cell Biol.* **2009**, *11*, 1473-1480.

(133) Boot-Handford, R. P.; Briggs, M. D. The unfolded protein response and its relevance to connective tissue diseases. *Cell Tissue Res.* **2010**, *339*, 197-211.

(134) Starborg, T.; Kalson, N. S.; Lu, Y.; Mironov, A.; Cootes, T. F.; Holmes, D. F.; Kadler, K. E. Using transmission electron microscopy and 3View to determine collagen fibril size and three-dimensional organization. *Nat. Protoc.* **2013**, *8*, 1433-1448.

(135) Bovolenta, M.; Neri, M.; Martoni, E.; Urciuolo, A.; Sabatelli, P.; Fabris, M.; Grumati, P.; Mercuri, E.; Bertini, E.; Merlini, L.; Bonaldo, P.; Ferlini, A.; Gualandi, F. Identification of a deep intronic mutation in the COL6A2 gene by a novel custom oligonucleotide CGH array designed to explore allelic and genetic heterogeneity in collagen VI-related myopathies. *BMC Med. Genet.* **2010**, *11*, 44.

(136) Lunstrum, G. P., Sakai, L.Y., Keene, D.R., Morris, N.P., Burgeson, R.E. Large complex globular domains of type VII procollagen contribute to the structure of anchoring fibrils. *J. Biol. Chem.* **1986**, *261*, 9042-9048.

(137) Hoffman, G. S.; Filie, J. D.; Schumacher, H. R., Jr.; Ortiz-Bravo, E.; Tsokos, M. G.; Marini, J. C.; Kerr, G. S.; Ling, Q. H.; Trentham, D. E. Intractable vasculitis, resorptive osteolysis, and immunity to type I collagen in type VIII Ehlers-Danlos syndrome. *Arthritis Rheum.* **1991**, *34*, 1466-1475.

(138) Godfrey, M.; Hollister, D. W. Type II achondrogenesis-hypochondrogenesis: identification of abnormal type II collagen. *Am. J. Hum. Genet.* **1988**, *43*, 904-913.

(139) Arita, M.; Fertala, J.; Hou, C.; Steplewski, A.; Fertala, A. Mechanisms of aberrant organization of growth plates in conditional transgenic mouse model of spondyloepiphyseal dysplasia associated with the R992C substitution in collagen II. *Am. J. Pathol.* **2015**, *185*, 214-229.

(140) Mizuno, K.; Boudko, S.; Engel, J.; Bachinger, H. P. Vascular Ehlers-Danlos syndrome mutations in type III collagen differently stall the triple helical folding. *J. Biol. Chem.* **2013**, *288*, 19166-19176.

(141) Bateman, J. F.; Chiodo, A. A.; Weng, Y. M.; Chan, D.; Haan, E. A type III collagen Gly559 to Arg helix mutation in Ehler's-Danlos syndrome type IV. *Hum. Mutat.* **1998**, *Suppl 1*, S257-259.

(142) Pousi, B.; Hautala, T.; Hyland, J. C.; Schroter, J.; Eckes, B.; Kivirikko, K. I.; Myllyla, R. A compound heterozygote patient with Ehlers-Danlos syndrome type VI has a deletion in one allele and a splicing defect in the other allele of the lysyl hydroxylase gene. *Hum. Mutat.* **1998**, *11*, 55-61.

(143) Hudson, B. G.; Reeders, S. T.; Tryggvason, K. Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. *J. Biol. Chem.* **1993**, *268*, 26033-26036.

(144) Symoens, S.; Syx, D.; Malfait, F.; Callewaert, B.; De Backer, J.; Vanakker, O.; Coucke, P.; De Paepe, A. Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. *Hum. Mutat.* **2012**, *33*, 1485-1493.

(145) Lamande, S. R.; Morgelin, M.; Selan, C.; Jobsis, G. J.; Baas, F.; Bateman, J. F. Kinked collagen VI tetramers and reduced microfibril formation as a result of Bethlem myopathy and introduced triple helical glycine mutations. *J. Biol. Chem.* **2002**, *277*, 1949-1956.

(146) Anthony, S.; Munk, R.; Skakun, W.; Masini, M. Multiple epiphyseal dysplasia. *J Am. Acad. Orthop. Surg.* **2015**, *23*, 164-172.

(147) Acke, F. R.; Malfait, F.; Vanakker, O. M.; Steyaert, W.; De Leeneer, K.; Mortier, G.; Dhooge, I.; De Paepe, A.; De Leenheer, E. M.; Coucke, P. J. Novel pathogenic COL11A1/COL11A2 variants in Stickler syndrome detected by targeted NGS and exome sequencing. *Mol. Genet. Metab.* **2014**, *113*, 230-235.

(148) Seppinen, L.; Pihlajaniemi, T. The multiple functions of collagen XVIII in development and disease. *Matrix Biol.* **2011**, *30*, 83-92.

Chapter 2:
A Molecular Code for Collagen C-Propeptide Assembly

DiChiara, A.S., Suen, P.H., Li, R.C., Weickhardt, A.F., Taylor, R.J., Malhotra, D., McCaslin, D.R., Shoulders, M.D. A molecular code for collagen C-propeptide assembly. *To Be Submitted.*

2.1 Author Contributions

A manuscript reporting this work is in the final stages of revision prior to re-submission. A.S.D. performed the majority of the experiments and analyses with assistance from the following research technicians (his mentees): P.H.S., A.F.W. and D.M. A.S.D. and R.C.L. collaboratively performed disulfide bond analyses. D.R.M. (UW–Madison) performed the analytical ultracentrifugation.

2.2 Introduction

Construction of the functionally diverse collagenous extracellular matrices, including bone, skin, cartilage, tendon, and beyond, is perhaps the most remarkable biological supramolecular assembly process.[1,2] The formation of these distinctive tissues begins with and depends upon proper folding of individual collagen triple helices in the endoplasmic reticulum (ER). There are 28 different types of collagen, which assume a wide range of extracellular structures that are necessary to support tissue and organismal health.[3] It is common for many different types of collagen to be synthesized and secreted from the same cell type. As such, improper assembly of different types of collagen monomers must be prevented, while concurrently ensuring that each type of collagen assembles into a trimer with the correct stoichiometric composition. This complex process is initiated and guided by the C-terminal non-collagenous (NC) domain, the globular, cysteine-rich, N-glycosylated ~30 kDa collagen C-propeptide (referred to as C-Pro for the remainder of this chapter) present on each nascent collagen polypeptide.[4,5]

Initial work aimed at understanding how the C-Pro domain addresses the ability of a given strand to form homotrimers used Colα1(III) and Colα2(I) as prototypical collagen strands. Colα1(III) can form homotrimers, while Colα2(I) only forms heterotrimers with Colα1(I). Mini-collagen genes consisting of each protein's endogenous N-propeptide, a shortened, ~ 200-

amino acid triple helical domain, and C-propeptide, were expressed in a cell-free translation system, and each protein's corresponding trimerization propensities were assessed by non-reducing and reducing SDS-PAGE. As expected, wild-type Colα1(III) formed homotrimers, while wild-type Colα2(I) remained monomeric under non-reducing conditions. When each sample was treated with reducing agent, both species migrated as monomers, indicating the disulfide-linked nature of the Colα1(III) homotrimer. Rationalizing that the propensity to form homotrimers lies in the C-Pro nucleation domain, and knowing that Colα1(III) formed disulfide-linked homotrimers, Lees, and co-workers[6] noted the difference in the cysteine composition of Colα1(III) and Colα2(I). Colα1(III) maintained eight highly conserved cysteine residues, while Colα2(I) only contained seven (**Figure 2.1**). Mutating the second position in Colα2(I) to a cysteine and monitoring assembly under non-reducing conditions, however, did not restore Colα2(I)'s ability to form a homotrimer in this system. In contrast, mutation of any of the first four cysteine residues to serine residues in Colα1(III) prevented disulfide-linked homotrimerization. The authors used this data to conclude that the cysteine network is not responsible for regulating homo- versus hetero-trimerization, and that there must be other factors in the sequence of the C-Pro domain.[6]

Building on the aforementioned experiments, later work created another set of mini-collagen genes with the same template (N-propeptide, short triple helical domain, C-Pro), but this time using complex chimeras of Colα1(III) and Colα2(I). The authors replaced regions of the Colα2(I) C-Pro with corresponding sequences derived from the homotrimerizing Colα1(III) C-Pro domain, in an effort to drive Colα2(I) homotrimerization. Each construct was expressed in cell-free lysate, and analyzed by reducing and non-reducing SDS-PAGE as well as proteolysis assays to assess the foldedness of the triple helix. (Triple helical collagen is resistant to protease treatment such as pepsin, while monomeric collagen is succeptible.) While the wild-type construct of Colα2(I) was completely degraded, the authors observed triple helical stability

of a Colα2(I) construct that had 23 amino acids in its C-Pro domain replaced with the corresponding 23 amino acid sequence from Colα1(III). This led to the identification of a region within the C-Pro domain that regulates type-specific collagen assembly, now known as the collagen recognition sequence (CRS).[7] Further work has confirmed a role for the CRS in ensuring that collagen polypeptides assemble only with their own collagen type,[8,9] but the CRS appears to have little or no role in defining whether a given collagen strand will form homo- versus hetero-trimers.

More recently, crystal structures of homotrimeric Colα1(III)[10] and Colα1(I)[8] C-Pro domains were solved (the Colα1(I) homotrimer structure is shown in **Figure 2.2A**). The disulfide-bonding pattern of collagen homotrimers was also revealed by the solved structure (**Figure 2.2B**). The structures are consistent with a role for the CRS in type-specific assembly, with salt bridges between adjacent monomers in the assembled trimer formed by residues within the CRS. Using site-directed mutagenesis and protein purification, the authors found that Lys129 and Glu130, which are both in the CRS, as well as Arg45 and Lys247 in Colα2(I) can promote assembly of Colα2(I) into a heterotrimer. Thus, this recent structural data provides significant insight into how the CRS guides type-specific collagen assembly and how Colα2(I) incorporates into a heterotrimer, but does not address the fundamental question of how homo- versus hetero-trimerization is encoded in the C-Pro domain.

None of the aforementioned work identifies an amino acid or sequence of amino acids that are responsible for driving homotrimerization of the collagen C-Pro domain. To date, no one has successfully converted a constitutively monomeric collagen C-Pro into one that can homotrimerize, indicating that we still do not understand the answer to this fundamental question. The early work of Lees and co-workers[6] concluding that cysteine residues are not the key feature has steered the field away from that possibility, even though the conclusions were based on complex protein chimeras and low-resolution experimental results. It is plausible that

fusion to an abiological, short triple-helical domain in the Lees and co-workers[6] work obfuscates the normal role of the C-Pro in initiating trimerization, as short triple-helical sequences are able to assemble in vitro even in the absence of a C-Pro.[11-13] Moreover, practical limitations at the time meant that purified proteins were neither obtained nor biophysically characterized to assess the proteins' oligomeric states. Furthermore, evaluating homo- versus hetero-trimerization depended largely on proteolysis assays, which interrogate the structural integrity of triple helices, not the assembly of the C-Pro domain. A subset of constructs studied may be able to form assembled homotrimers (directed by elements in the C-Pro domains) with proteolytically sensitive triple helices.

In this chapter, I describe our efforts to revisit the fundamental question of the elements that drive homo- versus hetero-trimerization of fibrillar collagen C-Pro domains. Here, I describe the resulting comprehensive suite of biochemical and biophysical studies that reveal the cysteine network as an ancient and the primary contributing factor governing collagen C-Pro trimerization propensity, which we ultimately show relies on the identity of just one single atom in the entire ~30 kDa protein domain.

```
Colα1(III)   SRKNPARNCRDLKFCHPELKSGEYWVDPNQGCKLDAIKVFCNMETGETCISANPLNVPRK
Colα2(I)     SRKNPARTCRDLRLSHPEWSSGYYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAK

Colα1(III)   HWWTDSSAEKKHVWFGESMDGGFQFSYGNPELPEDVLDVQLAFLRLLSSRASQNITYHCK
Colα2(I)     NWYRS-SKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCK

Colα1(III)   NSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTYTVLEDGCTKHTGEWSKTVFEYRT
Colα2(I)     NSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKT

Colα1(III)   RKAVRLPIVDIAPYDIGGPDQEFGVDVGPVCFL
Colα2(I)     NKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK
```

Figure 2.1 | Alignment of the C-Pro domains of Colα1(III) and Colα2(I)

Alignment of Colα1(III) and Colα2(I) C-Pro domains was done using Clustal Omega.[28,29] This alignment highlights a high degree of similarity between the two domains. Each cysteine residue is bolded in red. Note that the second cysteine is missing in Colα2(I).

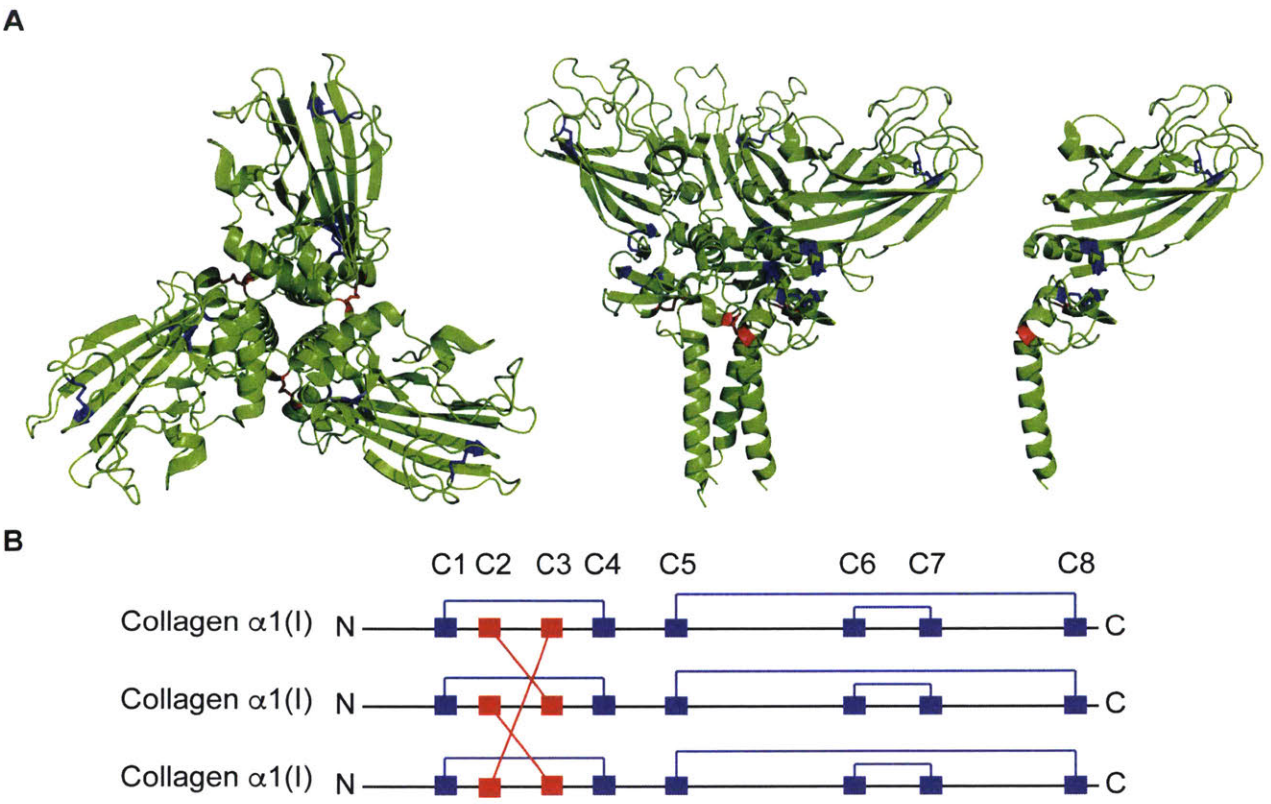


Figure 2.2 | Crystal structure highlights the disulfide bonds within the assembled homotrimer

(A) Three views of the crystal structure of Colα1(I) previously published (PDB: 5K31).[8] Left to right: top-down view to show the interstrand disulfide bonds (red); side view to show the overall shape of the C-Pro domain, as well as the intrastrand disulfide bonds (blue); and side view of a single monomer for simplicity. **(B)** Schematic representation of the disulfide bonding pattern in the Colα1(I) homotrimer. Each cysteine is represented by a colored square in **(B)**. In both **(A)** and **(B)** the intrastrand disulfide bonds are shown in blue, and interstrand disulfide bonds are shown in red.

2.3 Results

2.3.1 Collagen-I C-Pro Sequence Analysis

The most abundant collagen, type-I,[14] is the most prominent example of otherwise quite similar collagen strands displaying unique trimerization propensities.[15] Human collagen-I is formed from two distinctive polypeptide strands, Colα1(I) and Colα2(I), which are encoded by the *Col1A1* and *Col1A2* genes, respectively. The predominant secreted form is a 2:1 heterotrimer of Colα1(I) to Colα2(I). Although homotrimers of Colα1(I) have been identified in carcinomas,[16] fetal tissue,[17] and fibrotic wounds,[18] homotrimers of Colα2(I) have never been detected (**Figure 2.3A**). Alignment of human Colα1(I) and Colα2(I) reveals strong similarities in the C-Pro domain, with 62% sequence identity and 78% similarity. One striking difference, however, is that the homotrimer-forming Colα1(I) C-Pro has eight cysteine residues, (like Colα1(III) in **Figure 2.1**), while the heterotrimer-forming only Colα2(I) C-Pro has only seven cysteine residues (**Figure 2.3B**).

We began by aligning ancestral collagen C-Pro domains to track the appearance of each factor known to play a role in collagen assembly through evolutionary history. Our goal was to identify a conserved factor that emerged at the same evolutionary time-point as collagen-I heterotrimers, which likely begin to appear as early as hydra.[19] First, we noted that the CRS is a recently evolved vertebrate mechanism of chain selection. This sequence did not emerge until the problem of type-specific collagen assembly became significantly complicated by the presence of many types of collagen in a single organism.[20-22] For example, collagen sequences older than sharks do not contain a CRS at all, suggesting the CRS is a relatively recent evolutionary development. Second, modeling by Sharma and co-workers[8] based on the crystal structure of a Colα1(I) homotrimer suggested the importance of Lys129 and Glu130 in the CRS of human Colα2(I) in forming salt bridges with the adjacent Colα1(I) of the heterotrimer.

However, Lys129 did not fix in the evolutionary tree until coelacanths and Glu130 is present only in humans (an aspartic acid is present in frogs). Species older than frogs do not contain a negatively charged amino acid in this location. It therefore seems unlikely that Lys129 and Glu130 are responsible for the inability of Colα2(I) to homotrimerize. For Colα1(I), the negatively charged amino acids Asp126 and Asp129 proposed to be important by Sharma and co-workers[8] were not yet fixed in sharks, one of the early-diverging vertebrates that later lost bone (relevant amino acids highlighted in bold red in **Figure 2.4**). Bearing in mind that heterotrimeric collagens possibly emerged as early as hydra, these observations suggest that the CRS, while certainly important for type-specific assembly in modern collagens, is not the defining factor controlling whether a given collagen strand homo- versus hetero-trimerizes.

Examination of the crystal structure followed by biochemical characterization identified several other amino acids that influenced the ability of Colα2(I) to incorporate into heterotrimers.[8] While there is considerable variability in the early portions of the evolutionary tree, Arg45 and Lys247 both fix at about the same time. Lamprey collagen has Arg45, but no Lys247, while tunicate collagen contains Lys247, but not Arg45. Without a CRS domain in lampreys or tunicates, however, the other half of the salt bridge (proposed to be Asp126 and Asp129 in Colα1(I)) does not exist. Arg45 and Lys247 therefore also seem unlikely to be the key feature controlling the ability to homotrimerize.

In contrast to the low evolutionary penetrance of the CRS and specific amino acids proposed to be important by modeling Colα2(I) into the homotrimeric Colα1(I) crystal structure, the C-Pro cysteine network emerges early and is extremely well-conserved across evolutionary time. [We note that for simplicity and clarity in this chapter, each conserved cysteine residue is numbered from 1 to 8, as shown in **Figure 2.4**. The numbering is in reference to the order of the cysteine residues, from the N to C direction, and is not related to the specific amino acid number in each collagen strand. For example C2 in Colα1(I) is amino acid number 1265, numbered from

the first amino acid translated, while C2 in Colα2(I) is amino acid number 1169.] The cysteine residues involved in intrastrand disulfide bonding, C1, C4 and C5-C8, are conserved through sponges and hydra, the evolutionary counterparts of bilateria, with no variability in their presence or sequence location. The interstrand disulfide bonding cysteine residues, however, C2 and C3, show considerable variability across the early portions of the phylogenetic tree (**Figures 2.4** and **2.5**). Protostomes (arthropods and mollusks) appear to have collagen genes without either C2 or C3, while all of the deuterostomes examined have variability in cysteine occupancy in these two positions. However, the alignments in **Figure 2.4** beginning with chordates strongly suggest that the common ancestor of all chordates locked in to a pattern where one collagen-I strand conserves all eight cysteine residues, while the other consistently contains only seven cysteine residues. The strand that contains only seven cysteine residues is always missing either C2 or C3, the proposed interstrand disulfide bonding cysteine residues. These observations motivated us to revisit the question of a fundamental role for intermolecular disulfide bonds in governing collagen homo- versus hetero-trimerization.

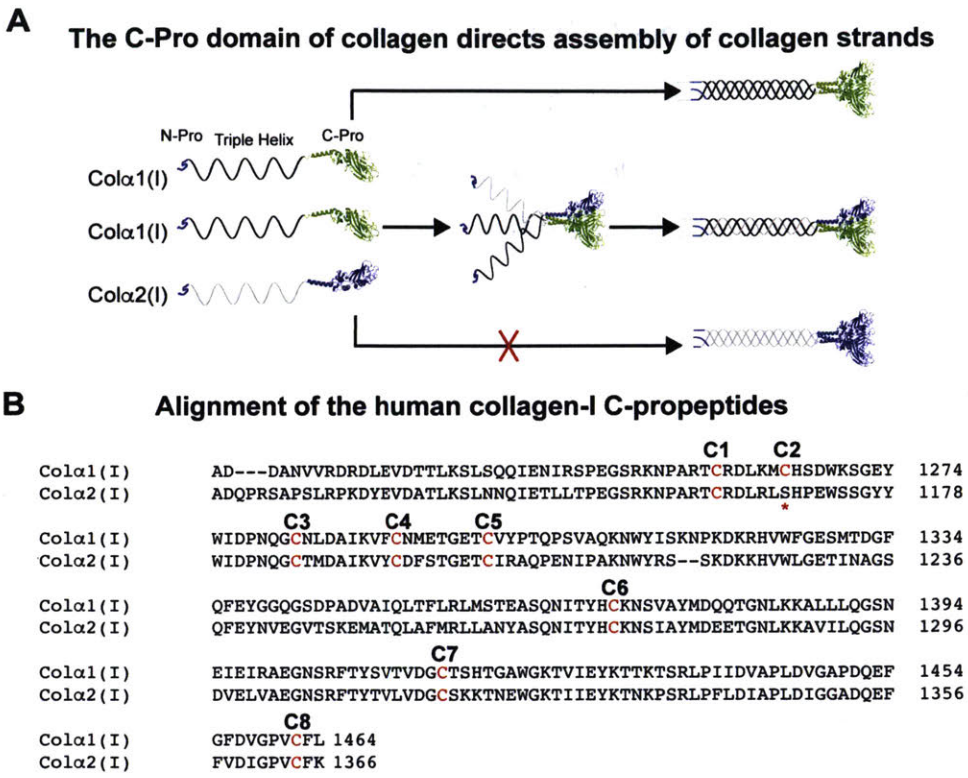
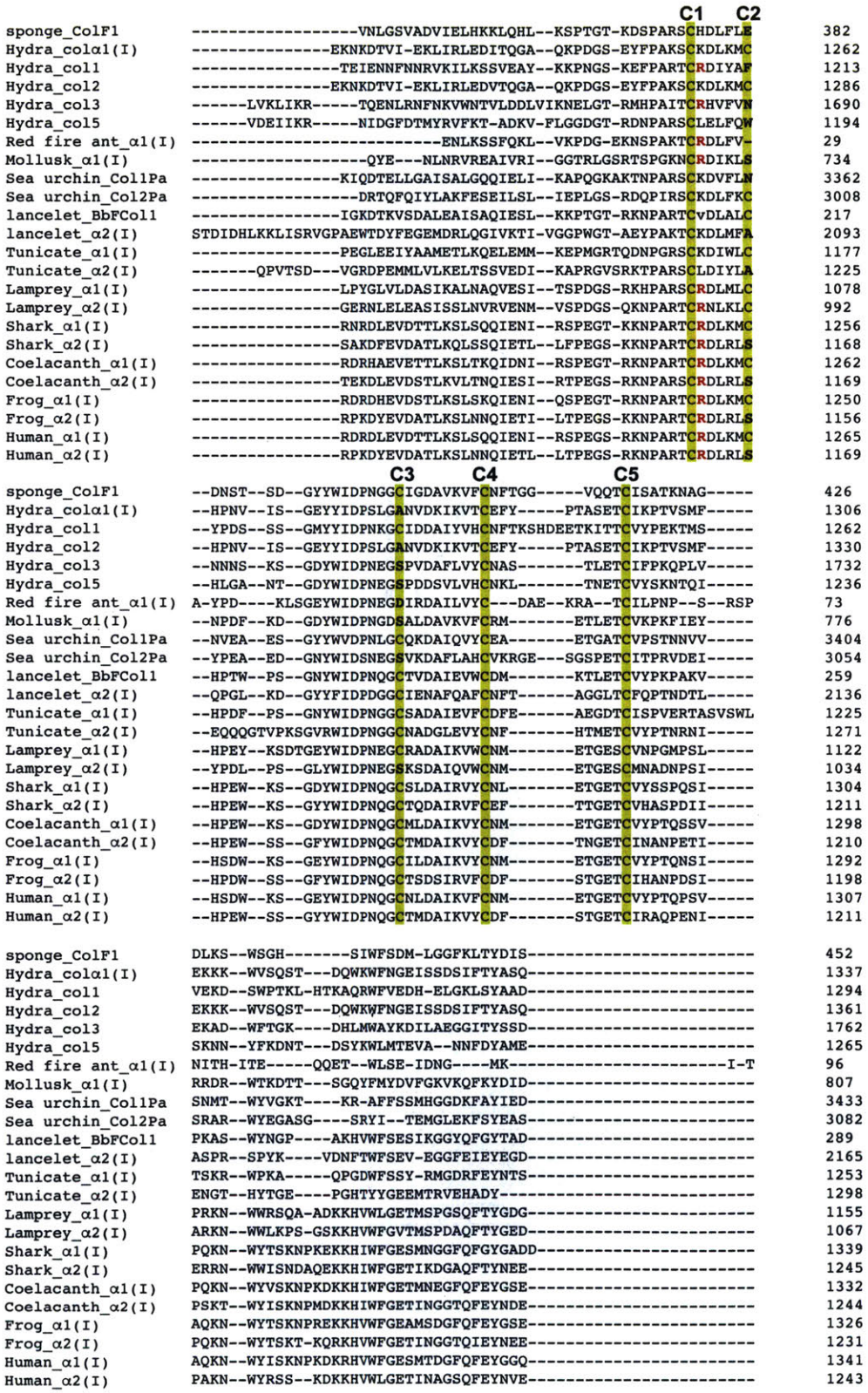


Figure 2.3 | C-Pro domain-mediated assembly of collagen type-I

(A) Schematic representation of collagen-I assembly. Two strands of Colα1(I) and one strand of Colα2(I) typically assemble into heterotrimers, a process that is initiated by their C-Pro domains. Colα1(I) is also known to form homotrimers, while Colα2(I) can only form heterotrimers. A crystal structure of a homotrimeric C-Pro domain is shown for illustration purposes (PDB-ID 5K31). **(B)** Alignment of Colα1(I) and Colα2(I) C-Pro domains highlights a high degree of similarity between the two domains. The cysteine network is numbered from C1–C8 from the N- to C- direction, with each cysteine residue colored in red, and the location of the missing cysteine residue in Colα2(I) shown with a red asterisk (*).

Figure 2.4


C1
C2
C3
C4
C5
sponge_ColF1
Hydra_colα1(I)
Hydra_col1
Hydra_col2
Hydra_col3
Hydra_col5
Red fire ant_α1(I)
Mollusk_α1(I)
Sea urchin_Col1Pa
Sea urchin_Col2Pa
lancelet_BbFCol1
lancelet_α2(I)
Tunicate_α1(I)
Tunicate_α2(I)
Lamprey_α1(I)
Lamprey_α2(I)
Shark_α1(I)
Shark_α2(I)
Coelacanth_α1(I)
Coelacanth_α2(I)
Frog_α1(I)
Frog_α2(I)
Human_α1(I)
Human_α2(I)

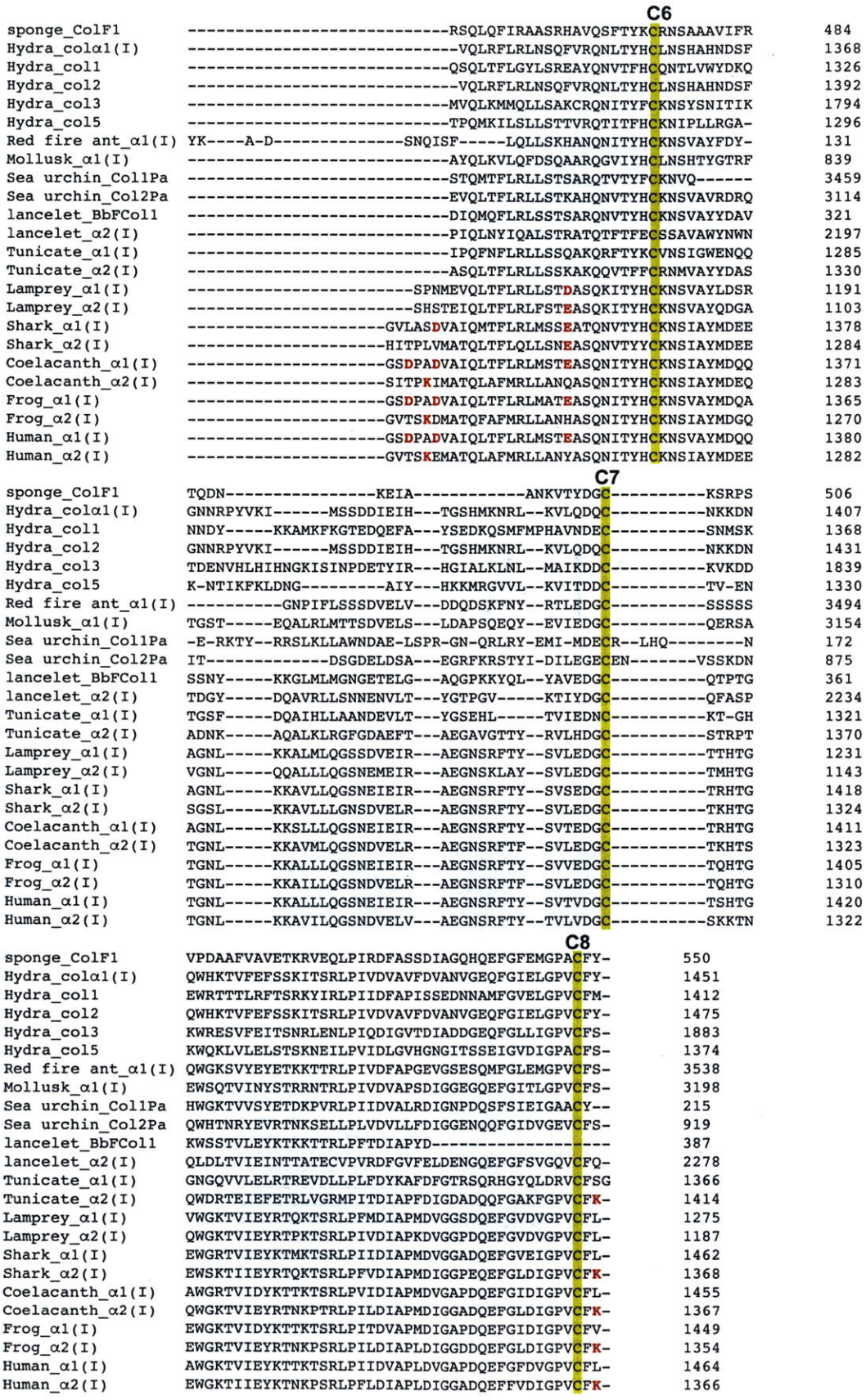

C6
C7
C8
sponge_ColF1
Hydra_colα1(I)
Hydra_col1
Hydra_col2
Hydra_col3
Hydra_col5
Red fire ant_α1(I)
Mollusk_α1(I)
Sea urchin_Col1Pa
Sea urchin_Col2Pa
lancelet_BbFCol1
lancelet_α2(I)
Tunicate_α1(I)
Tunicate_α2(I)
Lamprey_α1(I)
Lamprey_α2(I)
Shark_α1(I)
Shark_α2(I)
Coelacanth_α1(I)
Coelacanth_α2(I)
Frog_α1(I)
Frog_α2(I)
Human_α1(I)
Human_α2(I)

Figure 2.4 | Ancestral collagen C-Pro sequence alignment suggests the cysteine network is one of the more ancient sequence elements of the C-Pro domain

Ancestral collagen C-Pro domains were aligned using Clustal Omega.[28,29] Locations of the conserved cysteine residues are highlighted in yellow, with the cysteine number indicated at the top of the alignment. Residues other than cysteine in the conserved locations are bolded. Red bolded residues are the proposed[8] essential residues for salt bridge formation tracked through the evolutionary tree.

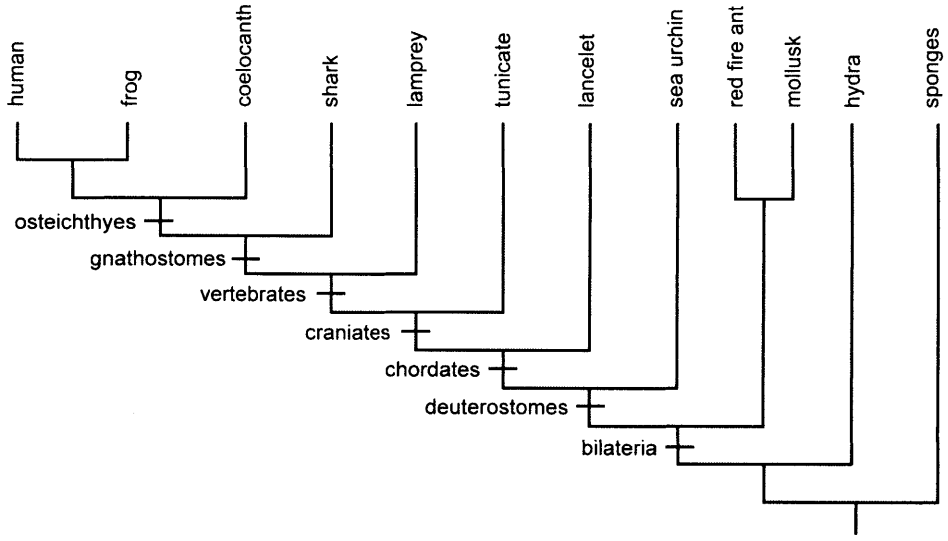


Figure 2.5 | Cladogram of ancestral collagen C-Pro domains as it relates to Figure 2.4

Evolution of the collagen-I C-Pro domain from sponges to humans, with the names of each clade indicated at each break point.

2.3.2 Wild-type C-Pro domains are secreted as expected from cultured cells

We began by creating plasmids for expression of HA-tagged C-Proα1(I) and FLAG-tagged C-Proα2(I) in human cells. The distinctive antibody epitopes were included to enable differential detection by immunoblotting. We incorporated a preprotrypsin signal sequence to target the proteins to the endoplasmic reticulum for folding and subsequent secretion. We found that, when expressed alone in HEK293 cells, both C-Proα1(I) and C-Proα2(I) were robustly secreted. SDS-PAGE immunoblot analysis of the media demonstrated the expected disulfide-dependent assembly patterns for these proteins. C-Proα1(I) migrated as a homotrimer and C-Proα2(I) migrated as a monomer under non-reducing conditions, while both migrated as monomers under reducing conditions (**Figure 2.6A**). Moreover, co-expression of both C-Proα1(I) and C-Proα2(I) rescued C-Proα2(I) into a disulfide-linked heterotrimer. This assembly pattern fully recapitulates the known assembly of full-length Colα1(I) and Colα2(I), indicating that these biochemically amenable constructs provide a valid system to study the molecular code for collagen homo- versus hetero-trimerization.

2.3.3 Cysteine swapping transposes oligomerization patterns of collagen-I C-Pro domains

We next tested the hypothesis that the homo- versus hetero-trimerization propensities of Colα1(I) and Colα2(I) C-Pro domains can be transposed simply by replacing C2 with a serine residue in C-Proα1(I) and restoring the C2 residue in C-Proα2(I) (**Figure 2.3B**). Strikingly, the C2S variant of C-Proα1(I) migrated as a monomer on a non-reducing SDS-PAGE gel when expressed alone in HEK293 cells, whereas the S2C variant of C-Proα2(I) migrated as a homotrimer (**Figure 2.6B**). Under reducing conditions, all of the protein samples migrated as monomers. These results suggest that changing just one amino acid is sufficient to transpose the functions of C-Proα1(I) and C-Proα2(I). To our knowledge, our results provide the first example of a homotrimerizing variant of Colα2(I) as a result of a only a single amino acid

change. Neither changes to the CRS nor to the Arg45 or Lys247 residues proposed to play a role by Sharma and co-workers[8] is required.

Complete function swapping requires, in addition, that S2C C-Proα2(I) can rescue C2S C-Proα1(I) into a heterotrimer upon co-expression. Indeed, we found that co-expression of both variants does rescue C2S C-Proα1(I) into a disulfide-linked heterotrimer with S2C C-Proα2(I) (**Figure 2.6B**). Thus, we can not only transpose the trimerization propensities of the C-Proα1(I) and C-Proα2(I) by modifying C2, but we can also retain the ability to rescue the variant C2S C-Proα1(I) domain that has lost the ability to form disulfide-linked homotrimers into heterotrimers.

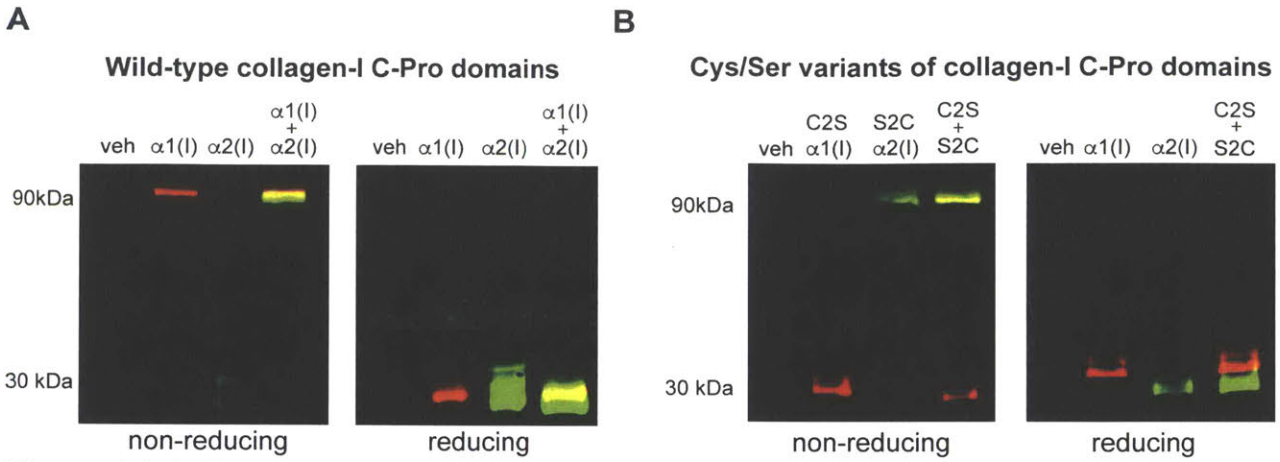


Figure 2.6 | The presence or absence of C2 controls disulfide-dependent collagen-I assembly

(A) Immunoblot analysis of individually expressed wild-type, HA-tagged C-Proα1(I) (red) and wild-type, FLAG-tagged C-Proα2(I) (green) proteins under non-reducing and reducing conditions demonstrated that wild-type C-Proα1(I) formed a disulfide-linked homotrimer whereas wild-type C-Proα2(I) did not, recapitulating the known disulfide-dependent assembly patterns of full length Colα1(I) and Colα2(I). Co-expression of wild-type C-Proα1(I) and C-Proα2(I) rescued C-Proα2(I) into a disulfide-linked heterotrimer (yellow, representing green and red overlap), as expected. **(B)** Immunoblot analysis of individually expressed HA-tagged Cys1265Ser (C2S) C-Proα1(I) (red) and FLAG-tagged Ser1169Cys (S2C) C-Proα2(I) (green) proteins under non-reducing and reducing conditions showed that the serine variant of C-Proα1(I) was no longer able to form a disulfide-linked homotrimer. In contrast, the cysteine variant of C-Proα2(I) was able to homotrimerize in a disulfide-dependent manner. Moreover, this single amino acid substitution transposed the functions of C-Proα1(I) and C-Proα2(I), as co-expression of both variants rescued C2S C-Proα1(I) into a disulfide-linked heterotrimer (yellow, representing green and red overlap). Amino acid numbering was derived from the corresponding full-length collagen-I proteins, as in **Figures 2.3** and **2.4**. Both **(A)** and **(B)** were developed using secondary antibodies conjugated to red or green dyes.

2.3.4 C-Pro domain purification and size exclusion chromatography analyses

These observations provide compelling evidence that the presence or absence of C2 in the collagen-I C-Pro domain is the defining feature controlling the ability to homotrimerize in a disulfide-dependent manner. However, results derived from non-reducing SDS-PAGE gels do not address the potential of C-Pro domains to trimerize independent of disulfide bond formation, as the analyses are necessarily performed under denaturing conditions. To address this critical issue, we purified mg quantities of His-tagged versions of all four C-Proα(I) variants, cleaving the 6×-His tag in the final step to yield C-Pro domains with only a two amino acid scar (Gly-Pro) on the N-terminus of the protein. **Figure 2.7A** outlines our purification scheme, **Figure 2.7B** displays a Coommassie-stained gel highlighting the purity of each recombinant protein under reducing gel conditions. Dynamic light scattering further suggested the purified proteins are relatively homogenous, with particles of similar radial size (**Figure 2.7C**). **Figure 2.8** shows corresponding MALDI-TOF characterization of the purified protein variants under reducing conditions.

Size exclusion chromatography analyses on the purified proteins suggested that wild-type C-Proα1(I) formed a homotrimer. In contrast, wild-type C-Proα2(I) was predominantly a monomer with some dimer present (**Figure 2.9A**). Removing or adding a cysteine transposed these assembly propensities even under these native conditions, with C2S C-Proα1(I) eluting primarily as a monomer with some dimer and S2C C-Proα2(I) predominantly forming a homotrimer (**Figure 2.9B**), as predicted based on the non-reducing, denaturing gel analysis in **Figure 2.6**.

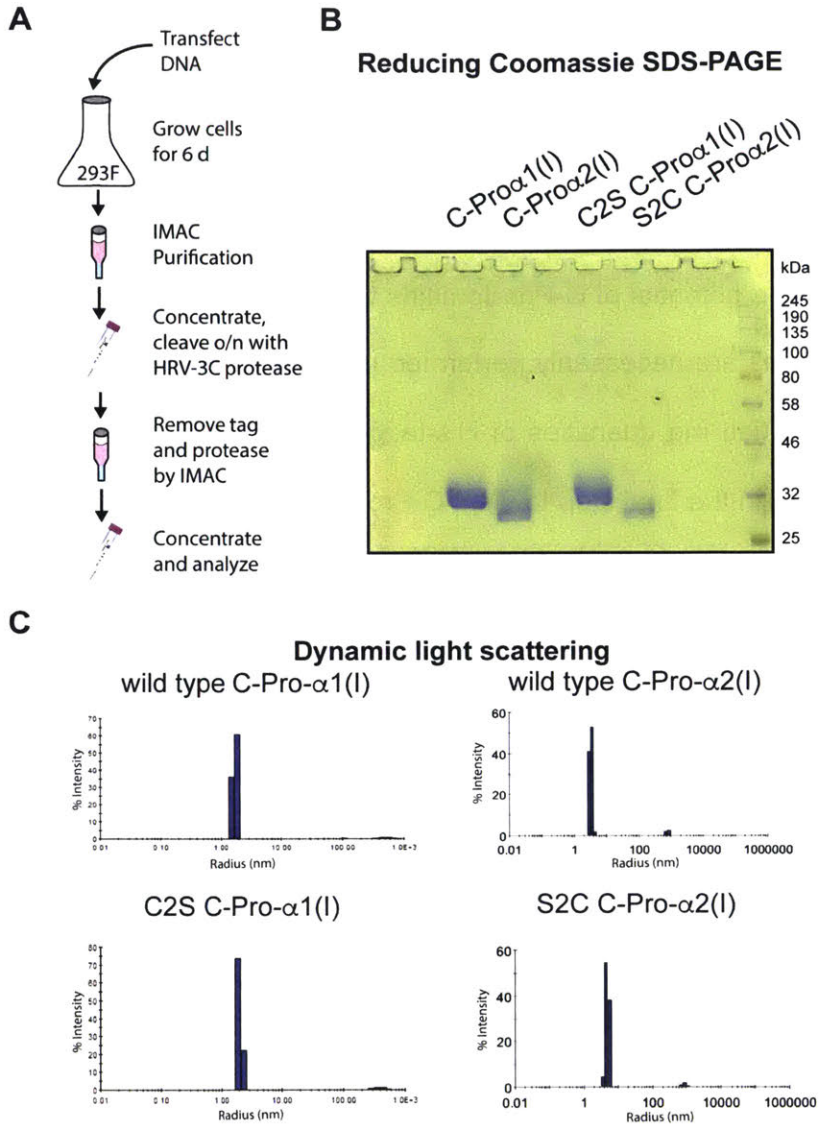


Figure 2.7 | C-Pro variants can be purified from culture media to a high degree of purity

(A) Schematic workflow of protein purification. Briefly, HEK293-Freestyle cells were transfected with each construct, and after 6 d, the conditioned media supernatant was collected and purified using Ni-NTA beads, eluted and cleaved overnight by HRV-3C protease. **(B)** Coomassie staining of a reducing SDS-PAGE gel analysis of recombinantly prepared samples of wild-type C-Proα1(I) and C-Proα2(I) as well as C2S C-Proα1(I) and S2C C-Proα2(I) demonstrated a high level of protein purity. **(C)** Dynamic light scattering analysis of protein samples in **(B)**, shown as % of total intensity, confirmed a high level of homogeneity.

