Mapping and Exploring the Collagen-I Proteostasis Network

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation: DiChiara, Andrew S. et al. “Mapping and Exploring the Collagen-I Proteostasis Network.” ACS Chemical Biology 11, 5 (March 2016): 1408–1421 © 2016 American Chemical Society

As Published: http://dx.doi.org/10.1021/ACSCHEMIBI0.5B01083

Publisher: American Chemical Society (ACS)

Persistent URL: http://hdl.handle.net/1721.1/113577

Version: Author’s final manuscript: final author’s manuscript post peer review, without publisher’s formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.
Mapping and Exploring the Collagen-I Proteostasis Network

Andrew S. DiChiara¹, Rebecca J. Taylor¹, Madeline Y. Wong¹, Ngoc Duc Doan¹, Amanda M. Del Rosario², and Matthew D. Shoulders¹,*

¹Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
²Koch Institute for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Collagen-I is the most abundant protein in the human body, yet our understanding of how the endoplasmic reticulum regulates collagen-I proteostasis (folding, quality control, and secretion) remains immature. Of particular importance, interactomics studies to map the collagen-I proteostasis network have never been performed. Such studies would provide insight into mechanisms of collagen-I folding and misfolding in cells, an area that is particularly important owing to the prominence of the collagen misfolding-related diseases. Here, we overcome key roadblocks to progress in this area by generating stable fibrosarcoma cells that inducibly express properly folded and modified collagen-I strands tagged with distinctive antibody epitopes. Selective immunoprecipitation of collagen-I from these cells integrated with quantitative mass spectrometry-based proteomics permits the first mapping of the collagen-I proteostasis network. Biochemical validation of the resulting map results in the assignment of numerous new players in collagen-I proteostasis, and the unanticipated discovery of aspartyl hydroxylation as a new post-translational modification in the N-propeptide of collagen-I. Furthermore, quantitative analyses reveal that Erp29, an abundant endoplasmic reticulum proteostasis machinery component with few known functions, plays a key role in collagen-I retention under ascorbate-deficient conditions. In summary, the work here provides fresh insights into the molecular mechanisms of collagen-I proteostasis, yielding a detailed roadmap for future investigations. Straightforward adaptations of the cellular platform developed will also enable hypothesis-driven, comparative research on the distinctive proteostasis mechanisms engaged by normal and disease-causing, misfolding collagen-I variants, potentially motivating new therapeutic strategies for the currently incurable collagenopathies.

Collagen-I serves as the predominant proteinaceous component of human tissues, including skin and bone.¹ The supramolecular properties of the collagen-I-based extracellular matrix

¹ Corresponding Author: Matthew D. Shoulders, Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 56-546, Cambridge, MA 02139, Phone: (617) 452-3525, Facsimile: (617) 324-0505, mshoulde@mit.edu.

The Supporting Information is available free of charge on the ACS Publications website. Supporting Information Figures S1–S3 and Supporting Information Tables S1–S4. Detailed listing of materials and reagents and complete experimental procedures for lentivirus production and cell line generation, qPCR experiments, immunoprecipitations and collagen-I secretion assays, MS sample preparation and analysis, pulse-chase analyses, and confocal microscopy (PDF) Polished MS data (XLS)

HHS Public Access
Author manuscript
ACS Chem Biol. Author manuscript; available in PMC 2017 May 20.

Published in final edited form as:
ACS Chem Biol. 2016 May 20; 11(5): 1408–1421. doi:10.1021/acschembio.5b01083.
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. 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. 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. 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 disulfide-linked C-terminal propeptide domain. 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. During triple-helix formation, peptidyl prolyl isomerases (PPIases) act to convert the hundreds of prolyl peptide bonds to the trans configuration required in folded triple helices.

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) or via the identification of a genetic defect that causes a collagenopathy (e.g., CRTAP). Despite these seminal discoveries, many questions still remain. For example, 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.

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, we report 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.