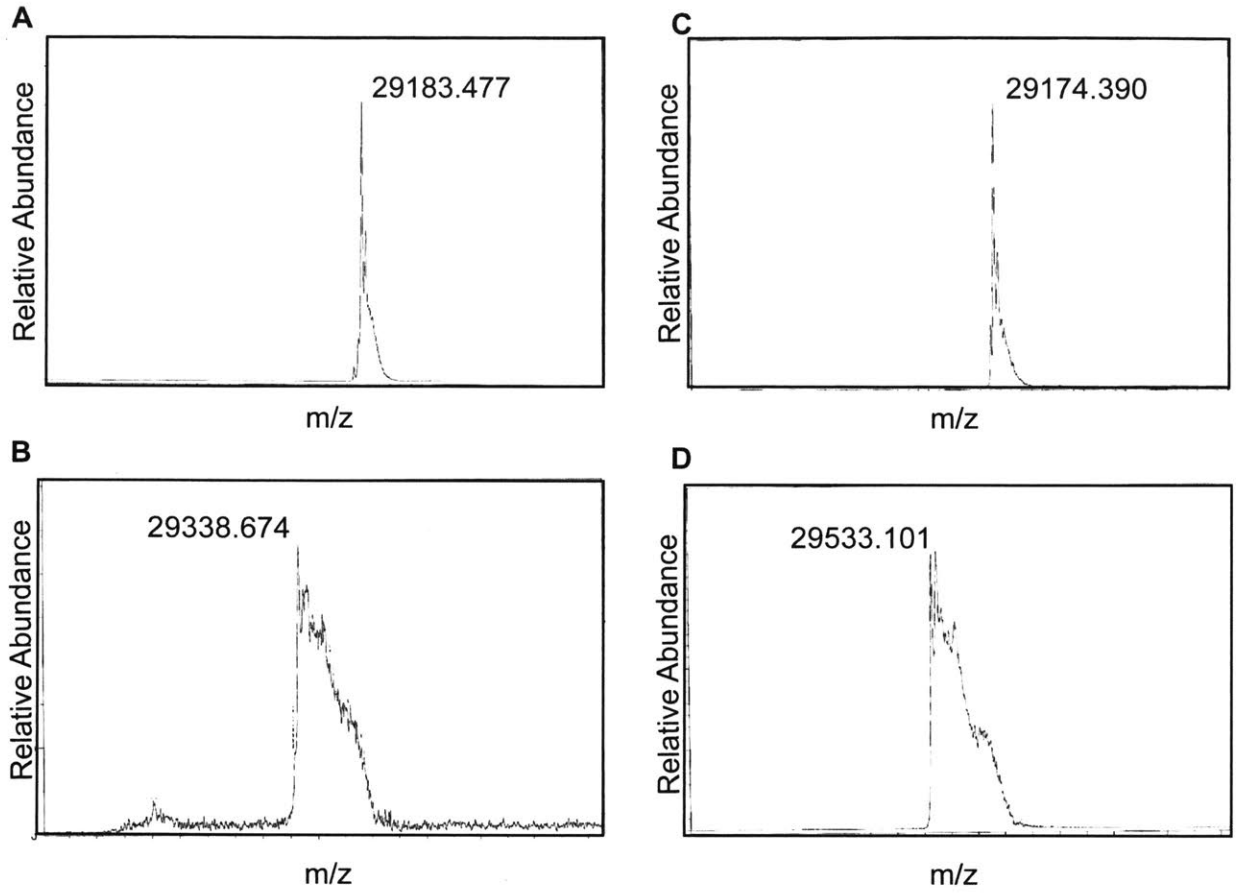


Figure 2.8 | MALDI-TOF analysis of the C-Pro variants

Each C-Proα(I) was analyzed by MALDI-TOF to determine the exact mass of the reduced, monomer. The masses are within error of the calculated mass. The mass change from mutating a cysteine to a serine or vice versa (< 20 Da) are within the error of the instrument, and therefore the masses identified cannot be compared too closely. Mass to charge ratios are given for the most intense peak in each spectrum. **(A)** C-Proα1(I) **(B)** C-Proα2(I) **(C)** C2S C-Proα1(I) and **(D)** S2C C-Proα2(I).

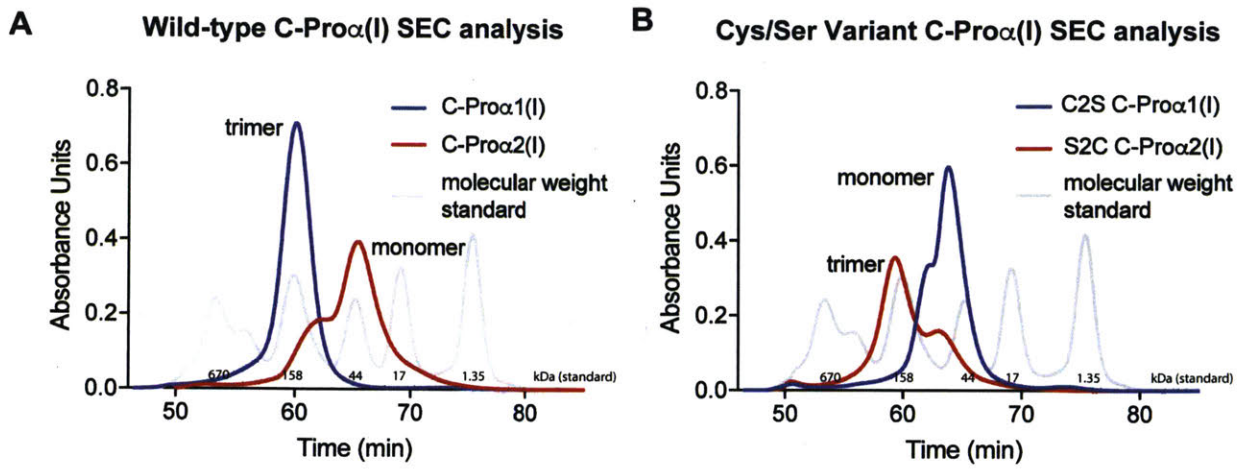


Figure 2.9 | Size exclusion chromatography demonstrates that constructs missing C2 cannot homotrimerize

(A) Size exclusion chromatography (SEC) analysis of purified wild-type C-Proα1(I) and C-Proα2(I) on a Superdex 200 10/30 GL column under native conditions revealed that C-Proα1(I) eluted as a homotrimer (blue), while C-Proα2(I) eluted predominantly as a monomer with some dimer also present (red). A molecular weight standard (grey) is shown for reference. **(B)** SEC analysis of purified C2S C-Proα1(I) and S2C C-Proα2(I) on a Superdex 200 10/30 GL column under native conditions revealed that C2S C-Proα1(I) eluted predominantly as a monomer with some dimer also present (blue), while S2C C-Proα2(I) eluted predominantly as a homotrimer (red). A molecular weight standard (grey) is shown for reference.

2.3.5 Sedimentation equilibrium analyses of C-Pro domain assembly

We next obtained sedimentation equilibrium data to more quantitatively evaluate assembly propensities under native conditions for each purified C-Pro protein sample. Data were analyzed using an approach similar to that described by Laue,[23] in which various models containing one or two species were fit to all the data collected for a given variant. To minimize the impact of errors in computed values of $\bar{v}$ (the partial specific volume of the particle) on the fits, each species was treated as a reduced molecular weight:

$$M_R = M_P(1 - \bar{v}\rho),$$

where M_P is the weight average molecular weight, which for a single species would be the polypeptide weight, or a multiple of it if a single oligomer. A single species model has one such term, a two species model has two such fitting terms which may or may not have an *a priori* relationship between the species (e.g., one species is a dimer of the other); the two species model can be elaborated to include an equilibrium among species. The slope of a plot of the logarithm of the absorbance as a function of the squared distance from the center of rotation is proportional to M_R and other measurement constants; when plotted as $\frac{2RT}{\omega^2}\ln(A)$ versus r^2, the slope is M_R and is directly related to the molecular weight. **Figure 2.10** provides complete sedimentation analysis of all four protein variants showing representative data and fits normalized for speed and temperature such that the slopes are directly proportional to the molecular weight.

Both wild-type C-Proα1(I) and S2C C-Proα2(I) were best fit as a single species with no evidence for a second species present in solution. From the reduced molecular weights and the $\bar{v}$ values from **Table 2.1**, the calculated molecular weights are 88,500 Da and 90,900 Da, respectively. These values are most consistent with both species being homotrimers, with molecular weights approximately 2000 Da larger than the sequence mass, consistent with some

degree of glycosylation. All the C-Pro constructs employed had a single N-glycosylation site in each polypeptide. Examples of fits to the observed gradients are shown in **Figure 2.11**.

For C2S C-Proα1(I) and wild-type C-Proα2(I), over the range of concentrations employed, both protein samples could be fit as a single species, but with molecular weights intermediate between a monomer and dimer of the respective polypeptide chains ($\frac{M_W}{M_S}$ = 1.597 and 1.38, respectively). As the MALDI-TOF data (**Figure 2.8**) do not show evidence of such a high molecular weight, the data are best viewed as an equilibrium between monomer and dimer. For C2S C-Proα1(I), an equilibrium constant of 76,300 M^{-1} provided the best fit, while for wild-type C-Proα2(I) an equilibrium between a species with $\frac{M_W}{M_S}$ = 1.17 and an equilibrium constant of 6500 M^{-1} provided the best fit. The small size of these equilibrium constants is consistent with the mixtures approximating a single species of intermediate molecular weight over the full concentration range employed. Examples of these fits are shown in **Figure 2.12** with the relative contributions of the monomeric and dimeric species to the overall gradients also computed and shown, highlighting that the monomer is the predominant species in both cases.

These observations are fully consistent with the size exclusion chromatography results in **Figure 2.9**. Thus, removal of a single cysteine residue in the conserved inter-strand disulfide bond-forming region of the collagen-I C-Pro does not simply result in a situation where disulfide linkages can no longer form between monomers. Instead, the proteins are not able to stably homotrimerize to any significant degree. We conclude that the presence or absence of C2 in the C-Pro domain is the critical nexus regulating the innate disulfide-linked trimerization propensity of collagen-I.

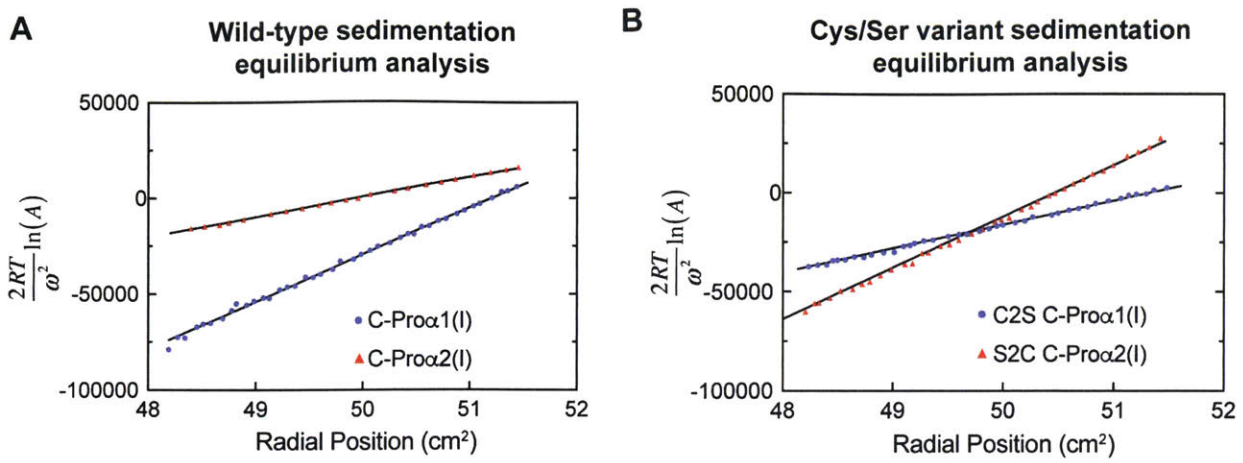


Figure 2.10 | Analytical ultracentrifugation confirms the presence of trimers only in the constructs with C2 intact

(A) Sedimentation equilibrium analysis showing that wild-type C-Proα1(I) (blue circles) is best fit as a single species homotrimer (see also **Figure 2.11**), while wild-type C-Proα2(I) (red triangles) is best fit as a monomer-dimer equilibrium (primarily monomer; see also **Figure 2.12**) with an association constant sufficiently small that over the range of concentrations in the study the protein sample could be approximated as a single species with a molecular weight intermediate between that of monomer and dimer. Data and fits shown are normalized for speed and temperature such that the slopes are directly proportional to the molecular weights. Only every third data point is shown for clarity. **(B)** Sedimentation equilibrium analysis showed that C2S C-Proα1(I) (blue circles) is best fit as a monomer-dimer equilibrium (primarily monomer; see also **Figure 2.12**) with an association constant sufficiently small to be approximated as a single species, while S2C C-Proα2(I) (red triangles) is best fit as a single species homotrimer (see also **Figure 2.11**). Data and fits shown are normalized for speed and temperature such that the slopes are directly proportional to the molecular weights. Only every third data point is shown for clarity.

Table 2.1. Molecular properties (calculated and measured) of collagen-I C-Pro variants studied.

	wild-type C-Proα1(I)	Cys1265Ser C-Proα1(I)	wild-type C-Proα2(I)	Ser1169Cys C-Proα2(I)
residues	251	251	252	252
M_S (kDa)	27,909	27,893	28,225	28,241
$\bar{\nu}$ (mL/g)	0.723	0.723	0.728	0.728
ε (M^{-1} cm^{-1})	39,420	39,420	43,890	43,890
MALDI-TOF (Da)	29,183	29,174	29,371	29,476

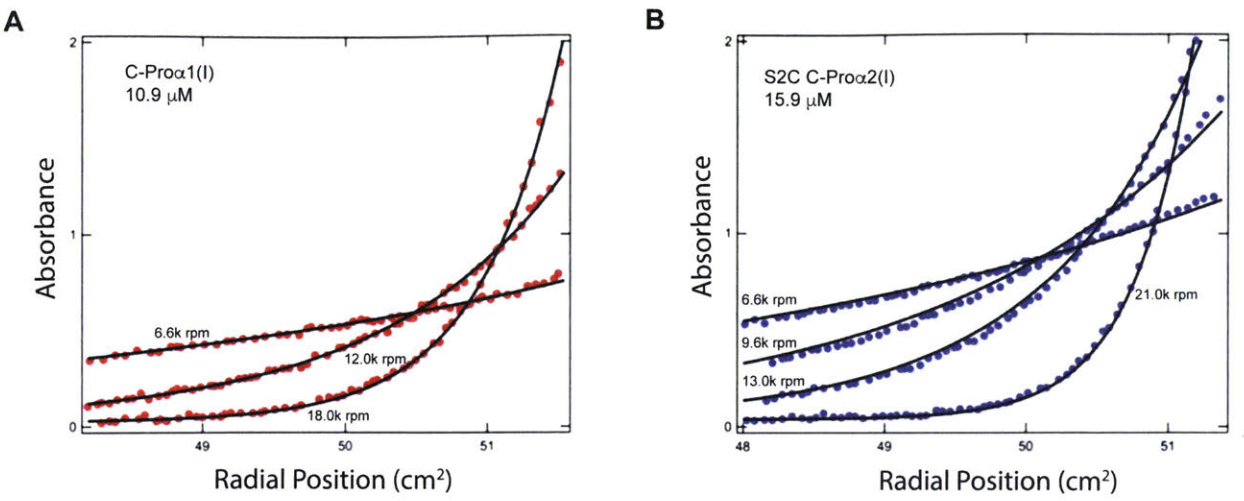


Figure 2.11 | Example global fits of the data collected for each homotrimerizing C-Pro variant during sedimentation equilibrium analyses

(A-B) Single species fits for wild-type C-Proα1(I) and S2C C-Proα2(I) at a single initial concentration and multiple speeds. The fitted single species molecular weights based on a global fit to all the data available for each variant were 88,500 Da and 90,900 Da, respectively (showing only every 2nd data point for clarity).

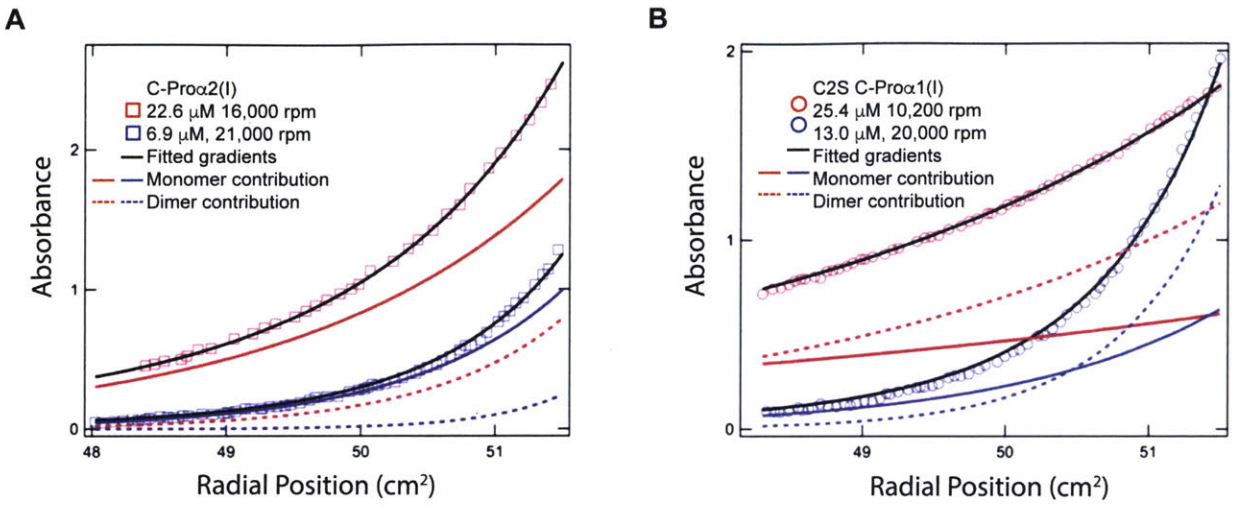


Figure 2.12 | Wild-type Colα2(I) and C2S Colα1(I) are predominantly monomeric, with only minor contributions from a dimeric species

(A-B) Fitted results for two of the speeds and initial concentrations used in the global fitting to a simple monomer-dimer equilibrium model for C2S C-Proα1(I) and wild-type C-Proα2(I). The fitted gradient superimposes on the measured data (showing only every 2nd data point for clarity). The contribution of the monomer and dimer species to the total gradient are shown as well, based on fitted equilibrium constants of 76,300 M^{-1} and 6500 M^{-1}, respectively. These distributions could be replaced by a single gradient with a molecular weight intermediate to the monomer and dimer.

2.3.6 Disulfide mapping of the homotrimerizing collagen-I C-Pro domains

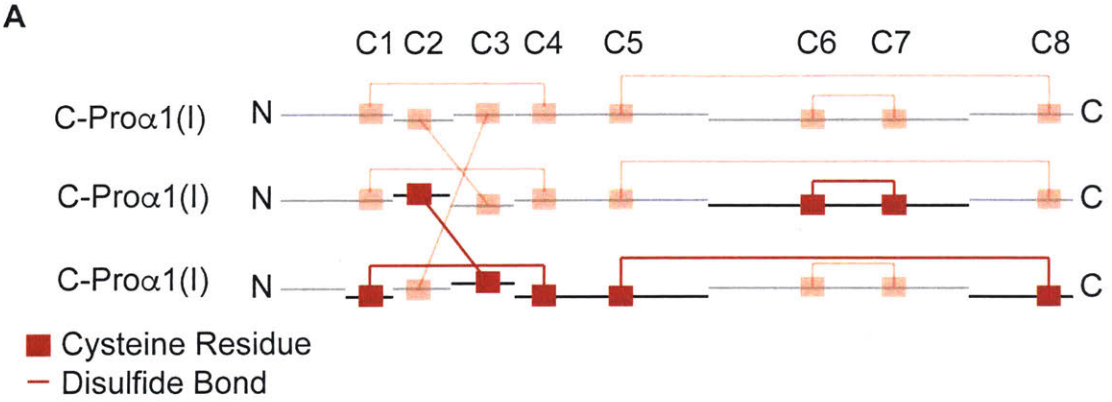
We next asked whether the homotrimeric S2C C-Proα2(I) domain is properly folded. In the absence of a crystal structure to assess foldedness and compare to the other homotrimeric C-Pro domain structures, we used mass spectrometry to interrogate the disulfide bonds present in the homotrimer. The recently published crystal structure (**Figure 2.2A**) provides the disulfide bonding pattern for C-Proα1(I) homotrimer **(Figure 2.2B)**. Thus, using the Colα1(I) crystal structure as a guide, we expected the S2C C-Proα2(I) C-Pro domain to show the same pattern of disulfide bonds.

We first trypsinized homotrimeric C-Proα1(I) protein and analyzed the digested peptides by mass spectrometry. A schematic of the expected disulfide-linked tryptic peptides is given in **Figure 2.13A**. Digesting with trypsin does not separate C4 from C5 due to a lack of arginine or lysine between the two cysteine residues. Therefore, we anticipated finding the mass corresponding to the sum the peptides containing C1, C4C5, and C8. After manually searching for the expected masses, we identified the C1-C4C5-C8 as well as the C2-C3 linked peptides. We were unable to detect the C6-C7 linked peptides, owing to the presence of the N-linked glycan on the peptide containing C6. To detect the C6-C7 disulfide linked peptide we digested the sample with PNGase F to remove the N-glycan prior to trypsinization. Indeed, when we removed the glycan we detected all of the expected disulfide bonded peptides (**Figure 2.13B** shows only data for PNGase F treated C-Proα1(I)). These data are consistent with the C-Proα1(I) homotrimer being properly folded, as expected, and suggest that our mass spectrometry-based disulfide mapping strategy can help to confirm that the S2C C-Proα2(I) is also properly folded.

Using the same protocol, we digested the S2C C-Proα2(I) homotrimer with trypsin and manually searched for the expected disulfides based on the confirmed Colα1(I) disulfide map. As with C-Proα1(I), C4 and C5 are not separated by a tryptic digest. Additionally for the C-

Proα2(I), C2 and C3 are also maintained on the same peptide when digested with trypsin. A schematic of the S2C C-Proα2(I) homotrimer digested with trypsin is given in **Figure 2.14A**. We identified masses corresponding to C1-C4C5-C8 and C2-C3 disulfide linkages, matching our expectation based on the Colα1(I) crystal structure (**Figure 2.14B**). These data suggest that the S2C C-Proα2(I) homotrimer is properly folded, and forms the same network of disulfide bonds as the wild-type C-Proα1(I) homotrimer.

A



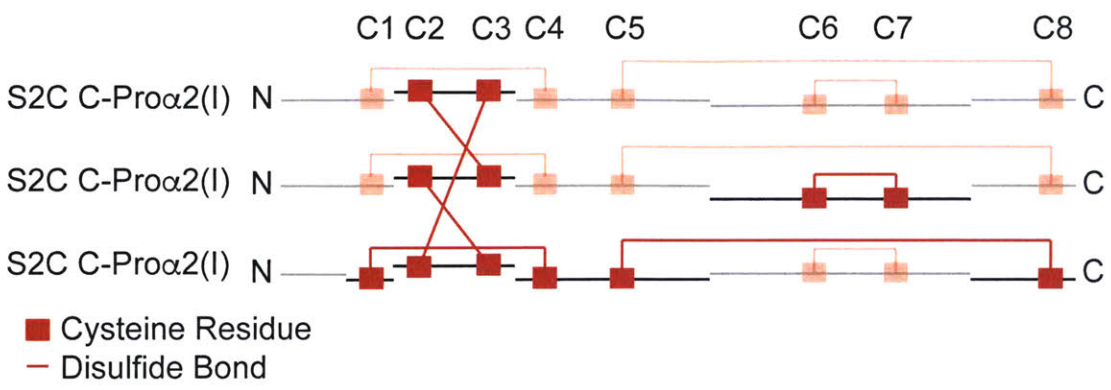
B

Disulfide linkage	Expected m/z	Observed m/z
C1-C4C5-C8	1150.9508	1151.1593
C2-C3	976.4274	976.0990
C6-C7	748.9442	748.9440

Figure 2.13 | C-Proα1(I) disulfide bonded peptides identified by mass spectrometry

(A) Expected disulfide bonding map based on tryptic digestion of C-Proα1(I), based on the published crystal structure. For simplicity, one of each type of disulfide bonded peptide is in bold, while the rest of the structure is lighter in color. **(B)** Expected and observed mass/charge ratio for disulfide bonded tryptic peptides.

A



B

Disulfide linkage	Expected m/z	Observed m/z
C1-C4C5-C8	1227.0876	1226.8440
C2-C3	1249.3996	1249.4039

Figure 2.14 | S2C C-Proα2(I) disulfide bonded peptides identified by mass spectrometry

(A) Expected disulfide bonding map based on tryptic digestion of S2C C-Proα2(I), based on the published crystal structure. For simplicity, one of each type of disulfide bonded peptide is in bold, while the rest of the structure is lighter in color. **(B)** Expected and observed mass/charge ratio for disulfide bonded tryptic peptides.

2.3.7 C2 and C3 are both essential for homotrimer formation

Our work described above outlines a detailed understanding of the essential role that C2 plays in homotrimerization. Based on the published crystal structures (see **Figure 2.2A**), we also expected C3 to participate in the interstrand disulfide bond, covalently linking the monomers together into a trimer, regardless of the collagen C-Pro domain in question (a schematic representation of the disulfide bonding is provided in **Figure 2.2B**). As above, we mutated C3S in both C-Proα1(I) and S2C C-Proα2(I) and monitored each protein's ability to homotrimerize by non-reducing SDS-PAGE. Both variants migrated as monomers under non-reducing conditions, perfectly mirroring the Colα1(I) and Colα2(I) variants with a serine in the place of the second cysteine (**Figure 15**). Paired together, these data highlight the importance of both positions 2 and 3 being occupied by Cys residues for a given strand to be able to form a disulfide-linked homotrimer.

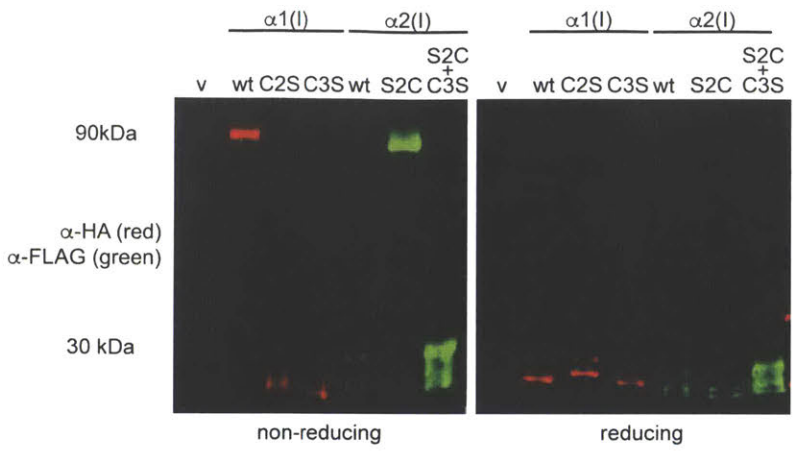


Figure 2.15 | C2 and C3 are required for trimerization

Immunoblot analysis of C2S and C3S variants of Colα1(I) and Colα2(I) under non-reducing and reducing conditions. Under non-reducing conditions, only C-Pro variants that contain both C2 and C3 cysteine residues can homotrimerize.

2.3.8 A generalizable rule for all fibrillar collagen C-Pro domains

The observation that a single amino acid substitution can swap the oligomerization propensities of the Colα1(I) and Colα2(I) C-Pro domains raises the provocative possibility that this simple cysteine-based code for collagen assembly could be generalizable beyond just collagen-I. Strikingly, alignment of the C-Pro domains of all of the human fibrillar collagens reveals that collagen strands known to form homotrimers consistently have all four cysteine residues in the interstrand disulfide-bonding region (**Figure 2.16A**). In contrast, those that are only known to form heterotrimers always lack one of the interstrand disulfide bonding cysteine residues, C2 or C3.

To test the generality of the cysteine code, we created HA-tagged constructs for expression of wild-type C-Pro domains for all of the fibrillar collagens and FLAG-tagged constructs for expression of appropriate variants. We observed that all of the wild-type fibrillar C-Pro domains that lack a single cysteine residue migrated as monomers on SDS-PAGE gels (**Figure 2.16B**), just as was observed for wild-type C-Proα2(I) (**Figure 2.6**). Moreover, re-introduction of the missing cysteine residue always conferred the ability to form disulfide-linked homotrimers. Similarly, all of the wild-type fibrillar C-Pro domains that naturally retain all four cysteine residues and are known to form homotrimers migrated as disulfide-linked homotrimers on SDS-PAGE gels (**Figure 2.16C**). Notably, for all of these C-Pro domains, we were able to identify a single cysteine residue whose conversion to a serine eliminates the ability to form disulfide-linked homotrimers. Finally, we observed that co-expression of a monomeric, C2S form of collagen-III's C-Pro domain could be rescued into a heterotrimer by co-expression with the wild-type C-Proα(III) containing all four cysteine residues (**Figure 2.17**). Thus, the cysteine-based molecular code for collagen trimerization appears to be conserved across a diverse suite of collagen types. We note that non-fibrillar collagens also rely on additional factors.

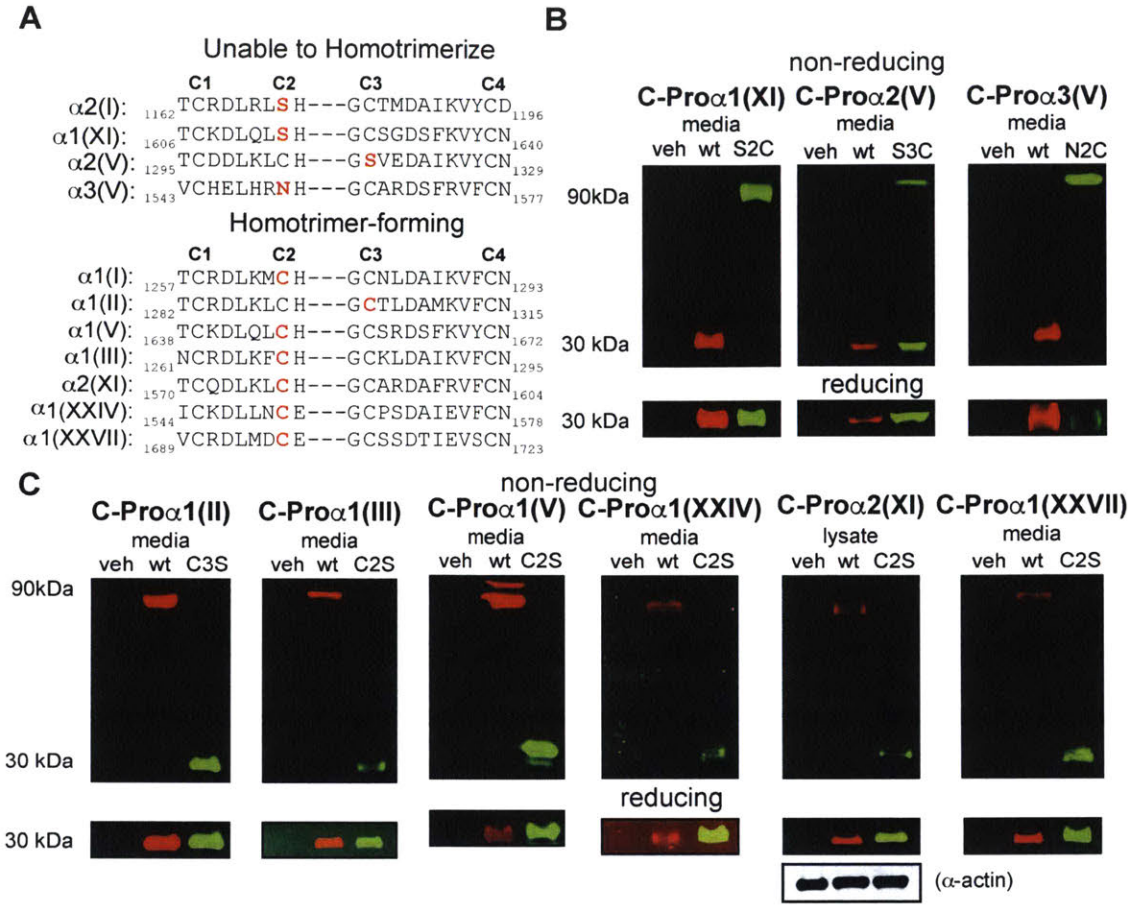


Figure 2.16 | The cysteine-based code for collagen C-Pro assembly is generalizable across all of the fibrillar collagens

(A) Alignment of the interstrand disulfide-bonding region of all of the human fibrillar collagen C-propeptides highlights that all C-Pro domains known to homotrimerize contain all four conserved cysteine residues, whereas all C-Pro domains known to only form heterotrimers lack a single cysteine. The residue colored red in each protein sequence corresponds to the mutated residue analyzed in **Figures 2.16B**, **2.16C**, and **2.17**. below. Amino acid numbering was derived from the corresponding full-length collagens. **(B)** Immunoblot analysis of individually expressed wild-type fibrillar collagen C-Pro domains known to only form heterotrimers. Assembly was analyzed under non-reducing and reducing conditions. Wild-type variants (HA-tagged; red) all migrated as monomers, while all variants in which the fourth cysteine residue was re-introduced (FLAG-

tagged; green) gained the ability to homotrimerize in a disulfide-dependent manner. **(C)** Immunoblot analysis of individually expressed wild-type fibrillar collagen C-Pro domains known to have the capacity to homotrimerize. Assembly was analyzed under non-reducing and reducing conditions. Wild-type variants (HA-tagged; red) all migrated as disulfide-dependent homotrimers, while all variants in which a single cysteine residue was mutated to Ser (FLAG-tagged; green) lost the capacity to form disulfide-linked homotrimers.

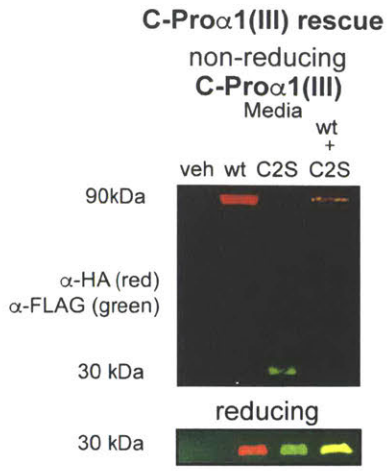


Figure 2.17 | Wild-type Colα1(III) can rescue C2S Colα1(III) into a heterotrimer

The wild-type collagen-III C-Pro domain homotrimerizes in a disulfide-dependent manner, but the C2S variant does not (see also **Figure 2.16C**). Immunoblot analysis of co-expressed wild-type C-Proα1(III) and C2S C-Proα1(III) revealed that the wild-type protein rescued the Ser variant into a disulfide-linked heterotrimer, mimicking the results obtained for C-Proα(I) in **Figure 2.6**.

2.3.9 Forming 1:1:1 Homotrimers

Both our mutational data and the disulfide mapping experiments indicate the importance of C2 and C3 for a given strand to form a homotrimer. Thus, we would expect the same residues to be involved in disulfide bonds in a heterotrimer. In collagen-I, we would hypothesize that C2 and C3 would form disulfides on adjacent chains, leaving one of the C3 residues of Colα1(I) free, due to the uneven number of cysteine residues in the molecule (**Figure 2.18A** consistent with the modeling of Colα2(I) into the Colα1(I) homotrimer).

We next considered the possibility of a 1:1:1 heterotrimer. If a set of collagen strands were to form a 1:1:1 heterotrimer, the cysteine requirements would be slightly different. It would require that one strand have cysteine residues in position 2 and 3, one strand with a cysteine in position 2, and the third with a cysteine in position 3 (shown schematically in **Figure 2.18B**). Looking to nature, this patter occurs in type-I collagen of both rainbow trout and zebrafish (**Figure 2.18C**). Gene expression analysis demonstrated that the α3(I) strand is expressed in discrete locations within each organism, suggesting it may play a distinct role in trimer formation from the other two strands.[24,25] Moreover, this same pattern also occurs in humans in the case of another fibrillar collagen, type-V, which is also known to form a 1:1:1 heterotrimer. In summary, it appears likely that the cysteine code for collagen trimerization extends to the formation of 1:1:1 heterotrimers.

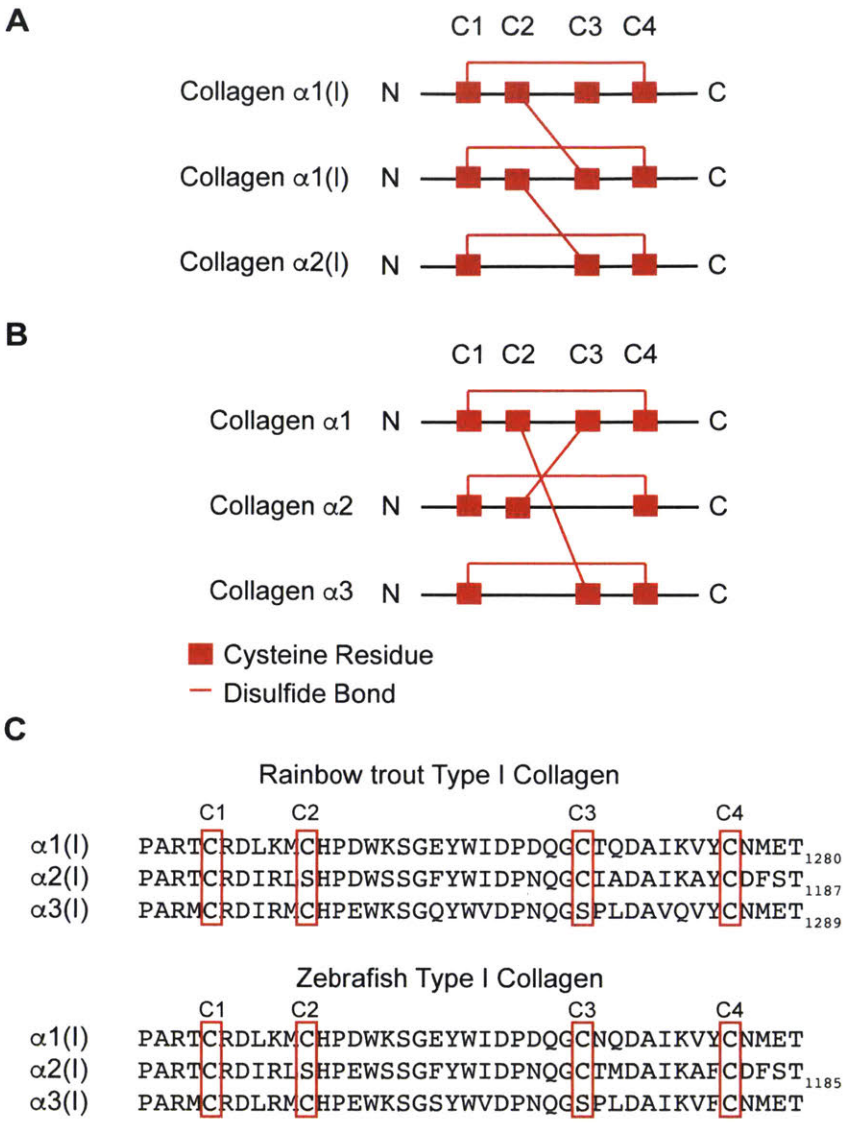


Figure 2.18 | Forming a 1:1:1 homotrimer requires that one strand contain cysteines in the C2 and C3 position, while the other two strands must contain one cysteine each

(A) Proposed disulfide-bonding network of a collagen heterotrimer that requires 2:1 assembly. This disulfide bonding network was also proposed by modeling C-Proα2(I) into the homotrimeric crystal structure of Colα1(I). **(B)** Proposed disulfide-bonding network of a collagen heterotrimer that requires 1:1:1 assembly. **(C)** Alignment of rainbow trout and zebrafish collagen-I C-Pro domains (C1-C4) only shown. The expected cysteine network for a 1:1:1 homotrimer in **Figure 2.18B** is present in both instances where collagen-I has three alpha strands in nature.

2.4 Concluding Remarks

Collectively, the work presented in this chapter takes a significant step towards defining the molecular code for collagen C-Pro assembly. Nature evolved different collagen strands that can homotrimerize, and others that can only heterotrimerize. The code for homo- versus heterotrimerization resides in the cysteine network of the C-Pro domain. Fibrillar collagen C-Pro domains require cysteine residues in both positions 2 and 3 to form disulfide-linked homotrimeric structures. The presence of these cysteine residues is both necessary and sufficient for collagen homotrimerization, in contrast to results suggested by prior work.

Our data are consistent with the following model for homotrimer assembly: fibrillar collagen C-Pro constructs of the same type of can transiently dimerize and trimerize intracellularly. Stable trimer formation, however, requires the ability to form disulfide bonds between C2 and C3 of neighboring strands to lock-in the structure for subsequent triple-helix formation.

Our model is not in conflict with a role for the CRS in determining type-specific collagen assembly. Further, the amino acids identified by Sharma and co-workers[8] as important for incorporating Colα2(I) into a heterotrimer are also consistent with our data, but we conclude they are not the critical residues as disulfide-linked homotrimers of Colα2(I) can be formed even in their presence. The next step is to understand why homotrimers of Colα1(I) are not more commonly observed in biology, as the cysteine network is intact and homotrimerization should be possible. Transient trimerization with Colα2(I) may be more prevalent, or localized expression of the strands in cells may play an important role. This question is the subject of ongoing investigation in our laboratory.

2.5 Experimental Methods and Supplies

2.5.1 Materials and Reagents

DMEM, 0.25% trypsin/EDTA solution, L-glutamine, penicillin/streptomycin were purchased from Corning/Cellgro. Serum-free media for HEK293-Freestyle cells was purchased from Life Technologies. Antibodies used were obtained from the following suppliers: Santa Cruz: HA probe (sc-7392); Agilent Technologies: Rat Anti-DYKDDDDK (200474); and Sigma: actin (A1978). Secondary antibodies were obtained from LiCor Biosciences: 800CW Goat Anti-Mouse, 800CW Goat Anti-Rat, 680LT Goat Anti-Mouse, and 680LT Goat Anti-Rat. All media and cell culture reagents were obtained from Corning/Cellgro. DNA/RNA preparation kits were obtained from Omega BioTech. *Col1A1* and *Col1A2* genes were obtained from the Origene True Clone Repository, Accession Numbers SC112997 and SC126717, respectively. All other fibrillar C-Pro domain sequences were designed and purchased from Genewiz. Lipofectamine 3000 was purchased from Life Technologies. Proteomics-grade dithiothreitol (DTT) and iodoacetamide was purchased from VWR. Sequencing-grade trypsin was purchased from Promega. PNGase F was purchased from New England BioLabs.

2.5.2 Plasmids

The C-Pro-encoding sequences (from the endogenous C-proteinase cleavage site to the C-terminus of the protein) were PCR-amplified using primers to incorporate the NotI and EcoRV sites and inserted into pcDNA3.1 vectors encoding the preprotrypsin signal sequence upstream of an HA or FLAG epitope tag, respectively. Variants were created by site-directed mutagenesis using the QuikChange XL II Kit from Agilent Technologies. For purification of recombinant collagen-I C-Pro domains, these vectors were modified to incorporate a cleavable 6×-His tag fused to the N-terminus of each C-Pro domain via an HRV-3C protease cleavage site such that both the 6×-His and HA/FLAG tags could be cleaved by a protease during protein purification. This cleavage only left a two-amino acid scar (Gly-Pro) on the recombinant protein.) Genes

encoding C-Pro domains for the other human fibrillar collagens were cloned into the same pcDNA3.1 vectors downstream of a preprotrypsin signal sequence and either an HA or FLAG epitope tag. The C-Pro gene for *Col5A2* was inserted using NotI and XbaI sites. All other collagen C-Pro-encoding genes were inserted between the NotI and EcoRV sites. FLAG-tagged type-II, -III, -V, -XI, -XXIV, and -XXVII genes were further modified by site-directed mutagenesis, as required.

2.5.3 Cell culture

Adherent HEK293 cells were cultured in DMEM supplemented with L-glutamine, penicillin/streptomycin, and 10% fetal bovine serum. HEK293 Freestyle suspension cells were cultured in Freestyle media. Transient transfections of pcDNA3.1 C-Pro domain-encoding plasmids were performed using Lipofectamine 3000 (Thermo Fisher Scientific). For all experiments not involving protein purification, transfection media was changed to fresh DMEM 24 h post-transfection. Media and lysates were harvested from cell culture 24–72 h post-transfection for analysis. When cell lysate samples were required, cells were harvested and then lysed on ice for 10 min in lysis buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM magnesium chloride, 1% Triton X-100, and protease inhibitor tablet (Thermo Fisher Scientific).

2.5.4 Immunoblotting

Prior to SDS-PAGE analyses, media samples were treated with 100 mM iodoacetamide in the dark for 1–2 h to prevent disulfide shuffling, or 100 mM dithiothrietol for 1 h at rt to reduce disulfides. All samples were then treated with 6× gel loading buffer (300 mM Tris, pH 6.8, 15% glycerol, 6% SDS and 10% (w/v) bromophenol blue) and boiled for 10 min prior to protein gel electrophoresis. Samples were then separated by SDS-PAGE using 12% polyacrylamide gels

and analyzed by immunoblotting on nitrocellulose membranes using a Licor Odyssey Scanner for detection.

2.5.5 Collagen C-Pro expression and purification

HEK293 Freestyle cells were transfected with appropriate plasmids using 293Fectin, according to the manufacturer's protocol (Thermo Fisher Scientific). Cells were expanded, as necessary, and the supernatant was harvested 6 d post-transfection after pelleting the cells by centrifugation. Clarified media was supplemented with 50 mM potassium phosphate at pH 7, 5 mM imidazole, 150 mM sodium chloride, and 10 mM Tris (final concentrations). The protein was bound to Ni-NTA resin by gravity flow, washed with four column volumes of 300 mM sodium chloride, 10 mM imidazole, and 150 mM potassium phosphate at pH 7, and then eluted with 300 mM sodium chloride, 250 mM imidazole, 150 mM potassium phosphate at pH 7, into a final EDTA concentration of 30 mM. Samples were eluted into EDTA to prevent free nickel ions from causing the protein to precipitate. The eluted proteins were then concentrated using a 10 kDa MWCO filter, buffer-exchanged into 20 mM *bis*-Tris Propane at pH 7,150 mM sodium chloride, and subjected to o/n cleavage with 6×His-tagged HRV-3C protease (Pierce). The cleaved protein mixture was applied to a new Ni-NTA column, such that the cleaved His epitope and the His-protease bound to the beads while the tag-less, purified protein was collected. The sample was then concentrated and further purified by a Superdex 200 10/30 GL column. Samples were separated using an isocratic flow 20 mM *bis*-Tris Propane at pH 7,150 mM sodium chloride at 0.5 mL/min. Fractions containing protein as indicated by the absorbance reading at 280 nm were collected and used for further analysis.

2.5.6 Dynamic light scattering analyses

Purified C-Pro domain proteins were analyzed using a DynaPro NanoStar from Wyatt Technology. Briefly, concentrated samples were inserted into disposable plastic cuvettes and

equilibrated to rt. Data were collected using a light source wavelength of 658 nm and a fixed scattering angle of 90°.

2.5.7 Analytical ultracentrifugation

All data were collected at 25 °C with gradients monitored at 275 nm in a Beckman XLA analytical ultracentrifuge. 1.2 cm pathlength double-sector, charcoal-filled Epon centerpieces were used for all samples with ~100 μL on the sample side and ~110 μL of buffer or water as the reference. In most instances, three samples of a construct were spun, diluted as required. Equilibrium data used in the analyses were collected at four or more speeds for each sample with equilibrium ascertained as superimposable gradients collected ≥ 3 h apart. A non-sedimenting contribution was measured after high-speed depletion of protein. This component was generally minor (< 0.04 au) and treated as a fixed, measured parameter in the data analyses. After attaining equilibrium at the highest speed, the speed was typically reduced to one of the earlier speeds and the sample allowed to re-equilibrate. In all tested cases, the gradients were nearly identical, suggesting no irreversible loss of material during the experiment.

For the variants, the sequence weights (M_S) and the partial specific volumes ($\bar{v}$) were computed from the sequences using tabulated values.[26] While the polypeptides were shown to be glycosylated, the extent and specific nature of the glycosylation was unknown and likely heterogeneous. However, MALDI-TOF mass spectra yielded masses (shown in **Table 2.1** and spectra shown in **Figure 2.8**) only slightly higher than the sequence weights. Therefore, the potential impact of glycosylation on $\bar{v}$ values was ignored. The buffer used for sedimentation equilibrium experiments was 20 mM *bis*-Tris propane at pH 7.0 with 150 mM sodium chloride. Buffer density was measured at 25 °C by an Anton Paar DMA 5000 to be 1.005 g/mL. The extinction coefficients at 280 nm were calculated based on average values for *W* and *Y*, ignoring potential contributions from disulfides.[27] In equilibrium constant calculations, this

extinction coefficient was used without adjustment for the 275 nm wavelength used in data collection. These molecular properties are summarized in **Table 2.1**.

3.5.8 Disulfide bond mapping by mass spectrometry

Protein samples were purified from HEK293 Freestyle cells as described above. The collagen-I C-Pro domain is N-glycosylated and the glycosylation site is located on one of the cysteine-containing tryptic fragments. Therefore, samples were treated with PNGase F overnight at 37° C, to remove the glycan (subsequently requiring us to search for an asparagine to an aspartic acid mutation: deamination). Samples were then applied to a 50 MWCO filter and spun, to remove the PNGase F enzyme. 1 μg of sequencing-grade trypsin (Promega) was added to each sample, and digested over night at 37° C. Peptides were injected onto an EASY-nLC 1000 LC-Q-Exactive Orbitrap. Peptides were separated on a 15 cm PepMap analytical column, over a 3 h gradient starting with 100% Buffer A / 0% Buffer B (Buffer A was 0.1% formic acid, Buffer B was 80% acetonitrile, 0.1% formic acid), and ending with 0% Buffer A / 100% Buffer B.

2.5.9 Proteomic searching and analysis

Oxidized peptides are not searchable using traditional methods of comparing to a database. Due to the possible variability in the disulfide linkages, the searches quickly become unwieldy for Proteome Discoverer to handle. Therefore, all disulfide-bonded peptides were searched for by hand in the raw LC-MS data file. Precursor masses of each peptide were calculated and summed accordingly, in addition to multiple charge states for each peptide. Samples were then manually searched for peptides with the expected masses.

2.6 References

(1) Ricard-Blum, S. The collagen family. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, a004978.

(2) Mouw, J. K.; Ou, G.; Weaver, V. M. Extracellular matrix assembly: a multiscale deconstruction. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 771-785.

(3) Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

(4) Boudko, S. P.; Engel, J.; Bachinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 21-32.

(5) Boudko, S. P.; Zientek, K. D.; Vance, J.; Hacker, J. L.; Engel, J.; Bachinger, H. P. The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J. Biol. Chem.* **2010**, *285*, 23721-23731.

(6) 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.* **1994**, *269*, 24354-24360.

(7) Lees, J. F.; Tasab, M.; Bulleid, N. J. Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *EMBO J.* **1997**, *16*, 908-916.

(8) Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R. N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671.

(9) Bourhis, J. M.; Mariano, N.; Zhao, Y.; Harlos, K.; Exposito, J. Y.; Jones, E. Y.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1031-1036.

(10) Bourhis, J. M.; Mariano, N.; Zhao, Y.; Walter, T. S.; El Omari, K.; Delolme, F.; Moali, C.; Hulmes, D. J.; Aghajari, N. Production and crystallization of the C-propeptide trimer from human procollagen III. *Acta. Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2012**, *68*, 1209-1213.

(11) Chiang, C. H.; Horng, J. C. Cation-pi Interaction Induced Folding of AAB-Type Collagen Heterotrimers. *J. Phys. Chem. B.* **2016**, *120*, 1205-1211.

(12) Jalan, A. A.; Hartgerink, J. D. Simultaneous control of composition and register of an AAB-type collagen heterotrimer. *Biomacromolecules* **2013**, *14*, 179-185.

(13) O'Leary, L. E.; Fallas, J. A.; Hartgerink, J. D. Positive and negative design leads to compositional control in AAB collagen heterotrimers. *J. Am. Chem. Soc.* **2011**, *133*, 5432-5443.

(14) Di Lullo, G. A.; Sweeney, S. M.; Korkko, J.; Ala-Kokko, L.; San Antonio, J. D. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.* **2002**, *277*, 4223-4231.

(15) Myllyharju, J.; Lamberg, A.; Notbohm, H.; Fietzek, P. P.; Pihlajaniemi, T.; Kivirikko, K. I. Expression of wild-type and modified proalpha chains of human type I procollagen in insect cells leads to the formation of stable [alpha1(I)]2alpha2(I) collagen heterotrimers and [alpha1(I)]3 homotrimers but not [alpha2(I)]3 homotrimers. *J. Biol. Chem.* **1997**, *272*, 21824-21830.

(16) Makareeva, E.; Han, S.; Vera, J. C.; Sackett, D. L.; Holmbeck, K.; Phillips, C. L.; Visse, R.; Nagase, H.; Leikin, S. Carcinomas contain a matrix metalloproteinase-resistant isoform of type I collagen exerting selective support to invasion. *Cancer Res.* **2010**, *70*, 4366-4374.

(17) Jimenez, S. A.; Bashey, R. I.; Benditt, M.; Yankowski, R. Identification of collagen alpha1(I) trimer in embryonic chick tendons and calvaria. *Biochem. Biophys. Res. Commun.* **1977**, *78*, 1354-1361.

(18) Narayanan, A. S.; Page, R. C.; Meyers, D. F. Characterization of collagens of diseased human gingiva. *Biochemistry* **1980**, *19*, 5037-5043.

(19) Zhang, X.; Boot-Handford, R. P.; Huxley-Jones, J.; Forse, L. N.; Mould, A. P.; Robertson, D. L.; Lili; Athiyal, M.; Sarras, M. P., Jr. The collagens of hydra provide insight into the evolution of metazoan extracellular matrices. *J. Biol. Chem.* **2007**, *282*, 6792-6802.

(20) Wada, H., Okuyama, M., Satoh, N., Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol. Dev.* **2006**, *8*, 370-377.

(21) Exposito, J. Y.; Cluzel, C.; Garrone, R.; Lethias, C. Evolution of collagens. *Anat. Rec.* **2002**, *268*, 302-316.

(22) Boot-Handford, R. P.; Tuckwell, D. S. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays* **2003**, *25*, 142-151.

(23) Laue, T. M. Sedimentation equilibrium as thermodynamic tool. *Methods Enzmol.* **1995**, *259*, 427-452.

(24) Gistelinck, C.; Gioia, R.; Gagliardi, A.; Tonelli, F.; Marchese, L.; Bianchi, L.; Landi, C.; Bini, L.; Huysseune, A.; Witten, P. E.; Staes, A.; Gevaert, K.; De Rocker, N.; Menten, B.; Malfait, F.; Leikin, S.; Carra, S.; Tenni, R.; Rossi, A.; De Paepe, A.; Coucke, P.; Willaert, A.; Forlino, A. Zebrafish Collagen Type I: Molecular and Biochemical Characterization of the Major Structural Protein in Bone and Skin. *Sci. Rep.* **2016**, *6*, 21540.

(25) Saito, M., Takenouchi, Y., Kunisaki, N., Kimura, S. Complete primary structure of rainbow trout type I collagen consisting of a1(I)a2(I)a3(I) heterotrimers. *Eur. J. Biochem.* **2001**, *268*, 2817-2827.

(26) Reynolds, J. A.; McCaslin, D. R. Determination of protein molecular weight in complexes with detergent without knowledge of binding. *Methods Enzymol.* **1985**, *117*, 41-53.

(27) Pace, N., Schmid, F.X. In *Protein Structure: A Practical Approach*; 2nd ed.; Creighton, T. E., Ed.; IRL Press: Oxford, 1997, p 253-259.

(28) Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **2010**, *38*, W695-699.

(29) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.

Chapter 3:
Mapping and Exploring the Collagen-I Proteostasis Network

This chapter is adapted from the following manuscript:
DiChiara, A.S., Taylor, R.J., Wong, M.Y., Doan, D.N., Del Rosario, A.M., Shoulders, M.D. *ACS. Chem. Biol.* **11**, 1408–1421 (2016).

3.1 Author Contributions

A.S.D. performed the majority of the experiments, analyses, and drafted the manuscript, all with assistance from his undergraduate mentee R.J.T. M.Y.W. performed the pulse chase experiment and analysis in **Figure 3.2** and the adenoviral expression of OI-variant collagen in **Figure 3.3**. D.N.D. performed the confocal imaging in **Figure 3.2** and generated the parent Saos-2-TRex cell line used in **Figures 3.9–3.13**. A.M.D. assisted with the mass spectrometry data acquisition and analysis.

3.2 Introduction

Collagen-I is the predominant proteinaceous component of human tissues, including skin and bone.[1] The supramolecular properties of the collagen-I-based extracellular matrix are determined, in large part, by the upstream, intracellular processes of collagen-I folding, modification, and quality control. These processes are governed by the endoplasmic reticulum's (ER's) proteostasis network.[2] Imperfections in collagen-I structure, most often caused by missense mutations in collagen-I genes, lead to debilitating diseases known as the collagenopathies, including osteogenesis imperfecta and Ehlers-Danlos Syndrome.[3,4] A key underlying problem in these typically autosomal dominant disorders is the failure of the ER's proteostasis network to properly fold and/or subject misfolding collagen-I variants to quality control.[5] Understanding the mechanisms of collagen-I proteostasis in the ER is therefore essential not just for our basic understanding of collagen biogenesis, but also for the long-term development of therapies for the collagenopathies.

The process of folding and secreting collagen-I is highly complex. Collagen-I triple helices are 2:1 heterotrimers of collagen-α1(I):collagen-α2(I) strands. Both this stoichiometry and the overall triple-helix register are governed by nucleation at the globular disulfide-linked C-terminal propeptide domain[6] (see Chapter 2 for progress towards understanding this process).

Consequently, nascent >1200 amino acid collagen-I polypeptide strands must remain in a monomeric, unfolded form until their C-termini enter the ER, fold into native disulfide-bonded globular states, and properly associate. Extensive co- and post-translational modifications, including glycosylation and hydroxylation, must also be completed prior to triple-helix folding.[7] During triple-helix formation, peptidyl prolyl isomerases (PPIases) act to ensure the hundreds of prolyl peptide bonds are in the trans configuration required in folded triple helices.[8]

In contrast to the marked complexity of collagen-I production, our knowledge of the proteostasis mechanisms that assist the process remains incomplete. Many of the known players were discovered as a result of either their relative abundance in collagen-producing cells (e.g., HSP47)[9] or via the identification of a genetic defect that causes a collagenopathy (e.g., CRTAP, FKBP10).[10,11] Despite these seminal discoveries, many questions still remain. For example, although the N-linked glycan is highly conserved across the various types of fibrillar collagen as well as across species, its function and the roles of the extensive lectin-based ER chaperone and quality control machineries in collagen-I folding are poorly understood. Additionally, proteins responsible for surveying nascent collagen-I structure and directing misfolding or aggregating strands to clearance or unfolding/refolding mechanisms remain unknown.[12-14] Beyond recognition of a misfolded triple helix, the method by which collagen is degraded has not been fully elucidated.

A detailed, well-validated map of the collagen-I proteostasis network is essential to begin to answer these and many other questions regarding collagen-I biogenesis. Unfortunately, appropriate reagents and cell culture model systems for molecular biology, genetics, and proteomics experiments on full-length collagen-I are either unavailable or unwieldy. A further complication is the lack of immunoprecipitation/mass spectrometry (IP/MS)-grade collagen-I antibodies, precluding systematic, proteomics-based characterization of the collagen-I proteostasis network.

Herein, I describe the development and application of an HT-1080 human cell-based platform for collagen-I studies that overcomes these challenges. Extensive characterization of the collagen-I produced confirms that native collagen-I folding and modification are faithfully recapitulated in our platform. We illustrate the value of the platform by performing quantitative, comparative proteomics characterization of the collagen-I interactome under biologically relevant experimental conditions. Biochemical studies validate the roles of identified novel interactome components in collagen-I production, and lead to the discovery of a novel collagen-I post-translational modification. Further mechanistic work delineates a role for the prominent, but largely unannotated, endoplasmic reticulum protein 29 (Erp29) in collagen-I quality control. Altogether, these applications of our platform yield both new insights and a compelling roadmap for continued studies of the mechanisms of production of this important extracellular matrix protein that is intimately associated with both normal and diseased states.

3.3 Results

3.3.1 Vector Construction

We sought to create a robust and experimentally flexible platform for biochemical studies of collagen-I folding and misfolding that would allow us to (1) readily distinguish between distinctive procollagen-I (referred to throughout as "collagen-I") strands and (2) perform robust and reliable IPs of intracellular collagen-I to enable accurate mapping of its proteostasis network. We began by cloning both the *COL1A1* and *COL1A2* genes into pTRE-Tight vectors for doxycycline (dox)-dependent control of gene transcription. Because collagen-I folding begins at the C-terminus,[6] we incorporated a short (<10 amino acid) antibody epitope tag at the amino-terminus of collagen-I's N-propeptide domain. We rationalized that this location was unlikely to disrupt collagen-I folding or structure because it is the final domain to fold. We used the HA epitope and the FLAG epitope for wild-type collagen-α1(I) and collagen-α2(I), respectively, to facilitate individual strand isolation and identification (**Figure 3.1A**).

Vector construction using the GC-rich collagen-I genes proved exceptionally challenging. Each collagen-I gene has about 65% GC-content overall, and local regions of each gene contain 85% GC content within a 50 base-pair span. The high local GC-content made PCR amplification of the collagen-I genes difficult, resulting in yields lower than expected and many small PCR fragments, likely due to incomplete synthesis of the genes. The yield of the full-length collagen-I genes was slightly boosted by increasing the final extension time by about 1.25 fold. Furthermore, each collagen-I gene is over 4000 base pairs in length, more than double the size of the pTRE-Tight expression vector, requiring an altered molar ratio when ligating the two linear pieces of DNA. Even with modified ratios, the most common ligation product observed was the pTRE-Tight vector ends being ligated together, despite the use of

incompatible NotI and EcoRV restriction enzymes to digest the vector. Sequencing confirmed that, in a rare event, the sticky end of the NotI site was destroyed, leaving a blunt end of DNA that ligated to the EcoRV blunt end. This rare intramolecular ligation was more favorable and occurred more frequently than insertion of the collagen-I genes. To avoid this problem, de-phosphorylation of the vector backbone with Antarctic phosphatase prior to ligating the two pieces of DNA was essential for successful insertion of the collagen genes.

We ultimately succeeded in ligating pTRE-Tight and the collagen-I PCR amplicons, confirmed by sequencing the resulting clones with a primer that binds to the vector, providing evidence for the presence of both the 5′-and the 3′-ends of the collagen-I genes. Unfortunately, in the case of *Col1A1*, sequencing of the full gene confirmed that a recombination event occurred, effectively removing > 1000 base pairs of the gene. To circumvent the propensity to recombine the Col1A1 gene, we thereafter transformed all *Col1A1*-containing vectors into recombination deficient Sure2 *E.coli* cells. When grown at 30° C, recombination was effectively shut down in these cells, finally allowing for the synthesis and propagation of intact collagen-I genes.

3.3.2 Inducible Expression of Orthogonally Tagged Collagen-I Strands in Human Cells

We next identified an appropriate cell line for heterologous expression of these collagen-I genes. HT-1080 fibrosarcoma cells secrete non-fibrillar collagen-IV, suggesting that they are capable of properly handling collagen-I without producing confounding endogenous forms of the protein. Indeed, HT-1080 cells were previously shown to permit the expression and secretion of thermostable, hydroxylated collagen-α1(I) homotrimers,[15] although native collagen-α1(I):collagen-α2(I) heterotrimer-producing cells have not been reported. We transfected HT-1080 Tet-Off cells that constitutively express the tetracycline transactivator with our wild-type HA-collagen-α1(I)- or FLAG-collagen-α2(I)-encoding pTRE-Tight vectors, selected for stable incorporation of the genes, and isolated genetically homogenous single colonies to analyze for collagen-I expression and secretion. Expression of the tetracycline transactivator allows for doxycycline (dox)-inducible expression of the collagen-I genes of interest. We optimized HT1080 culture conditions by varying the concentration of dox for 48 h, and then split each concentration into two conditions: continued culture in the same concentration of dox, or no dox in tet-free FBS containing media. **Figure 3.1B** demonstrates that 1 ng/mL dox, 1000-fold lower than the recommended concentration, efficiently suppresses both collagen-I genes, and removal of the dox-containing media and replacement with dox-free media induces both genes, and the resultant proteins are secreted into the media.

As previously observed,[15] genetically homogeneous single colony cell lines inducibly expressing collagen-I proved difficult to obtain in the HT1080 cells. After screening over 24 colonies for expression of either Colα1(I) or Colα2(I), we were eventually able to identify unique cell lines that displayed significant lysate and secreted levels of either protein (**Figure 3.1C**). We note that the complex banding patterns we observe are common for collagen-I owing to extensive proteolytic processing and post-translational modificatoins (for example, see immunoblots of native collagen-α2(I) secreted from primary fibroblasts in **Figure 3.4B** below).

Native collagen-I is a heterotrimer of collagen-α1(I) and collagen-α2(I). Therefore, with cell lines inducibly expressing either wild-type collagen-α1(I) or collagen-α2(I) in hand, we next transfected the FLAG-collagen-α2(I) expressing cells with the corresponding inducible HA-collagen-α1(I) plasmid, and again selected single colonies. Ultimately, we identified a cell line, termed HT-1080$^{Col\text{-}I}$ cells, that inducibly expresses moderate levels of both collagen-α1(I) and collagen-α2(I) observable in the lysate and media using their respective antibody epitope tags (**Figure 3.2A**). We observe no apparent activation of unfolded protein response-regulated genes[16] in HT-1080$^{Col\text{-}I}$ cells upon collagen-I induction (**Figure 3.2B** with primers from **Table 3.1**), indicating that collagen-I expression at these levels in HT-1080 cells does not cause ER stress.

To ensure that the collagen-I is trafficking properly through HT-1080$^{Col\text{-}I}$ cells, we employed confocal microscopy. We observe strong co-localization of collagen-I with ER markers (R_{mean} = 0.45 ± 0.12), as well as partial co-localization with Golgi (R_{mean} = 0.22 ± 0.06) and lysosomal markers (R_{mean} = 0.19 ± 0.05) (**Figure 3.2C**) in the absence of ascorbate (ER-retention conditions), consistent with previous studies.[17] We assessed collagen-I secretion kinetics using metabolic labeling and found that the majority of the collagen-I is secreted in < 3 h with minimal degradation, recapitulating endogenous collagen-I secretion kinetics in primary fibroblasts[18] (**Figures 3.2D** and **3.2E**) and further confirming that the HT-1080 cells are properly handling our collagen-I constructs.

Although beyond the scope of the work outlined in this chapter, one of our goals is to map the proteostasis network for disease variants of collagen, and compare the results to the wild type counter part. The work outlined here is aimed at establishing the wild type proteostasis network, as a benchmark for future work on the disease variants. Mutations to collagen-I genes cause an autosomal dominant form of osteogenesis imperfecta, an incurable brittle bone disease. Therefore, a cell-based platform for studying OI requires the cell line to synthesize wild

type Colα1(I) and Colα2(I), in addition to the OI-variant of interest. We were able to express c-Myc tagged Cys1299Trp Colα1(I)[19] in the HT-1080^{Col-I} cells using replication incompetent adenovirus, thereby creating an autosomal dominant model system for studying OI with each collagen strand differentially tagged (**Figure 3.3**). This work is ongoing in the lab, and the remainder of this chapter will focus solely on the HT-1080^{Col-I} cells.

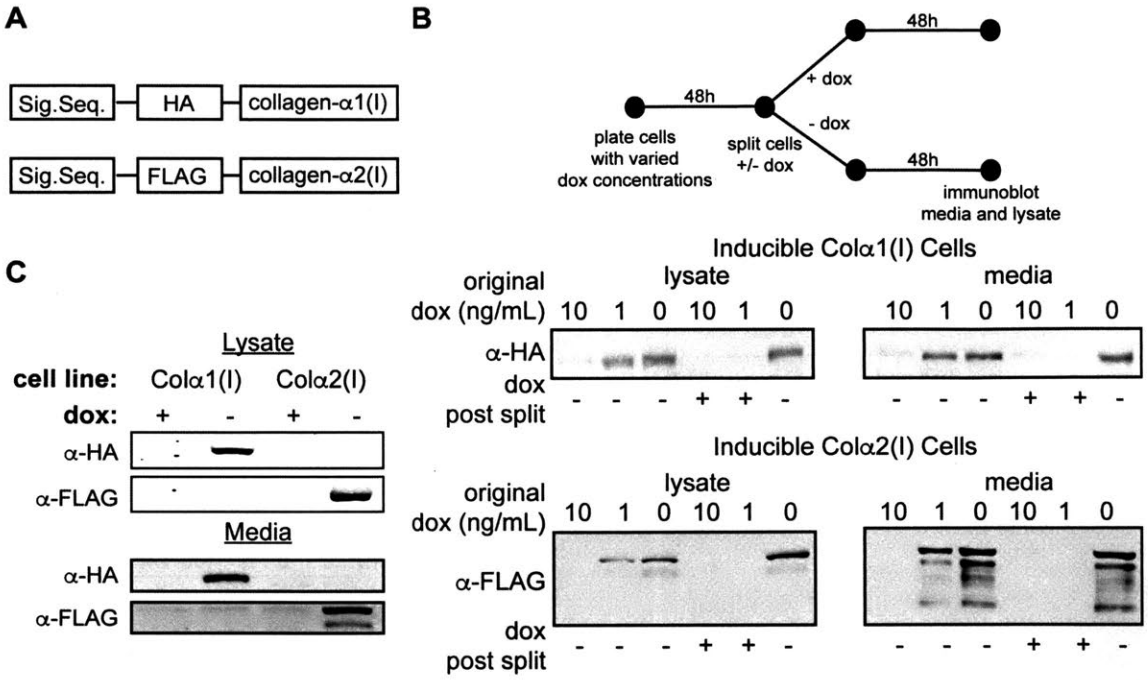


Figure 3.1 | Dox-inducibility of the collagen-I genes

(A) Schematic of collagen-I expression constructs. **(B)** Experimental timeline for dox treatments on collagen-I producing cells. Secretion analysis of Colα1(I) and Colα2(I) under varied dox concentrations, after two days of treatment. **(C)** Immunoblotting analysis of inducible HA-collagen-α1(I) or FLAG-collagen-α2(I) levels in the lysates and media of HT-1080 cells expressing either construct.

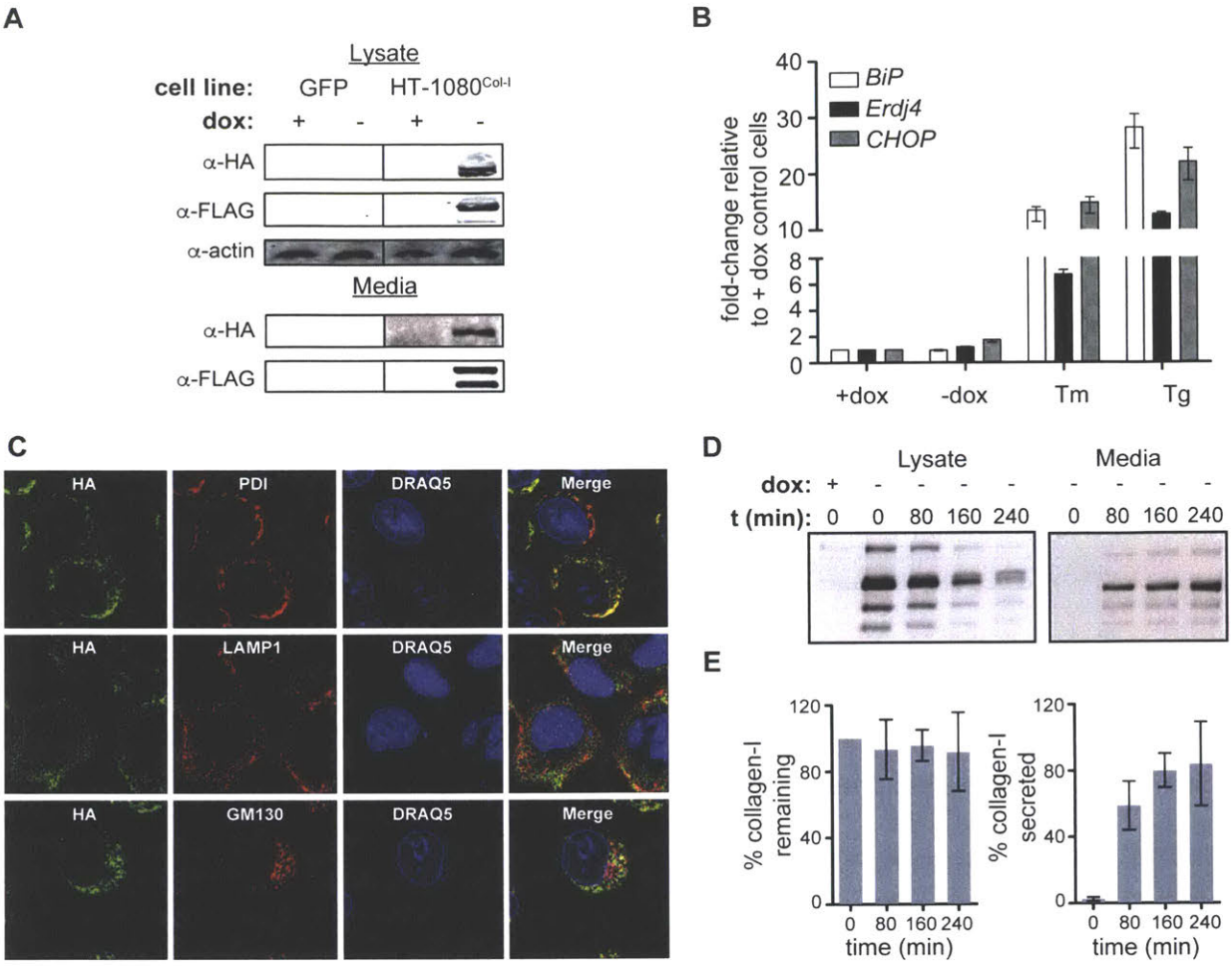


Figure 3.2 | Expression of Orthogonally Tagged Collagen-I Strands in HT-1080 Cells

(A) Immunoblotting analysis of inducible HA-collagen-α1(I) and FLAG-collagen-α2(I) levels in the lysate and media of HT-1080^{Col-I} cells expressing both constructs. **(B)** qPCR analysis of unfolded protein response-regulated genes in HT-1080^{Col-I} cells upon induction of collagen-I expression. Tunicamycin (Tm; 5 μg/mL, 12 h)- and thapsigargin (Tg; 1 μM, 12 h)-mediated unfolded protein response activation are shown as positive controls. qPCR data are reported relative to untreated HT-1080 cells as the mean ±95% confidence interval. **(C)** Confocal microscopy imaging of collagen-I trafficking in HT-1080^{Col-I} cells under ER-retention conditions. PDI is an ER marker, LAMP1 is an early lysosome marker, GM130 is a Golgi marker, and

DRAQ5 is a nuclear marker. **(D)** Representative autoradiograms of [^{35}S]-labeled HA-collagen-α1(I) immunoisolated from HT-1080$^{Col\text{-}I}$ media and lysates following induction of collagen-I expression. Control media and lysate were harvested from uninduced HT-1080$^{Col\text{-}I}$ cells. **(E)** Quantification of autoradiograms in **Figure 2.1D**. Collagen-I % remaining was calculated by normalizing the secreted and lysate collagen-I signals at the stated times to the total amount of labeled collagen-I observed at time = 0 h. Collagen-I % secreted was calculated by normalizing the secreted collagen-I signal to the total amount of collagen-I present at time = 0 h. Error bars represent SEM from biological replicates (n = 3).

Table 3.1. Compilation of primers for qPCR used in **Figure 3.2** and **Figure 3.10**.

Transcript	Forward	Reverse
RPLP2	5′-CCATTCAGCTCACTGATAACCTT-3′	5′-CGTCGCCTCCTACCTGCT-3′
COL1A1 (triple helix)	5′-TGGTAGCCGTGGTTTCCCTG-3′	5′-TCCAGTCAGACCCTTGGCAC-3′
COL1A2 (triple helix)	5′-TGGCTCGAGAGGTGAACGTG-3′	5′-AGCACCGTTGACTCCAGGAC-3′
COL1A1 (C-propeptide)	5′-GCAACAGCCGCTTCACCTAC-3′	5′-AGCCGAATTCCTGGTCTGGG-3′
COL1A2 (C-propeptide)	5′-TCGCTCAGCACCTTCTCTCAG-3′	5′-TGGGTGGCTGAGTCTCAAGTC-3′
CHOP	5′-GGAGCTGGAAGCCTGGTATG-3′	5′-GCCAGAGAAGCAGGGTCAAG-3′
BiP	5′-GCCTGTATTTCTAGACCTGCC-3′	5′-TTCATCTTGCCAGCCAGTTG-3′
Erdj4	5′-GGAAGGAGGAGCGCTAGGTC-3′	5′-ATCCTGCACCCTCCGACTAC-3′
ANXA2	5′-CCTTATCTGGCCACCTGGAG-3′	5′-GCTCCTGGTTGGTTCTGGAG-3′
ASPH	5′-GGTTCCTGTGGAGGCAGAAC-3′	5′-GGTTCTCCTGTGGGTCCATC-3′
CKAP4	5′-ACGTGGAGGAGCTGAAGAGG-3′	5′-AAGTCCTGAGGAGGCAGACG-3′
CRTAP	5′-TCCGGTTGAGAAATTTGTGG-3′	5′-GGTTCTGCTGCATGACCTTG-3′
DNAJB11	5′-GAACCCCTCGTCAGCAAGAC-3′	5′-TCTCTTGCCGACAATTGCAC-3′
ERO1L	5′-AAGAGGCCGTGTCCTTTCTG-3′	5′-TCCACTGCTCCAAGTCGTTC-3′
ERP29	5′-CTTCCCCTGGATACGGTCAC-3′	5′-CTGCCACCAAGAGATCATCG-3′
P4HB	5′-AAATCAAGCCCCACCTGATG-3′	5′-TATCCCAAATGGGAGCCAAC-3′
RCN1	5′-CCACTGGATCCTCCCTCAAG-3′	5′-CCCCGTAATTGGTAGCTTGG-3′
SERPINH1	5′-CAGCCTCATCATCCTCATGC-3′	5′-TTCTGCAGGTCATGGGTCAC-3′
SPARC	5′-AGAAGCTGCGGGTGAAGAAG-3′	5′-GGAGAGGTACCCGTCAATGG-3′
TGM2	5′-TGGCATGGTCAACTGCAACG-3′	5′-GCACTGGCCATACTTGACGC-3′
GOLIM4	5′-AACGAGAAGCAGCCAACCTC-3′	5′-GCAAAGCTTCCTGGTGTTCC-3′
CALU	5′-TTATGTGCCTGTCCCTGTGC-3′	5′-TTGCTTCTTCAGCACCCAAG-3′
PRKCSH	5′-CGACTGCAAAGATGGCTCTG-3′	5′-GCCGCTGTTGTACTCGTCTG-3′

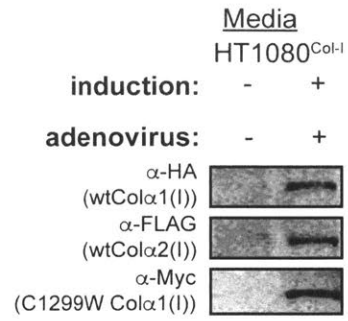


Figure 3.3 | Expression of Cys1299Trp in the HT1080^{Col-I} Cells

Cells were transduced with replication-incompetent human adenovirus-V carrying a constitutively expressed Colα1(I) Cys1299Trp full length collagen-I gene tagged with c-myc epitope tag. Each protein was differentially identified by the corresponding epitope tags. The results shown are immunoblot analysis of secreted media, demonstrating our ability to make an OI-like, autosomal dominant mimicking cell line.

3.3.3 Molecular Properties of Collagen-I Produced by HT-1080^{Col-I} Cells

We next asked whether the collagen-I produced by HT-1080^{Col-I} cells displays the expected molecular properties of endogenous collagen-I. We first tested for a direct intracellular interaction between the collagen-α1(I) and collagen-α2(I) polypeptides. We observe that an IP of collagen-α1(I) using HA-antibody beads co-IPs collagen-α2(I), and also that the reverse is true when using FLAG-antibody beads (**Figure 3.4A**). These results confirm the heteromeric assembly of collagen-α1(I) and collagen-α2(I) strands produced by HT-1080^{Col-I} cells.

The triple-helical domain of properly folded, extracellular collagen-I is resistant to proteases such as trypsin and chymotrypsin,[20] whereas poorly folded collagen-I is highly sensitive to proteolytic digestion. Indeed, we observe that collagen-I secreted from primary dermal fibroblasts that endogenously secrete the protein displays only a small shift in molecular weight upon protease treatment (**Figure 3.4B**), consistent with propeptide sensitivity and triple-helical resistance. Collagen-I secreted from HT-1080^{Col-I} cells displays an identical pattern upon protease digestion (**Figures 3.4B**), confirming that the collagen-I produced is folded into a stable triple helix.

Another key modification of collagen-I is N-glycosylation within the C-propeptide of collagen-I. Although the biological function of this conserved N-glycan remains unclear,[21] it is likely to enable interactions with the important lectin-based components of the ER proteostasis network. To test whether the collagen-I produced by our HT-1080^{Col-I} cells is N-glycosylated, we immunoprecipitated collagen-α1(I) and treated half of the eluate with PNGase-F, an endoglycosidase that hydrolytically cleaves N-glycans. The treated and untreated samples were separated by SDS-PAGE and the subsequent immunoblot was probed with concanavalin A (ConA), a lectin that recognizes N-glycans. In the absence of PNGase-F treatment, we observe a ConA signal that overlaps with the signal for collagen-α1(I), while treatment with PNGase-F eliminates ConA reactivity, confirming that the collagen-I produced by our HT-1080^{Col-I} cells is

indeed N-glycosylated (**Figure 3.4C**).

Native collagen-I undergoes other essential co- and post-translational modifications beyond just N-glycosylation during its folding and maturation in the ER.[7] Most important among these are the 4*R*-hydroxylation of Yaa-position Pro residues in the Gly-Xaa-Yaa repeats of the triple-helical domain,[1] 3*S*-hydroxylation of Xaa-position Pro residues,[22] and Lys hydroxylation.[7] To quantify the extent of collagen-I hydroxylation in HT-1080$^{Col\text{-}I}$ cells, we used the stable isotope labeling by amino acids in cell culture (SILAC) technique[23] to generate light, medium, and heavy Lys- and Arg-labeled HT-1080$^{Col\text{-}I}$ cells. Because ascorbate is an essential hydroxylase cofactor,[24] we induced collagen-I expression in the medium-labeled cells by removing dox without ascorbate (ER-retention conditions[25]) and in the heavy-labeled cells by removing dox with added ascorbate (secretion-promoting conditions), using the light cells as a control that expresses collagen-I lacking an HA tag. After IP of intracellular collagen-α1(I) using HA-antibody beads, MS analyses showed that any given hydroxylated collagen-I peptide is much more abundant in the ascorbate-treated sample, while non- or minimally-hydroxylated collagen-I peptides are relatively more abundant in the absence of ascorbate (**Figure 3.5A** presents representative data). As expected for an analysis of intracellular collagen-I, we observe a wide range of hydroxylation states for any given collagen-I peptide, ranging from no hydroxylation to apparently complete hydroxylation of Yaa-position Pro residues. None of the observed peptides ever display a larger number of hydroxylation events than should be possible based on the total number of Yaa-position Pro residues and Lys residues in a given peptide. Most notably, Pro986 in collagen-α1(I) and Pro707 in collagen-α2(I) are the only two Xaa-position Pro residues where 3*S*-hydroxylation is known to occur in native collagen-I.[22] We observe both of these hydroxylation events in our MS analyses and the hydroxylation is again ascorbate-dependent, as expected (**Figures 3.5B** and **3.5C** for MS2 scans). These results demonstrate that collagen-I produced by HT-1080$^{Col\text{-}I}$ cells displays the known hydroxylation

patterns of endogenous collagen-I.

Cumulatively, the data in **Figures 3.4** and **3.5** indicate that our HT-1080^{Col-I} cells produce heteromeric, proteolytically stable, and properly modified collagen-I, faithfully recapitulating key molecular properties of the native protein. These findings strongly motivate the application of HT-1080^{Col-I} cells as a convenient platform both for mapping the collagen-I proteostasis network and for more detailed mechanistic biochemistry studies.

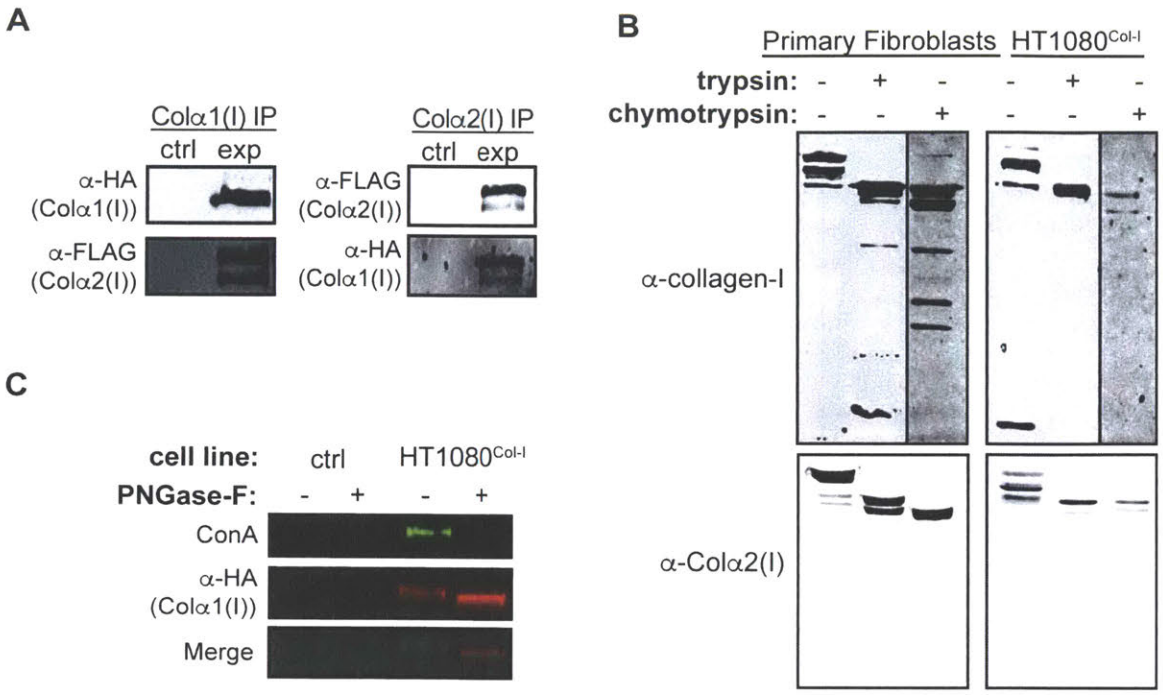
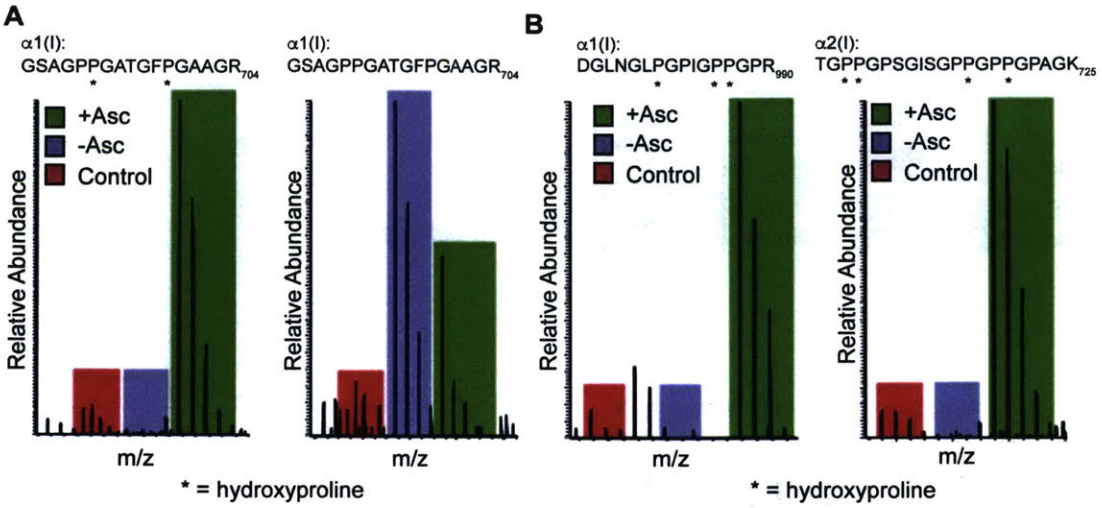


Figure 3.4 | Molecular Properties of Collagen-I Produced by HT-1080^{Col-I} Cells

(A) Collagen-α2(I) co-IPs with collagen-α1(I) and vice versa, demonstrating their intracellular association. The control sample represents HT-1080 cells that do not express HA/FLAG-tagged collagen-I upon induction. **(B)** Trypsin and chymotrypsin digests of collagen-I secreted from primary fibroblasts and HT-1080^{Col-I} cells demonstrate the presence of a stable, protease-resistant triple helix. **(C)** Analysis of the N-glycosylation of intracellular HA-tagged collagen-α1(I) immunoprecipitates. In the absence of PNGase-F treatment, HA-antibody reactivity overlaps with ConA reactivity, while treatment with PNGase-F eliminates ConA reactivity. The control sample represents HT-1080 cells that do not express HA-tagged collagen-I upon induction.