RESULTS AND DISCUSSION

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

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, we incorporated a short (<10 amino acid) antibody epitope tag at the amino-terminus of collagen-I’s N-propeptide domain, rationalizing that this location was unlikely to disrupt collagen-I folding or structure. 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 1A).

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, 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. We were thereby able to identify unique cell lines that display significant lysate and secreted levels of either protein (Figure 1B). We note that the complex banding patterns we observe are common for collagen-I owing to extensive proteolytic processing (for example, see immunoblots of native collagen-α2(I) secreted from primary fibroblasts in Figure 2B 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. We identified a cell line, termed HT-1080Col-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 1C). We observe no apparent activation of unfolded protein response-regulated
genes\textsuperscript{15} in HT-1080\textsuperscript{Col-I} cells upon collagen-I induction (Figure 1D), 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\textsuperscript{Col-I} cells, we employed confocal microscopy. We observe strong co-localization of collagen-I with ER markers ($R_{\text{mean}} = 0.45 \pm 0.12$), as well as partial co-localization with Golgi ($R_{\text{mean}} = 0.22 \pm 0.06$) and lysosomal markers ($R_{\text{mean}} = 0.19 \pm 0.05$) (Figure 1E) in the absence of ascorbate (ER-retention conditions), consistent with previous studies.\textsuperscript{16} We assessed collagen-I secretion kinetics using metabolic labeling and found that the majority of the collagen-I is secreted in <3 hours with minimal degradation, recapitulating endogenous collagen-I secretion kinetics in primary fibroblasts\textsuperscript{17} (Figures 1F and 1G) and further confirming that the HT-1080 cells are properly handling our collagen-I constructs.

**Molecular Properties of Collagen-I Produced by HT-1080\textsuperscript{Col-I} Cells**

We next asked whether the collagen-I produced by HT-1080\textsuperscript{Col-I} cells displays the expected molecular properties of endogenous collagen-I. We first tested for a direct intracellular interaction between the collagen-\(\alpha_1(I)\) and collagen-\(\alpha_2(I)\) strands. We observe that an IP of collagen-\(\alpha_1(I)\) using HA-antibody beads co-IPs collagen-\(\alpha_2(I)\), and also that the reverse is true when using FLAG-antibody beads (Figure 2A). These results confirm the heteromeric assembly of collagen-\(\alpha_1(I)\) and collagen-\(\alpha_2(I)\) strands produced by HT-1080\textsuperscript{Col-I} cells.

The triple-helical domain of properly folded, extracellular collagen-I is resistant to proteases such as trypsin and chymotrypsin,\textsuperscript{18} 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 2B), consistent with propeptide sensitivity and triple-helical resistance. Collagen-I secreted from HT-1080\textsuperscript{Col-I} cells displays an identical pattern upon protease digestion (Figures 2B and S1A), confirming that the collagen-I produced is folded into a stable triple helix.

Native collagen-I undergoes essential co- and post-translational modifications during its folding and maturation in the ER.\textsuperscript{7} Most important among these are the 4R-hydroxylation of Yaa-position Pro residues in the Gly-Xaa-Yaa repeats of the triple-helical domain,\textsuperscript{1} 3S-hydroxylation of Xaa-position Pro residues,\textsuperscript{19} and Lys hydroxylation.\textsuperscript{7} To quantify the extent of collagen-I hydroxylation in HT-1080\textsuperscript{Col-I} cells, we used the stable isotope labeling by amino acids in cell culture (SILAC) technique\textsuperscript{20} to generate light, medium, and heavy Lys- and Arg-labeled HT-1080\textsuperscript{Col-I} cells. Because ascorbate is an essential hydroxylase cofactor,\textsuperscript{21} we induced collagen-I expression in the medium-labeled cells without ascorbate (ER-retention conditions\textsuperscript{22}) and in the heavy-labeled cells with ascorbate (secretion-promoting conditions), using the light cells as a control that expresses collagen-I lacking an HA tag. After IP of intracellular collagen-\(\alpha_1(I)\) 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 2C 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 (Table S1). Most notably, Pro986 in collagen-α1(I) and Pro707 in collagen-α2(I) are the two Xaa-position Pro residues where 3S-hydroxylation is known to occur in native collagen-I.19 We observe both of these hydroxylation events in our MS analyses (see Figure S1B for MS2 scans). The hydroxylation is again ascorbate-dependent, as expected (Figure 2D). These results demonstrate that collagen-I produced by HT-1080Col-I cells displays the known hydroxylation patterns of endogenous collagen-I.