A
α1(I):
GSAGPPGATGFPGAAGR704
α1(I):
GSAGPPGATGFPGAAGR704
+Asc
-Asc
Control
Relative Abundance
m/z
* = hydroxyproline
B
α1(I):
DGLNGLPGPIGPPGPR990
α2(I):
TGPPGPSGISGPPGPPGPAGK725
+Asc
-Asc
Control
Relative Abundance
m/z
* = hydroxyproline

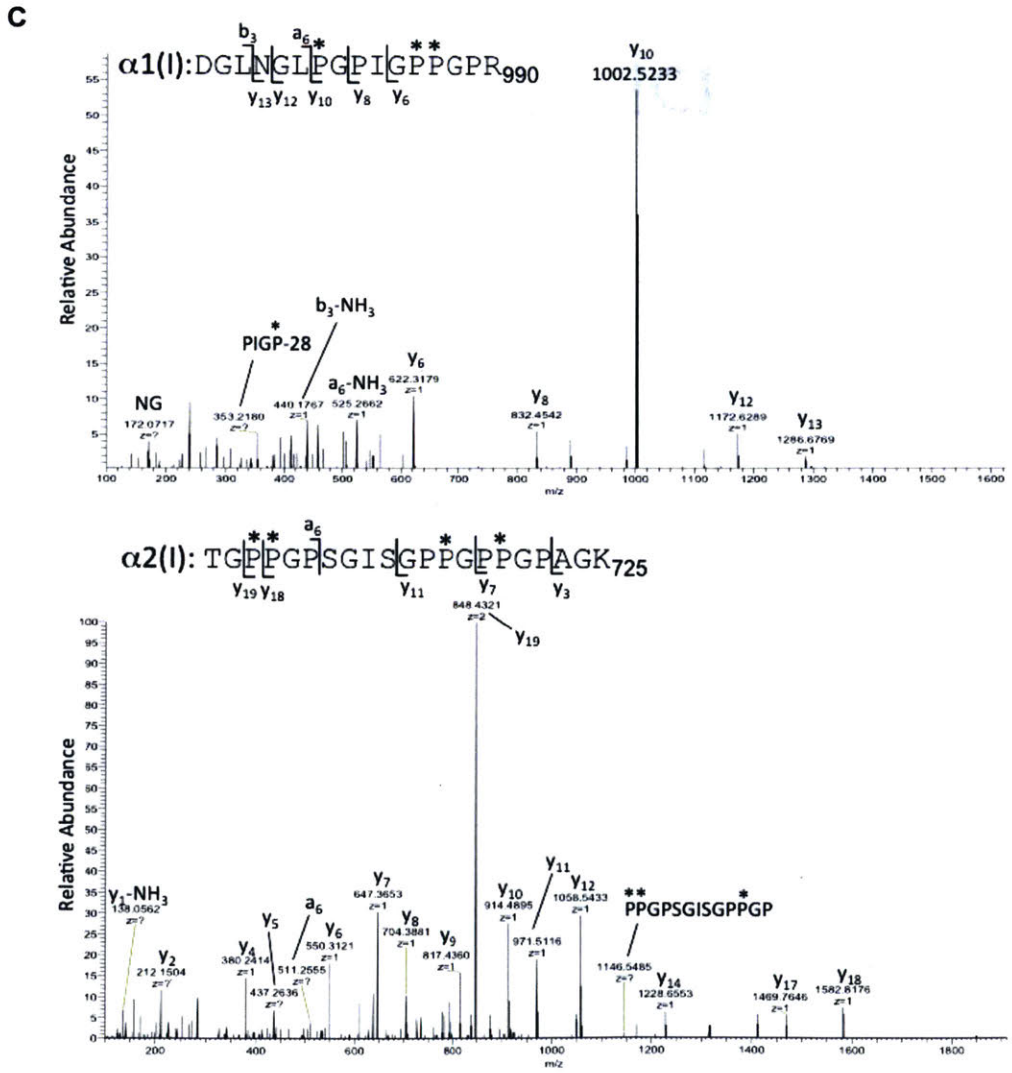

C
α1(I):DGLNGLPGPIGPPGPR990
b3
a6
y13 y12 y10 y8 y6
y10
1002.5233
Relative Abundance
b3-NH3
PIGP-28
y6
a6-NH3
NG
y8
y12
y13
m/z
α2(I): TGPPGPSGISGPPGPPGPAGK725
y19 y18
y11
y7
y3
848.4321
y19
y1-NH3
y2
y5
a6
y6
y4
y7
y8
y9
y10
y11
y12
PPGPSGISGPPGP
y14
y17
y18
m/z

Figure 3.5 | Hydroxylation of Collagen-I Produced by HT-1080$^{Col\text{-}I}$ Cells

(A) MS1 scans showing the enrichment of the hydroxylated peptide in the heavy (ascorbate-treated) sample, and enhanced abundance of the unmodified peptide in the medium (ascorbate-deficient) sample. **(B)** MS1 scans showing that Xaa-position Pro residues 986 in collagen-α1(I) and 707 in collagen-α2(I) (numbering beginning at the first Gly-Xaa-Yaa repeat) are hydroxylated in HT-1080$^{Col\text{-}I}$ cells. **(C)** Top Panel: MS2 data showing hydroxylation on proline 986 of collagen-α1(I) upon ascorbate treatment of collagen-I expression induced HT-1080$^{Col\text{-}I}$ cells. The internal fragment labeled as PIGP*-28 and y_6 together provide clear evidence for the Xaa-position Pro being hydroxylated. Bottom Panel: MS2 data showing hydroxylation on proline 707 of collagen-α2(I) upon ascorbate treatment of collagen-I expression induced HT-1080$^{Col\text{-}I}$ cells. The a_6 peak provides clear evidence for the Xaa-position Pro being hydroxylated. For both spectra, numbering of amino acids begin at the first Gly-Xaa-Yaa repeat. Asterisks (*) indicate the site of modification. In the case of prolyl modification, the * indicates hydroxylation. In the case of cysteine modification, the * indicates alkylation during the mass spectrometry workflow to prevent re-oxidation of cysteines. For labeling of spectra, the b-ion is termed a fragment of the parent ion, originating at the N-terminus of the parent peptide, and extending the number of amino acids into the peptide indicated. y-Ions are the same, but originating at the C-terminus of the peptide. An a-ion is a b-ion that has lost a carboxyl group during the fragmentation. For clarity, not all y and b ions are labeled on the peptide sequence, but all ions identified are labeled in each spectrum.

3.3.4 Covalent Crosslinking for Robust Co-Immunoprecipitation of the Collagen-I Proteostasis Network

We found that we can reliably IP collagen-α1(I) using HA-antibody beads or collagen-α2(I) using FLAG-antibody beads (**Figure 3.4A**). However, a pilot MS study of the co-immunoprecipitated interactome identified a very limited suite of interactors (**Figure 3.6A**). The failure to observe even well-established components of the collagen-I proteostasis network, including the lysyl hydroxylases and the PPIases, suggested to us that most proteostasis network interactions with collagen-I are too transient to be maintained during a traditional IP, as has also been observed for other ER proteostasis network client proteins.[26,27]

To overcome this obstacle, we immortalized interactions with collagen-I in live HT-1080$^{Col\text{-}I}$ cells by incubating with the cell-permeable, lysine-reactive, and reversible crosslinker dithiobis(succinimidyl propionate)[28] (DSP; **Figure 3.6B**). Collagen-α1(I) displays 57 lysine residues distributed along the length of the protein, so it is in principle possible to identify interactors that engage each collagen-I domain using this crosslinking strategy. We found that a 30 min treatment with 200 μM DSP followed by rapid quenching with Tris buffer is optimal to stably co-IP known, well-established collagen-I proteostasis network components (**Figures 3.6C** and **3.6D**) that are not otherwise observable via traditional IP workflows.

A

Protein	Experimental IP		Control IP	
	Total Peptides	*Unique Peptides*	*Total Peptides*	*Unique Peptides*
Colα1(I)	272	58	0	0
Colα2(I)	60	26	8	8
Prolyl-4-hydroxylase-α1	37	21	0	0
Protein disulfide isomerase	27	27	0	0
Prolyl-4-hydroxylase-α2	10	9	0	0
CRTAP	7	6	0	0
Procollagen galactosyltransferase 1	4	4	0	0

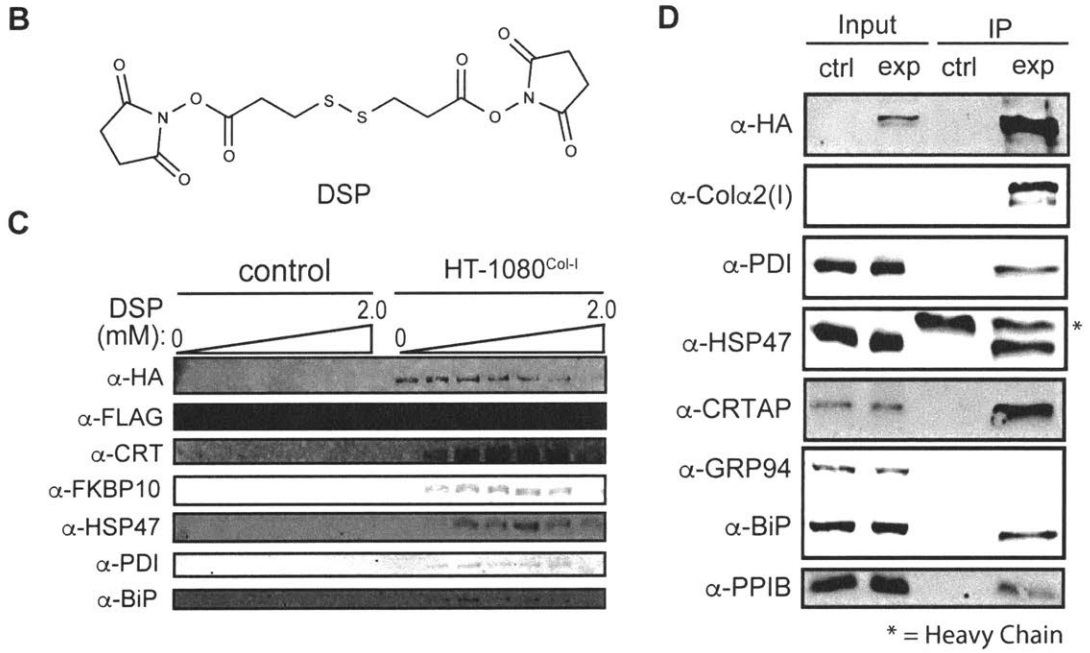


Figure 3.6 | Covalent Crosslinking to Enable Robust Co-Immunoprecipitation of the Collagen-I Proteostasis Network

(A) LC-MS/MS-mediated analysis of the stable collagen-I interactome in the absence of covalent crosslinker. The control sample represents HT-1080 cells that do not express HA-tagged collagen-I upon induction. **(B)** Structure of the covalent crosslinker employed, dithiobis(succinimidyl propionate) (DSP). **(C)** Optimization of crosslinking conditions for robust co-IP of the collagen-I interactome using HA-antibody beads. The control sample represents HT-1080 cells that do not express HA-tagged collagen-I upon induction. **(D)** Immunoblot showing the ability to co-IP an array of known components of the collagen-I proteostasis

network using the covalent crosslinking protocol, compared to HT-1080$^{Col\text{-}I}$ cells that produce collagen-I that is not HA-tagged.

3.3.5 Quantitative Proteomic Mapping of the Collagen-I Proteostasis Network

Having demonstrated that well-established collagen-I interactions can be identified by immunoblotting after covalent crosslinking, we shifted to a quantitative MS proteomics approach to enable unbiased mapping of the collagen-I proteostasis network. Using the light, medium, and heavy SILAC-labeled HT-1080$^{Col\text{-}I}$ cells described above, we induced collagen-I expression, performed covalent crosslinking with DSP, and then used HA-antibody beads to IP collagen-α1(I) and its interactors in biological triplicate (**Figure 3.7**). The light-labeled negative control HT-1080 cells inducibly expressed collagen-I lacking the HA epitope. The medium- and heavy-labeled HT-1080$^{Col\text{-}I}$ cells (both expressing HA-tagged collagen-α1(I)) were treated without or with ascorbate, creating ER-retention and secretion-promoting conditions, respectively. Ascorbate is an essential co-factor for the hydroxylases required for proper collagen-I biogenesis.[24] Therefore, we hypothesized that comparing the collagen-I interactome in the presence or absence of ascorbate would provide insight into proteostasis network components that differentially engage collagen-I under ER-retention versus secretion-promoting conditions. After elution from stringently washed antibody beads, samples were further processed and tryptic peptides were injected on a Thermo QExactive LC-MS/MS. Peptides were then identified and quantified using the Mascot Database and Proteome Discoverer.[29]

One salient feature of the MS data collected was the sheer abundance of the bait protein, HA-tagged Colα1(I) and its counterpart, Colα2(I). Collagen peptides accounted for 75–85% of the total peptides observed in a given run. While desirable for the bait protein to be the most abundant protein identified, thus indicating the efficiency of the IP, the size of collagen and total number of tryptic peptides generated by digestion dominated the signal detected by mass spectrometry, minimizing the signal from the interacting proteins. Therefore, prior to digestion, the IP-eluate was filtered through a 100-kDa molecular weight cutoff filter to remove the collagen-I protein. The solution that passed through the filter was then digested and analyzed by

MS. It is likely that filtering the elutions through a 100-kDa molecular weight cut off filter would preclude us from detecting interacting proteins larger than 100 kDa. We chose the risk of not detecting very high molecular weight interactors, as most proteostasis network components of the secretory pathway are smaller than 100 kDa.

Filtration did not completely extract collagen-I from the sample, as 60% of the peptides identified by this workflow in the heavy and medium samples correspond to either collagen-α1(I) or collagen-α2(I). The presence of collagen-I peptides in the flow through sample is likely due to the presence of reducing and denaturing agents as a method of elution from the beads. Although still larger than 100 kDa, the collagen-I monomers likely pass through the membrane more readily in a reduced and denatured state. However the increased depth of coverage proved sufficient to identify a more comprehensive collagen-I interactome. Further optimization of the elution and filtration technique could further illuminate less abundant collagen-I interactions.

Overall, a total of 171 proteins were identified across all samples, and 91 of those proteins were quantified with heavy:light and/or medium:light ratios. In **Table 3.2**, we present the results for all high-confidence interactors defined by meeting the following additional criteria: (1) ≥ 2 unique peptides identified in ≥ 2 biological replicates and (2) an average enrichment in either the heavy or medium sample ≥ 2-fold relative to the light control sample (or complete absence from the light sample). We categorize these high-confidence components of the collagen-I interactome based on their expected functions in the ER, grouping complexes where appropriate, and rank the proteins within each category based on their relative enrichment in the heavy-labeled (ascorbate-treated) sample.

The result is the first map of the collagen-I proteostasis network (**Table 3.2** and **Figure 3.8A**), comprising a total of 48 proteins and encompassing interactors with previously unknown, poorly characterized, and already well-characterized roles in collagen-I maturation. Importantly,

several features of the data support that our workflow identifies the *bona fide* collagen-I interactome. First, we note that >80% of the interactors are unequivocally localized to the secretory pathway (**Figure 3.8B**), where collagen-I is targeted for folding and modification. Although interactors that are not localized to the secretory pathway may be artifacts, considering the overall reliability of our data set (further verified below) it is also possible that annotations are incomplete or that these interactors engage collagen-I that has been transported from the ER for quality control purposes. Second, we are only identifying a small percentage of the known ER proteome,[30] indicating that we are maintaining high specificity for actual collagen-I interactors. Third, of the 48 proteins identified, approximately 30% already have well-characterized roles in collagen-I maturation. Indeed, we successfully identified all the most prominent known players in collagen-I proteostasis, particularly notable in the collagen-specific chaperones, PPIase, and collagen modification groups (**Table 3.2** and **Figure 3.8A**). Two well-established collagen-I chaperones, HSP47[9] and PPIB,[31] are the most highly enriched of all the proteins identified. We also identify FKBP10, an ER PPIase whose absence causes autosomal recessive osteogenesis imperfecta.[32] The robust enrichment of these known collagen-I proteostasis network components enhances our confidence in the entirety of the dataset. Of the 34 remaining high-confidence proteins in our interactome, 29 putative interactors have, to the best of our knowledge, ill-defined or previously unknown roles in collagen-I proteostasis (indicated by "*" in **Table 3.2**). The other five interactors have been implicated in collagen-I proteostasis-related functions, although mechanistic characterization remains incomplete.

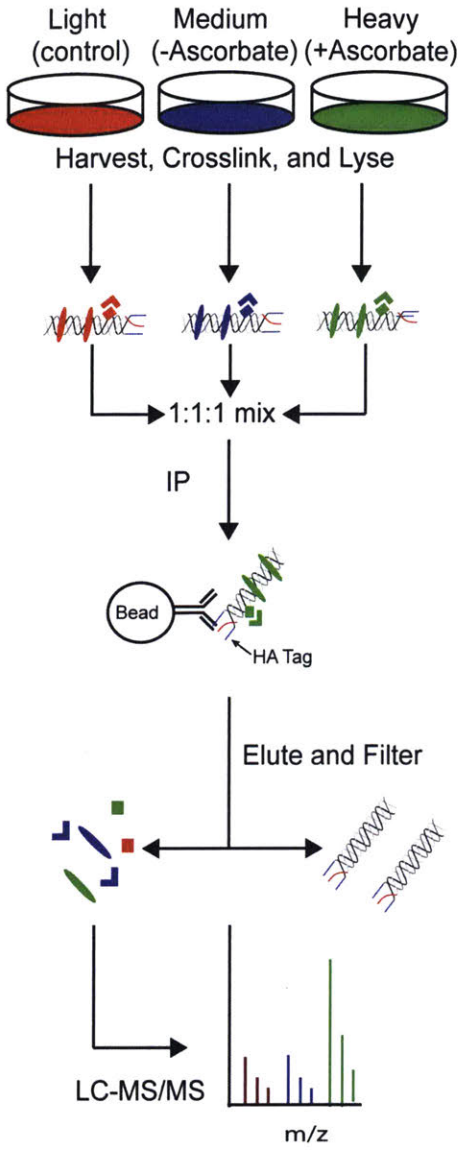


Figure 3.7 | Mass spectrometry workflow

(A) Schematic representation of mass spectrometry workflow employed. Collagen was eluted from the antibody beads, and applied to a 100 kDa cut off filter. The low molecular weight fraction was trypsinized and analyzed by mass spectrometry.

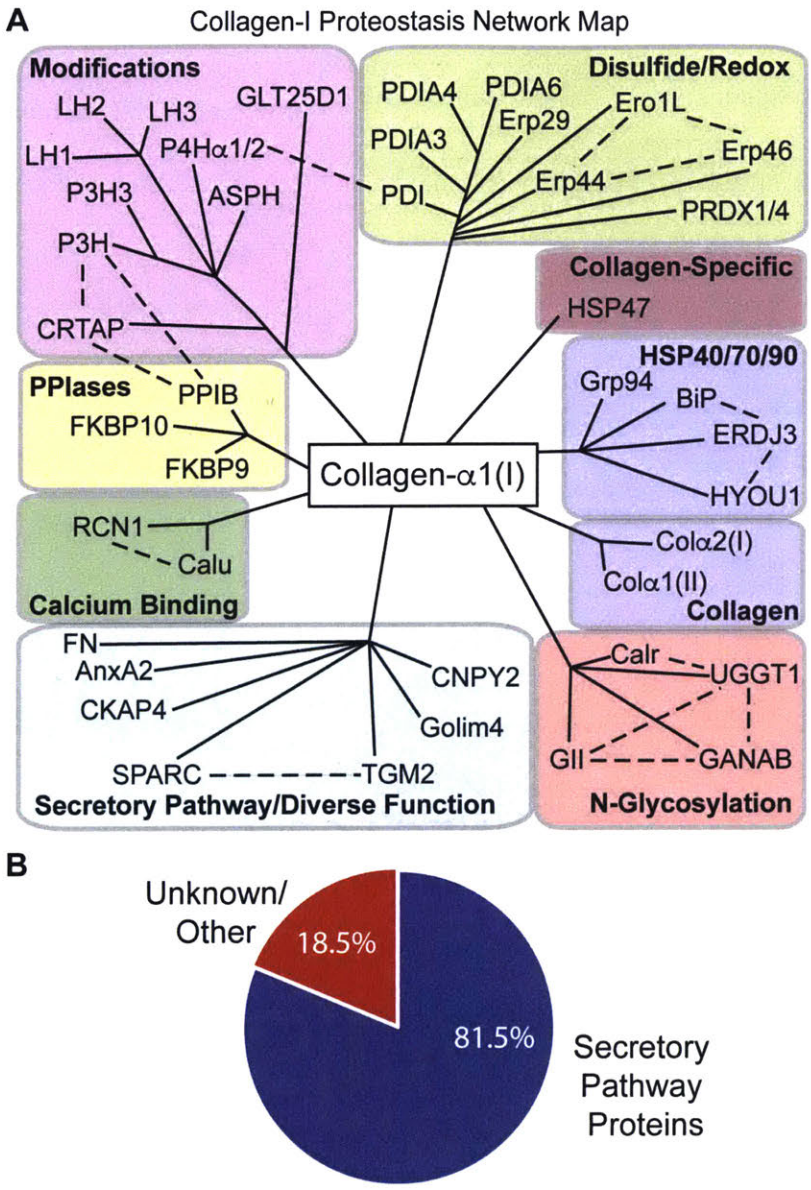


Figure 3.8 | Map of the Collagen-I Proteostasis Network

(A) Interactomics map of the collagen-I proteostasis network (illustrated by solid black lines) and respective complex components known to interact with each other (illustrated by dotted black lines). **(B)** Pie chart showing enrichment of secretory pathway proteins in the collagen-I interactome.

Table 3.2. Mass Spectrometry-Based Mapping of the Collagen-I Proteostasis Network in the Presence or Absence of Ascorbate

Protein (Common Name)	Gene Name	+ Asc Fold-Enrichment[a]	– Asc Fold-Enrichment[a]	Unique Peptides
Collagen-Specific Chaperones				
Heat shock protein 47 (HSP47)	*SERPINH1*	30.5	37.6	4
Peptidyl Prolyl Isomerases				
Peptidyl-prolyl cis-trans isomerase B (CyPB)	*PPIB*	41.2	22.0	9
Peptidyl-prolyl cis-trans isomerase FKBP65	*FKBP10*	4.0	1.4	2
Peptidyl-prolyl cis-trans isomerase FKBP9*	*FKBP9*	2.4	1.7	2
Co- and Post-Translational Collagen Modifications				
Protein disulfide isomerase (PDI)	*P4HB*	19.4	29.3	7
Prolyl-4-hydroxylase-α1	*P4HA1*	9.0	14.9	5
Prolyl-4-hydroxylase-α2	*P4HA2*	4.5	7.0	2
Procollagen galactosyltransferase 1 (GLT251)	*COLGALT1*	14.0	8.5	3
Peptidyl-prolyl cis-trans isomerase B (CyPB)	*PPIB*	41.2	22.0	9
Cartilage-associated protein (CRTAP)	*CRTAP*	13.2	23.5	4
Prolyl-3-hydroxylase 1 (P3H1)	*LEPRE1*	7.9	14.6	3
Aspartyl-asparaginyl beta hydroxylase*[,b]	*ASPH*	7.4	9.4	2
Prolyl-3-hydroxylase 3 (P3H3)	*LEPREL2*	4.8	9.4	3
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3 (LH3)	*PLOD3*	4.3	8.8	2
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (LH2)	*PLOD2*	2.3	2.4	2
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1 (LH1)	*PLOD1*	2.0	5.3	3
Disulfide Bond Formation / Shuffling and Cellular Redox Chemistry				
Protein disulfide isomerase (PDI)	*P4HB*	19.4	29.3	7
Thioredoxin domain-containing protein 5 (Erp46)*	*TXNDC5*	8.4	13.5	3
Protein disulfide isomerase A4 (Erp72)*	*PDIA4*	8.2	15.4	3
Protein disulfide isomerase A3 (Erp57)*	*PDIA3*	7.4	14.3	10
Endoplasmic reticulum protein 29*	*ERP29*	4.8	8.3	2
ERO1-like protein-α[c]	*ERO1L*	4.3	6.6	2
Protein disulfide isomerase A6 (Erp5)*	*PDIA6*	4.1	8.3	5
Peroxiredoxin-1 or peroxiredoxin-4 (PAG or Prx-IV)*	*PRDX1 or PRDX4*	0.8	2.9	3
Endoplasmic reticulum resident protein-44*[,b]	*ERP44*	0.2	2.0	3
Hsp40/70/90				
BiP (Grp78)	*HSPA5*	13.6	22.9	20
Endoplasmin (Grp94)*	*HSP90B1*	4.4	6.4	11
Hypoxia up-regulated protein 1*[,b]	*HYOU1*	2.2	6.1	3
DnaJ homolog subfamily B member 11 (Erdj3)*	*DNAJB11*	2.0	2.6	3
N-Glycosylation and Lectin-Assisted Folding				
Glucosidase-2 β subunit (GII)*	*PRKCSH*	6.9	7.2	4
Neutral alpha-glucosidase AB*	*GANAB*	2.8	2.9	3

UDP-glucose:glycoprotein glucosyltransferase 1*	*UGGT1*	4.1	6.9	3
Calreticulin*	*CALR*	3.4	5.5	4
Calcium Binding				
Calumenin*	*CALU*	20.4	25.3	5
Reticulocalbin-1*	*RCN1*	7.5	8.1	2
Secretory Pathway-Localized / Diverse Functions				
Golgi integral membrane protein 4 (Golim-4/GPP130)*,[c]	*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
Secreted protein acidic and rich in cysteine (SPARC)	*SPARC*	9.8	5.3	2
Annexin A2*	*ANXA2*	6.7	0.9	2
Fibronectin	*FN1*	5.2	4.5	9
Protein-glutamine gamma-glutamyltransferase 2*	*TGM2*	2.0	1.6	10
Collagen				
Collagen α-1(II)	*COL2A1*	5.3	1.5	3
Collagen α-2(I)	*COL1A2*	2.6	5.4	55
Not Known to be Secretory Pathway-Localized				
ATP synthase subunit α, mitochondrial*	*ATP5A1*	4.6	0.5	3
Major vault protein (MVP)*	*MVP*	2.6	1.9	19
Heat shock protein HSP 90-β*	*HSP90AB1*	2.5	0.4	4
Poly[ADP-ribose] polymerase 4*	*PARP4*	2.3	1.1	4
40S ribosomal protein S3*	*RPS3*	2.3	1.3	2
40S ribosomal protein SA*	*RPSA*	0.4	20.1	2

[a]All protein quantifications calculated for proteins with > 2-fold enrichment across at least two biological replicates with at least two unique peptides used to positively identify the protein, unless otherwise noted.
[b]Peptides corresponding to this protein were identified in all three replicates, but could not be quantified in all replicates owing to the control (light) peptide levels falling below the limit of detection.
[c]Peptides corresponding to this protein were identified in two replicates, but could not be quantified in both replicates owing to the control (light) peptide levels falling below the limit of detection.
Note: Blue lines along the left-hand side of the chart denote complexes. Some proteins are repeated throughout the table due to multiple subgroup classifications or to indicate a complex. Asterisk (*) represents proteins with ill-defined or previously unknown roles in collagen-I proteostasis.

3.3.6 Establishing new roles for putative interactions

Prior to our MS studies, we established an efficient protocol for ensuring well-known collagen-I proteostasis network components co-IP with the HA-collagen-α1(I) (see **Figures 3.6C** and **3.6D**). To ensure that the other ~35 proteins with ill-defined roles in collagen-I proteostasis that we identified were not artifacts, we performed the same co-IP/immunoblot analysis on a panel of the new putative interactors. Of the ten new interactors tested, we observe robust, immunoblot-detectable co-IP of nine: Erp44, Golim-4, reticulocalbin-1, protein disulfide isomerase-3 (PDIA3), protein disulfide isomerase-4 (PDIA4), Erdj3, HYOU1, Erp29, and ER oxidase-like protein 1 (Ero1L) (**Figure 3.9A**). Calumenin was the only tested interactor that could not be positively identified, possibly because the calumenin protein runs at the same location as the HA-antibody light chain on an SDS-PAGE gel, rendering detection difficult as both our HA antibody and the calumenin primary antibody were produced in mice.

Although the HT-1080$^{Col\text{-}I}$ cell line is ideal for MS studies, as it does not express potentially confounding endogenous collagen-I, it could theoretically introduce false positives. We next employed IPs and immunoblotting to explore whether interactions are recapitulated in Saos-2 osteosarcoma cells that natively express collagen-I.[33] We prepared stable Saos-2 cells that inducibly express our HA-tagged collagen-α1(I) construct under control of the tetracycline repressor and probed the crosslinked co-IP for putative interactors from **Figure 3.9A.** The majority of the interactions tested are clearly recapitulated in Saos-2 cells (**Figure 3.9B**), although a few proved difficult to detect likely owing to the fact that low levels of the HA-tagged collagen-α1(I) are competing for proteostasis network interactions with abundant endogenous collagen-I in Saos-2 cells. This high reproducibility across multiple cell lines of the novel putative collagen-I interactions observed in our HT-1080$^{Col\text{-}I}$ cells engenders further confidence in our map of the collagen-I proteostasis network (**Table 3.2** and **Figure 3.8A**).

Next, we obtained lentiviruses expressing shRNAs (see **Table 3.3**) against a select set of the putative collagen-I interactors we identified. We generated fifteen stable Saos-2 cell lines in which a single interactor was uniquely knocked down (see **Figure 3.10A** for qPCR validation of knockdown cell lines). To discern effects of each knockdown on collagen-I secretion, we differentiated Saos-2 cells for 3 d to stimulate endogenous collagen-I secretion,[34] and then analyzed collagen-I levels in the corresponding media samples using immunoblotting (**Figure 3.11**). Our results confirm that many of these interactors are relevant to collagen-I proteostasis. For example, RCN1, Ero1L, SPARC, and TGM2 knockdowns considerably enhance collagen-I secretion. RCN1 is particularly intriguing because it is an abundant ER protein that currently has no well-defined function beyond calcium binding.[35] However, recent work identified another member of the reticulocalbin family, RCN3, as a novel negative regulator of collagen production.[36] These findings further validate our collagen-I interactome and, most importantly, very strongly motivate follow-up studies to delineate the molecular mechanisms of action of these interactors in collagen-I proteostasis. We also note that there are minimal effects on collagen-I secretion induced by knockdown of several of the novel collagen-I interactors identified by MS, including Erdj3 and calumenin. Considering the overall demonstrated reliability of our interactome (see results both above and below), it is perhaps most likely that these interactors play roles in collagen-I production that are simply not possible to discern in a secretion assay—or that they are most important under other conditions, such as when collagen-I is misfolding (e.g., in the collagenopathies).

A particularly surprising finding in our proteomics data is the observation that collagen-I robustly interacts with aspartyl-asparaginyl β-hydroxylase (ASPH). Asp hydroxylation in collagen has not been previously reported, to the best of our knowledge. Supporting an actual role for ASPH in collagen-I biosynthesis, our MS data on collagen-I immunoprecipitated from HT-$1080^{Col\text{-}I}$ cells is consistent with hydroxylation of both Asp71 and Asp72 in the N-propeptide

domain of collagen-α1(I) (**Figure 3.12A**). As expected, the MS1 spectrum indicates that the modified peptide is more prominent in the ascorbate-treated, heavy sample. To address the possibility that the result is an artifact of working in HT-1080 cells, we also performed MS analyses on collagen-I immunoprecipitated from Saos-2 media, and again identified apparent Asp hydroxylation specifically on Asp71 of collagen-α1(I) (**Figure 3.12B**), providing further confirmation that Asp hydroxylation is a native collagen-I N-propeptide modification. Asp hydroxylation in collagen-I has likely been previously overlooked because traditional MS protocols for analysis of collagen begin with proteolytic removal of the propeptides. Intriguingly, *ASPH* mutations were very recently identified as causative factors in facial dysmorphism,[37] supporting a possible role in the maturation of extracellular matrix proteins like collagen-I and motivating further mechanistic work in this area.

Cumulatively, the observations that a diverse set of putative collagen-I interactors do indeed engage collagen-I across multiple cell lines and that numerous members of a panel of shRNA-mediated knockdowns of those interactors modulate collagen-I secretion provides compelling validation of our map of the collagen-I proteostasis network. Equally important, we are able to identify a probable functional consequence of an unanticipated collagen-Iinteraction with ASPH—the apparent hydroxylation of at least one collagen-I N-propeptide Asp residue. Such a discovery is made possible by our proteomics-guided approach to understanding collagen-I proteostasis.

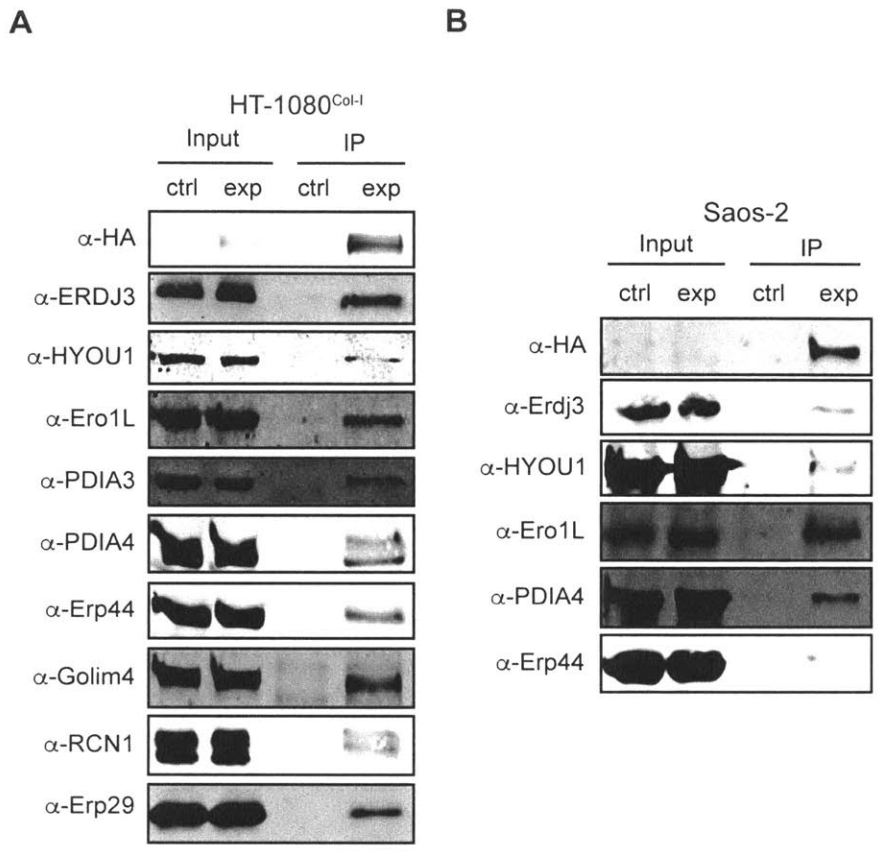


Figure 3.9 | Validation and Characterization of Putative Collagen-I Proteostasis Network Components

(A) HA antibody-mediated IP of HA-collagen-α1(I) from HT-1080^{Col-I} cells coimmunoprecipitates numerous novel or poorly characterized collagen-I interactions identified by mass spectrometry. The control sample represents HT-1080 cells that do not express HA-tagged collagen-I upon induction. **(B)** Validation of the biological relevance of the interactions identified in the HT-1080 cells by immunoprecipitating HA-tagged wild type collagen-α1(I) from osteosarcoma Saos-2 cells and then immunoblotting. Collagen-I expression was induced by treatment with 1 μg/mL dox and 50 μM ascorbate. The control sample represents untreated Saos-2 cells.

Table 3.3. Compilation of shRNA constructs used in **Figures 3.10–3.13**.

Gene Name	Oligo Sequence (5′ to 3′)
Non-Mammalian	CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT
Erp29	CCGGCAAGTTCGTCTTGGTGAAGTTCTCGAGAACTTCACCAAGACGAACTTGTTTTT
P4HB	CCGGAGGTGAAATCAAGACTCACATCTCGAGATGTGAGTCTTGATTTCACCTTTTTTG
SerpinH1	CCGGCCTCTACAACTACTACGACGACTCGAGTCGTCGTAGTAGTTGTAGAGGTTTTT
AnxA2	CCGGCGGGATGCTTTGAACATTGAACTCGAGTTCAATGTTCAAAGCATCCCGTTTTTG
CRTAP	CCGGCTTTACTCTCCAAAGTGAAAGCTCGAGCTTTCACTTTGGAGAGTAAAGTTTTTTG
Ero1L	CCGGCCTCATAAATGACCCATAATTCTCGAGAATTATGGGTCATTTATGAGGTTTTTG
Golim4	CCGGGAAGATATAAACCCAGCAGATCTCGAGATCTGCTGGGTTTATATCTTCTTTTTTG
RCN1	CCGGCCGCAGAGTTTCATGATTCTTCTCGAGAAGAATCATGAAACTCTGCGGTTTTT
Calu	CCGGGTTAGAGATGAGCGGAGGTTTCTCGAGAAACCTCCGCTCATCTCTAACTTTTTG
ASPH	CCGGCGTGGTTTATGGTGATTGCATCTCGAGATGCAATCACCATAAACCACGTTTTTG
CKAP4	CCGGGCAGGATTTGAAAGCCTTAAACTCGAGTTTAAGGCTTTCAAATCCTGCTTTTTG
DNAJB11	CCGGGTTGGCTTTGAGATGGATATTCTCGAGAATATCCATCTCAAAGCCAACTTTTT
PRKCSH	CCGGCAGCCTTCAAAGATGGGTAAACTCGAGTTTACCCATCTTTGAAGGCTGTTTTTG
SPARC	CCGGCGGTTGTTCTTTCCTCACATTCTCGAGAATGTGAGGAAAGAACAACCGTTTTT
TGM2	CCGGACAGCAACCTTCTCATCGAGTCTCGAGACTCGATGAGAAGGTTGCTGTTTTTTG
PDIA3	CCGGGCTTACTATGATGTGGACTATCTCGAGATAGTCCACATCATAGTAAGCTTTTTTG
PDIA4	CCGGCCTGAGAGAAGATTACAAATTCTCGAGAATTTGTAATCTTCTCTCAGGTTTTTG
Erp44	CCGGGCACCCAGTGAATATAGGTATCTCGAGATACCTATATTCACTGGGTGCTTTTTG
Erp46	CCGGGCCAAGCGAAAGACGAACTTTCTCGAGAAAGTTCGTCTTTCGCTTGGCTTTTTG

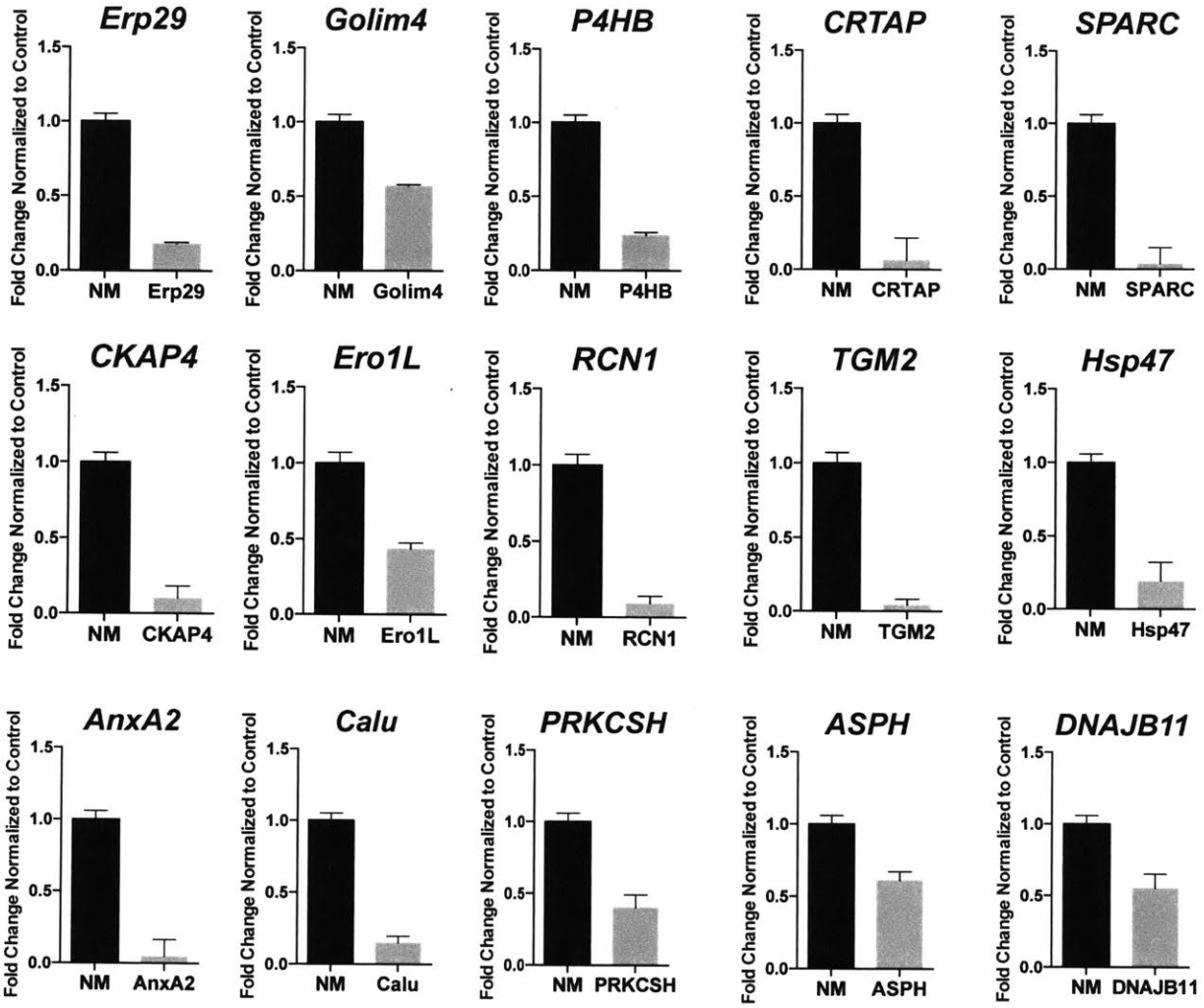


Figure 3.10 | Stable shRNA Knockdown Saos-2 Cell Lines Analyzed By qPCR

Bar charts showing the extent of knockdown for each individual stable Saos-2 cell line assayed for collagen-I secretion in **Figures 11** and **13**. Samples were normalized to the given gene's expression in a corresponding non-mammalian shRNA control sample. The shRNA construct is listed along the x-axis, and the corresponding gene analyzed is listed as the title of the graph. Results are presented as the mean ±95% confidence interval.

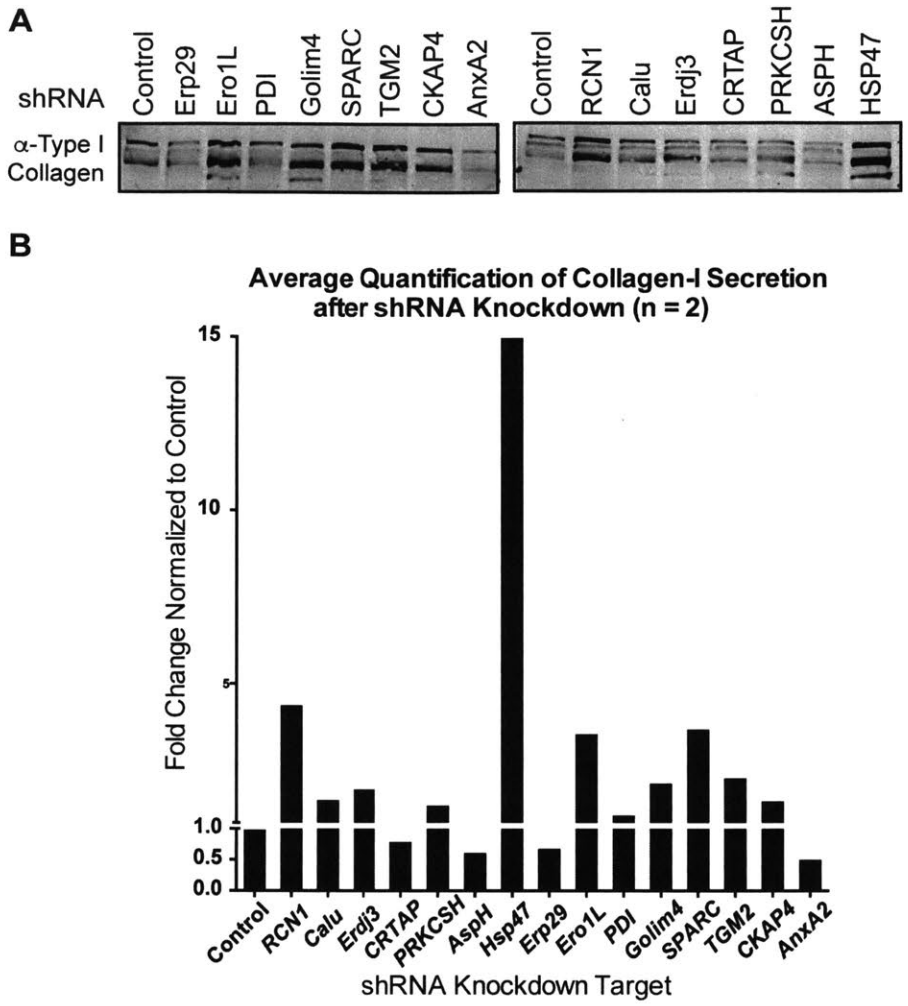


Figure 3.11 | Alterations in Secretion in the Presence of Different shRNA constructs

(A) Secreted collagen-I levels analyzed by immunoblot in the context of a panel of shRNA stable knockdowns of the indicated proteins in Saos-2 cells. **(B)** Quantification of the average fold change in collagen-I secretion compared to control cells (n = 2).

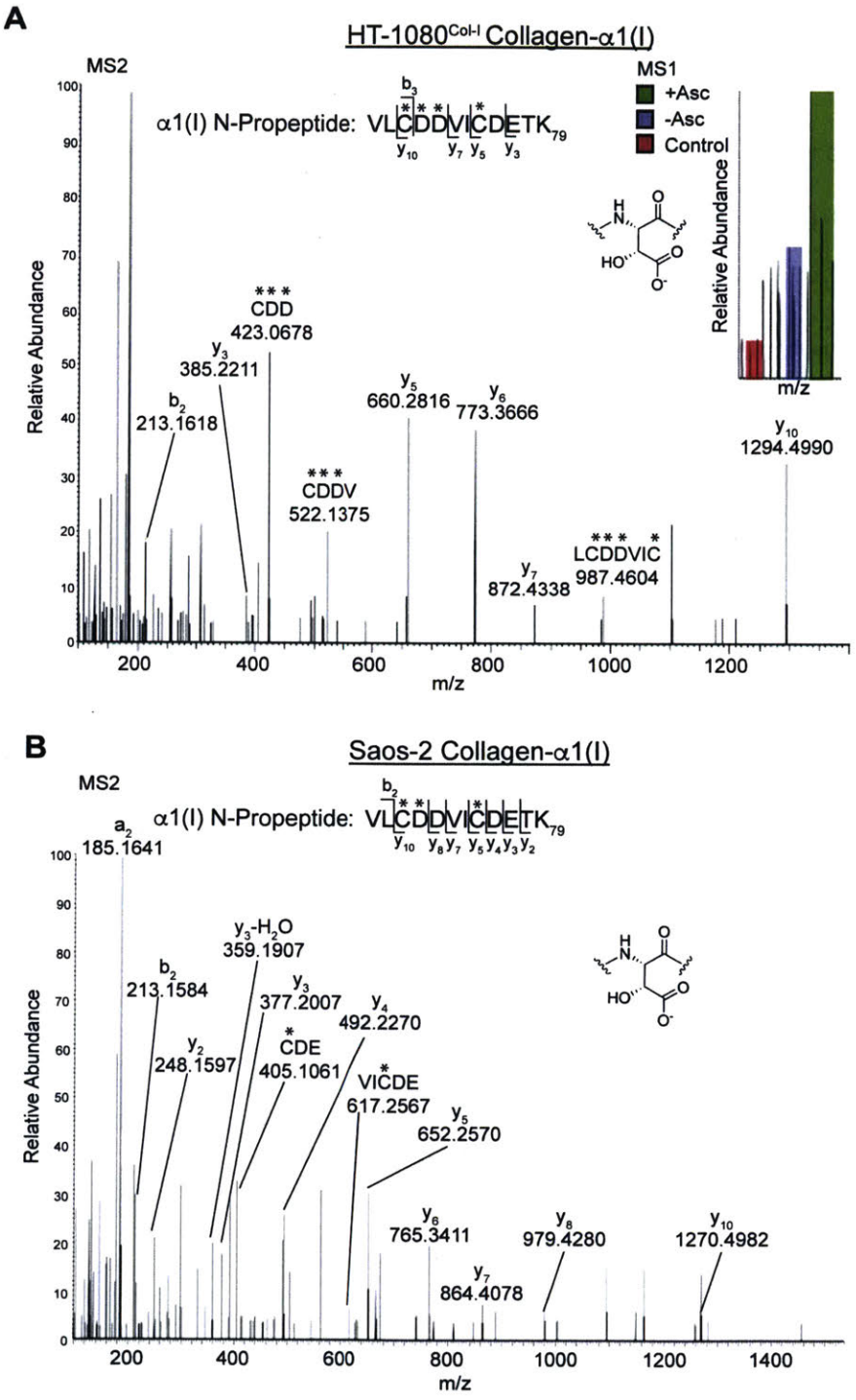


Figure 3.12 | Aspartyl Hydroxylation of Colα1(I) N-propeptide

(A) MS1 and MS2 spectra with respective peaks assigned for a peptide whose fragmentation pattern is consistent with aspartyl-hydroxylation at residues Asp71 and Asp72 (numbering

beginning at collagen-I's start codon) in the N-propeptide of collagen-α1(I) immunoisolated from HT-1080$^{Col\text{-}I}$ cells. **(B)** MS2 spectrum with respective peaks labeled that is consistent with aspartyl-hydroxylation at residue Asp71 of collagen-α1(I) immunoisolated from Saos-2 media. For both **(A)** and **(B)**, asterisks (*) indicate the site of a hydroxylated aspartate or alkylated cysteine. b-Ions are fragments of the parent ion, originating at the N-terminus and extending the number of amino acids into the peptide, as indicated. y-Ions are the same, but originating at the C-terminus of the peptide.

3.3.7 Mechanistic Studies Suggest a Role for Erp29 in Collagen-I Proteostasis

Collagen-I secretion from cells that endogenously produce the protein (including Saos-2 cells) is significantly reduced in the absence of ascorbate.[38] Therefore, we anticipated that quality control mechanisms might engage collagen-I to a greater extent in the absence of ascorbate (ER-retention conditions). A $\log_2$ scatter plot of collagen-I interactor enrichment under secretion-promoting conditions versus ER-retention conditions (**Figure 3.13A**) initially suggests that the interactomes in the presence or absence of ascorbate are similar, consistent with ascorbate-mediated regulation of collagen-I secretion occurring primarily at the level of transcription/translation.[39] Nonetheless, under-hydroxylated collagen-I produced in the absence of ascorbate is retained in the ER to a significant extent,[25] implying that there are likely to be quality control or ER retention factors that can recognize immature collagen-I and help prevent its secretion. A closer examination of functional subsets of the collagen-I proteostasis network under collagen-I secretion-promoting versus ER-retention conditions (**Figure 3.13B**) reveals quantitatively enhanced collagen-I engagement under ER-retention conditions both for the HSP40/70/90 class of chaperones and for proteostasis network components functionally or structurally related to disulfide bond formation/shuffling and cellular redox chemistry, including PDIA3, PDIA4, Erp44, Erp46, Erp29,[40] and the peroxiredoxins.[41] The observation of new ER proteins related to disulfide bond formation/shuffling in the collagen-I proteostasis network was anticipated by us, because both the N- and C-terminal propeptides of collagen-I are cysteine-rich. However, the discovery that these collagen-I interactors engage collagen-I to a greater extent under ER-retention conditions prompted us to further investigate their mechanistic roles in collagen-I proteostasis.

Following the experimental protocol employed in **Figures 3.10** and **3.11**, we began by using shRNA-expressing lentiviruses (see **Table 3.3**) to attempt to stably knockdown the new putative collagen-I interactors PDIA3, PDIA4, Erp44, Erp46, and Erp29 in Saos-2 cells.

Interestingly, we found that *PDIA3*-, *PDIA4*-, *Erp46*- and *ERP44*-targeting shRNAs are highly toxic to Saos-2 cells, potentially implicating them as having a critical function in cells that endogenously express collagen-I. However, as noted above, we were able to deplete Erp29 to ~20% of normal levels in stable Saos-2 cells (**Figure 3.10**). Erp29 depletion does not obviously impact collagen-I secretion under secretion-promoting (ascorbate-treated) conditions (see **Figure 3.11A**), nor does it inordinately stress Saos-2 cells, as indicated by a lack of chronic unfolded protein response activation (**Figure 3.13C**). Intriguingly, upon analyzing collagen-I intracellular steady-state levels and secretion under ER-retention conditions, we observe that Erp29 knockdown has the dual effects of decreasing intracellular collagen-I levels while simultaneously increasing collagen-I secretion (**Figure 3.13D**). Because collagen-I cannot be properly hydroxylated in the absence of ascorbate, this observation implicates Erp29 as playing a critical role in collagen-I maturation and retaining immature collagen-I in the ER (**Figure 3.14**).

Erp29 is an abundant and widespread component of the ER proteostasis network,[42] displaying a thioredoxin (PDI-like) fold but lacking any catalytic activity for disulfide bond isomerization.[40] In the past 20 years, a number of groups have explored possible roles for Erp29 in the proteostasis of CFTR, thyroglobulin, and viral proteins.[43-46] These studies have led to various suggestions that Erp29 may accompany nascent proteins out of the cell,[44] that it may have a role in ER client protein maturation,[45] or that it could assist the folding of certain complex plasma membrane proteins.[46] Our findings here, informed by MS-based proteomics in our new collagen-I expression platform, suggest a novel function for Erp29—retaining immature collagen-I in the ER to prevent secretion of potentially damaging, improperly modified, and unstable forms of the protein into the extracellular matrix (**Figure 3.11D**).

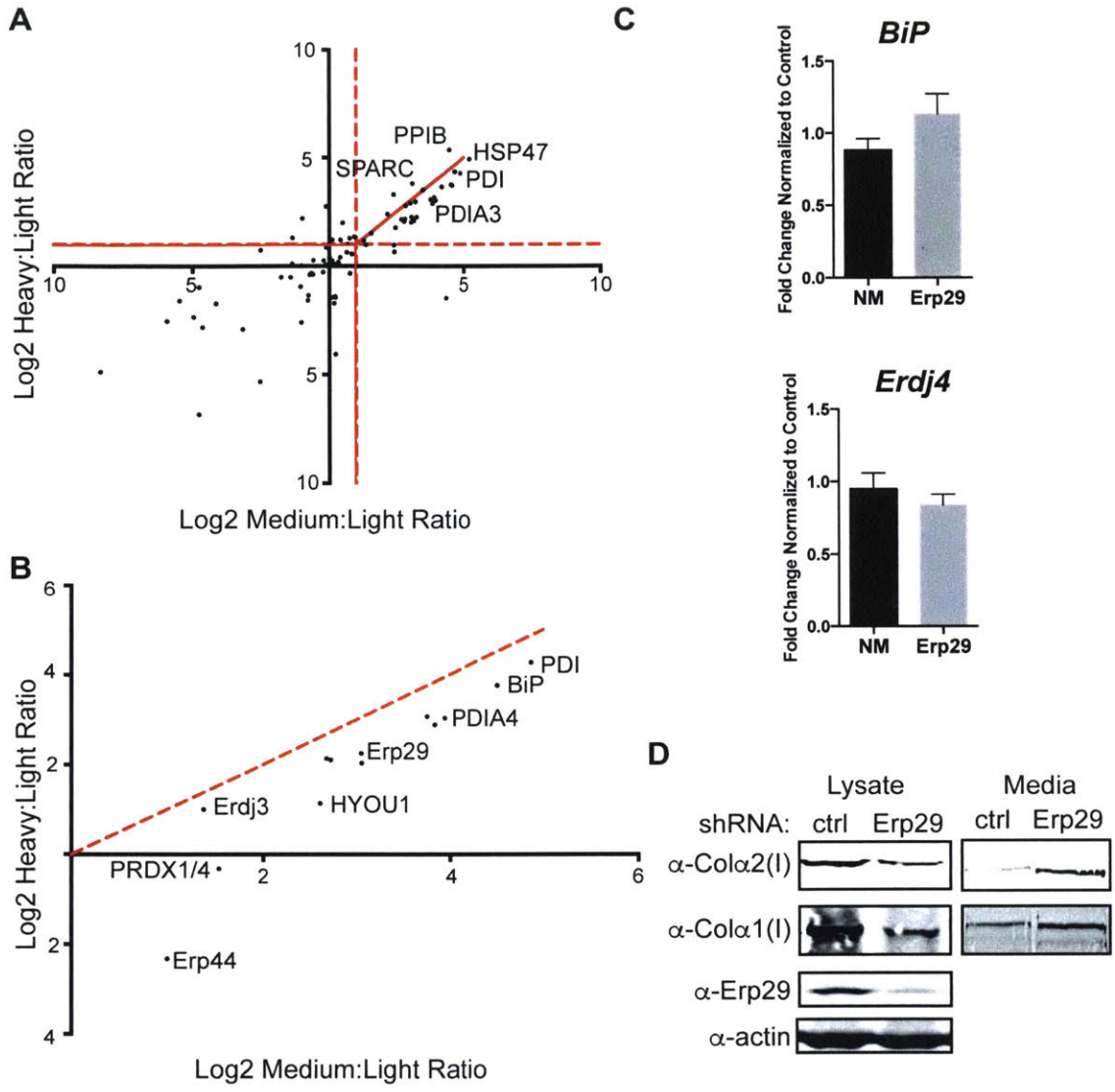


Figure 3.13 | A Role for Erp29 in Collagen-I Proteostasis

(A) Scatter plot of the $\log_2$ SILAC ratios for proteins identified in two or more biological replicates with heavy (ascorbate-treated):light (control) and medium (ascorbate-deficient):light (control) ratios. **(B)** Scatter plot of the $\log_2$ SILAC ratios for the Disulfide Redox and HSP40/70/90 functional groups showing that these collagen-I interactors are quantitatively enriched under ER-retention (ascorbate-deficient) conditions. **(C)** Relative *BiP* and *Erdj4* mRNA levels upon stable Erp29 knockdown in Saos-2 cells compared to control Saos-2 cells showing no

detectable UPR activation upon Erp29 knockdown. **(D)** Collagen-I secretion from Saos-2 cells under ER-retention (ascorbate-deficient) conditions is increased by Erp29 knockdown, while intracellular collagen-I levels are correspondingly reduced.

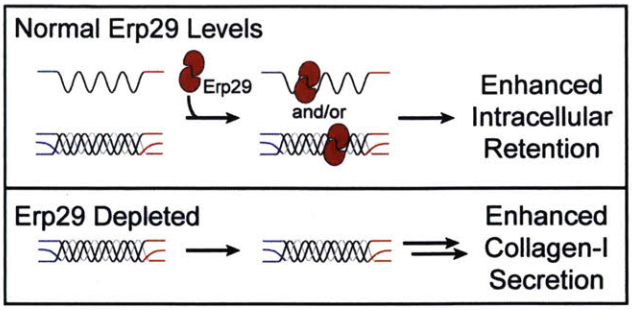


Figure 3.14 | Schematic of Collagen Secretion with and without Erp29

Model illustrating the role of Erp29 in collagen-I retention/quality control under ascorbate-deficient conditions.

3.4 Concluding Remarks.

Altogether, this work provides numerous valuable resources for continued studies of collagen-I biogenesis. First, we have generated a panel of cell lines inducibly expressing antibody epitope-tagged collagen-I strands with validated molecular properties. Key advantages of our cell-based platform include the abilities to orthogonally visualize and quantify different collagen-I strands in a single experiment, as well as to reliably IP collagen-I with MS-grade antibodies. This platform can be easily leveraged in the future to model the autosomal dominant collagenopathies in a manner that allows differential quantification of mutant and wild-type collagen-I strands. Second, we have generated the first map of the collagen-I proteostasis network (**Figure 3.8A**) using our antibody epitope-tagged collagen-I expressing cell lines and quantitative MS-based proteomics. Validation efforts not only confirm a number of putative new interactors, but also unveil an unanticipated collagen-I post-translational modification (Asp hydroxylation). The results yield new insights into the ER proteostasis network components likely involved in collagen-I biogenesis, providing a roadmap for ongoing and future studies of collagen-I folding, quality control, trafficking, and post-translational modification. Third, we delineate a previously unknown role for the abundant but ill-characterized Erp29 protein in collagen-I surveillance under ascorbate-deficient, ER-retention conditions.

Cumulatively, the work here highlights the usefulness of our platform for biochemical studies of collagen-I proteostasis, both for discovery and for hypothesis testing. Continued development of the platform and mechanistic studies on the newly identified collagen-I interactors will provide further important insights into how cells handle this critical scaffold for animal life. Furthermore, we anticipate that our findings here will prove relevant for the twenty-seven structurally-related collagen types and other fibrous extracellular matrix proteins, as well as lead to potential new targets to help resolve disease-associated collagen misfolding.

3.5 Experimental Methods and Supplies

3.5.1 Materials and Reagents

Antibodies used were obtained from the following suppliers: Santa Cruz: HA probe (sc-7392), PDI (sc-20132), CRTAP (sc-99367), CyPB (sc-20361); ProteinTech: DNAJB11 (15484-1-AP); GeneTex: GFP (GTX113617), HYOU1 (GTX102255); Sigma: Colα2(I) (SAB4500363), LAMP1 (SAB3500285), GM130 (G7295), actin (A1978); MyBioSource: Colα1(I) (MBS502153); Cedarlane: Human Collagen Type I (CL50111AP-1); Agilent Technologies: Rat Anti-DYKDDDDK (200474); Enzo Life Sciences: HSP47 (ADI-SPA-470), KDEL (ADI-SPA-827), Calreticulin (ADI-SPA-601), Golim4 (ALX-804-603-C100); IBL America: Reticulocalbin-1 (10367); Abcam: Erp29 (ab11420); Abgent: FKBP10 (AP7383b); Cell Signaling: Ero1L (3264), Erp44 (3798), Erp57 (2881), Erp72 (5033), Calumenin (11991). Antibody beads were obtained from Sigma: FLAG M1 Agarose Affinity Gel (A4596), FLAG M2 Affinity Gel (A2220), and HA-Agarose beads (A2095). Secondary antibodies were obtained from LiCor Biosciences: 800CW Goat Anti-Rabbit, 800CW Goat Anti-Mouse, 800CW Goat Anti-Rat, 800 CW Donkey Anti-Goat, 800CW Streptavidin, 680LT Goat Anti-Rabbit, 680LT Goat Anti-Mouse, 680LT Goat Anti-Rat. Hygromycin and G418 were obtained from Enzo Life Sciences. Puromycin was obtained from Corning. All media and cell culture reagents were obtained from Corning/Cellgro, unless otherwise noted. Restriction enzymes, ligases and polymerases were obtained from New England BioLabs. DNA/RNA preparation kits were obtained from Omega BioTech. *Col1A1* and *Col1A2* genes were obtained from the Origene True Clone Repository, Accession Numbers SC112997 and SC126717, respectively.

3.5.2 Plasmids

The human *Col1A1* and *Col1A2* genes were obtained from Origene. Collagen-I genes were

PCR-amplified to insert into the pTRE-Tight vectors. Q5 Polymerase (New England BioLabs) was used according to the manufacturer's protocol, with the addition of the High GC Enhancer. Sequencing of the collagen genes received from Origene revealed a multitude of missense mutations that were repaired by site directed mutagenesis. Collagen-I genes were PCR-amplified using the following primers:

Col1A1: Forward – 5′-ACATCAGCGGCCGCACAAGAGGAAGGCCAAGTCGAG-3′

Reverse – 5′- AAAAAAGTCGACTTACAGGAAGCAGACAGGG-3′

Col1A2: Forward – 5′-AAAAAAGCGGCCGCAACATGCCAATCTTTACAAGAGGAAAC-3′

Reverse – 5′-AAAAAAGATATCTTATTTGAAACAGACTGGGCCAATG-3′

After 35–40 PCR cycles with a 2 min extension time each, samples were gel purified using Omega Gel Purification Kits. Preprotrypsin signal sequences with either HA or FLAG epitope tags were inserted into pTRE-Tight first, using BamHI and EcoRV, and the collagen-I genes were then inserted into the respective vectors. Collagen-I vector propagation was always performed in *recB* and *recJ*-deficient Sure2 *E. coli*, which we found to be essential to prevent recombination of the GC-rich collagen genes.

Once sequenced, vectors were transfected with XFect (ClonTech) into HT-1080 TetOff cells (ClonTech) along with either a puromycin or hygromycin linear selection marker (ClonTech). 48 h post-transfection, cells were treated with the appropriate antibiotic (0.25 μg/mL puromycin, or 150 μg/mL hygromycin) for 10–12 d for stable selection. Heterostable colonies were amplified, and then split to establish genetically homogenous single colonies for the collagen-I genes. The process was then repeated to create cell lines that express both *Col1A1* and *Col1A2* gene transcripts.

3.5.3 Cell Culture and Transfections

HT-1080 Tet-Off cells (ClonTech) were cultured in complete DMEM supplemented with 10%

FBS. Saos-2 cells were cultured in complete DMEM supplemented with 15% heat-inactivated FBS. Transfections of *Col1A1* and *COL1A2*-encoding plasmids along with linear markers for puromycin or hygromycin resistance were performed using XFect (ClonTech). Stable HT-1080 lines were selected by culturing in 150 μg/mL hygromycin or 0.25 μg/mL puromycin, as appropriate, as well as continuously cultured in 100 μg/mL of G418 to maintain the tetracycline transactivator and 1 ng/mL dox to suppress collagen-I expression. Culturing cells in the presence of the recommended 1 μg/mL of dox was too high of a concentration, as we found that removal of the dox-containing media did not induce high enough expression of the collagen-I genes at 48 hours to be detected by immunoblot of conditioned media or cellular lysate. We optimized the culture conditions by sub-culturing the cells in varied concentrations of dox for 48 h. Collagen-I expression in HT-1080 cells was induced by plating cells in 10% Tet-Approved FBS (ClonTech) complete DMEM supplemented with 50 μM ascorbate for 48 h, with ascorbate replenishment every 24 h, unless otherwise noted. Healthy primary patient fibroblasts (GM05294 from Coriell) were cultured in complete EMEM containing 15% FBS. Collagen-I production by the primary fibroblasts was stimulated by treating with 200 μM ascorbate for 48 h, supplementing every 24 h.

3.5.4 Lysis, Protein Preparation, and Immunoprecipitations

Collagen-I expression by HT-1080 cells, unless noted otherwise, was induced by removing dox and adding 50 μM sodium ascorbate (as indicated) for 48 h, with fresh sodium ascorbate supplementation at the 24 h mark, then cells were harvested and washed three times with 1x-phosphate buffered saline (PBS). Samples for stable (no covalent crosslinking) IPs were lysed using the following lysis buffer (LB): 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 Scientific) and 1.5 mM phenylmethylsulfonyl fluoride (PMSF; Amresco). Cells used for cross-linking IPs were

cross-linked in 1x-PBS in the presence of 0.2 mM dithiobis(succinimidyl propionate) (DSP; Lomant's Reagent), rotating at rt for 30 min. Cross-linking was quenched by addition of 0.1 M Tris at pH 8.0 (final concentration), rotating for 15 min at rt. Cells were then lysed in RIPA (150 mM sodium chloride, 50 mM Tris-HCl, pH 7.5, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate), protease inhibitor tablets and 1.5 mM PMSF. All lysed samples were centrifuged at 21,100 × g for 15 min at 4 °C. Supernatants were collected, quantified for protein content, normalized, and incubated for 16 h with the appropriate antibody-conjugated beads. Samples were then washed with the corresponding lysis buffer (LB or RIPA) and eluted using 300 mM Tris at pH 7.5 with 6% SDS by boiling for 10 minutes. Eluates were then subjected to SDS-PAGE and Western blotting, or prepared for mass spectrometry analysis.

3.5.5 Immunoblotting

SDS-PAGE gels were transferred to a nitrocellulose membrane for 1.5 h at a constant 100 V. Membranes were rinsed, and blocked with 5% w/v non-fat milk/TBS with 0.01% sodium azide. Primary antibodies were diluted in 5% BSA according to the manufacturer's instructions, and incubated 2 h – overnight. Membranes were washed with TBS and incubated with secondary antibody in 5% w/v non-fat milk/TBS for 45 m, washed with TBS and imaged using the LI-COR Biosciences Odyssey System.

3.5.6 Quantitative RT-PCR

Relative mRNA expression levels of genes of interest were assessed by quantitative RT-PCR. Cells were harvested by trypsinization, washed with PBS, and total RNA was extracted using the Omega RNA Purification Kit according to the manufacturer's instructions. RNA concentrations were quantified and normalized to 1 μg total RNA for cDNA reverse transcription. Using the Applied Biosystems Reverse Transcriptase cDNA Kit, cDNA was synthesized in a

BioRad Thermocycler. LightCycler 480 Sybr Green Master Mix (Roche), appropriate primers (purchased from Integrated DNA Technologies, Life Technologies and Sigma Aldrich, **Table 3.1**) and cDNA were used for amplification in a Light Cycler 480 II Real Time PCR Instrument in the MIT BioMicro Center. Primer integrity was assessed by thermal melt and agarose gel analysis to ensure that a single gene was amplified. Transcripts were normalized to the housekeeping gene *Rplp2,* and all measurements were performed in technical and biological triplicate.

3.5.7 Confocal Microscopy

HT-1080$^{Col\text{-}I}$ cells, suspended in complete DMEM with dox (1 ng/mL), were plated on a 24-well plate with coverslips (Chemglass Life Sciences, NJ) coated in poly-D-lysine and incubated for 24 h at 37 °C in a humidified 5% CO_2 atmosphere. The media was changed to Tet-approved media (ClonTech) and the cells were incubated in the 37 °C incubator for 48 h without dox or ascorbate. Media was removed and coverslips were washed three times with PBS. The cells were then fixed with 4% formaldehyde for 30 min and permeabilized with 0.1% Triton in PBS for another 30 min. Coverslips were incubated for 1 h at rt in a blocking buffer containing 1% BSA in TBS (pH 7.5). Double-labeling was performed by incubating coverslips in TBS (5% BSA, 0.01% sodium azide) containing mouse anti-HA (1:200) and then rabbit anti-PDI (1:200), anti-LAMP1 (1:500), or anti-GM130 (1:500) for 2 h at rt or overnight at 4 °C. Secondary antibodies (Alexa Fluor 488-conjugated anti-mouse or Alexa Fluor 568-conjugated anti-rabbit) were diluted (1:1000) in TBS (5% BSA, 0.01% sodium azide) and then applied to the coverslips for 1 h at rt in the dark. After each incubation, cells were rinsed with TBS at least three times. Nuclei were stained for 15 min at rt with DRAQ5 (1 μM). After the final washing with PBS, the sections were then mounted with ProLong to prevent photobleaching. Negative controls for non-specific binding of the secondary antibodies obtained by omitting primary antibodies in the staining protocol were included for each experiment. The subcellular localization was directly analyzed

with an HCX PL APO 63X/1.40-0.60 oil objective mounted onto a Leica TCS-SP2 confocal microscope (Buffalo Grove, IL, USA) equipped with a Leica microsystem. Co-localization quantification was performed using ImageJ software, and calculated for seven different cells per slide in biological triplicate.

3.5.8 Pulse-Chase Analyses

HT-1080$^{Col\text{-}I}$ cells were plated on poly-D-lysine-coated plates 36 h before initiating the pulse-chase, under the appropriate conditions (+/- dox, + ascorbate). Before the pulse with ^{35}S-methionine/cysteine, cells were starved of methionine and cysteine by treating for 30 min with Cys/Met-free DMEM (FBS was dialyzed for 16–18 h against PBS prior to making the media). After starving, cells were pulsed for 30 min with ^{35}S-containing media. Pulsing media was removed, and full DMEM replaced it for the indicated chase times. At the conclusion of each time point, cells and media were harvested and immunoprecipitated for HA-tagged collagen-I overnight, using HA-antibody beads. The next day, samples were washed with RIPA buffer, eluted from the beads using 6x-loading buffer, and separated by SDS-PAGE. Gels were then dried using a gel slab dryer for 1 h at 80 °C. Dried gels were applied to the mounting screen, developed using a GE Healthcare Phospho Screen, and imaged on a Typhoon 3–4 d later.

3.5.9 Protease Digestions

Collagen-I was precipitated from media by treatment with 176 mg/mL ammonium sulfate, 100 mM Tris-HCl pH 7.4, and 1 mM PMSF, at 4 °C overnight. Samples were then centrifuged at 3000 rpm for 15 min to pellet the precipitate. The supernatant was then removed and the precipitated pellets were resuspended in 400 mM NaCl, 150 mM Tris (pH 7.4). Samples were aliquoted and treated with trypsin or chymotrypsin (0.1 mg/mL final concentration)[20] prior to analysis by immunoblotting to determine the extent of stable triple-helix formation.

3.5.10 MS Sample Preparation and Analysis

HT-1080$^{Col\text{-}I}$ cells were propagated for >6 passages in light, medium, or heavy SILAC media (light media was supplemented with Lys ($^{12}C_6$, 99%; $^{14}N_2$, 99%) and Arg ($^{12}C_6$, 99%; $^{14}N_4$, 99%); medium media was supplemented with Lys ($^{13}C_6$, 99%; $^{14}N_2$, 99%) and Arg ($^{13}C_6$, 99%; $^{14}N_2$, 99%); heavy media was supplemented with Lys ($^{13}C_6$, 99%; $^{15}N_2$, 99%) and Arg ($^{13}C_6$, 99%; $^{15}N_4$, 99%); Cambridge Isotopes). All media were supplemented with 2 mM unlabeled proline to reduce the probability of labeled Arg being converted into proline intracellularly. Collagen-I co-IPs were prepared and MS analyses performed. Labeled cells were lysed, trypsinized and the subsequent peptides were analyzed by mass spectrometry to ensure >95% incorporation of isotopically labeled amino acids into the proteome. Upon confirmation, collagen-I IPs were performed, and after elution from the antibody beads, samples were then treated with 100 μM DTT to release the interactome from the collagen-I bait protein (cleaving the crosslinker, DSP). Each sample was then centrifuged over a 100 kDa molecular weight cut off filter (100 MWCO, Millipore) to ensure deeper coverage of the interactome by removing high levels of collagen-I proteins. The flow-through (i.e., the interactome) samples were precipitated by adding 450 μL of MeOH to a lysate sample of <150 μL and vortexed. 150 μL of chloroform was then added and the resulting solution was vortexed. Finally, 450 μL of water was added and samples were vortexed and then centrifuged at 10,000 × g for 3 min. The upper aqueous phase was removed while the precipitate at the solvent interface was preserved. The collected precipitate was then washed 3 × 1 mL with MeOH, centrifuging between washes. The washed pellet was dried and resuspended in 8 M urea, 50 mM ammonium bicarbonate. 10 mM DTT was added, samples were mixed, and then incubated in a 56 °C water bath for 45 min. Samples were then cooled and incubated with 55 mM iodoacetamide for 1 h in the dark while rotating. Samples were next incubated with 1 μg of sequencing-grade trypsin (Promega) overnight at rt, while rotating. Proteolyzed samples were acidified to a final concentration of 5% formic acid and subjected to

C18 Stage Tips for desalting. Prior to sample addition, Stage Tips were washed with 0.1% TFA, then 0.1% TFA with 90% acetonitrile, then again with 0.1% TFA. Samples were then loaded, ensuring all the volume passed through the column. Each column was then washed with 0.1% formic acid, and eluted with 0.1% formic acid with 80% acetonitrile. Elutions were dried by speedvac, resuspended up to 20 μL of 0.1% formic acid, and injected onto a nanoflow HPLC, with MS data acquired on a Thermo QExactive mass spectrometer (LC-MS/MS). Protein identification was carried out using the Mascot database search software. Database search results were assembled using Proteome Discoverer.[29] Mascot search parameters were as follows: mass tolerance for precursor ions was 10 ppm; fragment ion mass tolerance was 0.8 Da; 2 missed cleavages of trypsin; fixed modifications were carbamidomethylation of cysteine; variable modifications were methionine oxidation, and hydroxylation of proline, lysine, aspartate, or asparagine. Peptides with Mascot scores greater than 25 and isolation inference less than 30 were considered identified, resulting in an average false discovery rate of 0.0077. SILAC quantitation was calculated by integrating the area under the curve of the light, medium, and heavy MS peaks. A fraction of the IP supernatant (proteins that did not bind to the beads during the initial incubation) was analyzed and used to normalize the quantitation from the collagen IP to account for potential variability, similar to a loading control.

We later searched our MS data for evidence of aspartyl hydroxylation. In the $HT1080^{Col\text{-}I}$ heavy-labeled sample, we identified 4 peptides with the sequence VLCDDVICDETK, of which one was modified as shown in **Figure 3.12A**, across two replicates. For collagen-I secreted from Saos-2 cells, we used an inclusion list to better enrich peptides with appropriate m/z ratios (searching for both the unmodified and hydroxylated VLCDDVICDETK peptides), and found that 1/25 peptides were modified as shown in **Figure 3.12B**.

3.5.11 Saos-2 Stable Cell Line Generation and Secretion Analysis

A Col1A1.pENTR1A vector was constructed by ligating *COL1A1* cut from the Col1A1.pTRE.Tight vector using the NotI and EcoRV restriction sites into a pENTR1A vector already containing an ER-targeting preprotrypsin signal sequence and an HA epitope tag. The resulting construct was inserted into the pLenti.CMV/TO.DEST Gateway destination vector (Life Technologies) via LR Clonase-mediated recombination (Life Technologies). pLenti vectors were sequence-confirmed before proceeding with lentivirus production. Lentiviral production was performed by co-transfecting 293FT cells with the lentiviral plasmids and packaging vectors encoding RRE, REV, and VSVG using Lipofectamine 2000 (Life Technologies). Briefly, the plasmid mixture (pLenti vector (15 μg), RRE (15 μg), REV (6 μg), and VSVG (3 μg)) was incubated with 60 μL of Lipofectamine 2000 in 3 mL of Opti-MEM media for 45 min at rt. The mixture was then added dropwise to a 10 cm dish of 293FT cells (8×10^6 cells) and incubated at 37 °C for 12 h. The media was then removed and replaced with DMEM (6 mL) and the plates were incubated for another 36 h. 48 h post-transfection, viral supernatant was collected and used immediately for transductions of Saos-2-TREx cells (see below).

To create Saos-2T-REx cells, Saos-2 cells (125,000 cells/well in a 12-well plate) were transduced with a range of volumes of lentivirus encoding a constitutively expressed tetracycline repressor protein prepared as described above. Polybrene (4 μg/mL) was added to increase the efficiency of viral infection. 12 h post-transduction, the media was removed and replaced by fresh DMEM media for another 24 h before selection. Cells were selected for using 2 μg/mL blasticidin, and single colonies were isolated and assayed for the colony most responsive to dox treatment (using Tet-responsive CFP to visualize activity of the tetracycline repressor protein). Once a genetically homogenous single colony was isolated, Saos-2T-REx cells were transduced as above with a range of volumes of crude viral supernatants of Tet-Responsive HA-collagen-α1(I), selected using 250 μg/mL hygromycin, and assayed for expression at the heterogenous cell population level. The resulting heterogenous cell line was used for IP analysis

of the collagen interactome and for shRNA knockdowns. For shRNA knockdowns, cells were transduced with commercially available lentiviruses (Sigma) at an MOI of 2, or with homemade lentiviruses produced as described above. Stable cells were selected using 2 μg/mL puromycin. The resulting heterogenous cell populations were used for further experiments.

3.5.12 Saos-2 Differentiation and Collagen-I Production

Cells were treated as previously described to induce Saos-2 differentiation into osteoblasts.[47] Briefly, cells were treated with McCoy's 5A media, supplemented with 10% heat inactivated FBS, penicillin/streptomycin, L-glutamine, 50 μM sodium ascorbate (refreshed daily) and 5 mM β-glycerophosphate for three days prior to soluble collagen-I secretion analysis.

3.5.13 Adenoviral Production and Transductions

Col1A1.pENTR1A was mutagenized to contain the Cys1299Trp variant known to cause osteogenesis imperfecta. After sequence confirmation, the resultant vector was recombined with Ad.DEST40 (Life Techonologies) using LR Clonase according to the manufacturer's protocol (Life Technologies). Ad.DEST40 vectors containing the OI-variant collagen gene were digested with PacI and transfected into 293A cells to produce replication-incompetent adenovirus-V. After complete cytopathic effect, the viral supernatant was collected, subjected to three freeze (-80° C)/thaw (37 °C) cycles, centrifuged and amplified on fresh 293A cells. This cycle was repeated once more, and the resultant viral supernatant was applied to the $HT1080^{Col\text{-}I}$ cells. 2 d post transduction, media was harvested and analyzed by immunoblot.

3.6 REFERENCES

(1) Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

(2) Ishikawa, Y.; Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta* **2013**, *1833*, 2479-2491.

(3) Myllyharju, J.; Kivirikko, K. I. Collagens and collagen-related diseases. *Ann. Med.* **2001**, *33*, 7-21.

(4) Tosi, L. L.; Warman, M. L. Mechanistic and therapeutic insights gained from studying rare skeletal diseases. *Bone* **2015**, *76*, 67-75.

(5) Forlino, A.; Cabral, W. A.; Barnes, A. M.; Marini, J. C. New perspectives on osteogenesis imperfecta. *Nat. Rev. Endocrinol.* **2011**, *7*, 540-557.

(6) Boudko, S. P.; Engel, J.; Bachinger, H. P. The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 21-32.

(7) Myllyharju, J. Intracellular post-translational modifications of collagens. *Top. Curr. Chem.* **2005**, *247*, 115-147.

(8) Buevich, A. V.; Dai, Q.-H.; Liu, X.; Brodsky, B.; Baum, J. Site-specific NMR monitoring of *cis–trans* isomerization in the folding of the proline-rich collagen triple helix. *Biochemistry* **2000**, *39*, 4299-4308.

(9) 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.* **1986**, *103*, 223-229.

(10) Morello, R.; Bertin, T. K.; Chen, Y.; Hicks, J.; Tonachini, L.; Monticone, M.; Castagnola, P.; Rauch, F.; Glorieux, F. H.; Vranka, J.; Bachinger, H. P.; Pace, J. M.; Schwarze, U.; Byers, P. H.; Weis, M.; Fernandes, R. J.; Eyre, D. R.; Yao, Z.; Boyce, B. F.; Lee, B. CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* **2006**, *127*, 291-304.

(11) Barnes, A. M.; Cabral, W. A.; Weis, M.; Makareeva, E.; Mertz, E. L.; Leikin, S.; Eyre, D.; Trujillo, C.; Marini, J. C. Absence of FKBP10 in recessive type XI osteogenesis imperfecta leads to diminished collagen cross-linking and reduced collagen deposition in extracellular matrix. *Hum. Mutat.* **2012**, *33*, 1589-1598.

(12) Forlino, A.; Kuznetsova, N. V.; Marini, J. C.; Leikin, S. Selective retention and degradation of molecules with a single mutant alpha 1(I) chain in the Brtl IV mouse model of OI. *Matrix Biol.* **2007**, *26*, 604-614.

(13) Lamandé, S. R.; Chessler, S. D.; Golub, S. B.; Byers, P. H.; Chan, D.; Cole, W. G.; Sillence, D. O.; Bateman, J. F. Endoplasmic reticulum-mediated quality-control of type-I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro-alpha-1(I) chain carboxyl-terminal propeptide which impair subunit assembly. *J. Biol. Chem.* **1995**, *270*, 8642-8649.

(14) Fitzgerald, J.; Lamandé, S. R.; Bateman, J. F. Proteasomal degradation of unassembled mutant type I collagen pro-alpha 1(I) chains. *J. Biol. Chem.* **1999**, *274*, 27392-27398.

(15) 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* **1993**, *13*, 399-405.

(16) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J.; Tu, P. G.; Wu, C. L.; Yates, J. R. I.; Su, A. I.; Kelly, J. W.; Wiseman, R. L. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013**, *3*, 1279-1292.

(17) Stephens, D. J.; Pepperkok, R. Imaging of procollagen transport reveals COPI-dependent cargo sorting during ER-to-Golgi transport in mammalian cells. *J. Cell Sci.* **2002**, *115*,

1149-1160.

(18) Diegelmann, R. F., Peterkofsky B. Inhibition of collagen secretion from bone and cultured fibroblasts by microtubular disruptive drugs. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 892-896.

(19) Pace, J. M., Kuslich, C.D., Willing, M.C. Byers, P.H. Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the pro-alpha1(I) chain of type I collagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta. *J. Med. Genet.* **2001**, *38*, 443-449.

(20) Bruckner, P.; Prockop, D. J. Proteolytic enzymes as probes for the triple-helical conformation of procollagen. *Anal. Biochem.* **1981**, *110*, 360-368.

(21) Lamandé, S. R.; Bateman, J. F. The type-I collagen pro-alpha-1(I) COOH-terminal propeptide N-linked oligosaccharide - functional-analysis by site-directed mutagenesis. *J. Biol. Chem.* **1995**, *270*, 17858-17865.

(22) Weis, M. A.; Hudson, D. M.; Kim, L.; Scott, M.; Wu, J. J.; Eyre, D. R. Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J. Biol. Chem.* **2010**, *285*, 2580-2590.

(23) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Prot.* **2002**, *1*, 376-386.

(24) Myllyharju, J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **2003**, *22*, 15-24.

(25) Tschank, G., Brocks, D.G., Engelbart, K., Mohr, J., Baader, E., Gunzler, V., Hanauske-Abel, H.M. Inhibition of prolyl hydroxylation and procollagen processing in chick-embryo calvaria by a derivative of pyridine-2,4-dicarboxylate. *Biochem. J.* **1991**, *275*, 469-476.

(26) Tan, Y. L.; Genereux, J. C.; Pankow, S.; Aerts, J. M.; Yates, J. R., 3rd; Kelly, J. W. ERdj3 is an endoplasmic reticulum degradation factor for mutant glucocerebrosidase variants linked to Gaucher's disease. *Chem. Biol.* **2014**, *21*, 967-976.

(27) Balch, W. E.; Yates, J. R. Application of mass spectrometry to study proteomics and interactomics in cystic fibrosis. *Methods Mol. Biol.* **2011**, *742*, 227-247.

(28) Lomant, A. J., Fairbanks, G. Chemical Probes of Extended Biological Structures: Synthesis and Properties of the Cleavable Protein Cross-linking Reagent [35S]Dithiobis(succinimidyl propionate). *J. Mol. Biol.* **1976**, *104*, 243-261.

(29) Perkins, D. N., Pappin, D.J.C., Creasy, D.M., Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.

(30) Chen, X. Q.; Karnovsky, A.; Sans, M. D.; Andrews, P. C.; Williams, J. A. Molecular characterization of the endoplasmic reticulum: Insights from proteomic studies. *Proteomics* **2010**, *10*, 4040-4052.

(31) Pyott, S. M.; Schwarze, U.; Christiansen, H. E.; Pepin, M. G.; Leistritz, D. F.; Dineen, R.; Harris, C.; Burton, B. K.; Angle, B.; Kim, K.; Sussman, M. D.; Weis, M.; Eyre, D. R.; Russell, D. W.; McCarthy, K. J.; Steiner, R. D.; Byers, P. H. Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum. Mol. Genet.* **2011**, *20*, 1595-1609.

(32) Barnes, A. M.; Cabral, W. A.; Weis, M.; Makareeva, E.; Mertz, E. L.; Leikin, S.; Eyre, D.; Trujillo, C.; Marini, J. C. Absence of FKBP10 in recessive type XI osteogenesis imperfecta leads to diminished collagen cross-linking and reduced collagen deposition in extracellular matrix. *Hum. Mut.* **2012**, *33*, 1589-1598.

(33) Rodan, S. B.; Imai, Y.; Thiede, M. A.; Wesolowski, G.; Thompson, D.; Barshavit, Z.; Shull, S.; Mann, K.; Rodan, G. A. Characterization of a human osteosarcoma cell line (Saos-2) with osteoblastic properties. *Cancer Res.* **1987**, *47*, 4961-4966.

(34) Cheng, S., Lai, C., Blystone, S.D., Avioli, L.V. Bone mineralization and osteoblast differentiation are negatively modulated by integrin alpha (V) beta (III). *J. Bone Miner. Res.* **2001**, *16*, 277-288.

(35) Suzuki, N.; Ban, S.; Itoh, E.; Chen, S.; Imai, F. L.; Sawano, Y.; Miyakawa, T.; Tanokura, M.; Yonezawa, N. Calcium-dependent structural changes in human reticulocalbin-1. *J. Biochem.* **2014**, *155*, 281-293.

(36) Martinez-Martinez, E.; Ibarrola, J.; Fernandez-Celis, A.; Santamaria, E.; Fernandez-Irigoyen, J.; Rossignol, P.; Jaisser, F.; Lopez-Andres, N. Differential Proteomics Identifies Reticulocalbin-3 as a Novel Negative Mediator of Collagen Production in Human Cardiac Fibroblasts. *Sci. Rep.* **2017**, *7*, 12192.

(37) Patel, N.; Khan, A. O.; Mansour, A.; Mohamed, J. Y.; Al-Assiri, A.; Haddad, R.; Jia, X. F.; Xiong, Y.; Megarbane, A.; Traboulsi, E. I.; Alkuraya, F. S. Mutations in ASPH cause facial dysmorphism, lens dislocation, anterior-segment abnormalities, and spontaneous filtering blebs, or Traboulsi Syndrome. *Am. J. Hum. Genet.* **2014**, *94*, 755-759.

(38) Franceschi, R. T., Iyer, B.S., Cui, Y. Effects of ascorbic acid on collagen matrix formation and osteoblast differentiation in murine MC3T3-E1 cells. *J. Bone Miner. Res.* **1994**, *9*, 843-854.