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,23 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-1080Col-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-1080Col-I cells is indeed N-glycosylated (Figure 2E).

Cumulatively, the results in Figure 2 indicate that our HT-1080Col-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-1080Col-I cells as a convenient platform both for mapping the collagen-I proteostasis network and for more detailed mechanistic biochemistry studies.

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 2A). However, a pilot MS study of the co-immunoprecipitated interactome identified a very limited suite of interactors (Figure 3A). 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 been observed for other ER proteostasis network client proteins.24,25

To resolve this challenge, we immortalized interactions with collagen-I in live HT-1080Col-I cells by incubating with the cell-permeable, lysine-reactive, and reversible crosslinker dithiobis(succinimidyl propionate)26 (DSP; Figure 3B). 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 3C and S2) that are not otherwise observable via traditional IP workflows.
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-1080Col-I cells described above, we induced collagen-I expression, performed covalent crosslinking with DSP, and then used HA-antibody beads to IP collagen-\(\alpha_1\)(I) and its interactors in biological triplicate (Figure 4A). The light-labeled negative control HT-1080 cells inducibly expressed collagen-I lacking the HA epitope. The medium- and heavy-labeled HT-1080Col-I cells (both expressing HA-tagged collagen-\(\alpha_1\)(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. 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.

Approximately 60% of the peptides identified by this workflow in the heavy and medium samples correspond to either collagen-\(\alpha_1\)(I) or collagen-\(\alpha_2\)(I), indicating that we are strongly enriching the bait. 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 (see Table S2 for the complete data set). In Table 1, we present the results for all high-confidence interactors defined by meeting the following additional criteria: (1) \(\geq\) 2 unique peptides identified in \(\geq\) 2 biological replicates and (2) an average enrichment in either the heavy or medium sample \(\geq\) 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 1 and Figure 4B), 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 4C), 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, 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 1 and Figure 4B). Indeed, two well-established collagen-I chaperones, HSP47 and PPIB, are the most highly enriched of all the proteins identified. We also identify FKBP10, an ER PPIase whose absence causes autosomal recessive osteogenesis imperfecta. 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 1). The other five interactors have been implicated in collagen-I proteostasis-related functions, although mechanistic characterization remains incomplete.

**Experimental Validation of the Collagen-I Proteostasis Network Map**

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 3C and S2). 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 5A). 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-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. We prepared stable Saos-2 cells that inducibly express our HA-tagged collagen-α1(I) construct under control of the tetracycline repressor (see Supporting Information) and probed the crosslinked co-IP for putative interactors from Figure 5A. The majority of the interactions tested are clearly recapitulated in Saos-2 cells (Figure 5B), 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-I cells engenders further confidence in our map of the collagen-I proteostasis network (Table 1 and Figure 4B).

Next, we obtained lentiviruses expressing shRNAs (see Table S3) 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 S3A for qPCR validation of knockdown cell lines). To discern effects of each knockdown on collagen-I secretion, we differentiated Saos-2 cells for three days to stimulate endogenous collagen-I secretion,
then analyzed collagen-I levels in the corresponding media samples using immunoblotting (Figures 5C and S3B). 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.33 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 discovery from 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-1080Col-1 cells is consistent with hydroxylation of both Asp71 and Asp72 in the N-propeptide domain of collagen-α1(I) (Figure 5D). 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 5E), 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,34 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-I interaction 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.

**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.35 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 6A) 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. Nonetheless, under-hydroxylated collagen-I produced in the absence of ascorbate is retained in the ER to a significant extent, 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 6B) 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, and the peroxiredoxins. 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 Figure 5C, we began by using shRNA-expressing lentiviruses (see Table S3) 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 S3A). Erp29 depletion does not obviously impact collagen-I secretion under secretion-promoting (ascorbate-treated) conditions (see Figure 5C), nor does it inordinately stress Saos-2 cells, as indicated by a lack of chronic unfolded protein response activation (Figure S3C). 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 6C). 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 6D).

Erp29 is an abundant and widespread component of the ER proteostasis network, displaying a thioredoxin (PDI-like) fold but lacking any catalytic activity for disulfide bond isomerization. In the past 20 years, a number of groups have explored possible roles for Erp29 in the proteostasis of CFTR, thyroglobulin, and viral proteins. These studies have led to various suggestions that Erp29 may accompany nascent proteins out of the cell, that it may have a role in ER client protein maturation, or that it could assist the folding of certain complex plasma membrane proteins. Our findings here, informed by MS-based proteomics in our new collagen-I expression platform, suggest a novel function for Erp29—

ACS Chem Biol. Author manuscript; available in PMC 2017 May 20.
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 6D).

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 4B) 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.

METHODS

Plasmids

The human \textit{Col1A1} and \textit{Col1A2} genes were obtained from Origene. \textit{Col1A1} was amplified using primers with NotI and SalI sites, while \textit{Col1A2} was amplified using primers with NotI and EcoRV sites. DNA encoding the pre-protrypsin signal sequence, appropriate antibody epitope tags, and the NotI, SalI, and EcoRV sites was inserted into a pTRE-Tight vector (ClonTech). Collagen-I genes were then inserted into the modified pTRE-Tight vectors.

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 \textit{Col1A1} and \textit{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. 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. Stable Saos-2 cell lines inducibly expressing HA-collagen-α1(I) were generated by lentiviral transduction as described in the Supporting Information. 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.

**Immunoblotting**

Lysate and media samples were separated by SDS-PAGE and analyzed by immunoblotting using the LI-COR Biosciences Odyssey System. Details of IP experiments and antibodies used are in the Supporting Information.

**Quantitative RT-PCR**

The relative mRNA expression levels of target genes were measured using quantitative RT-PCR (see the Supporting Information and Table S4 for details).

**Confocal Microscopy**

Detailed protocols are provided in the Supporting Information.

**Pulse-Chase Analyses**

Detailed protocols are provided in the Supporting Information.

**Protease Digestions**

Collagen-I was precipitated from media as previously described. Precipitated pellets were resuspended in 400 mM NaCl, 150 mM Tris (pH 7.5). Samples were aliquoted and treated with trypsin or chymotrypsin (0.1 mg/mL final concentration) prior to analysis by immunoblotting to determine the extent of stable triple-helix formation.

**MS Sample Preparation and Analysis**

HT-1080Col-I cells were propagated for >6 passages in light, medium, or heavy SILAC media (light media was supplemented with Lys (12C6, 99%; 14N2, 99%) and Arg (12C6, 99%; 14N4, 99%); medium media was supplemented with Lys (13C6, 99%; 14N2, 99%) and Arg (13C6, 99%; 14N2, 99%); heavy media was supplemented with Lys (13C6, 99%; 15N2, 99%) and Arg (13C6, 99%; 15N4, 99%); Cambridge Isotopes). All media were supplemented with 2 mM unlabeled proline. Collagen-I co-IPs were prepared and MS analyses performed as described in the Supporting Information.
Saos-2 Stable Cell Line Generation and Secretion Analysis

Cells were transduced with shRNA-encoding lentiviruses, stably selected using hygromycin or puromycin, and analyzed for collagen-I secretion under differentiation conditions. See the Supporting Information for a detailed protocol.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the Mallinckrodt Foundation, the Smith Family Foundation, the NIH/NIAMS (1R03AR067503), and MIT for financial support. A.S.D. was supported by the NIAMS (1F31AR067615). M.Y.W. was supported by an NSF Graduate Fellowship and a Prof. Amar G. Bose Research Grant. N.D.D. was supported by a postdoctoral fellowship from the Fonds de la Recherche en Santé du Québec. This work was also supported in part by the NIH/NIEHS (P30-ES002109). We thank J. Genereux (University of California–Riverside) and E. Weerapana (Boston College) for fruitful discussions.