(39) Tajima, S.; Pinnell, S. R. Regulation of collagen synthesis by ascorbic acid: Ascorbic acid increases type-I procollagen messenger RNA. *Biochem. Biophys. Res. Comm.* **1982**, *106*, 632-637.

(40) Mkrtchian, S., Sandalova, T. ERp29, an unusual redox-inactive member of the thioredoxin family. *Antioxid. Redox. Sign.* **2006**, *8*, 325-337.

(41) Rhee, S. G.; Woo, H. A.; Kil, I. S.; Bae, S. H. Peroxiredoxin functions as a peroxidase and a regulator and sensor of local peroxides. *J. Biol. Chem.* **2012**, *287*, 4403-4410.

(42) Mkrtchian, S., Fang, C., Hellman, U., Ingelman-Sundberg, M. A stress-inducible rat liver endoplasmic reticulum protein, ERp29. *Eur. J. of Biochem.* **1998**, *251*, 304-313.

(43) Magnuson, B.; Rainey, E. K.; Benjamin, T.; Baryshev, M.; Mkrtchian, S.; Tsai, B. ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. *Mol. Cell* **2005**, *20*, 289-300.

(44) Sargsyan, E.; Baryshev, M.; Szekely, L.; Sharipo, A.; Mkrtchian, S. Identification of ERp29, an endoplasmic reticulum lumenal protein, as a new member of the thyroglobulin folding complex. *J. Biol. Chem.* **2002**, *277*, 17009-17015.

(45) Grumbach, Y.; Bikard, Y.; Suaud, L.; Chanoux, R. A.; Rubenstein, R. C. ERp29 regulates epithelial sodium channel functional expression by promoting channel cleavage. *Am. J. Physiol. Cell Physiol.* **2014**, *307*, C701-709.

(46) Suaud, L.; Miller, K.; Alvey, L.; Yan, W.; Robay, A.; Kebler, C.; Kreindler, J. L.; Guttentag, S.; Hubbard, M. J.; Rubenstein, R. C. ERp29 regulates DeltaF508 and wild-type cystic fibrosis transmembrane conductance regulator (CFTR) trafficking to the plasma membrane in cystic fibrosis (CF) and non-CF epithelial cells. *J. Biol. Chem.* **2011**, *286*, 21239-21253.

(47) Cheng, S. L.; Lai, C. F.; Blystone, S. D.; Avioli, L. V. Bone mineralization and osteoblast differentiation are negatively modulated by integrin alpha(v)beta3. *J. Bone Miner. Res.* **2001**, *16*, 277-288.

Chapter 4:
The XBP1s arm of the unfolded protein response can selectively resolve collagen-I secretion defects in osteogenesis imperfecta primary cells

4.1 Author Contributions

A.S.D. designed and performed the majority of the experiments. Dr. Ngoc-Duc Doan performed and analyzed the Ad.ATF6 experiments in **Figure 4.4**.

4.2 Introduction

The collagenopathies are a class of diseases that affect nearly every tissue in the body. Most commonly caused by autosomal dominant mutations to one of the 45 different collagen genes, the collagenopathies are caused by the resulting proteostasis defect that leads to either (1) insufficient production of folded collagen,[1-3] (2) excessive accumulation of toxic collagen aggregates in cells leading to ER stress-induced apoptosis,[4-6] (3) or failed quality control allowing misfolded collagen molecules to be deposited in the extracellular matrix (ECM).[7,8] The relative contribution of these possibilities to pathology for different collagenopathy genotypes is likely variant-dependent. Unfortunately only palliative treatments are currently available to patients suffering from a collagenopathy.

The prototypical collagenopathy is osteogenesis imperfecta (OI).[9] OI is a bone disease that most commonly manifests at an early age. Patients with OI can display a wide range of symptoms ranging from virtually undetectable with slightly below average height, to perinatal lethality as a result of malformed, dysfunctional bones.[10-12] Mild cases of OI often present with possibly hundreds of fractures before the patient reaches puberty, curvature in their long bones and spine, difficulty hearing, and dwarfism. More severe cases present symptoms including shortening of the long bones, abnormal ribcage and skull formation, multiple vertebra crush fractures and decreased mobility.[13]

Bisphosphonates traditionally used to treat osteoporosis showed some promise in the early stages of treating OI. The mechanism of action of this class of drugs, however, inhibits bone resorption, by ultimately inhibiting osteoclasts and inducing apoptosis in these cells.[14,15] The long-term side effects of treating with bisphosphonates, especially in pediatrics where

symptoms of the collagenopathies often present, are currently unknown.[16-19] Furthermore, bisphosphonates indirectly address only one of the two main causes of the collagenopathies, that of decreased bone mass. These treatments do not improve the cellular stress induced by the collagen disease variants. If treatment options are to be improved for patients with a collagenopathy, we must gain a better understanding of the molecular underpinnings that cause collagen to be retained in the presence of a disease-causing variant.

OI is most commonly caused by autosomal dominant missense mutations in one of the collagen-I genes, *Col1A1* or *Col1A2*. To date, over 1500 OI-causing mutations in these genes have been identified.[20-22] Most commonly, glycine residues in the triple helical domain are mutated to serine residues. Removal of a Gly residue in the Gly-Xaa-Yaa repeats of the triple-helical domain disrupts one of the essential interstrand hydrogen bonds that stitch the triple helix together, as any amino acid side chain larger than a hydrogen atom cannot fit inside the sterically constricted triple helix.[23,24] Beyond Gly to Ser mutations, many other mutations that remove essential Gly residues and/or disrupt collagen-I assembly in the C-propeptide domain can also cause OI.[5,25-30] Interestingly, there is considerable variability in the patient phenotypes displayed for a given variant.[31-33] As such, a clear genotype-phenotype remains elusive, making the prediction of disease prognoses extremely difficult.[20,34]

While the underlying reason for autosomal dominant pathologic consequences in OI remains unclear, one hallmark of OI is a sharp reduction in the total amount of collagen-I secreted with a corresponding increase in the retention of collagen-I intracellularly.[35,36] The stress response pathway typically responsible for addressing the accumulation of misfolded proteins in the ER is the unfolded protein response (UPR).[37] The UPR transcriptionally upregulates ER chaperones and quality control machinery to assist in resolving protein folding stress in the ER lumen. The pro-folding and pro-quality control aspects of the UPR are predominantly accomplished by activation of two of the three transcriptionally regulated arms of the UPR: ATF6 and XBP1s.[38] If upregulation of chaperone and quality control factors is

insufficient to clear the unfolded protein from the ER lumen, the third arm of the UPR, PERK, will induce cellular apoptosis.[37]

In an effort to devise new treatment strategies for other ER proteostasis defect-related diseases, extensive work suggested that stress independent activation of XBP1s and/or ATF6 can enhance quality control, improve folding, and reduce chronic ER stress in a variety of disease model systems.[39-44] Specifically in the case of light chain amyloidosis, UPR activation selectively decreased the amount of amyloidogenic protein secreted from cells, while showing no effect on the secretion of a more stable, non-amyloidogenic variant. Furthermore, stress independent, arm-selective activation of ATF6 was sufficient to cause the same selective retention of the disease-variant protein in the absence of cytotoxicity engendered by global UPR activation.[45,46] These data suggest the possibility that resculpting the ER proteostasis landscape could also be a viable strategy for the collagenopathies.

At least in principle, improving collagen-I quality control could both reduce toxic intracellular collagen-I accumulation and reduce deposition of misfolded collagen-I in the ECM. Enhancing collagen-I folding could similarly reduce intracellular accumulation and correct the OI-associated collagen-I secretion defect. Perhaps surprisingly, the vast majority of OI-variants that cause collagen-I misfolding do not activate the UPR.[47] Even though the ER quality control machinery does not normally recognize the misfolded OI variants, it is possible that remodeling the proteostasis network would engender a recognition event that could assist in handling the misfolded molecules. While perturbation of the UPR-chaperone system is one viable option, it is also possible to consider specific chaperone targeting by small molecules,[37] or using recently developed Cas9 technology to target individual nodes within the ER that are essential loci of quality control decisions for the fate of an OI-variant collagen-I.[48]

Herein, I describe our efforts towards resolving OI-associated collagen-I proteostasis defects via chemical and genetic manipulation of the UPR. We used a stress-independent induction of the UPR's ATF6 and XBP1s arms to probe effects on OI-causing disease variants

of collagen-I. While we observed no change upon ATF6 activation, tightly regulated forced expression of XBP1s selectively increases collagen-I secretion from a subset of OI patient fibroblasts, with no change on wild-type fibroblast collagen-I secretion. Mechanistic follow-up suggests the surprising possibility that XBP1s-mediated improvements in collagen-I secretion require both enhanced quality control and improved folding. These studies set the stage for future work in animal models to define the potential therapeutic benefits of resolving OI-associated collagen-I proteostasis defects.

4.3 Results

4.3.1 Collagen-I secretion is defective and the protein accumulates inside OI patient primary fibroblasts

One prominent reported hallmark of OI is a collagen-I proteostasis defect that manifests as an inability to secrete normal amounts of properly folded collagen-I. For at least some OI variants, reduced secretion levels may also correlate with intracellular accumulation of collagen-I, and consequent chronic ER stress. Resulting decreases in cell health would further reduce collagen-I secretion levels, creating a potentially devastating negative feedback loop culminating in the bone pathology clinicians associate with OI.

To investigate this reported collagen-I secretion defect, we obtained primary patient dermal fibroblasts from the Coriell Cell Repository corresponding to a wild-type line and several OI-causing collagen-I variants lines: Gly247Ser and Gly352Ser in Colα1(I), and Gly502Ser, Gly610Cys and Gly907D in Colα2(I). Each cell line's collagen-I secretion profile was analyzed in the presence and absence of ascorbate for 24 h.[49] Most of the lines tested have comparably low collagen-I levels in the absence of ascorbate in the media (note that collagen-I should be only minimally secreted under such conditions). In contrast, all four OI cell lines secreted significantly less collagen-I upon ascorbate treatment when compared to a wild-type primary fibroblast line (**Figure 4.1**).

Intriguingly, our data and that of others suggest that the OI patient lines synthesize normal amounts of collagen-I, they just cannot secrete it.[50] Instead, the collagen-I either accumulates or is degraded. Consistent with accumulation, while wild-type fibroblasts show a strong decrease in Colα1(I) and Colα2(I) in the lysate when secretion is stimulated by treatment with ascorbate, the OI lines maintain high intracellular levels of collagen-I. Such accumulation of collagen-I in OI patient lines might be toxic, especially when constantly stimulated to synthesize collagen-I. Alternatively, accumulation could have other deleterious effects. One such effect

could be that the proteostasis network becomes effectively inaccessible to other ECM proteins due to excessive engagement with accumulated collagen-I, resulting in a global secretion defect. Strikingly, in one atypical OI mutation, Warman and co-workers showed that enhancing secretion of OI collagen-I by expression of an LRP5 protein variant that increases bone mass resolves many aspects of disease pathology,[51-53] highlighting that if we learn how to resolve the secretion defect it could present a new opportunity to treat the disease.

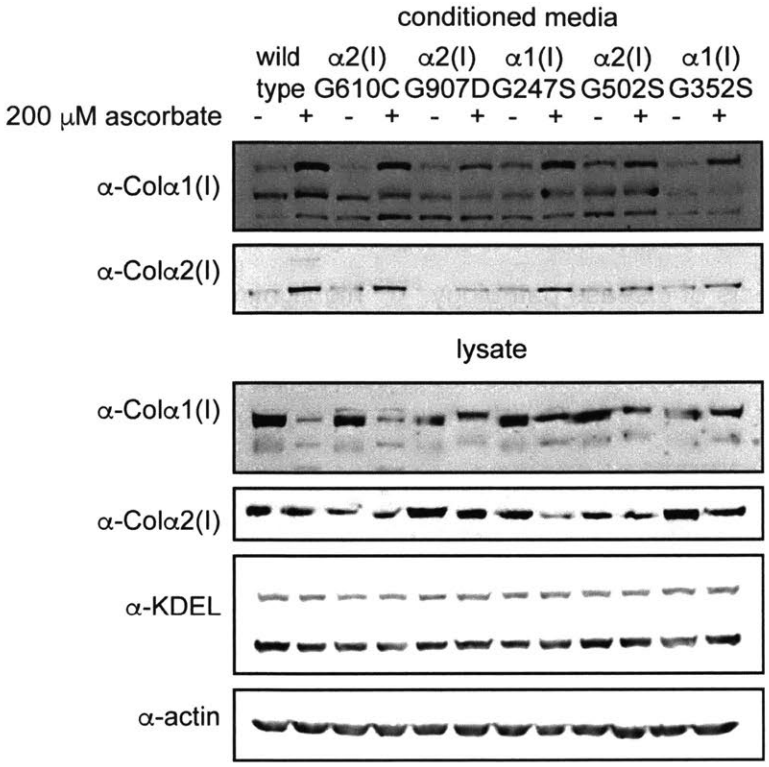


Figure 4.1 | A unifying feature of osteogenesis imperfecta cell lines is the reduction in collagen-I secretion

Equal numbers of cells were plated and treated with or without 200 μM sodium ascorbate for 24 h. Conditioned media was analyzed by SDS-PAGE immunoblot. Cells were harvested and lysed and analyzed by SDS-PAGE immunoblot.

4.3.2 Global UPR activation does not improve collagen-I secretion

The work outlined in Chapter 3 provides a list of known and putative collagen-I interacting proteins that likely play a role in collagen-I maturation and folding (**Table 3.2**). A subset of the interacting proteins identified is transcriptional targets of the UPR (fold change enrichment over negative control in HT1080 cells shown in **Figure 4.2A**). For typical globular and membrane proteins, as unfolded or misfolded proteins build up in the ER, the UPR is activated to ameliorate the stress. This trend is somewhat surprisingly not observed in OI patient cells, despite apparent collagen-I misfolding and accumulation. For example, in all the cell lines tested here the UPR is consistently not activated, as indicated by equal protein levels of GRP94 and BiP, as compared to the levels in wild-type fibroblasts (**Figure 4.1**). Although the cell apparently cannot sense the misfolding of triple helix collagen-I variants (perhaps because an unfolded or misfolded triple helix does not display prototypical hydrophobic regions for sensing) given the strong overlap between UPR targets and the collagen-I proteostasis network it is still possible that remodeling the ER proteostasis network by artificial UPR activation could resolve the collagen-I secretion defect.

We tested whether global UPR activation using thapsigargin (Tg, inhibits calcium flux) could similarly improve collagen-I secretion from OI fibroblast lines. Unfortunately, at all concentrations of Tg that activated the UPR, we observed a strong decrease in collagen-I secretion from not just OI patient lines but also wild-type cells (**Figure 4.2B**; note that qPCR data confirms dosable upregulation of the UPR-target genes *BiP, Erdj4,* and *CHOP* by Tg treatment, see **Figure 4.2C**; **Table 4.1** provides the list of primers used). The decrease in collagen-I secretion was not specific to Tg treatment, as another UPR activator tunicamycin (Tm, N-glycosylation inhibitor) also reduced collagen-I secretion in all the cell lines tested (**Figure 4.3**). Given that wild-type fibroblasts responded identically to the OI fibroblasts when treated with a global UPR stressor, we concluded that this was not a viable method for correcting the collagen-I secretion defect in OI patient primary fibroblasts. Collagen-I is both N-

glycosylated,[54] and putatively binds calcium in the C-propeptide domain, based on the crystal structure of homotrimeric Colα1(I).[55] Furthermore, collagen-I is known to engage chaperones in the ER that require calcium in order to function properly.[56-58]Therefore, perturbations to either of these processes would likely decrease the folding efficiency collagen-I. Moreover, Tg and Tm are ER stressors that globally induce ER protein misfolding, so their ineffectiveness for resolving a protein folding at even low concentrations is perhaps unsurprising.

A

Protein	Fold enrichment
BiP	13.6
PDIA4	8.2
PDIA3	7.4
GRP94	4.4
Ero1L	4.3
UGGT1	4.1
PDIA6	4.1
CRT	3.4
HYOU1	2.2
Erdj3	2.0

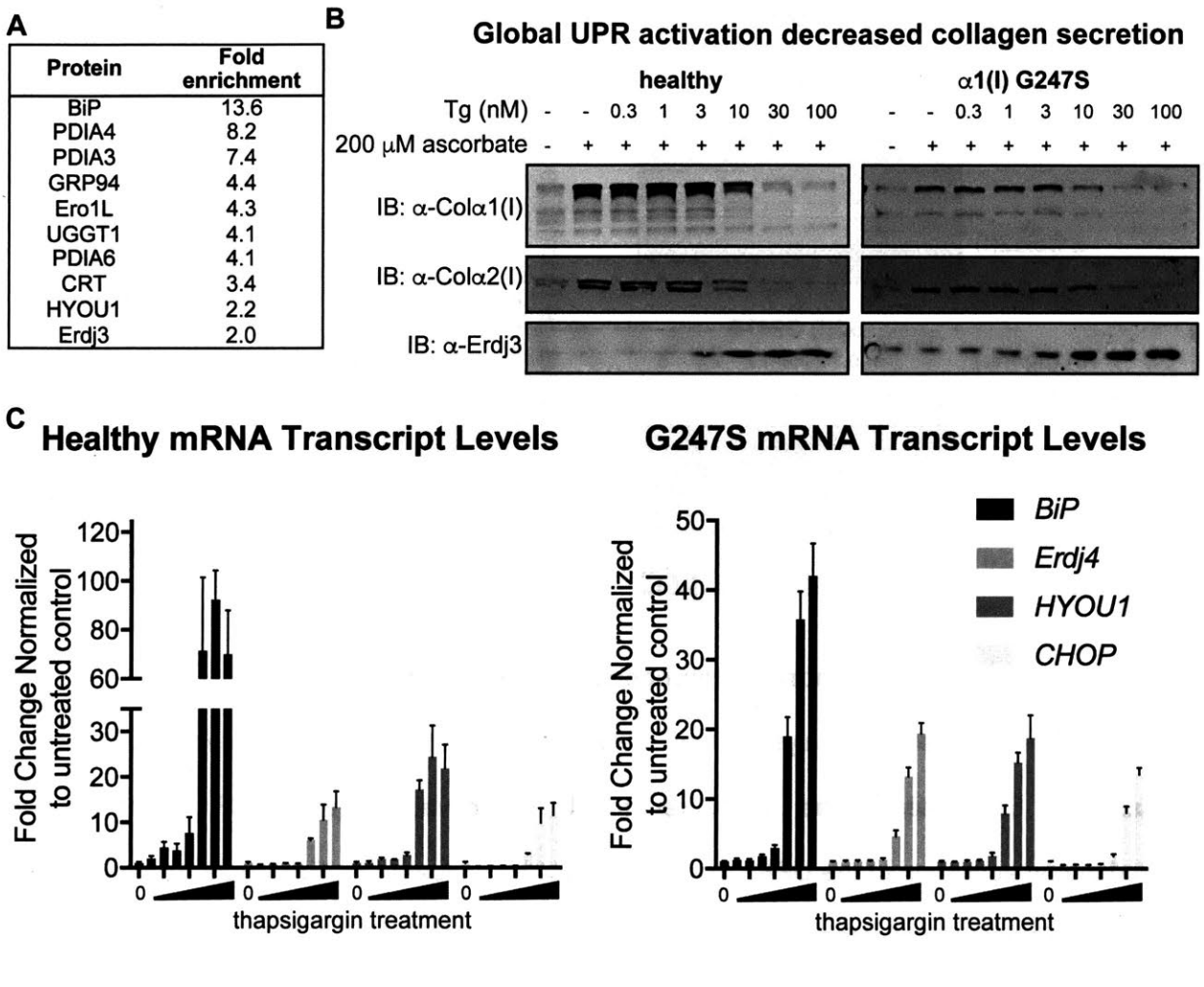


Figure 4.2 | Thapsigargin treatment decreases collagen-I secretion

(A) UPR transcriptional targets that were identified in Chapter 3 as part of the proteostasis network of collagen I. Their corresponding fold enrichments over a negative control IP in HT1080 cells is given in the chart. **(B)** Secreted collagen-I analysis for wild-type and Gly247Ser cells in the presence of increasing concentrations of thapsigargin. Erdj3 secretion increases as the concentration of thapsigargin increases, and serves as a marker for UPR activation.[85] **(C)** qPCR data validating the transcriptional response to increasing concentrations of thapsigargin for both wild-type and Gly247Ser cells. Error bars represent SEM from three biological replicates.

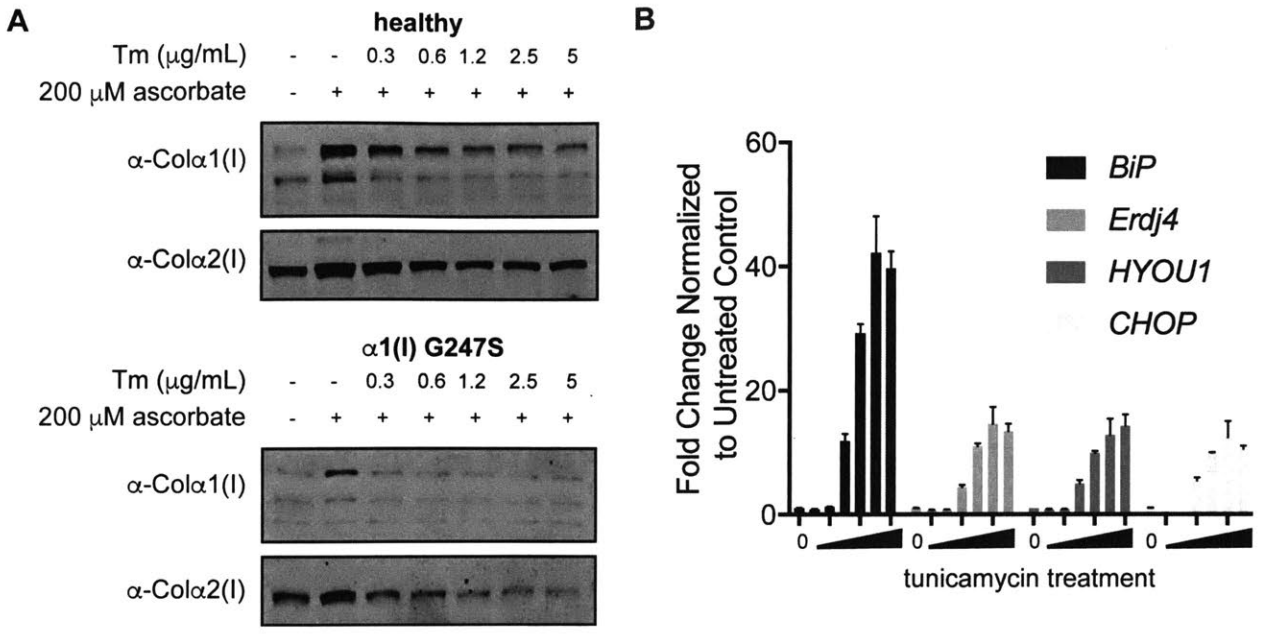


Figure 4.3 | Tunicamycin treatment decreases collagen-I secretion

(A) Collagen-I secretion analysis from cells treated with increasing concentrations of tunicamycin. Erdj3 serves as a marker for UPR activation. **(B)** qPCR validating the transcriptional activation of the UPR upon treatment with tunicamycin. Error bars represent SEM from three biological replicates.

Table 4.1 List of qPCR primers

Transcript	Forward	Reverse
RPLP2	5′-CCATTCAGCTCACTGATAACCTT-3′	5′-CGTCGCCTCCTACCTGCT-3′
CHOP	5′-GGAGCTGGAAGCCTGGTATG-3′	5′-GCCAGAGAAGCAGGGTCAAG-3′
BiP	5′-GCCTGTATTTCTAGACCTGCC-3′	5′-TTCATCTTGCCAGCCAGTTG-3′
Erdj4	5′-GGAAGGAGGAGCGCTAGGTC-3′	5′-ATCCTGCACCCTCCGACTAC-3′
COL1A1 (triple helix)	5′-TGGTAGCCGTGGTTTCCCTG-3′	5′-TCCAGTCAGACCCTTGGCAC-3′
COL1A2 (triple helix)	5′-TGGCTCGAGAGGTGAACGTG-3′	5′-AGCACCGTTGACTCCAGGAC-3′

4.3.3 Stress-independent activation of ATF6 does not improve OI collagen-I secretion

The UPR has three arms that coordinate a transcriptional response to an unfolded protein stress in the ER: ATF6, IRE1 and PERK. PERK attenuates translation and ultimately induces apoptosis in the face of a chronic protein folding stress. It seemed, therefore, unlikely that PERK activation would be a productive method to enhance collagen-I secretion. In contrast, ATF6 and IRE1 generally target and upregulate genes responsible for increasing the protein folding capacity of the ER by directly upregulating chaperones and quality control machinery. Therefore, ATF6 and IRE1 are attractive targets for ameliorating misfolding protein stress in the ER[37,38] (**Figure 1.7**).

We began by testing the potential of the ATF6 arm of the UPR for addressing collagen-I secretion defects. The most well-established method for chemical genetic, arm-selective activation of ATF6 is fusion of the active form of ATF6 to a destabilized domain (DD; schematic shown in **Figure 4.4A**).[59] The destabilized domain (in this case, *E.coli* dihydrofolate reductase (ecDHFR)) is targeted for proteasomal degradation in the absence of the pharmacological chaperone, trimethoprim (TMP). When TMP is present, the fusion protein is rescued from degradation and allows the active transcription factor to upregulate its target genes. This approach allows for dosable control of ATF6 activity, without inducing an actual stress on the cells.

To test whether stress-independent ATF6 activation increased collagen-I secretion by OI patient cells, we generated replication-incompetent adenovirus-V for delivery of a constitutively active form of ATF6 (ATF6, amino acids 1–373) fused to ecDHFR.[38] Cells were transduced, split, and subsequently treated with ascorbate, in the presence or absence of the TMP. After 24 h of activation, collagen-I secretion was induced for an additional 24 h, and media samples were analyzed by immunoblotting. We found that the DHFR-ATF6 construct was a potent, TMP-dependent ATF6 activator in the primary fibroblast cells. As expected, when the MOI of Ad.ATF6 was too high, all three arms of the UPR were activated in a TMP-independent

manner.[38] While we were able to find an amount of virus that selectively upregulated only ATF6 targets (**Figure 4.4B**), ATF6 activation did not enhance collagen-I secretion (**Figure 4.4C**).

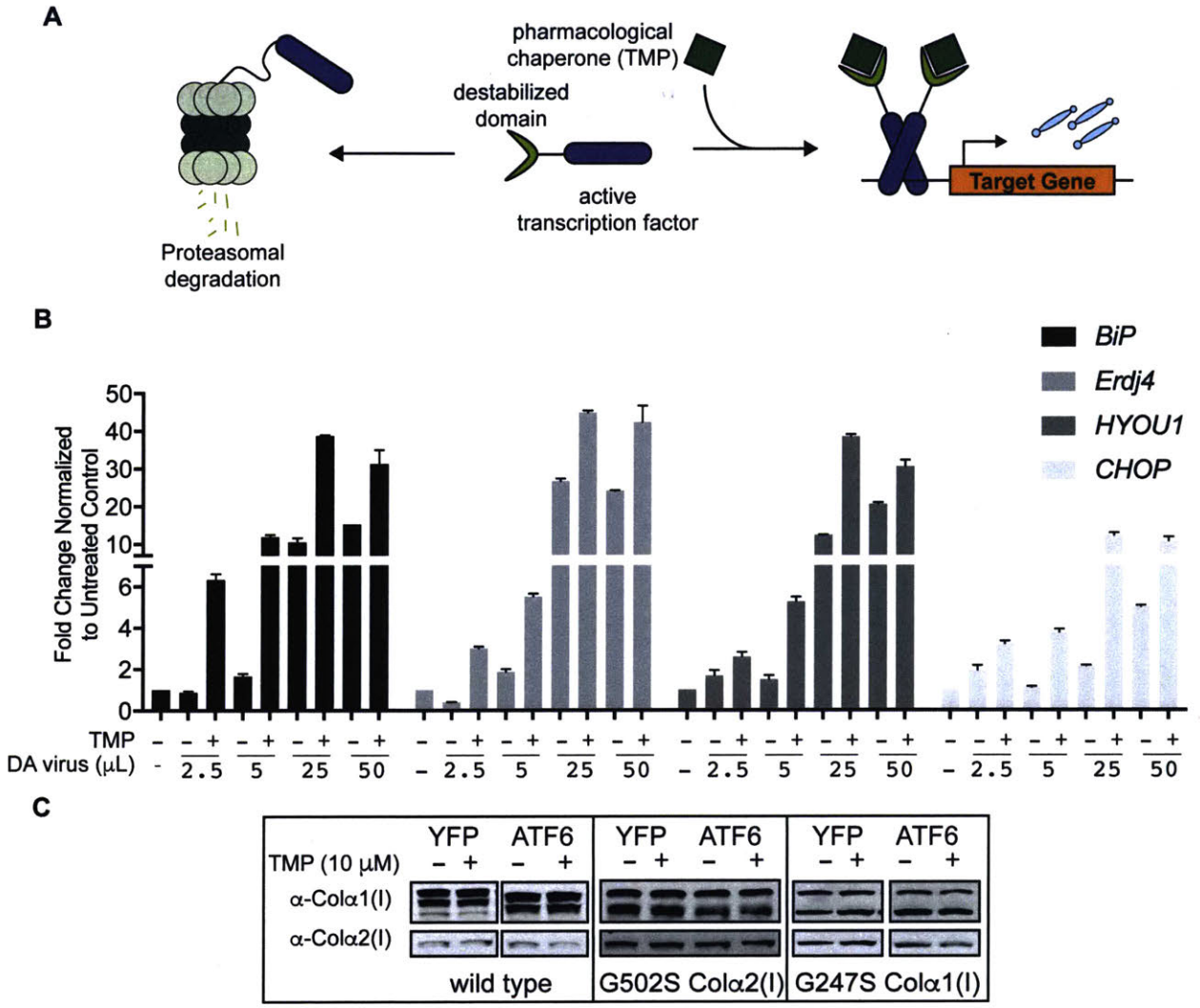


Figure 4.4 | ATF6 activation does not alter collagen-I secretion from OI cell lines

(A) Schematic representation of destabilized domain technology. In the absence of a pharmacologic chaperone, the fusion protein is degraded by the proteasome. When the pharmacological chaperone is present, the destabilized domain is no longer targeted for degradation, and the fusion protein is maintained and can perform the function desired. **(B)** qPCR analysis of cells treated with Ad.ATF6 in the presence and absence of trimethoprim (TMP). At higher MOIs, the transcriptional activity of ATF6 was not dosable and activated the entire UPR non-selectively. We found concentrations of virus that efficiently activated the ATF6 arm of the UPR while leaving the others largely untouched. Error bars represent SEM from

three biological replicates. **(C)** Immunoblot analysis of secreted collagen-I from each wild-type and OI cell line, transduced with Ad.YFP or Ad.ATF6 shows no change in collagen-I secretion profiles for any of the OI lines tested.

4.3.4 XBP1s selectively increases collagen-I secretion from OI patient primary fibroblasts

We next tested whether stress-independent activation of the IRE1 arm of the UPR could increase collagen-I secretion from OI-variant cell lines. IRE1 is a transmembrane protein, spanning the ER membrane. When an unfolded protein stress is sensed on the ER luminal side, IRE1 self-olimogerizes and through a phosphorylation-dependent signaling cascade activates an RNase domain in its cytosolic domain. The RNase domain specifically excises 26 nucleotides from *XBP1u* mRNA to form *XBP1s*. XBP1s is the active form of the transcription factor associated with the IRE1 arm of the UPR.[60] Here, we used replication-incompetent human adenovirus-V to deliver a constitutively expressed XBP1s gene (Ad.XBP1s) to fibroblast cells. We also transduced cells with a constitutively expressed GFP (Ad.GFP) as a negative control transduction, and compared the secretion of collagen-I from each cell line for each sample. Remarkably, we observed a selective increase from two OI-variant fibroblasts that express the most common type of missense mutation in OI, glycine to serine (**Figure 4.5**), with no change in the amount of collagen-I secreted from wild-type fibroblasts. The collagen-I increase from cells expressing Gly502Ser Colα2(I) could not be quantified, as the Ad.GFP transduction shows no detectable collagen-I secretion. We note that the extent of increased secretion from this cell line appears to be sensitive to age in culture and the details of conditions, and although the results are compelling we did not study the Gly502Ser Colα2(I) variant further. Collagen-I secretion from cells expressing Gly247Ser Colα1(I) consistently increases 2–3 fold when treated with Ad.XBP1s compared to Ad.GFP. This result was highly consistent and robust, and we therefore used these cells in the remainder of our analyses.

Notably, we discovered that a very tight window of Ad.XBP1s MOIs was essential for producing the secretion results. To define the MOI at which Ad.XBP1s needed to be transduced, we titrated viral supernatant on a fixed number of wild-type or Gly247Ser Colα1(I)-producing cells. Collagen-I secretion was then induced for 24 h in each treated cell line, and we

analyzed the conditioned media by SDS-PAGE immunoblot (**Figure 4.6A**). Too low of an MOI did not activate XBP1s transcriptional targets sufficiently to increase collagen-I secretion, while too high of an MOI decreased collagen-I secretion likely owing to global UPR activation (**Figure 4.6B**) as also observed when *ec*DHFR.ATF6 was induced to high levels (see above).

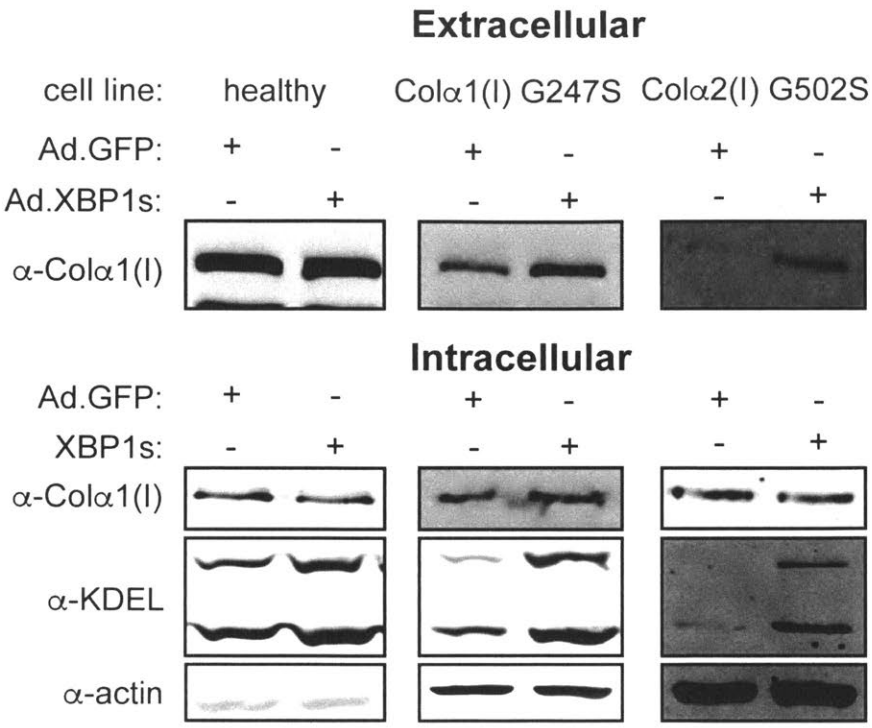


Figure 4.5 | XBP1s activation selectively increases collagen-I secretion from Gly247Ser and Gly502Ser OI patient primary cells

Representative immunoblot analysis of secreted collagen-I from the wild-type, Gly247Ser Colα1(I) and Gly502Ser Colα2(I) cells in the presence of either Ad.GFP or Ad.XBP1s. Top immunoblots are probed for secreted collagen-I in the conditioned media. Bottom immunoblots are intracellular lysate levels of collagen-I, GRP94 and BiP (probed for using the KDEL antibody) as validation of the XBP1s activity, and actin as a loading control.

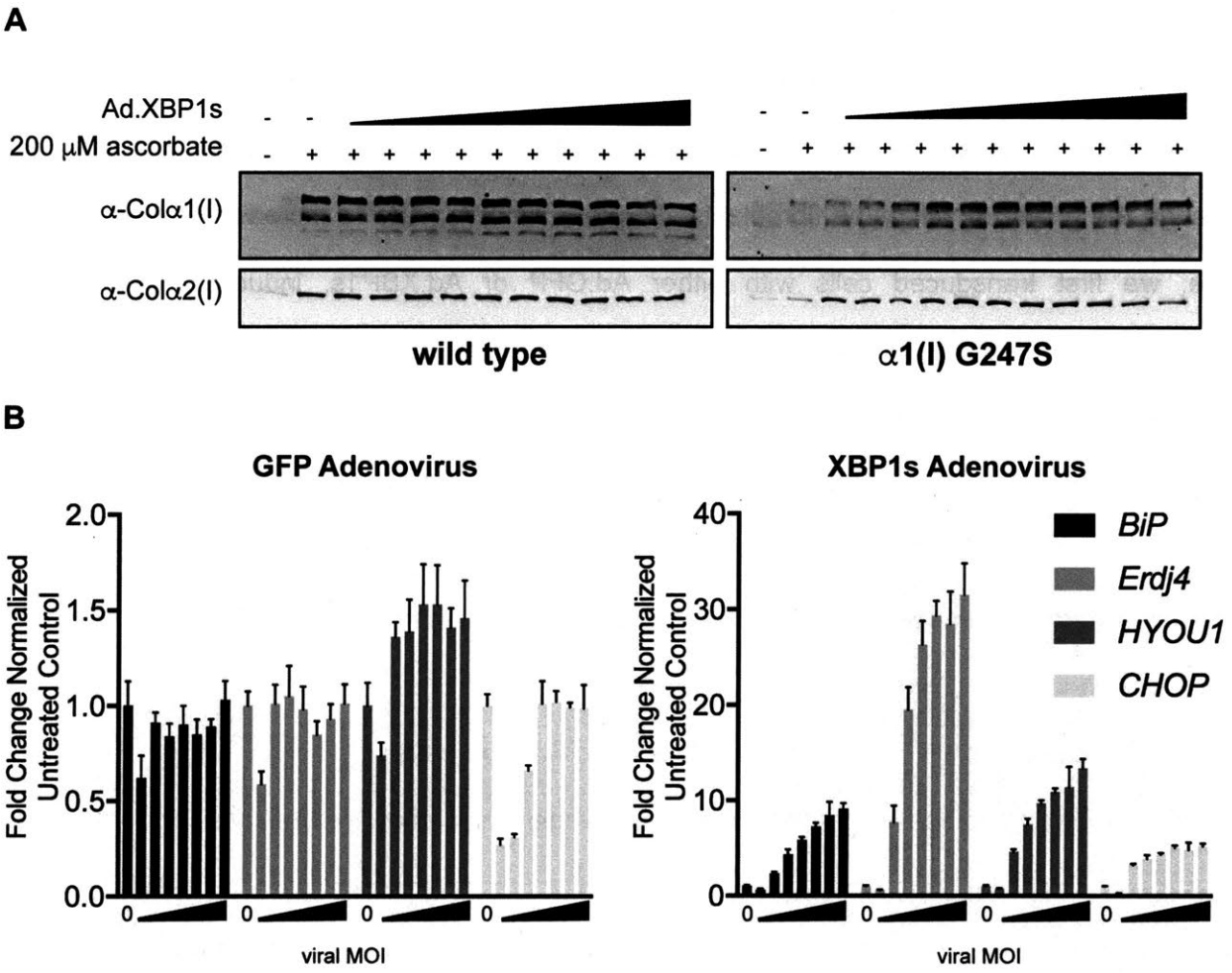


Figure 4.6 | Optimization of adenoviral titering in wild-type and Gly247Ser cells

(A) Cells were transduced with increasing volumes of Ad.XBP1s, induced for collagen-I synthesis, and the media was analyzed by immunoblot. Shown are the results for increasing volumes of Ad.XBP1s, left to right. **(B)** qPCR analysis of the viral titrations. Left panel shows transcriptional upregulation of UPR genes in the presence of Ad.GFP, compared to the right panel when treated with Ad.XBP1s. Error bars represent SEM from three biological replicates.

4.3.5 XBP1s induction does not alter cell viability

A simple explanation for observing an increase in collagen-I secretion is that XBP1s transduction increased the growth rate of the Gly247Ser-producing fibroblasts, but does not change the growth of wild-type fibroblasts. To probe whether XBP1s increases cellular growth rate, we first transduced cells with either Ad.GFP or Ad.XBP1s, induced for collagen-I expression by treating with ascorbate, and assayed cell viability with both resazurin and CellTiter Glo™ assays (**Figure 4.7**). Each sample was normalized to the corresponding negative control treatment of the same cell line. We found that Ad.XBP1s does not change the growth rate of the cells over the time course of our experiments, ruling out altered cell viability/growth as a cause of the observed increased secretion levels.

4.3.6 XBP1s does not alter transcript levels of the collagen-I genes

Another possible explanation for the increased collagen-I secretion we observe upon Ad.XBP1s treatment could be that XBP1s transcriptionally upregulates collagen-I mRNA, thereby increasing the amount of collagen-I synthesized during ascorbate treatment. We used qPCR to compare the levels of *Col1A1* and *Col1A2* mRNA under Ad.GFP and Ad.XBP1s transduction conditions. Neither viral transduction changed the mRNA levels of *Col1A1* or *Col1A2* compared to cells that were not transduced with virus (**Figure 4.8**). More importantly, the mRNA levels between the two viral transductions showed no statistical difference in the Gly247Ser cell line, based on a Student's *t*-test. Therefore, the increase in collagen-I secretion is likely not due to a change in transcription of the collagen-I genes.

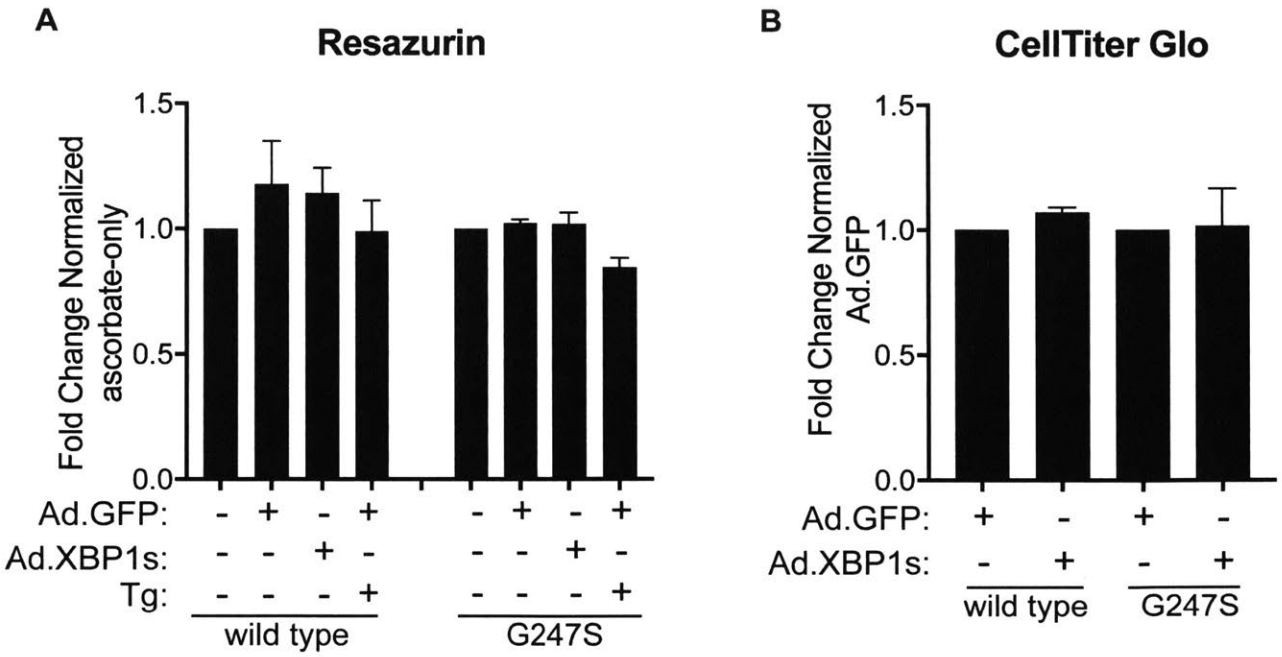


Figure 4.7 | Adenoviral transduction does not change cell viability

(A) Resazurin assay for cell viability comparing different adenoviral transductions, normalized within each cell line, to the ascorbate-only treated control. All cells were treated with ascorbate. **(B)** CellTiter Glo assay for cell viability comparing Ad.XBP1s treatment to Ad.GFP. Data are normalized to Ad.GFP transductions. Error bars for both the CellTiter Glo assay and resazurin represent SEM from the average of six biological replicates.

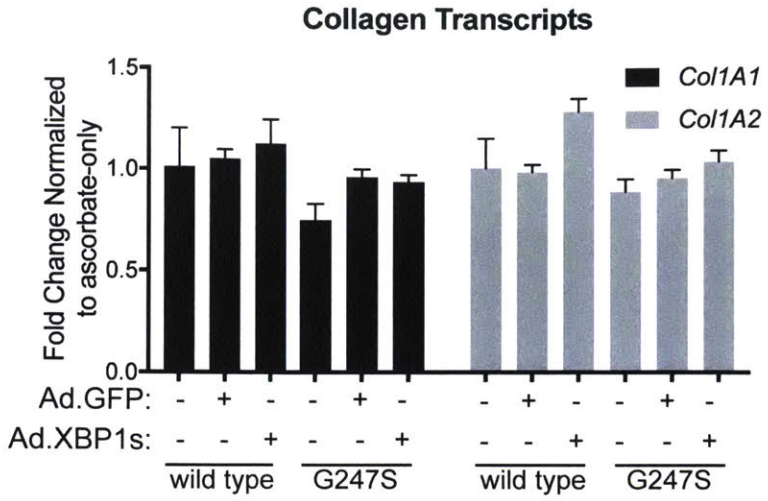


Figure 4.8 | XBP1s does not increase collagen-I transcript levels

qPCR analysis of *Col1A1* (black bars) and *Col1A2* (grey bars), normalized to ascorbate-only control cells. Error bars represent SEM from three biological replicates.