REFERENCES

(1). Shoulders MD, Raines RT. Collagen structure and stability. Annu. Rev. Biochem. 2009; 78:929–958. [PubMed: 19344236]
(2). Ishikawa Y, Bächinger HP. A molecular ensemble in the rER for procollagen maturation. Biochim. Biophys. Acta. 2013; 1833:2479–2491. [PubMed: 23602968]
(3). Myllyharju J, Kivirikko KI. Collagens and collagen-related diseases. Ann. Med. 2001; 33:7–21. [PubMed: 11310942]
(4). Tosi LL, Warman ML. Mechanistic and therapeutic insights gained from studying rare skeletal diseases. Bone. 2015; 76:67–75. [PubMed: 25819040]
(5). Forlino A, Cabral WA, Barnes AM, Marini JC. New perspectives on osteogenesis imperfecta. Nat. Rev. Endocrinol. 2011; 7:540–557. [PubMed: 21670757]
(6). Boudko SP, Engel J, Bächinger HP. The crucial role of trimerization domains in collagen folding. Int. J. Biochem. Cell Biol. 2012; 44:21–32. [PubMed: 22001560]
(7). Myllyharju J. Intracellular post-translational modifications of collagens. Top. Curr. Chem. 2005; 247:115–147.
(8). Buevich AV, 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. [PubMed: 10757978]
(9). Nagata K, Saga S, Yamada KM. A major collagen-binding protein of chick-embryo fibroblasts is a novel heat-shock protein. J. Cell Biol. 1986; 103:223–229. [PubMed: 3722264]
(10). Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M, Castagnola P, Rauch F, Glorieux FH, Vranka J, Bächinger HP, Pace JM, Schwarze U, Byers PH, Weis M, Fernandes RJ, Eyre DR, Yao Z, Boyce BF; Lee B. CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. Cell. 2006; 127:291–304. [PubMed: 17055431]
(11). Forlino A, Kuznetsova NV, Marini JC, 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. [PubMed: 17662583]
(12). Lamandé SR, Chessler SD, Golub SB, Byers PH, Chan D, Cole WG, Silence DO, Bateman JF. 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. [PubMed: 7721766]
(13). Fitzgerald J, Lamandé SR, Bateman JF. Proteasomal degradation of unassembled mutant type I collagen pro-alpha 1(I) chains. J. Biol. Chem. 1999; 274:27392–27398. [PubMed: 10488070]
(14). Geddis AE, Prockop DJ. 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. [PubMed: 8246835]

(15). Shoulders MD, Ryno LM, Genereux JC, Moresco J, Tu PG, Wu CL, Yates JRI, Su AI, Kelly JW, Wiseman RL. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. Cell Rep. 2013; 3:1279–1292. [PubMed: 23583182]

(16). Stephens DJ, 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. [PubMed: 11884515]

(17). Diegelmann RF, Peterkofsky B. Inhibition of collagen secretion from bone and cultured fibroblasts by microtubular disruptive drugs. Proc. Natl. Acad. Sci. USA. 1972; 69:892–896. [PubMed: 4502941]

(18). Bruckner P, Prockop DJ. Proteolytic enzymes as probes for the triple-helical conformation of procollagen. Anal. Biochem. 1981; 110:360–368. [PubMed: 7015914]

(19). Weis MA, Hudson DM, Kim L, Scott M, Wu JJ, Eyre DR. 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. [PubMed: 19940144]

(20). Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, 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.

(21). Myllyharju J. Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biol. 2003; 22:15–24. [PubMed: 12714038]

(22). Tschant G, Brocks DG, Engelbart K, Mohr J, Baader E, Gunzler V, Hanuske-Abel HM. 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. [PubMed: 1850989]

(23). Lamandé SR, Bateman JF. 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. [PubMed: 7629078]

(24). Tan YL, Genereux JC, Pankow S, Aerts JM, Yates JR. ERdj3 is an endoplasmic reticulum degradation factor for mutant glucocerebrosidase variants linked to Gaucher’s disease. Chem. Biol. 2014; 21:967–976. [PubMed: 25126989]

(25). Balch WE, Yates JR. Application of mass spectrometry to study proteomics and interactomics in cystic fibrosis. Methods Mol. Biol. 2011; 742:227–247. [PubMed: 21547736]

(26). Lomant AJ, 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. [PubMed: 957432]

(27). Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–3567. [PubMed: 10612281]

(28). Chen XQ, Karnovsky A, Sans MD, Andrews PC, Williams JA. Molecular characterization of the endoplasmic reticulum: Insights from proteomic studies. Proteomics. 2010; 10:4040–4052. [PubMed: 21080494]

(29). Pyott SM, Schwarze U, Christiansen HE, Pepin MG, Leistritz DF, Dineen R, Harris C, Burton BK, Angle B, Kim K, Sussman MD, Weis M, Eyre DR, Russell DW, McCarthy JK, Steiner RD, Byers PH. 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. [PubMed: 21281288]

(30). Barnes AM, Cabral WA, Weis M, Makareeva E, Mertz EL, Leikin S, Eyre D, Trujillo C, Marini JC. 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. [PubMed: 22718341]

(31). Rodan SB, Imay T, Thiede MA, Wesolowski G, Thompson D, Barshavit Z, Shull S, Mann K, Rodan GA. Characterization of a human osteosarcoma cell line (Saos-2) with osteoblastic properties. Cancer Res. 1987; 47:4961–4966. [PubMed: 3040234]

ACS Chem Biol. Author manuscript; available in PMC 2017 May 20.
(32). Cheng S, Lai C, Blystone SD, Avioli LV. Bone mineralization and osteoblast differentiation are negatively modulated by integrin alpha (V) beta (III). J. Bone Miner. Res. 2001; 16:277–288. [PubMed: 11204428]

(33). Suzuki N, Ban S, Itoh E, Chen S, Imai FL, Sawano Y, Miyakawa T, Tanokura M, Yonezawa N. Calcium-dependent structural changes in human reticulocalbin-1. J. Biochem. 2014; 155:281–293. [PubMed: 24451493]

(34). Patel N, Khan AO, Mansour A, Mohamed JY, Al-Assiri A, Haddad R, Jia XF, Xiong Y, Megarbane A, Traboulsi EI, Alkuraya FS. 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. [PubMed: 24768550]

(35). Franceschi RT, Iyer BS, 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. [PubMed: 8079660]

(36). Tajima S, Pinnell SR. Regulation of collagen synthesis by ascorbic acid: Ascorbic acid increases type-I procollagen messenger RNA. Biochem. Biophys. Res. Comm. 1982; 106:632–637. [PubMed: 7104012]

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

(38). Rhee SG, Woo HA, Kil IS, Bae SH. Peroxiredoxin functions as a peroxidase and a regulator and sensor of local peroxides. J. Biol. Chem. 2012; 287:4403–4410. [PubMed: 22147704]

(39). Mkrtchian S, Fang C, Hellman U, Ingelman-Sundberg M. A stress-inducible rat liver endoplasmic reticulum protein, ERp29. Eur. J. Biochem. 1998; 251:304–313. [PubMed: 9492298]

(40). Magnuson B, Rainey EK, 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. [PubMed: 16246730]

(41). 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. [PubMed: 11884402]