4.3.7 XBP1s does not alter global protein or collagen-I synthesis

While XBP1s did not change the transcript level of collagen-I, it is still possible the translation rate of the collagen-I genes could be altered and result in an increase in secretion. Indeed, translational attenuation is an established UPR mechanism for relieving a protein folding stress.[61] We performed metabolic labeling experiments to determine if the protein synthesis rates are different between Ad.GFP and Ad.XBP1s treated samples. Using ^{35}S-Cys/Met to label all nascent chains synthesized in a ten-minute span, we measured the rates of protein synthesis and analyzed the result by gel electrophoresis. **Figure 4.9A** shows a representative ^{35}S-labeled gel, with the quantitation for 3 biological replicates in **Figure 4.9B**. The total signal in each lane corresponds to the total amount of proteins synthesized during the ^{35}S pulse. The signal for each condition within each cell line tested does not change, regardless of treatment, suggesting that global protein synthesis was not altered by the addition of Ad.XBP1s.

To determine which signal on the gel corresponds to collagen-I synthesized during the pulse, samples were split in half, and treated with or without bacterial collagenase. Collagenase treatment identifies which signal is due to newly synthesized collagen-I. By this analysis, collagen-I synthesis itself also does not change with Ad.XBP1s treatment, compared to control transductions (**Figure 4.9C**). These results indicate that the OI collagen-I secretion enhancement mediated by XBP1s cannot be attributed to either collagen-I-specific or proteome-wide translational effects, acting instead by a post-translational mechanism.

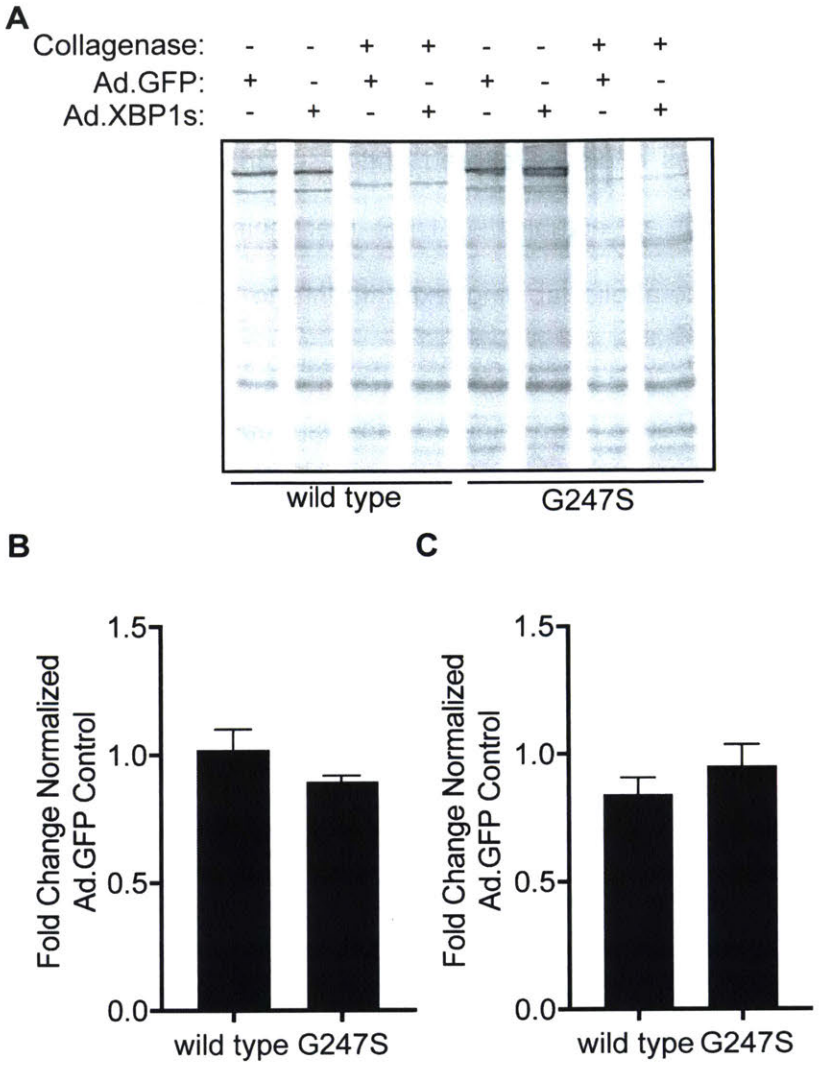


Figure 4.9 | Metabolic labeling reveals no change in proteome or collagen-I synthesis upon XBP1s expression

(A) Representative SDS-PAGE gel of ^{35}S-Cys/Met-labeled protein samples. Samples were split in half and treated with and without bacterial collagenase for 3 d to determine which protein bands were derived from collagen-I signal. **(B)** Quantification of the total signal per lane, averaged over three biological replicates and normalized to the corresponding Ad.GFP for each cell line. Error bars represent SEM. **(C)** Quantification of the collagen-I signal in each cell line transduced with Ad.XBP1s, normalized to the signal in the Ad.GFP transduced sample, averaged over three biological replicates. Error bars represent SEM. Bar charts in both **(B)** and

(C) are normalized to Ad.GFP signal, which was omitted from the plot. Normalized Ad.XBP1s is plotted for each cell line.

4.3.8 XBP1s does not significantly alter triple-helical stability of collagen-I secreted by Gly247Ser OI patient primary fibroblasts

Malformed triple helices are thought contribute to the OI symptoms of brittle, structurally unsound bones, potentially compromising the overall strength of bone.[62,63] One of the unique characteristics of collagen-I is that folded triple helices are resistant to protease digestion.[36,64] Measuring protease stability of triple helices as a function of temperatures is one common assay to assess the stability of collagen-I molecules. We applied this assay to assess whether the enhanced collagen-I secretion by OI patient cells was resulting in production of increasing amounts of malformed collagen-I or collagen-I of similar quality. We began with a mild proteolysis protocol, treating purified procollagen-I with trypsin or chymotrypsin at rt for 2 h, and then analyzed degradation of the triple helices by SDS-PAGE immunoblotting. Prior work indicated the need to test resistance to multiple proteases, because each protease probes different misfolding events.[64] Regardless of whether XBP1s is present or not, collagen-I helices harvested from each cell line are stable to prolonged treatment with trypsin and chymotrypsin (**Figure 4.10A**).

These initial experiments prompted us to more rigorously assess the thermal stability of the secreted collagen-I helices. We used a temperature gradient to probe the temperature at which each set of helices unfolds and thus becomes susceptible to protease digestion. After a 2 min incubation at a range of temperatures between 25–44 °C, procollagen-I samples were cooled to rt and treated with a mixture of trypsin and chymotrypsin for 2 min, after which the reaction was quenched and analyzed by SDS-PAGE immunoblot (**Figure 4.10B**). The data suggest that the helices isolated from each cell line with or without XBP1s activation have roughly the same stability, unfolding within the range of temperatures that wild-type triple helices typically begin to unfold.[36,64-66] XBP1s activity therefore does not decrease the quality of the collagen-I helices secreted.

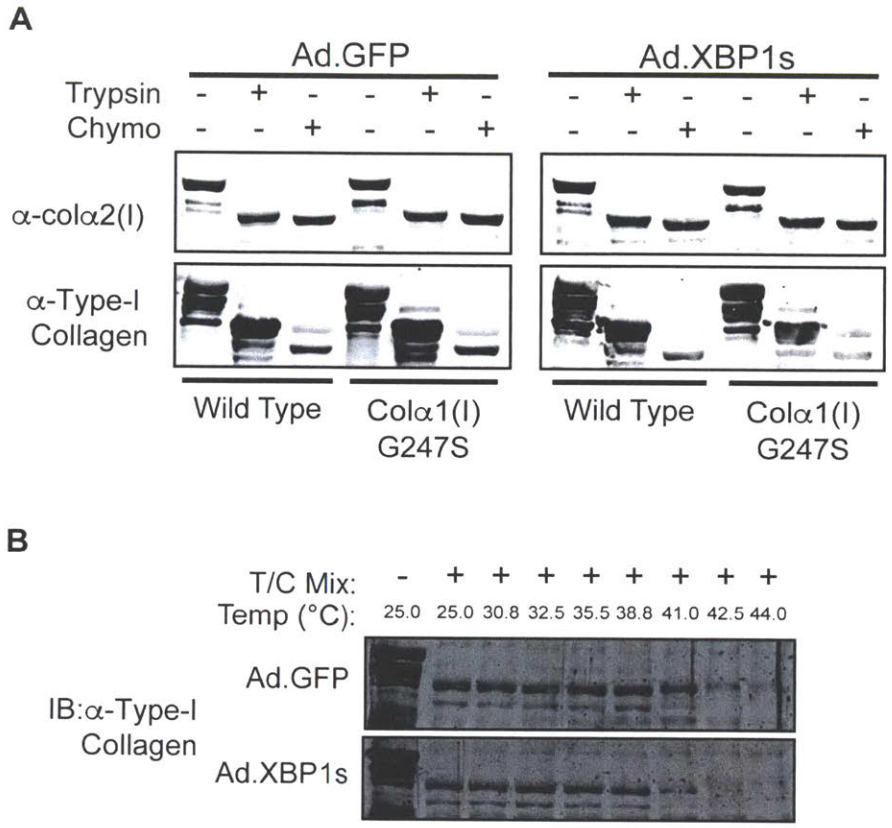


Figure 4.10 | Triple helices produced under XBP1s activated conditions are protease-resistant

(A) Room temperature treatment of purified procollagen-I molecules with trypsin or chymotrypsin for 2 h demonstrates protease resistance in all samples tested. **(B)** Temperature gradient and subsequent treatment with a trypsin and chymotrypsin mixture to assess stability. Both representative immunoblots are representative of collagen-I secreted form the Gly247Ser cells after treatment with Ad.GFP (control) or Ad.XBP1s.

4.3.9 XBP1s activation changes the ratio of wild-type to Gly247Ser variant secreted

We next asked what the ratio of wild-type to Gly247Ser protein was altered upon treatment with Ad.XBP1s compared to Ad.GFP treatment. We first induced collagen-I synthesis and secretion by addition of ascorbate for 24 h with and without XBP1s activation in the Gly247Ser cell line. Media was collected after 24 h and the collagen-I was precipitated, then resolubilized and treated with pepsin for 36–48 h at 4 °C. Stable triple helices were then precipitated with NaCl, solubilized, run on a 6% SDS-PAGE gel, and stained with Coomassie[67] (**Figure 4.11**). These pepsin-treated collagen-I samples demonstrate that the helices are resistant to digestion by yet a third protease over time. Colα1(I) and Colα2(I) separate well, and the Colα1(I) band was excised, reduced, alkylated and trypsinized. The resultant peptides were analyzed by LC-MS/MS analysis to determine the abundance of the wild-type- and Gly247Ser-containing peptides.

The Gly247Ser-containing tryptic peptide and its wild-type counter part eluted from the column with roughly the same retention time (**Figure 4.12A**). **Figure 4.12B** provides a sample MS1 scan demonstrating that we can detect the precursor m/z ratios of the correct mass. To quantify the difference between the wild-type and OI-variant peptides we analyzed the area under the curve of the LC trace for each peptide in **Figure 4.12A**. When transduced with Ad.GFP, the wild-type allelic product is about four-fold more abundant than the serine containing peptides. In contrast, transducing with Ad.XBP1s shifts the ratio to almost 1:1 wild-type to Gly247Ser variant (**Figure 4.12C**). An increase in the OI-variant being secreted from the cell suggests that XBP1s activation rescues the serine variant from an intracellular fate of degradation or aggregation, promoting its assembly instead into folded triple helices that can be secreted.

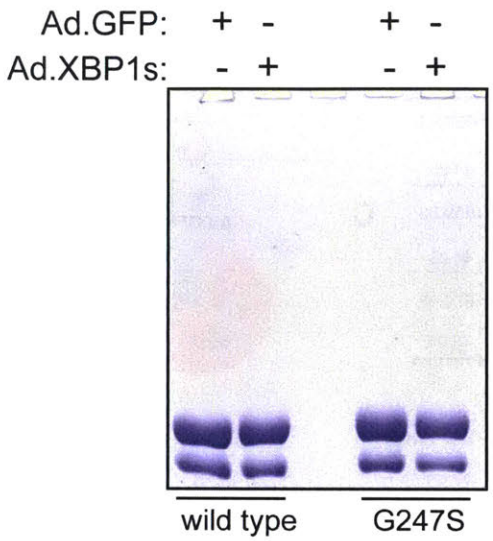


Figure 4.11 | Pepsin resistance of collagen-I triple helices

Coomassie stained SDS-PAGE gel demonstrating stability of collagen-I helices from wild-type and Gly247Ser cells after treatment with Ad.GFP or Ad.XBP1s. Colα1(I) has a slower electrophoretic mobility than Colα2(I). Colα1(I) was extracted from the gel, and analyzed by mass spectrometry in **Figure 4.12**.

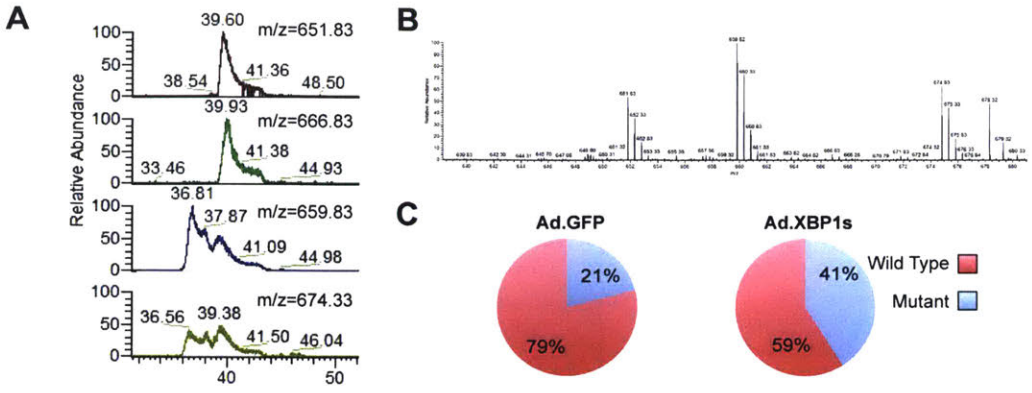


Figure 4.12 | Mass spectrometry analysis of secreted collagen-I demonstrates a change in the ratio of wild-type to mutant allelic product when XBP1s is activated

(A) Representative LC traces for each tryptic peptide of interest. Mass/charge ratios of 651.83 and 666.83 correspond to the peptide without lysine hydroxylation, while the ratios of 659.82 and 674.33 account for the hydroxylated lysine. **(B)** Representative MS1 scan highlighting the four relevant m/z ratios. **(C)** Quantification of the wild-type and mutant allelic products in the presence of Ad.GFP or Ad.XBP1s.

4.3.10 Treatment with Ad.XBP1s rescues collagen-I from autophagic degradation

To address the intracellular fate of collagen-I, we hypothesized that collagen-I was likely being degraded in the absence of XBP1s activation. Some data suggest that collagen-I can be targeted to either endoplasmic reticulum associated degradation (ERAD) or autophagy.[68,69] Thus we inhibited both of these processes to identify the fate of collagen-I in the presence and absence of XBP1s. Unfortunately, proteasome inhibitors cannot be used in fibroblasts when studying collagen-I because they decrease transcript levels of *Col1A1* and *Col1A2*.[70] Therefore, we inhibited ERAD upstream of the proteasome using the VCP (p97) inhibitor eeyarestatin-I, a compound that prevents ER client protein retrotranslocation.[71] We used bafilomycin A to inhibit autophagy.[72] Cells were co-treated with ascorbate and either of these ERAD and autophagy inhibitors to determine if and by what mechanism OI-variant collagen-I was degraded.

All samples were treated in the presence of ascorbate and small molecule inhibitor for 24 h before harvesting and analyzing the conditioned media as well as the lysate. **Figure 4.13** shows representative data for the inhibition studies in wild-type and Gly247Ser-producing cells. In the case of Ad.GFP treatment, VCP inhibition did not alter the levels of collagen-I in lysate compared to DMSO-treated controls. Bafilomycin A, on the other hand, increased the collagen-I signal in the lysate in both wild-type and Gly247Ser lysates under control transduction conditions. In contrast, when either cell line was treated with Ad.XBP1s, the amount of collagen-I rescued by bafilomycin A treatment was drastically reduced, suggesting that the collagen-I was no longer being targeted for autophagy when XBP1s was present. Our results suggest that XBP1s is likely affecting collagen-I folding in such a way that less of the nascent protein is targeted for autophagy.

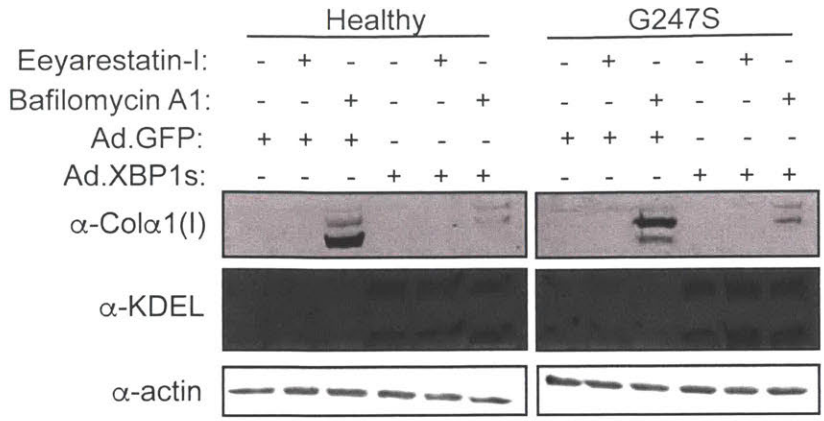


Figure 4.13 | Small molecule inhibitors of ERAD and autophagy demonstrate that XBP1s is rescuing collagen-I from autophagy

Cells were transduced with Ad.GFP or Ad.XBP1s and co-treated with ascorbate and either eeyarestatin-I or bafilomycin. The eeyarestatin concentration was 750 ng/mL and the bafilomycin concentration was 100 nM. Cells were lysed and immunoblotted for KDEL as a marker for XBP1s activation and actin as a loading control.

4.3.11 XBP1s perturbs autophagy in OI patient primary cells

Autophagy is a complex intracellular degradation process. The combination of LC3 lipidation (formation of LC3-II) and p62 degradation are considered the most concrete evidence that autophagy is activatied.[73,74] In contrast, LC3-II formation coinciding with an increase in p62 levels indicates that autophagy is inhibited, unless levels of p62 are perturbed in some other way. We began by testing levels of LC3-II and p62 in lysates of cells treated with either Ad.GFP or Ad.XBP1s. In both wild-type and Gly247Ser cells, we observed a large increase in both LC3-II and p62 levels when Ad.XBP1s was present (**Figure 4.14A**). Given that XBP1s is a transcription factor, a protein-level increase cannot be adequately interpreted until transcript-level changes are also assessed. qPCR confirmed that XBP1s does upregulate both LC3 and p62 at the transcript level (**Figure 4.14B**). Therefore, our data suggest that Ad.XBP1s perturbs autophagic flux, but we cannot specify the direction of that perturbation based on our current data.

4.3.12 Gly247Ser cells exhibit characteristics of defective autophagic flux

An increase in the size and or number of p62-positive protein bodies inside cells is indicative of defective or inhibited autophagy.[75] Positing that perturbations to autophagy by XBP1s played a role in the observed increase in collagen-I secretion, we asked whether autophagic flux was stalled in the presence of the Gly247Ser variant of collagen-I, compared to cell expressing wild-type collagen-I. We transduced cells at the same MOI with baculovirus carrying RFP-p62[76] and monitored fluorescence intensity, size and localization after 72 h of expression. Gly247Ser cells show a massive increase in fluorescence intensity as well as number and size of the fluorescent puncta when compared to wild-type cells (**Figure 4.15**). With a larger amount of puctated RFP signal than the wild-type, this data suggests that basal autophagic flux in the Gly247Ser cells is decreased.

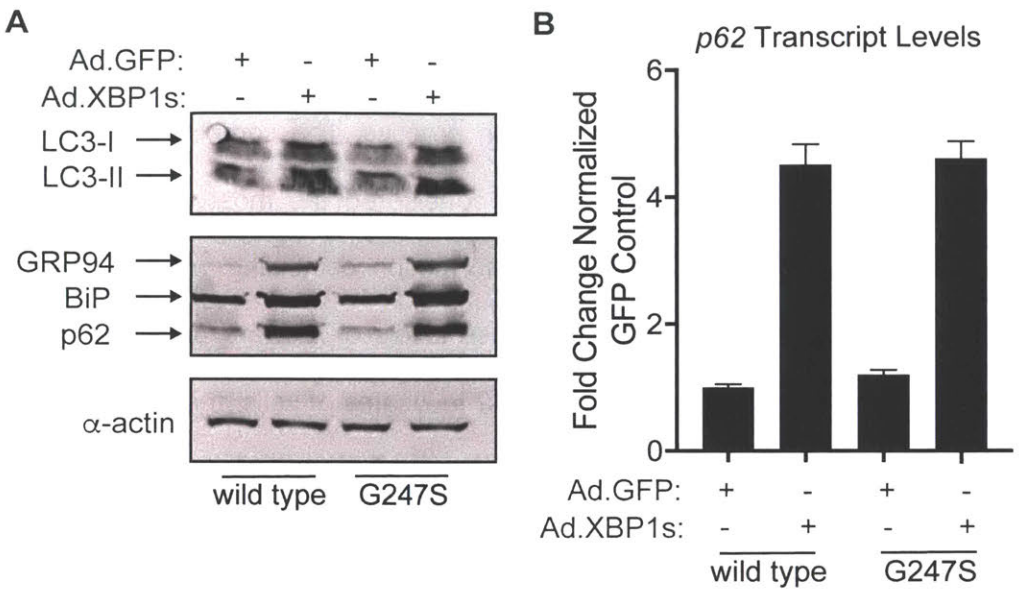


Figure 4.14 | XBP1s upregulates autophagic targets

(A) Protein level changes in LC3 and p62 upon Ad.XBP1s treatment compared to Ad.GFP treatment. GRP94 and BiP upregulation validate XBP1s was active. Immunoblot was probed with the α-KDEL antibody, and then with α-p62. The signal from both antibodies is shown in one image due to their proximity on the membrane. **(B)** Transcript levels of *p62* increase when cells are treated with Ad.XBP1s, normalized to Ad.GFP. Error bars represent SEM from three biological replicates.

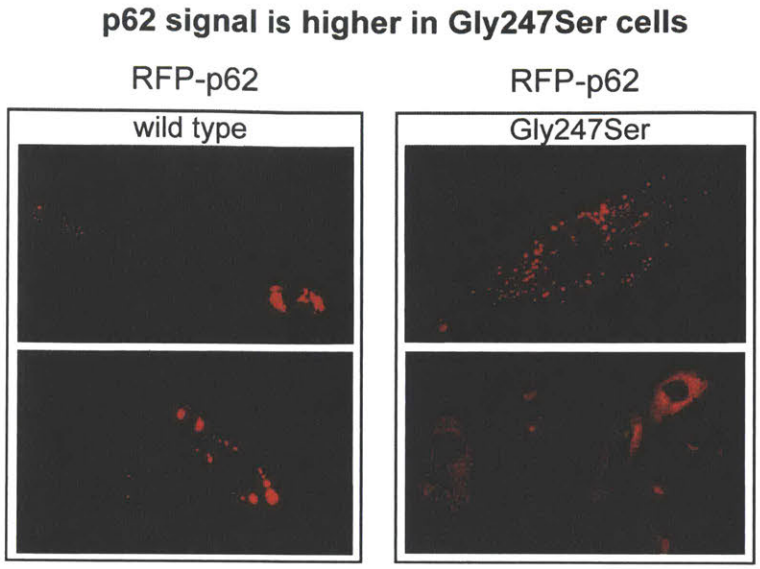


Figure 4.15 | Gly247Ser cells likely exhibit defective autophagic flux

Representative images of wild-type and Gly247Ser cells expressing an RFP-p62 fusion protein. The Gly247Ser cells display many more p62 puncta, as well as much larger staining throughout the cells than that of the wild-type line.

4.3.13 XBP1s-dependent collagen-I secretion relies on the ability to activate autophagy

Our data suggest that autophagy may play an important role in XBP1s-dependent secretion from Gly247Ser cells. We next asked whether autophagy is essential to increase collagen-I secretion from the Gly247Ser cell line, or whether it is a secondary, unrelated effect of XBP1s. We used a genetic approach to uncouple XBP1s perturbations to autophagy from its other proteostasis network effects. In this way, we could determine if autophagy activation was required for the increase in collagen-I secretion to be observed, or whether changes to autophagy were a secondary effect of XBP1s, unrelated to collagen-I secretion. Beclin-1 is a cystosolic protein that regulates and initiates macroautophagy.[73] Beclin-1 is a transcriptional target of XBP1s in endothelial cells,[77-79] and thus we began by knocking it down and monitoring collagen-I secretion in its absence. Using lentivirus we delivered control or beclin-1 targeting-shRNA constructs to wild-type and Gly247Ser cells prior to adenoviral transduction to deliver the XBP1s active construct. Media was then analyzed by SDS-PAGE immunoblotting, and cells were lysed to analyze RNA transcript levels for the targeted proteins.

We first validated beclin-1 knockdown by qPCR, finding only 10–20% percent of the *BECN-1* transcript remaining in the knockdown samples (**Figure 4.16A**). **Figure 4.16B** shows the immunoblot analysis for secreted collagen-I in the presence or absence of XBP1s, with different shRNA constructs transduced into each cell population. There was no detectable change in collagen-I secretion from the Ad.GFP transductions treated with control or beclin-1 targeting shRNAs. In contrast, the control shRNA transduced into Gly247Ser cells with Ad.XBP1s showed the expected increase in collagen-I secretion, as compared to the paired Ad.GFP transduced cells. However, when beclin-1 was knocked down in the presence of Ad.XBP1s, collagen-I secretion decreased to levels comparable to the control transduction. Both beclin-1-targeting shRNA constructs tested produce the same, XBP1s-dependent decrease in collagen-I secretion. Without beclin-1, collagen-I secretion is reduced to basal

levels in the Gly247Ser cells, suggesting that autophagy activation is a necessary part of the XBP1s-dependent result.

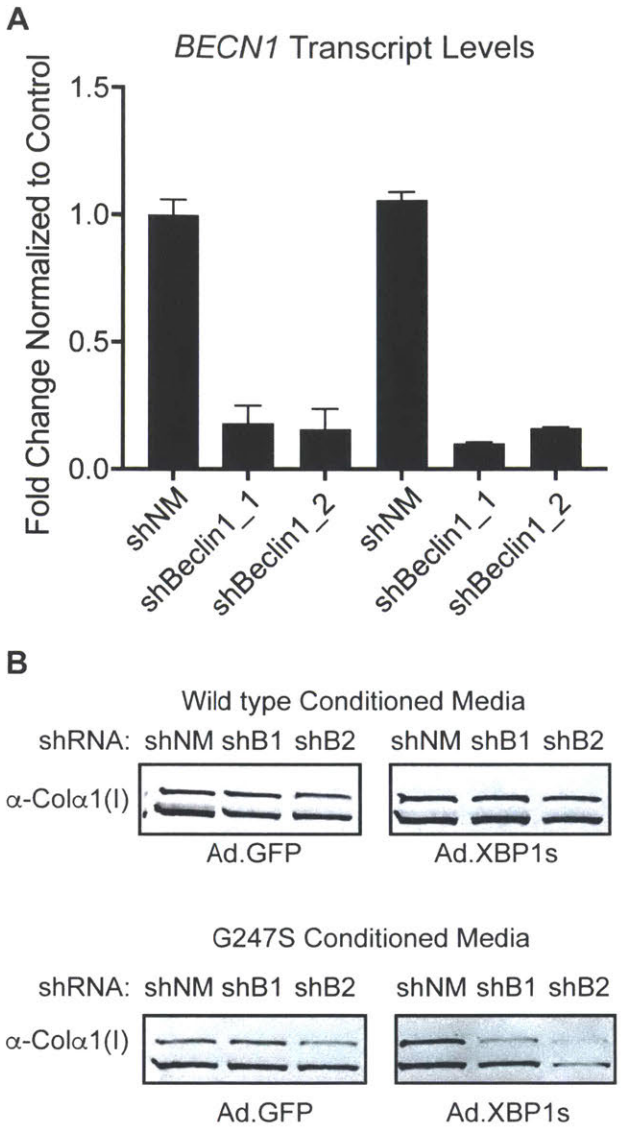


Figure 16 | Autophagic activation is necessary for the XBP1s-dependent increase in secretion

(A) qPCR validating the knockdown of *BECN-1* using two different shRNA constructs. All samples are normalized to the non-mammalian (NM) transduction control, targeting GFP mRNA. Error bars represent SEM from three biological replicates. **(B)** Representative immunoblot data of secreted collagen-I in the presence of each shRNA construct.

4.3.14 Autophagy alone is insufficient to increase collagen-I secretion

Knowing that the XBP1s-dependent secretion result relied on autophagic activation, we next probed whether autophagic activation was sufficient to increase collagen-I secretion from the Gly247Ser cells in the absence of XBP1s. To address this question, we pretreated cells with small molecule activators of autophagy, and assayed for collagen-I secretion after 48 h of activation. Wild-type and Gly247Ser cells were treated with rapamycin,[80] tat-beclin-1,[81] and carbamazepine[82] to activate autophagy, with media changes and drug supplementation every 24 h to ensure autophagy was constitutively activated. After 48 h, media was replaced with the corresponding drug one more time and supplemented with 200 μM ascorbate to stimulate collagen-I synthesis and secretion. After a final 24 h, media was then harvested and analyzed by immunoblot, and cells were harvested to validate that autophagy was activated.

We found that autophagy activation by small molecules is insufficient to increase collagen-I secretion (**Figure 4.17**) with rapamycin actually causing a decrease in collagen-I secretion. Intracellular levels of p62 levels decrease, validating the activation of autophagy. Therefore, autophagy activation alone is insufficient to cause an increase in OI-variant primary cell lines. Resculpting of the ER chaperone landscape by XBP1s is also required.

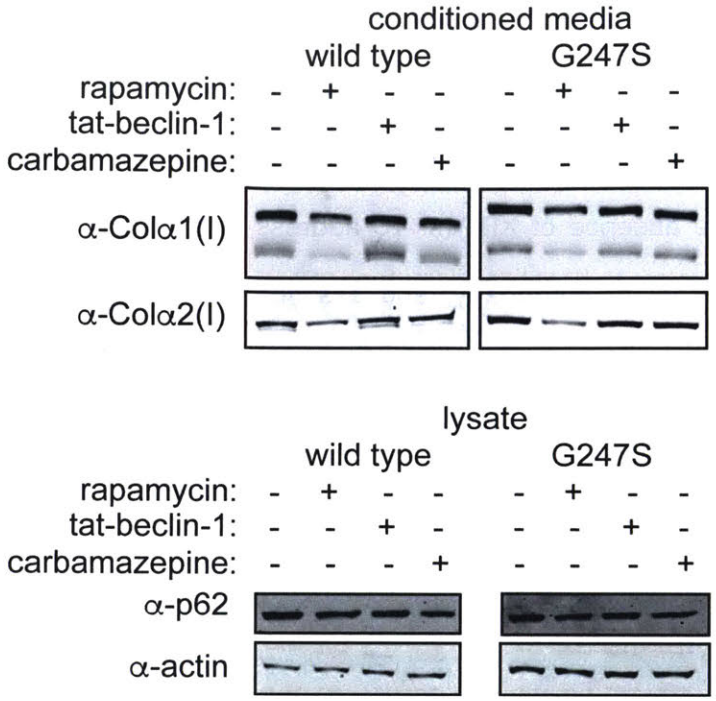


Figure 17 | Autophagic activation alone is insufficient to cause an increase in collagen-I secretion

Immunoblot analysis of secreted collagen-I from cells treated with autophagy activators. The rapamycin concentration was 500 nM. The tat-beclin-I concentration was 10 μM. Carbamazepine concentration was 100 μM. Lysate levels of p62 show a decrease compared to the control sample, confirming that autophagy was activated.

4.4 Concluding Remarks

Our data provide the first example of XBP1s addressing a loss of function defect, here by selectively increasing collagen-I secretion from OI patient-derived primary fibroblasts. The data suggest the following model: Gly247Ser cells accumulate collagen-I intracellularly due to a folding defect. In the absence of XBP1s, collagen-I perpetually accumulates with minimal degradation, and effectively overcrowds the folding environment of the ER. When secretion of collagen-I is induced with ascorbate treatment, the newly synthesized collagen-I strands are blocked from being secreted as a result of the misfolded collagen-I accumulation in the ER. It is tempting to speculate that the chaperones necessary for collagen-I folding are likely saturated by the misfolded variants, leaving nascent chains even more likely to misfold without the assistance from necessary chaperones. XBP1s ameliorates the stress caused by constitutive Gly247Ser misfolding by two mechanisms. Autophagy induction likely clears the blockage of misfolded collagen-I in the ER. Concurrently, XBP1s reloads the ER with an upregulated chaperone network that is no longer saturated by misfolding collagen-I variants. In this manner, when collagen-I synthesis is induced, nascent chains experience an ER with fewer misfolded collagen-I variants, and a higher chaperone capacity to handle the challenge of folding disease-variant forms of collagen-I.

This model suggests a new mechanism of improving proteostasis for a disease variant: a combination of enhanced quality control and improved folding, rather than just one or the other. Activation of a degradation pathway might suggest several characteristics of the misfolded collagen-I variants: (1) these variants are terminal and cannot be rescued to any sufficient extent by disaggregases, holdases or foldases; and (2) the basal expression of the ER proteostasis network does not provide adequate support for recognizing and degrading misfolding variants of a complex protein such as collagen-I.

Our work provides a step forward in our understanding of how cells handle the complex folding problem of collagen-I. Collagen-I is a complex multidomain protein to fold, and as such

even wild-type collagen-I likely misfolds to some degree. If we consider the Gly247Ser variant as an exacerbated version of wild-type collagen-I, it is likely that autophagic elimination of even wild-type collagen-I occurs to some extent. The Gly247Ser variant likely overwhelms the proteostasis system and causes a severe blockage, but similar mechanisms are likely at play with the wild-type form of the protein as well. However, with a lower flux of protein into the misfolded state, the process is likely more efficient.

Future work aims at better understanding this mechanism and investigating the role XBP1s could play in improving organismal health. Ideally, we would validate our findings in cells using small molecule activators of the endogenous IRE1 arm of the UPR. Unfortunately, currently available small molecule activators generally lack specificity and/or potency for IRE1 activation.[83,84] Development of an IRE1 activator would be useful not only for our systems, but for applications beyond the scope of OI, including the other collagenopathies. While testing our results in animal models would be ideal, we note that the currently available OI mouse models are not particularly relevant as they do not display the prototypical glycine to serine mutation in the triple helical domain. One of the main questions with increasing OI-variant secretion is the viability of the organism and how increasing mutant secretion would affect the bone structure, stability and health. Creation of an appropriate OI mouse model is necessary to assess the answers to these questions with greater rigor.

4.5 Experimental Methods and Supplies

4.5.1 Materials and Reagents

EMEM, trypsin, L-glutamine, Penicillin-streptomycin and FBS were all purchased from Corning/Cellgro. Eeyarestatin-I and bafilomycin were purchased from Sigma. Premo-Autophagy RFP-p62 was purchased from Life Technologies.

4.5.2 Cell culture

Primary dermal fibroblasts were obtained from Coriell Cell Repository, with no method to identify the patients' identities. Cells were maintained in EMEM supplemented with 15% FBS, penicillin/streptomycin, and L-glutamine. When sub-cultured, cells were trypsinized and resuspended in ample fresh media, centrifuged at 1000 rpm for 5 min, and resuspended in complete fresh media to count the cells. Cells were counted using the Countess II FL (Life Technologies).

4.5.3 Adenoviral amplification and transduction

1.0×10^5 cells were plated the day before transduction in a 6 well plate. After 24 h, media was replaced and supplemented with adenoviral supernatant. After 24 h of incubation with virus, the media was removed, and the cells were either split accordingly for the appropriate experiments, or supplemented with fresh media. 48 h post media change or split, media was removed and replaced with 1 mL of media with 200 μM ascorbate supplemented every 24 h as necessary. 6–24 h post later, media and cells were harvested for downstream analysis. Cells and viral MOI was scaled accordingly as necessary. Baculoviruses were transduced identically to that of our homemade adenoviruses.

4.5.4 Quantitative PCR

Relative mRNA expression levels of genes of interest were assessed by quantitative RT-PCR. Cells were harvested by trypsinization, washed with PBS, and total RNA was extracted using the Omega RNA Purification Kit according to the manufacturer's instructions. RNA concentrations were quantified and normalized to 0.5–1 μg total RNA for cDNA reverse transcription. Using the Applied Biosystems Reverse Transcriptase cDNA Kit, cDNA was synthesized in a BioRad Thermocycler. LightCycler 480 Sybr Green Master Mix (KAPA), appropriate primers (purchased from Integrated DNA Technologies, Life Technologies and Sigma Aldrich, **Table 1**) and cDNA were used for amplification in a Light Cycler 480 II Real Time PCR Instrument in the MIT BioMicro Center. Primer integrity was assessed by thermal melt and agarose gel analysis to ensure that a single gene was amplified. Transcripts were normalized to the housekeeping gene *Rplp2,* and all measurements were performed in technical and biological triplicate.

4.5.5 Metabolic labeling

72 h post transduction, media was removed from cells and replaced with fresh media with 200 μM ascorbate for 2 h. Media was then replaced with fresh media containing 200 μM ascorbate and 100 mCi/mL of ^{35}S-Cys/Met (Perkin Elmer) for 10 min. Media was removed, cells were washed 2×PBS, and lysed immediately in the plate. After 20 min on ice, samples were collected and centrifuged at 16,400 rpm, 4 °C for 15 minutes. Supernatants were collected and split equally into two tubes, both supplemented with 8× collagenase buffer (50 mM HEPES, pH 7.5, 360 μM calcium chloride). Half of the samples were treated with 5 μg/mL final concentration bacterial collagenase (Worthington Biochemical) for 3 d at 37 °C. Samples were then loaded onto a 10% gel, dried and imaged using the Phosphor screen and Typhoon 7800 after developing for 2 d.

4.5.7 Cell lysis and SDS-PAGE analysis

Cells were trypsinized, the trypsin was neutralized and lysed in unconditioned DMEM. Samples were then centrifuged at 1,500 rpm, for 5 min at rt. Cell pellets were washed 1×PBS, centrifuged again, and lysed in RIPA buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1% Triton-X 100, 0.5% SDS, and 0.1% sodium deoxycholate, a protease inhibitor tablet (Pierce) and 1 mM PMSF). Samples were lysed on ice for 20-30 min, centrifuged at 16,800 rpm at 4 °C, quantified and normalized for SDS-PAGE analysis. Protein gels were transferred to nitrocellulose membranes using the TransBlot Turbo system from BioRad, blocked in 5% non-fat milk w/v and probed with the corresponding primary antibody for 2 h–o/n. Images were analyzed using an Odyssey Scanner.

4.5.8 Procollagen-I purification and proteolysis

2.50×10^5 cells were plated in a 6cm dish, and transduced 24 h later with Ad.GFP or Ad.XBP1s. Media was changed on the cells after 24 h, and 72 h post-transduction the cells were induced with 200 μM ascorbate for 24–48 h. Media samples were then collected, chilled and treated with 1 mM PMSF, 100 mM Tris-HCl, pH 7.4, and 176 mg/mL ammonium sulfate, o/n end over end mixing at 4 °C. The next day, samples were centrifuged and the supernatant removed, with the pellet being resolubilized in 400 mM NaCl, 100 mM Tris-HCl, pH 7.4. Procollagen-I samples were then aliquotted accordingly, and treated with 0.1 mg/mL trypsin and/or 0.2 mg/mL chymotrypsin, or 0.2 mg/mL pepsin.

4.5.9 Autophagy visualization

Cells were first plated, and transduced with the corresponding viruses for 24 h, when the media was replaced with fresh media. We used an MOI of about 30 for the Premo Autophagy RFP-p62. After 24 h, media was replaced with fresh media, and cells were imaged.

4.6 References

(1) Bateman, J. F.; Wilson, R.; Freddi, S.; Lamande, S. R.; Savarirayan, R. Mutations of COL10A1 in Schmid metaphyseal chondrodysplasia. *Hum. Mutat.* **2005**, *25*, 525-534.

(2) Bouma, P.; Cabral, W. A.; Cole, W. G.; Marini, J. C. COL5A1 exon 14 splice acceptor mutation causes a functional null allele, haploinsufficiency of alpha 1(V) and abnormal heterotypic interstitial fibrils in Ehlers-Danlos syndrome II. *J. Biol. Chem.* **2001**, *276*, 13356-13364.

(3) Lamande, S. R.; Bateman, J. F.; Hutchison, W.; McKinlay Gardner, R. J.; Bower, S. P.; Byrne, E.; Dahl, H. H. Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum. Mol. Genet.* **1998**, *7*, 981-989.

(4) Jones, F. E.; Bailey, M. A.; Murray, L. S.; Lu, Y.; McNeilly, S.; Schlotzer-Schrehardt, U.; Lennon, R.; Sado, Y.; Brownstein, D. G.; Mullins, J. J.; Kadler, K. E.; Van Agtmael, T. ER stress and basement membrane defects combine to cause glomerular and tubular renal disease resulting from Col4a1 mutations in mice. *Dis. Model Mech.* **2016**, *9*, 165-176.

(5) Mirigian, L. S.; Makareeva, E.; Mertz, E. L.; Omari, S.; Roberts-Pilgrim, A. M.; Oestreich, A. K.; Phillips, C. L.; Leikin, S. Osteoblast Malfunction Caused by Cell Stress Response to Procollagen Misfolding in alpha2(I)-G610C Mouse Model of Osteogenesis Imperfecta. *J. Bone Miner. Res.* **2016**, *31*, 1608-1616.

(6) Gawron, K. Endoplasmic reticulum stress in chondrodysplasias caused by mutations in collagen types II and X. *Cell Stress Chaper.* **2016**, *21*, 943-958.

(7) Jobling, R.; D'Souza, R.; Baker, N.; Lara-Corrales, I.; Mendoza-Londono, R.; Dupuis, L.; Savarirayan, R.; Ala-Kokko, L.; Kannu, P. The collagenopathies: review of clinical phenotypes and molecular correlations. *Curr. Rheumatol. Rep.* **2014**, *16*, 394.

(8) Barnes, A. M.; Cabral, W. A.; Weis, M.; Makareeva, E.; Mertz, E. L.; Leikin, S.; Eyre, D.; Trujillo, C.; Marini, J. C. Absence of FKBP10 in recessive type XI osteogenesis imperfecta leads to diminished collagen cross-linking and reduced collagen deposition in extracellular matrix. *Hum. Mutat.* **2012**, *33*, 1589-1598.

(9) Marini, J. C.; Forlino, A.; Bachinger, H. P.; Bishop, N. J.; Byers, P. H.; Paepe, A.; Fassier, F.; Fratzl-Zelman, N.; Kozloff, K. M.; Krakow, D.; Montpetit, K.; Semler, O. Osteogenesis imperfecta. *Nat. Rev. Dis. Primers* **2017**, *3*, 17052.

(10) Webb, E. A.; Balasubramanian, M.; Fratzl-Zelman, N.; Cabral, W. A.; Titheradge, H.; Alsaedi, A.; Saraff, V.; Vogt, J.; Cole, T.; Stewart, S.; Crabtree, N. J.; Sargent, B. M.; Gamsjaeger, S.; Paschalis, E. P.; Roschger, P.; Klaushofer, K.; Shaw, N. J.; Marini, J. C.; Hogler, W. Phenotypic Spectrum in Osteogenesis Imperfecta Due to Mutations in TMEM38B: Unraveling a Complex Cellular Defect. *J. Clin. Endocrinol. Metab.* **2017**, *102*, 2019-2028.

(11) Lietman, C. D.; Rajagopal, A.; Homan, E. P.; Munivez, E.; Jiang, M. M.; Bertin, T. K.; Chen, Y.; Hicks, J.; Weis, M.; Eyre, D.; Lee, B.; Krakow, D. Connective tissue alterations in Fkbp10-/- mice. *Hum. Mol. Genet.* **2014**, *23*, 4822-4831.

(12) Valli, M.; Barnes, A. M.; Gallanti, A.; Cabral, W. A.; Viglio, S.; Weis, M. A.; Makareeva, E.; Eyre, D.; Leikin, S.; Antoniazzi, F.; Marini, J. C.; Mottes, M. Deficiency of CRTAP in non-lethal recessive osteogenesis imperfecta reduces collagen deposition into matrix. *Clin. Genet.* **2012**, *82*, 453-459.