(42). Grumbach Y, Bikard Y, Suaud L, Chanoux RA, Rubenstein RC. ERp29 regulates epithelial sodium channel functional expression by promoting channel cleavage. Am. J. Physiol. Cell. Physiol. 2014; 307:C701–709. [PubMed: 24944201]

(43). Suaud L, Miller K, Alvey L, Yan W, Robay A, Kebler C, Kreindler JL, Guttentag S, Hubbard MJ, Rubenstein RC. 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. [PubMed: 21525008]

(44). Makareeva E, Mertz EL, Kuznetsova NV, Sutter MB, DeRidder AM, Cabrall WA, Barnes AM, McBride DJ, Marini JC, Leikin S. Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. J. Biol. Chem. 2008; 283:4787–4798. [PubMed: 18073209]
Figure 1. Expression of Orthogonally Tagged Collagen-I Strands in HT-1080 Cells
(A) Schematic of collagen-I expression constructs.
(B) 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.
(C) Immunoblotting analysis of inducible HA-collagen-α1(I) and FLAG-collagen-α2(I) levels in the lysate and media of HT-1080Col-I cells expressing both constructs.
(D) qPCR analysis of unfolded protein response-regulated genes in HT-1080Col-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.
(E) Confocal microscopy imaging of collagen-I trafficking in HT-1080Col-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.
(F) Representative autoradiograms of [35S]-labeled HA-collagen-α1(I) immunoisolated from HT-1080Col-I media and lysates following induction of collagen-I expression. Control media and lysate were harvested from uninduced HT-1080Col-I cells.
(G) Quantification of autoradiograms in Figure 1F. 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).
Figure 2. Molecular Properties of Collagen-I Produced by HT-1080Col-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-1080Col-I cells demonstrate the presence of a stable, protease-resistant triple helix. See also Figure S1A for complete immunoblots.
(C) 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.
(D) 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-1080Col-I cells. See Figure S1B for MS2 scans.
(E) 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.
Figure 3. 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. See also Figure S2.
Figure 4. Quantitative, Mass Spectrometry-Based Mapping of the Collagen-I Proteostasis Network

(A) Schematic representation of mass spectrometry workflow employed.

(B) 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).

(C) Pie chart showing enrichment of secretory pathway proteins in the collagen-I interactome.
Figure 5. Validation and Characterization of Putative Collagen-I Proteostasis Network Components

(A) HA antibody-mediated IP of HA-collagen-α1(I) from HT-1080Col-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.

(C) 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. See also Figure S3.

(D) 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-1080Col-I cells.

(E) 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 (D) and (E), 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.
Figure 6. 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) 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.

(D) Model illustrating the role of Erp29 in collagen-I retention/quality control under ascorbate-deficient conditions.
Table 1

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 | − Asc Fold-Enrichment | 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 a,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) a,b       | TXNDC5    | 8.4                    | 13.5                   | 3               |
| Protein disulfide isomerase A4 (Erp72) a,b                | PDIA4     | 8.2                    | 15.4                   | 3               |
| Protein disulfide isomerase A3 (Erp57) a,b                | PDIA3     | 7.4                    | 14.3                   | 10              |
| Endoplasmic reticulum protein 29 a,b                      | ERP29     | 4.8                    | 8.3                    | 2               |
| ERO1-like protein-α a                                     | ERO1L     | 4.3                    | 6.6                    | 2               |
| Protein disulfide isomerase A6 (Erp5) a,b                 | PDIA6     | 4.1                    | 8.3                    | 5               |
| Peroxiredoxin-1 or peroxiredoxin-4 (PAG or Prx-IV) a,b    | PRDX1 or PRDX4 | 0.8                  | 2.9                    | 3               |
| Endoplasmic reticulum resident protein-                    | ERP44     | 0.2                    | 2.0                    | 3               |