(13) Van Dijk, F. S.; Sillence, D. O. Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. *Am. J. Med. Genet. A.* **2014**, *164A*, 1470-1481.

(14) Drake, M. T.; Clarke, B. L.; Khosla, S. Bisphosphonates: mechanism of action and role in clinical practice. *Mayo Clin. Proc.* **2008**, *83*, 1032-1045.

(15) Castillo, H.; Samson-Fang, L.; American Academy for Cerebral, P.; Developmental Medicine Treatment Outcomes Committee Review, P. Effects of bisphosphonates in children with osteogenesis imperfecta: an AACPDM systematic review. *Dev. Med. Child. Neurol.*. **2009**, *51*, 17-29.

(16) Marini, J. C. Bone: Use of bisphosphonates in children-proceed with caution. *Nat. Re. Endocrinol.* **2009**, *5*, 241-243.

(17) Marini, J. C. Should children with osteogenesis imperfecta be treated with bisphosphonates? *Nat. Clin. Pract. Endocrinol. Metab.* **2006**, *2*, 14-15.

(18) Letocha, A. D.; Cintas, H. L.; Troendle, J. F.; Reynolds, J. C.; Cann, C. E.; Chernoff, E. J.; Hill, S. C.; Gerber, L. H.; Marini, J. C. Controlled trial of pamidronate in children with types III and IV osteogenesis imperfecta confirms vertebral gains but not short-term functional improvement. *J. Bone Miner. Res.* **2005**, *20*, 977-986.

(19) Marini, J. C. Do bisphosphonates make children's bones better or brittle? *N. Engl. J. Med.* **2003**, *349*, 423-426.

(20) Marini, J. C.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; San Antonio, J. D.; Milgrom, S.; Hyland, J. C.; Korkko, J.; Prockop, D. J.; De Paepe, A.; Coucke, P.; Symoens, S.; Glorieux, F. H.; Roughley, P. J.; Lund, A. M.; Kuurila-Svahn, K.; Hartikka, H.; Cohn, D. H.; Krakow, D.; Mottes, M.; Schwarze, U.; Chen, D.; Yang, K.; Kuslich, C.; Troendle, J.; Dalgleish, R.; Byers, P. H. 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. *Hum. Mutat.* **2007**, *28*, 209-221.

(21) Dalgleish, R. The human type I collagen mutation database. *Nucleic Acids Res.* **1998**, *25*, 181-187.

(22) Dalgleish, R. The human collagen mutation database 1998. *Nucleic Acids Res.* **1998**, *26*, 253-255.

(23) Bella, J.; Brodsky, B.; Berman, H. M. Disrupted collagen architecture in the crystal structure of a triple-helical peptide with a Gly-->Ala substitution. *Connect. Tissue Res.* **1996**, *35*, 401-406.

(24) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **1994**, *266*, 75-81.

(25) Mundlos, S.; Chan, D.; McGill, J.; Bateman, J. F. An alpha 1(II) Gly913 to Cys substitution prevents the matrix incorporation of type II collagen which is replaced with type I and III collagens in cartilage from a patient with hypochondrogenesis. *Am. J. Med. Genet.* **1996**, *63*, 129-136.

(26) Kuznetsova, N. V.; Forlino, A.; Cabral, W. A.; Marini, J. C.; Leikin, S. Structure, stability and interactions of type I collagen with GLY349-CYS substitution in alpha 1(I) chain in a murine Osteogenesis Imperfecta model. *Matrix Biol.* **2004**, *23*, 101-112.

(27) Cabral, W. A.; Chernoff, E. J.; Marini, J. C. G76E substitution in type I collagen is the first nonlethal glutamic acid substitution in the alpha1(I) chain and alters folding of the N-terminal end of the helix. *Mol. Genet. Metab.* **2001**, *72*, 326-335.

(28) Forlino, A.; Porter, F. D.; Lee, E. J.; Westphal, H.; Marini, J. C. Use of the Cre/lox recombination system to develop a non-lethal knock-in murine model for osteogenesis imperfecta with an alpha1(I) G349C substitution. Variability in phenotype in BrtIIV mice. *J. Biol. Chem.* **1999**, *274*, 37923-37931.

(29) Dawson, P. A.; Kelly, T. E.; Marini, J. C. Extension of phenotype associated with structural mutations in type I collagen: siblings with juvenile osteoporosis have an alpha2(I)Gly436 --> Arg substitution. *J. Bone Miner. Res.* **1999**, *14*, 449-455.

(30) Bateman, J. F.; Chiodo, A. A.; Weng, Y. M.; Chan, D.; Haan, E. A type III collagen Gly559 to Arg helix mutation in Ehler's-Danlos syndrome type IV. *Hum. Mutat.* **1998**, *Suppl 1*, S257-259.

(31) Mackay, K.; Byers, P. H.; Dalgleish, R. An RT-PCR-SSCP screening strategy for detection of mutations in the gene encoding the alpha 1 chain of type I collagen: application to four patients with osteogenesis imperfecta. *Hum. Mol. Genet.* **1993**, *2*, 1155-1160.

(32) Konstantinidou, A. E.; Agrogiannis, G.; Sifakis, S.; Karantanas, A.; Harakoglou, V.; Kaminopetros, P.; Hatzaki, A.; Petersen, M. B.; Karadimas, C.; Velissariou, V.; Velonis, S.; Papantoniou, N.; Antsaklis, A.; Patsouris, E. Genetic skeletal disorders of the fetus and infant: pathologic and molecular findings in a series of 41 cases. *Birth Defects Res. A. Clin. Mol. Teratol.* **2009**, *85*, 811-821.

(33) Bodian, D. L.; Chan, T. F.; Poon, A.; Schwarze, U.; Yang, K.; Byers, P. H.; Kwok, P. Y.; Klein, T. E. Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype-phenotype relationships. *Hum. Mol. Genet.* **2009**, *18*, 463-471.

(34) Ben Amor, I. M.; Glorieux, F. H.; Rauch, F. Genotype-phenotype correlations in autosomal dominant osteogenesis imperfecta. *J. Osteoporos.* **2011**, *2011*, 540178.

(35) Pace, J. M.; Wiese, M.; Drenguis, A. S.; Kuznetsova, N.; Leikin, S.; Schwarze, U.; Chen, D.; Mooney, S. H.; Unger, S.; Byers, P. H. Defective C-propeptides of the proalpha2(I) chain of type I procollagen impede molecular assembly and result in osteogenesis imperfecta. *J. Biol. Chem.* **2008**, *283*, 16061-16067.

(36) Pace, J. M., Kuslich, C.D., Willing, M.C. Byers, P.H. Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the pro-alpha1(I) chain of type I collagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta. *J. Med. Genet.* **2001**, *38*, 443-449.

(37) Wong, M. Y.; DiChiara, A. S.; Suen, P. H.; Chen, K.; Doan, N. D.; Shoulders, M. D. Adapting Secretory Proteostasis and Function Through the Unfolded Protein Response. *Curr. Top. Microbiol. Immunol.* **2017**.

(38) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R., 3rd; Su, A. I.; Kelly, J. W.; Wiseman, R. L. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013**, *3*, 1279-1292.

(39) Pankow, S.; Bamberger, C.; Calzolari, D.; Martinez-Bartolome, S.; Lavallee-Adam, M.; Balch, W. E.; Yates, J. R., 3rd F508 CFTR interactome remodelling promotes rescue of cystic fibrosis. *Nature* **2015**, *528*, 510-516.

(40) Mu, T. W.; Ong, D. S.; Wang, Y. J.; Balch, W. E.; Yates, J. R., 3rd; Segatori, L.; Kelly, J. W. Chemical and biological approaches synergize to ameliorate protein-folding diseases. *Cell* **2008**, *134*, 769-781.

(41) Wang, X.; Venable, J.; LaPointe, P.; Hutt, D. M.; Koulov, A. V.; Coppinger, J.; Gurkan, C.; Kellner, W.; Matteson, J.; Plutner, H.; Riordan, J. R.; Kelly, J. W.; Yates, J. R., 3rd; Balch, W. E. Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* **2006**, *127*, 803-815.

(42) Pratt, W. B.; Gestwicki, J. E.; Osawa, Y.; Lieberman, A. P. Targeting Hsp90/Hsp70-based protein quality control for treatment of adult onset neurodegenerative diseases. *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 353-371.

(43) Hulleman, J. D.; Kaushal, S.; Balch, W. E.; Kelly, J. W. Compromised mutant EFEMP1 secretion associated with macular dystrophy remedied by proteostasis network alteration. *Mol. Biol. Cell* **2011**, *22*, 4765-4775.

(44) Ryno, L. M.; Wiseman, R. L.; Kelly, J. W. Targeting unfolded protein response signaling pathways to ameliorate protein misfolding diseases. *Curr. Opin. Chem. Biol.* **2013**, *17*, 346-352.

(45) Chen, J. J.; Genereux, J. C.; Qu, S.; Hulleman, J. D.; Shoulders, M. D.; Wiseman, R. L. ATF6 activation reduces the secretion and extracellular aggregation of destabilized variants of an amyloidogenic protein. *Chem. Biol.* **2014**, *21*, 1564-1574.

(46) Cooley, C. B.; Ryno, L. M.; Plate, L.; Morgan, G. J.; Hulleman, J. D.; Kelly, J. W.; Wiseman, R. L. Unfolded protein response activation reduces secretion and extracellular aggregation of amyloidogenic immunoglobulin light chain. *Proc. Natl. Acad. Sci. U S A* **2014**, *111*, 13046-13051.

(47) Boot-Handford, R. P.; Briggs, M. D. The unfolded protein response and its relevance to connective tissue diseases. *Cell Tissue Res.* **2010**, *339*, 197-211.

(48) Maji, B.; Moore, C. L.; Zetsche, B.; Volz, S. E.; Zhang, F.; Shoulders, M. D.; Choudhary, A. Multidimensional chemical control of CRISPR-Cas9. *Nat. Chem. Biol.* **2017**, *13*, 9-11.

(49) Chan, D.; Lamande, S. R.; Cole, W. G.; Bateman, J. F. Regulation of procollagen synthesis and processing during ascorbate-induced extracellular matrix accumulation in vitro. *Biochem. J.* **1990**, *269*, 175-181.

(50) Forlino, A.; Kuznetsova, N. V.; Marini, J. C.; Leikin, S. Selective retention and degradation of molecules with a single mutant alpha1(I) chain in the Brtl IV mouse model of OI. *Matrix Biol.* **2007**, *26*, 604-614.

(51) Masci, M.; Wang, M.; Imbert, L.; Barnes, A. M.; Spevak, L.; Lukashova, L.; Huang, Y.; Ma, Y.; Marini, J. C.; Jacobsen, C. M.; Warman, M. L.; Boskey, A. L. Bone mineral properties in growing Col1a2(+/G610C) mice, an animal model of osteogenesis imperfecta. *Bone* **2016**, *87*, 120-129.

(52) Jacobsen, C. M.; Schwartz, M. A.; Roberts, H. J.; Lim, K. E.; Spevak, L.; Boskey, A. L.; Zurakowski, D.; Robling, A. G.; Warman, M. L. Enhanced Wnt signaling improves bone mass and strength, but not brittleness, in the Col1a1(+/mov13) mouse model of type I Osteogenesis Imperfecta. *Bone* **2016**, *90*, 127-132.

(53) Jacobsen, C. M.; Barber, L. A.; Ayturk, U. M.; Roberts, H. J.; Deal, L. E.; Schwartz, M. A.; Weis, M.; Eyre, D.; Zurakowski, D.; Robling, A. G.; Warman, M. L. Targeting the LRP5 pathway improves bone properties in a mouse model of osteogenesis imperfecta. *J Bone Miner. Res.* **2014**, *29*, 2297-2306.

(54) Lamande, S. R.; Bateman, J. F. The type I collagen pro alpha 1(I) COOH-terminal propeptide N-linked oligosaccharide. Functional analysis by site-directed mutagenesis. *J. Biol. Chem.* **1995**, *270*, 17858-17865.

(55) Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R. N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671.

(56) Martinez-Martinez, E.; Ibarrola, J.; Fernandez-Celis, A.; Santamaria, E.; Fernandez-Irigoyen, J.; Rossignol, P.; Jaisser, F.; Lopez-Andres, N. Differential Proteomics Identifies Reticulocalbin-3 as a Novel Negative Mediator of Collagen Production in Human Cardiac Fibroblasts. *Sci. Rep.* **2017**, *7*, 12192.

(57) DiChiara, A. S.; Taylor, R. J.; Wong, M. Y.; Doan, N. D.; Rosario, A. M.; Shoulders, M. D. Mapping and Exploring the Collagen-I Proteostasis Network. *ACS Chem. Biol.* **2016**, *11*, 1408-1421.

(58) 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.* **2010**, *285*, 7067-7078.

(59) Rakhit, R.; Edwards, S. R.; Iwamoto, M.; Wandless, T. J. Evaluation of FKBP and DHFR based destabilizing domains in Saccharomyces cerevisiae. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4965-4968.

(60) Back, S. H.; Schroder, M.; Lee, K.; Zhang, K.; Kaufman, R. J. ER stress signaling by regulated splicing: IRE1/HAC1/XBP1. *Methods* **2005**, *35*, 395-416.

(61) Trusina, A.; Papa, F. R.; Tang, C. Rationalizing translation attenuation in the network architecture of the unfolded protein response. *Proc. Natl. Acad. Sci. U S A* **2008**, *105*, 20280-20285.

(62) Torre-Blanco, A.; Adachi, E.; Romanic, A. M.; Prockop, D. J. Copolymerization of normal type I collagen with three mutated type I collagens containing substitutions of cysteine at different glycine positions in the alpha 1 (I) chain. *J. Biol. Chem.* **1992**, *267*, 4968-4973.

(63) Sweeney, S. M.; Orgel, J. P.; Fertala, A.; McAuliffe, J. D.; Turner, K. R.; Di Lullo, G. A.; Chen, S.; Antipova, O.; Perumal, S.; Ala-Kokko, L.; Forlino, A.; Cabral, W. A.; Barnes, A. M.; Marini, J. C.; San Antonio, J. D. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* **2008**, *283*, 21187-21197.

(64) Bruckner, P.; Prockop, D. J. Proteolytic enzymes as probes for the triple-helical conformation of procollagen. *Anal. Biochem.* **1981**, *110*, 360-368.

(65) Han, S.; Makareeva, E.; Kuznetsova, N. V.; DeRidder, A. M.; Sutter, M. B.; Losert, W.; Phillips, C. L.; Visse, R.; Nagase, H.; Leikin, S. Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *J. Biol. Chem.* **2010**, *285*, 22276-22281.

(66) Makareeva, E.; Mertz, E. L.; Kuznetsova, N. V.; Sutter, M. B.; DeRidder, A. M.; Cabral, W. A.; Barnes, A. M.; McBride, D. J.; Marini, J. C.; Leikin, S. Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J. Biol. Chem.* **2008**, *283*, 4787-4798.

(67) Weis, M. A.; Hudson, D. M.; Kim, L.; Scott, M.; Wu, J. J.; Eyre, D. R. Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J. Biol. Chem.* **2010**, *285*, 2580-2590.

(68) Ishida, Y.; Kubota, H.; Yamamoto, A.; Kitamura, A.; Bachinger, H. P.; Nagata, K. Type I collagen in Hsp47-null cells is aggregated in endoplasmic reticulum and deficient in N-propeptide processing and fibrillogenesis. *Mol. Biol. Cell* **2006**, *17*, 2346-2355.

(69) Lamande, S. R.; Chessler, S. D.; Golub, S. B.; Byers, P. H.; Chan, D.; Cole, W. G.; Sillence, D. O.; Bateman, J. F. Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro alpha 1 (I) chain carboxyl-terminal propeptide which impair subunit assembly. *J. Biol. Chem.* **1995**, *270*, 8642-8649.

(70) Pujols, L.; Fernandez-Bertolin, L.; Fuentes-Prado, M.; Alobid, I.; Roca-Ferrer, J.; Agell, N.; Mullol, J.; Picado, C. Proteasome inhibition reduces proliferation, collagen expression, and inflammatory cytokine production in nasal mucosa and polyp fibroblasts. *J. Pharmacol. Exp. Ther.* **2012**, *343*, 184-197.

(71) Wang, Q.; Shinkre, B. A.; Lee, J. G.; Weniger, M. A.; Liu, Y.; Chen, W.; Wiestner, A.; Trenkle, W. C.; Ye, Y. The ERAD inhibitor Eeyarestatin I is a bifunctional compound with a membrane-binding domain and a p97/VCP inhibitory group. *PLoS One* **2010**, *5*, e15479.

(72) Mauvezin, C.; Nagy, P.; Juhasz, G.; Neufeld, T. P. Autophagosome-lysosome fusion is independent of V-ATPase-mediated acidification. *Nat. Commun.* **2015**, *6*, 7007.

(73) Ichimura, Y.; Komatsu, M. Selective degradation of p62 by autophagy. *Semin. Immunopathol.* **2010**, *32*, 431-436.

(74) Tanida, I.; Ueno, T.; Kominami, E. LC3 conjugation system in mammalian autophagy. *Int. J .Biochem. Cell Biol.* **2004**, *36*, 2503-2518.

(75) Bjorkoy, G.; Lamark, T.; Brech, A.; Outzen, H.; Perander, M.; Overvatn, A.; Stenmark, H.; Johansen, T. p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *J.. Cell Biol.* **2005**, *171*, 603-614.

(76) Zhang, Y. B.; Gong, J. L.; Xing, T. Y.; Zheng, S. P.; Ding, W. Autophagy protein p62/SQSTM1 is involved in HAMLET-induced cell death by modulating apotosis in U87MG cells. *Cell. Death Dis.* **2013**, *4*, e550.

(77) Sinha, S.; Levine, B. The autophagy effector Beclin 1: a novel BH3-only protein. *Oncogene* **2008**, *27 Suppl 1*, S137-148.

(78) Tian, P. G.; Jiang, Z. X.; Li, J. H.; Zhou, Z.; Zhang, Q. H. Spliced XBP1 promotes macrophage survival and autophagy by interacting with Beclin-1. *Biochem. Biophys. Res. Commun.* **2015**, *463*, 518-523.

(79) Margariti, A.; Li, H.; Chen, T.; Martin, D.; Vizcay-Barrena, G.; Alam, S.; Karamariti, E.; Xiao, Q.; Zampetaki, A.; Zhang, Z.; Wang, W.; Jiang, Z.; Gao, C.; Ma, B.; Chen, Y. G.; Cockerill, G.; Hu, Y.; Xu, Q.; Zeng, L. XBP1 mRNA splicing triggers an autophagic response in endothelial cells through BECLIN-1 transcriptional activation. *J. Biol. Chem.* **2013**, *288*, 859-872.

(80) Ishida, Y.; Yamamoto, A.; Kitamura, A.; Lamande, S. R.; Yoshimori, T.; Bateman, J. F.; Kubota, H.; Nagata, K. Autophagic elimination of misfolded procollagen aggregates in the endoplasmic reticulum as a means of cell protection. *Mol. Biol. Cell* **2009**, *20*, 2744-2754.

(81) Shoji-Kawata, S.; Sumpter, R.; Leveno, M.; Campbell, G. R.; Zou, Z.; Kinch, L.; Wilkins, A. D.; Sun, Q.; Pallauf, K.; MacDuff, D.; Huerta, C.; Virgin, H. W.; Helms, J. B.; Eerland, R.; Tooze, S. A.; Xavier, R.; Lenschow, D. J.; Yamamoto, A.; King, D.; Lichtarge, O.; Grishin, N. V.; Spector, S. A.; Kaloyanova, D. V.; Levine, B. Identification of a candidate therapeutic autophagy-inducing peptide. *Nature* **2013**, *494*, 201-206.

(82) Hidvegi, T., Ewing, M., Hale, P., Dippold, C., Beckett, C., Kemp, C., Maurice, N., Mukherjee, A., Goldbach, C., Watkins, S., Michalopoulos, G., Perlmutter An autophagy-enhancing drug promotes degradation of mutant alpha-antitrypsin Z and reduces hepatic fibrosis. *Science* **2010**, *329*, 229-232.

(83) Tufanli, O.; Telkoparan Akillilar, P.; Acosta-Alvear, D.; Kocaturk, B.; Onat, U. I.; Hamid, S. M.; Cimen, I.; Walter, P.; Weber, C.; Erbay, E. Targeting IRE1 with small molecules counteracts progression of atherosclerosis. *Proc. Natl. Acad. Sci. U S A* **2017**, *114*, E1395-E1404.

(84) Mendez, A. S.; Alfaro, J.; Morales-Soto, M. A.; Dar, A. C.; McCullagh, E.; Gotthardt, K.; Li, H.; Acosta-Alvear, D.; Sidrauski, C.; Korennykh, A. V.; Bernales, S.; Shokat, K. M.; Walter, P. Endoplasmic reticulum stress-independent activation of unfolded protein response kinases by a small molecule ATP-mimic. *Elife* **2015**, *4*.

(85) Genereux, J. C.; Qu, S.; Zhou, M.; Ryno, L. M.; Wang, S.; Shoulders, M. D.; Kaufman, R. J.; Lasmezas, C. I.; Kelly, J. W.; Wiseman, R. L. Unfolded protein response-induced ERdj3 secretion links ER stress to extracellular proteostasis. *EMBO J.* **2015**, *34*, 4-19.

Appendix A:
Further investigation of the cysteine network in the collagen-I C-Pro domain

A.1 Introduction

In Chapter 2, we redefined the essential role of the C-terminal cysteine network in regulating any fibrillar collagen C-Pro domain's ability to form a disulfide-linked homotrimer. We determined that C2 and C3 (the same nomenclature derived from Chapter 2) are essential, highly conserved cysteine residues required to form a homotrimer. If one of these two cysteine residues is not present in a fibrillar collagen strand's C-Pro domain, that C-Pro domain will be unable to form stable homotrimers. Two remaining questions that follow our work from Chapter 2 are: (1) What role do the six other cysteine residues play in collagen assembly and (2) Does this cysteine code extend to full-length collagen-I?

Based on the crystal structure, the other cysteine residues in each C-Pro domain all participate in intrastrand disulfide bonds.[1] The function of the intrastrand disulfide bonds, however, is unknown. Several disease variants of collagen-I result in mutations to the cysteine network, and cause osteogenesis imperfecta. In particular, removal of C1 or C8 of Colα2(I) (due to a missense mutation or an early stop codon, respectively) slowed collagen assembly and promoted formation and secretion of Colα1(I) homotrimers.[2] Variants that replace C5 of Colα1(I) decreased collagen secretion, causing overmodified, yet stable triple helices to be secreted.[3] Thus, OI disease phenotypes demonstrate the importance of the cysteine network of the collagen C-Pro, but we still lack a biochemical understanding of the role each evolutionarily conserved cysteine plays in collagen-I assembly. The only systematic interrogation of the role that each cysteine played in collagen-I assembly was performed using C-Pro domains fused to short (~200 amino acid) triple helical domains.[4,5] As Chapter 2 demonstrated, studying the C-Pro domain in isolation has distinct advantages that enable a better understanding of the regulating factors guiding collagen-I assembly. Therefore, in this Appendix I describe our work toward understanding the essential role of each cysteine in collagen assembly.

Our data presented in Chapter 2 demonstrates that a single amino acid change in a 30 kDa protein alters the protein's oligomeric assembly. It is tempting to hypothesize that same single amino acid change could alter the assembly propensity of a 130 kDa protein, in the case of full-length collagen-I. In this Appendix, I also describe the assembly propensities of each full-length collagen-I strand with and without C2. Our results suggest that there are features of the Colα1(I) triple-helical domain sequence that may further promote homotrimeric triple helix assembly in the absence of C2, at least in this artificial expression system used. Additionally, our work with full-length S2C Colα2(I) demonstrates characteristic unfolded protein response activation, suggesting that cells have quality control mechanisms to recognize aberrantly assembled collagen-I trimers and potentially explaining why homotrimers of Colα2(I) have not previously been identified in nature.

A.2 Results

A.2.1 The C1-C4 disulfide bond is essential for C-Pro secretion

The crystal structure of homotrimerizing Colα1(I) demonstrated the following intrastrand disulfide bonding pairs: C1-C4, C5-C8, and C6-C7.[1] To understand how each of the intrastrand disulfide bonds participates in homotrimer formation, we individually mutagenized each cysteine residue to serine, transfected the resulting constructs into HEK293 cells, and monitored expression, secretion and assembly of each construct. To date we have only analyzed C1 and C4, but future work aims at understanding the roles of C5, C6, C7 and C8.

Our results revealed that C1S and C4S of both Colα1(I) and Colα2(I) (in the context of S2C Colα2(I), because without C2 wild type Colα2(I) cannot form homotrimers) are synthesized but retained intracellularly (**Figure A.1**). With no change in the protein level expression of GRP94 or BiP, two well-established markers for the unfolded protein response (UPR),[6,7] we do not believe that any of the tested variants activate the transcriptional response to unfolded proteins in the ER. We speculate that the intrastrand disulfide bond formed between C1 and C4 in a given monomer is an essential step in the folding of the C-Pro domain. When this linkage is not formed, the cell recognizes the C-Pro as a misfolded protein, prevents it from being secreted, and likely shuttles it to degradation pathways. Ongoing work in our lab is aimed at understanding the mechanism by which the cell recognizes the C1S or C4S C-Pro domains as misfolded, and identifying where these variants are targeted for degradation.

Our results recapitulate phenotypes observed in C-Pro variants that cause osteogenesis imperfecta (OI). In one prior study, the authors noted that full-length Colα2(I) C1R (an OI-causing mutation) does not incorporate into heterotrimers efficiently, and is retained intracellularly.[2] Remarkably, the retention of the full-length protein is fully recapitulated when we only express the C1S Colα2(I) C-Pro domain. Thus, our results suggest that this autosomal

recessive form of OI is likely caused by a failure of the Colα2(I) C-Pro domain to fold properly when it cannot form the C1-C4 intrastrand disulfide bond.

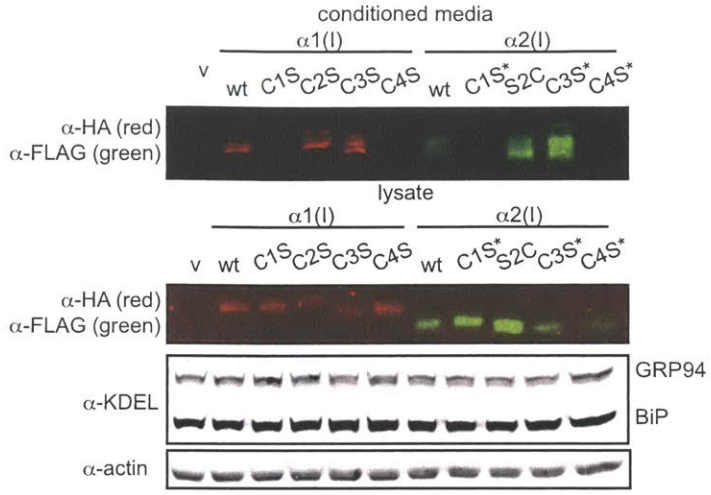


Figure A.1 | Serine variants preventing an intrastrand disulfide bond are retained intracellularly

Immunoblot analysis of transient transfections of wild-type, C1S, C2S, C3S, and C4S of C-Proα(I) constructs from Chapter 2. Top immunoblot demonstrates that C1S and C2S of both C-Proα1(I) and C-Proα2(I) are not secreted. Bottom blots are the corresponding lysate samples, demonstrating that all the transfected constructs are synthesized. None of the transfected constructs activate the UPR, as shown by no change in GRP94 or BiP levels compared to the control transfection (v).

A.2 C2 variants in full length collagen-I

In order to analyze a given collagen strand's propensity to form and secrete stable triple helices, we required a cell line that basally expresses the necessary machinery to fold collagen, but does not produce potentially confounding, endogenous collagen-I. Our prior work as well as others in HT1080 cells (Chapter 3) established this fibrosarcoma cell line as a viable model for studying collagen-I assembly and folding.[8-10] Using replication-incompetent human adenovirus type V, we delivered each full-length collagen-I gene of interest into HT1080 cells under a CMV promoter to drive expression. Three days post-transduction, the culture media was supplemented every 24 h, for 2 d total with ascorbate, a known co-factor for the prolyl-4-hydroxylases essential for proper triple helical folding.[11,12] Media was then harvested and collagen-I was precipitated with ammonium sulfate and prepared for downstream analysis.

Wild-type Colα1(I) is efficiently secreted from the cell and displays a slower electrophoretic mobility when analyzed by non-reducing versus reducing SDS-PAGE, suggesting the monomers are disulfide bonded to one another. In contrast, full length C2S Colα1(I) migrates as a monomer on an SDS-PAGE gel regardless of the presence of reducing agent, as we would expect based on the published disulfide bonding map and our own assembly data in the C-Pro domain above (**Figure A.2A**). Contrary to what we would expect, however, both wild type and C2S Colα1(I) form stable triple helices, as assayed by protease resistance and immunoblot analysis (**Figure A.2B**).

Performing similar analyses on full length wild type and S2C Colα2(I) sheds light on one of the possible reasons that homotrimers of Colα2(I) have never before been observed. Analyzing conditioned media for secreted collagen showed that, while wild type Colα2(I) was secreted as avidly as the Colα1(I) protein variants, S2C Colα2(I) was not secreted at all. Instead, the S2C Colα2(I) variant was retained intracellularly (**Figure A.3A**). Moreover, expression of the S2C Colα2(I) construct induced the unfolded protein response, as shown by

qPCR and RT-PCR amplifying XBP1 indicating it is more heavily spliced in the presence of the S2C variant (**Figure A.3B** and **A.3C**). This UPR activation was induced by the full length S2C Colα2(I), whereas the wild type strand did not activate the UPR. These data suggest that cells recognize homotrimeric Colα2(I) as a misfolded protein, and may provide a convenient handle for the future discovery of collagen quality control mechanisms.

Taken together, assembly of heterotrimeric full length collagen is more complex than was previously understood. Prior to our work, the literature discredited the role that the cysteine residues play in forming a disulfide linked C-Pro domains. However, we show that the cysteine network of the C-Pro domain is essential in regulating a strand's propensity to form homotrimers. Full-length collagen-I data presented here also supports the hypothesis that the sequence of Colα1(I)'s triple-helical domain can also promote its homotrimerization. In the case of Colα2(I), the misfolded, likely homotrimerized protein complex is recognized by the UPR and retained intracellularly, likely one of the main reasons that homotrimers of Colα2(I) have never been observed in nature or in cellulo.

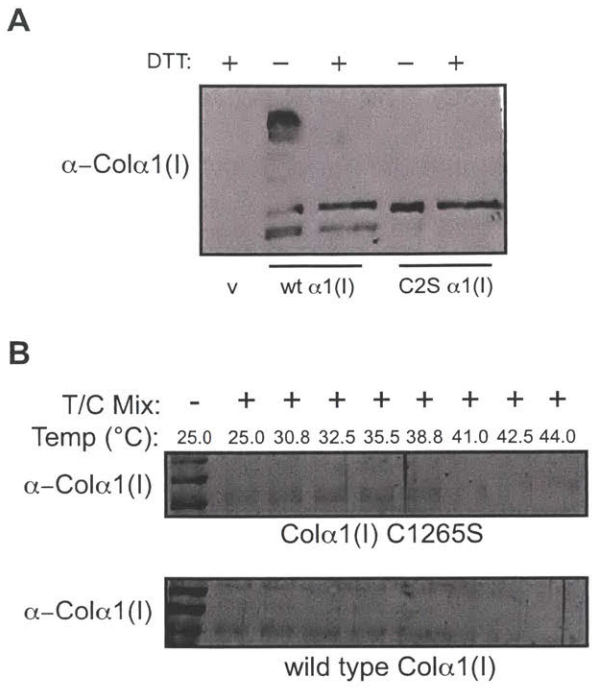


Figure A.2 | Full-length cysteine variants of Colα1(I) suggest there are factors in the triple helical domain that promote trimerization

(A) Precipitated media samples of full length wild-type Colα1(I) and C2S Colα1(I) were treated with and without reducing agent and analyzed by immunoblot. The wild type protein assembles into a disulfide dependent trimer, based on the higher signal being succeptible to reducing agent, while the C2S variant does not assembly into higher order oligomers. **(B)** Preciptiated wild type and C2S Colα1(I) were heated to the temperatures indicated, and treated with a mixture of trypsin and chymotrypsin to assess the folded state of the triple helix. Remarkably both wild type and the C2S variant are equally resistant to protease treatment.

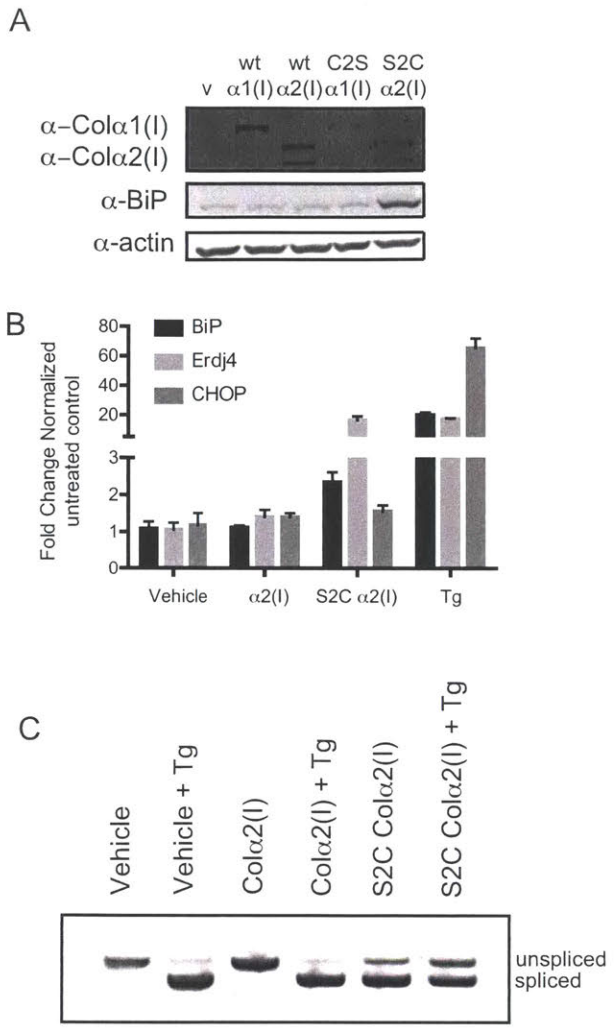


Figure A.3 | Retention of full length S2C Cola2(I) and subsequent UPR activation

(A) Intracellular retention of S2C Colα2(I) activates the UPR, as indicated by the upregulation of BiP compared to the control lane (v). **(B)** qPCR data demonstrating UPR activation in the presence of the S2C variant, but not the wild type Colα2(I). **(C)** RT-PCR analysis of XBP1 unspliced to spliced ratio. The proportion of XBP1s is higher in the presence of the S2C variant compared to the wild-type Colα2(I). S2C induces comparable amounts of XBP1 splicing to Tg treatment.

Table A.1 | Primers used for qPCR.

Transcript	Forward	Reverse
RPLP2	5′-CCATTCAGCTCACTGATAACCTT-3′	5′-CGTCGCCTCCTACCTGCT-3′
CHOP	5′-GGAGCTGGAAGCCTGGTATG-3′	5′-GCCAGAGAAGCAGGGTCAAG-3′
BiP	5′-GCCTGTATTTCTAGACCTGCC-3′	5′-TTCATCTTGCCAGCCAGTTG-3′
Erdj4	5′-GGAAGGAGGAGCGCTAGGTC-3′	5′-ATCCTGCACCCTCCGACTAC-3′

A.3 Concluding Remarks and Future Work

For the C-Pro domain, the cysteine code is essential and clearly regulates trimerization of individual chains. The work I describe provides the first systematic approach to determine each evolutionarily conserved cysteine residue's role in regulating trimer assembly. Studying the C-Pro domain in the absence of triple helical domains provides a unique platform to directly interrogate both roles in assembly and mechanistic information regarding the molecular cause of the more severe cases of OI. Future work will continue to study each cysteine residue in the collagen-I C-Pro domains, assessing their assembly propensities, secretion kinetics and ability to incorporate into heterotrimers.

Our C-Pro data demonstrates the importance of the cysteine network in regulating collagen C-Pro domain assembly. Analyzing the same C2 variants in the context of full length collagen suggests that even though removal of the second cysteine from Colα1(I) prevents disulfide-dependent assembly, there are further factors that we are currently investigating that guide triple helix formation. On the contrary, in the case of Colα2(I), our data suggests that the lack of a cysteine in the wild type sequence is the only thing preventing Colα2(I) from homotrimerizing. Furthermore, in the event that Colα2(I) does homotrimerize (in our S2C construct), nature has engineered a method for detection of such assembled homotrimers, preventing them from being secreted.

In conclusion, we have unequivocally shown that the cysteine network in the C-Pro domain of fibrillar collagen is essential for C-Pro trimerization, but it is clearly not the entire story. Further experimental work will help to illuminate the function the triple helix plays in trimer assembly, and what role it may play in vivo.

A.4 Experimental Methods and Supplies

A.4.1 Adenoviral vector construction and viral amplification

Full-length collagen-I genes were cloned from the vectors in pTRE-Tight of Chapter 3, into pENTR1A vectors. Full-length wild-type Colα1(I) and Colα2(I) were then mutagenized, replacing the cysteine in position 2 of Colα1(I) with a serine, and replacing the serine in position 2 of Colα2(I) with a cysteine. Samples were then recombined using LR Clonase into Ad-Dest5 (Invitrogen), o/n at rt, according to the manufacturer's instructions. Resultant vectors were transformed into Stbl3 cells, amplified and sequence-confirmed for the presence of the expected collagen genes. Vectors were then treated with PacI to expose the LTRs of the Adenoviral vector, and transfected into 293A cells using Lipofectamine 2000. Cells were monitored daily for the formation of plaques, and the viral supernatant was harvested upon full cytopathic effect (CPE). Viral supernatants were subjected to three freeze/thaw cycles at -80° C and 37° C, respectively. Cell debris was removed by centrifugation at 1,500 rpm for 15 min at rt, and the supernatant was collected, aliquotted and used for subsequent amplification. Each virus was amplified 2–3 more times on 293A cells to create p3 or p4 virus, freeze/thawed, and stored in single use aliquots to minimize loss of titer due to freeze/thaw.

A.4.2 HT1080 adenoviral transductions

6.0×10^5 HT1080 cells were plated in each well of a 6-well plate. 24 h later, the media was changed and supplemented with p3 or p4 viral supernatant. After another 24 h, the media was changed one final time, or cells were subcultured as necessary. Cells were grown for an additional 2–3 d to accumulate collagen constructs in the media and lysate samples.

A.4.3 Cell lysis and immunoblot analysis

6.0×10^5 HEK293 cells were plated. After 24 h cells were transfected with Lipofectamine 3000 (Thermo Fisher Scientific). After 24 h the media was changed, and 72 h post-transfection media and cells were harvested for immunoblot analysis. Cells were trypsinized, washed 2× with PBS, and lysed in buffer containing 50 mM Tris-HCl, pH 7.4, 1% Triton-X 100, 150 mM NaCl, 0.5% sodium dodecyl sulfate, 0.1% sodium deoxycholate, 1 mM EDTA, a protease inhibitor tablet (Pierce) and 1.5 mM PMSF, on ice for 20 min. Samples were then centrifuged at 4° C for 15 min at 16,4000 × g. Supernatants were collected, quantified and normalized for SDS-PAGE analysis. Samples were treated with 6× gel loading dye (300 mM Tris-HCl, pH 6.8, 15% glycerol, 6% SDS, 0.01% bromophenol blue, and 1 M dithiothrietol), boiled for 10 min, and loaded on a gel for separation.

Media samples were harvested, treated with 6× gel loading dye, with or without reducing agent (dithiothreitol), boiled and run on a gel. All gels were transferred using the TransBlot Turbo (BioRad) according to the manufacturer's instructions. Nitrocellulose membranes were temporarily Ponceau stained, blocked with 5% non-fat milk, and probed with primary antibody in 5% BSA overnight. Membranes were washed with TBS + 0.1% Tween-20, and probed with secondary antibody in 5% non-fat milk for 45–60 min. Secondary antibody was removed, the membrane was washed with TBS + 0.1% Tween-20, rinsed with water and scanned on a LiCor Odyssey Scanner.

A.4.4 Protease digestion of collagen-I

Protease digestion was adapted from prior work.[13] HT1080 cells transduced for each collagen-I variant were grown to confluence in the presence of 50 μM ascorbate. At 96 h post transduction, media was harvested and cell debris was removed by centrifugation. The supernatant was then chilled, supplemented with 100 mM Tris-HCl pH 7.4 (final concentration), 1 mM PMSF, and 176 mg/mL ammonium sulfate. Samples were incubated in an end-over-end rotator at 4° C o/n. Media was then centrifuged and discarded, collecting the insoluble pellet.

The pellet was resuspended in 200 μL of 100 mM Tris-HCl, pH 7.4, 400 mM sodium chloride. 20 μL aliquots were distributed in PCR tubes, and incubated in a thermocycler at the range of temperatures indicated, for 5 min, then cooled to rt in a water bath. Samples were then treated with 6 μL of a mixture of 0.1 mg/mL trypsin and 0.2 mg/mL chymotrypsin for 2 minutes. The reaction was quenched with 6 μL of 2.5 mg/mL soybean trypsin inhibitor, and then 6.5 μL of 6x gel loading dye, boiled and analyzed by SDS-PAGE. Immunoblots were quantified and plotted using Image Studio Lite.

A.4.5 Quantitative PCR analysis

Relative mRNA expression levels of genes of interest were assessed by quantitative RT-PCR. Cells were harvested by trypsinization, washed with PBS, and total RNA was extracted using the Omega RNA Purification Kit according to the manufacturer's instructions. RNA concentrations were quantified and normalized to 0.5 - 1 μg total RNA for cDNA reverse transcription. Using the Applied Biosystems Reverse Transcriptase cDNA Kit, cDNA was synthesized in a BioRad Thermocycler. LightCycler 480 Sybr Green Master Mix (KAPA), appropriate primers (purchased from Integrated DNA Technologies, Life Technologies and Sigma Aldrich, **Table A.1**) and cDNA were used for amplification in a Light Cycler 480 II Real Time PCR Instrument in the MIT BioMicro Center. Primer integrity was assessed by thermal melt and agarose gel analysis to ensure that a single gene was amplified. Transcripts were normalized to the housekeeping gene *Rplp2,* and all measurements were performed in technical and biological triplicate.

A.4.6 RT-PCR amplification of XBP1

cDNA was synthesized as described above. XBP1 transcript levels were assessed using the following primers:

5′-CCTTGTAGTTGAGAACCAGG-3′ and

5′-GGGGCTTGGTATATATGTGG-3′

spanning the splicing site. RT-PCR was performed using NEB Q5 Polymerase, according to manufacturer protocols. The resultant products were loaded onto a 14% polyacrylamide gel and separated by 100V for three hours. The gel was then incubated with a solution of 0.5% ethidium bromide solution in TBE buffer for 5 minutes, rinsed with TBE and imaged using 365 nm light.

A.5 References

(1) Sharma, U.; Carrique, L.; Vadon-Le Goff, S.; Mariano, N.; Georges, R. N.; Delolme, F.; Koivunen, P.; Myllyharju, J.; Moali, C.; Aghajari, N.; Hulmes, D. J. Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **2017**, *8*, 14671.

(2) Pace, J. M.; Wiese, M.; Drenguis, A. S.; Kuznetsova, N.; Leikin, S.; Schwarze, U.; Chen, D.; Mooney, S. H.; Unger, S.; Byers, P. H. Defective C-propeptides of the proalpha2(I) chain of type I procollagen impede molecular assembly and result in osteogenesis imperfecta. *J. Biol. Chem.* **2008**, *283*, 16061-16067.

(3) Pace, J. M., Kuslich, C.D., Willing, M.C. Byers, P.H. Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the pro-alpha1(I) chain of type I collagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta. *J. Med. Genet.* **2001**, *38*, 443-449.

(4) Bulleid, N. J.; Wilson, R.; Lees, J. F. Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation. *Biochem. J.* **1996**, *317 (Pt 1)*, 195-202.

(5) 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.* **1994**, *269*, 24354-24360.

(6) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R., 3rd; Su, A. I.; Kelly, J. W.; Wiseman, R. L. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013**, *3*, 1279-1292.

(7) Walter, P.; Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **2011**, *334*, 1081-1086.

(8) 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* **1993**, *13*, 399-405.

(9) DiChiara, A. S.; Taylor, R. J.; Wong, M. Y.; Doan, N. D.; Rosario, A. M.; Shoulders, M. D. Mapping and Exploring the Collagen-I Proteostasis Network. *ACS Chem. Biol.* **2016**, *11*, 1408-1421.

(10) Fertala, A.; Sieron, A. L.; Ganguly, A.; Li, S. W.; Ala-Kokko, L.; Anumula, K. R.; Prockop, D. J. Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). *Biochem. J.* **1994**, *298 (Pt 1)*, 31-37.

(11) Kivirikko, K. I.; Myllyharju, J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol.* **1998**, *16*, 357-368.

(12) Chan, D.; Lamande, S. R.; Cole, W. G.; Bateman, J. F. Regulation of procollagen synthesis and processing during ascorbate-induced extracellular matrix accumulation in vitro. *Biochem. J.* **1990**, *269*, 175-181.

(13) Makareeva, E.; Mertz, E. L.; Kuznetsova, N. V.; Sutter, M. B.; DeRidder, A. M.; Cabral, W. A.; Barnes, A. M.; McBride, D. J.; Marini, J. C.; Leikin, S. Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J. Biol. Chem.* **2008**, *283*, 4787-4798.