ACS Chem Biol. Author manuscript; available in PMC 2017 May 20.
| Protein (Common Name)                                              | Gene Name | + Asc Fold-Enrichment | − Asc Fold-Enrichment | Unique Peptides |
|-------------------------------------------------------------------|-----------|------------------------|------------------------|----------------|
|                                                                   |                                                      |                         |                        |                |
| 44<sup>a,b</sup>                                                  |                                                      |                         |                        |                |
| Hsp40/70/90                                                       |                                                      |                         |                        |                |
| BiP (Grp78)                                                       | HSPA5     | 13.6                   | 22.9                   | 20              |
| Endoplasmic (Grp94)<sup>a</sup>                                   | HSP90B1   | 4.4                    | 6.4                    | 11              |
| Hypoxia up-regulated protein 1<sup>a,b</sup>                      | HYOU1     | 2.2                    | 6.1                    | 3               |
| DNAJ homolog subfamily B member 11 (Erdj3)<sup>a</sup>            | DNAJB11   | 2.0                    | 2.6                    | 3               |
| N-Glycosylation and Lectin-Assisted Folding                       |                                                      |                         |                        |                |
| Glucosidase-2 β subunit (GII)<sup>y</sup>                         | PRKCSH    | 6.9                    | 7.2                    | 4               |
| Neutral alpha-glucosidase AB<sup>e</sup>                          | GANAB     | 2.8                    | 2.9                    | 3               |
| UDP-glucose:glycoprotein glucosyltransferase 1<sup>e</sup>         | UGGT1     | 4.1                    | 6.9                    | 3               |
| Calreticulin<sup>e</sup>                                           | CALR      | 3.4                    | 5.5                    | 4               |
| Calcium Binding                                                   |                                                      |                         |                        |                |
| Calumenin<sup>e</sup>                                             | CALU      | 20.4                   | 25.3                   | 5               |
| Reticulocalbin-1<sup>e</sup>                                      | RCN1      | 7.5                    | 8.1                    | 2               |
| Secretory Pathway-Localized / Diverse Functions                   |                                                      |                         |                        |                |
| Golgi integral membrane protein 4 (Golim-4/GPP130)<sup>e,c</sup>  | GOLIM4    | 14.3                   | 19.0                   | 2               |
| Protein canopy homolog 2<sup>e</sup>                              | CNPY2     | 12.6                   | 17.9                   | 3               |
| Secreted protein acidic and rich in cysteine (SPARC)              | SPARC     | 9.8                    | 5.3                    | 2               |
| Annexin A2<sup>e</sup>                                            | ANXA2     | 6.7                    | 0.9                    | 2               |
| Fibronectin                                                       | FN1       | 5.2                    | 4.5                    | 9               |
| Protein-glutamine gamma-glutamyltransferase 2<sup>e</sup>          | 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                       |                                                      |                         |                        |                |
| Cytoskeleton-associated protein 4<sup>e</sup>                      | CKAP4     | 11.5                   | 11.1                   | 5               |
| ATP synthase subunit α, mitochondrial<sup>e</sup>                 | ATP5A1    | 4.6                    | 0.5                    | 3               |
| Major vault protein (MVP)<sup>e</sup>                             | MVP       | 2.6                    | 1.9                    | 19              |
| Heat shock protein HSP 90-β<sup>a</sup>                           | HSP90AB1  | 2.5                    | 0.4                    | 4               |
| Poly(ADP-ribose) polymerase 4<sup>e</sup>                         | PARP4     | 2.3                    | 1.1                    | 4               |
| 40S ribosomal protein S3<sup>e</sup>                              | RPS3      | 2.3                    | 1.3                    | 2               |
| 40S ribosomal protein SA<sup>e</sup>                              | RPSA      | 0.4                    | 20.1                   | 2               |
Note: Blue lines along the left-hand side of the chart denote complexes.

superscript aAll 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.

superscript bPeptides 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.

superscript cPeptides 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.

superscript *Asterisk represents proteins with ill-defined or previously unknown roles in collagen-I proteostasis.