Elucidation of proteostasis defects caused by osteogenesis imperfecta mutations in the collagen-α2(I) C-propeptide domain

Ngoc-Duc Doan[1,‡], **Azade S. Hosseini**[1,‡], **Agata A. Bikovtseva**[1,‡], **Michelle S. Huang**[1], **Andrew S. DiChiara**[1] ,
Louis J. Papa III[1], **Antonius Koller**[2], and **Matthew D. Shoulders**[1,*]

From the [1]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA and* [2]*Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*

Edited by Ursula Jakob

Intracellular collagen assembly begins with the oxidative folding of ~30-kDa C-terminal propeptide (C-Pro) domains. Folded C-Pro domains then template the formation of triple helices between appropriate partner strands. Numerous C-Pro missense variants that disrupt or delay triple-helix formation are known to cause disease, but our understanding of the specific proteostasis defects introduced by these variants remains immature. Moreover, it is unclear whether or not recognition and quality control of misfolded C-Pro domains is mediated by recognizing stalled assembly of triple-helical domains or by direct engagement of the C-Pro itself. Here, we integrate biochemical and cellular approaches to illuminate the proteostasis defects associated with osteogenesis imperfecta-causing mutations within the collagen-α2(I) C-Pro domain. We first show that "C-Pro-only" constructs recapitulate key aspects of the behavior of full-length Colα2(I) constructs. Of the variants studied, perhaps the most severe assembly defects are associated with C1163R C-Proα2(I), which is incapable of forming stable trimers and is retained within cells. We find that the presence or absence of an unassembled triple-helical domain is not the key feature driving cellular retention *versus* secretion. Rather, the proteostasis network directly engages the misfolded C-Pro domain itself to prevent secretion and initiate clearance. Using MS-based proteomics, we elucidate how the endoplasmic reticulum (ER) proteostasis network differentially engages misfolded C1163R C-Proα2(I) and targets it for ER-associated degradation. These results provide insights into collagen folding and quality control with the potential to inform the design of proteostasis network-targeted strategies for managing collagenopathies.

Proper folding and assembly of collagen is critical for maintaining the structural integrity of extracellular matrices (1–3). Mature collagen-I is a heterotrimeric, nearly 1000-amino-acid-long triple helix composed of two collagen-α1(I) strands and one collagen-α2(I) strand [abbreviated Colα1(I) and Colα2(I), respectively] (4, 5). Both collagen-I and the other fibrillar collagens are initially folded and assembled inside cells as procollagens, in which the triple-helical domain is sandwiched by ~15-kDa N-propeptide and ~30-kDa C-propeptide (C-Pro) domains (6, 7). Upon secretion, these propeptide domains are proteolytically cleaved to enable deposition of mature collagen-I triple helices in the extracellular matrix (6, 8).

Procollagen assembly occurs within the endoplasmic reticulum (ER). The process begins with folding of individual C-Pro domains. C-Pro is an *N*-glycosylated, globular domain stabilized by three key intramolecular disulfide bonds (9). Folded collagen-I C-Pro domains recognize appropriate partner C-Pro domains to nucleate the formation of 2:1 Colα1(I):Colα2(I) triple helices. This recognition process is mediated by a cysteine code that relies on a combination of Ca^{2+}-binding and intermolecular disulfide bond formation to stabilize the required 2:1 heterotrimers (10). Folding of the nearly 1000-amino-acid-long triple-helical domain then proceeds in the C → N direction (11–13). When triple-helix folding is complete, the soluble procollagen molecule is secreted and the C-Pro and N-propeptide domains are cleaved (6), releasing the mature collagen molecule for fibril assembly (14).

As the first committed step in collagen folding, missteps at the C-Pro folding and assembly stages can have devastating consequences for collagen proteostasis. Unsurprisingly, several C-Pro domain missense mutations cause autosomal-dominant collagenopathies, including osteogenesis imperfecta (OI) in the case of collagen type I (15–18). Disease-causing C-Pro variants delay C-Pro folding and assembly or, in the worst cases, likely induce severe collagen misfolding or aggregation (18–20). These defects in the C-Pro domain can then cause problems both inside the cell (*e.g.* collagen accumulation, cellular dysfunction, and/or ER stress) and outside the cell (*e.g.* deposition of overmodified collagen in bone) (18).

Extensive studies have significantly advanced our understanding of the structure and biochemistry of collagen-I C-Pro domains. Most recently, Hulmes and coworkers obtained a high-resolution crystal structure of homotrimeric C-Proα1(I) (9). The structure is reminiscent of a bouquet of flowers in which triple-helix assembly is templated by the trimerized C-Pro stem (Fig. 1A). Unfortunately, the field still lacks a structure of the biologically relevant 2:1 C-Proα1(I):C-Proα2(I) heterotrimer and, therefore, is restricted to modeling in the C-Proα2 (I) domain (9). The homotrimeric C-Proα1(I) [and C-Proα1 (III)] structures confirm the existence of three intramolecular disulfide bonds in each monomer and of a single intermolecular
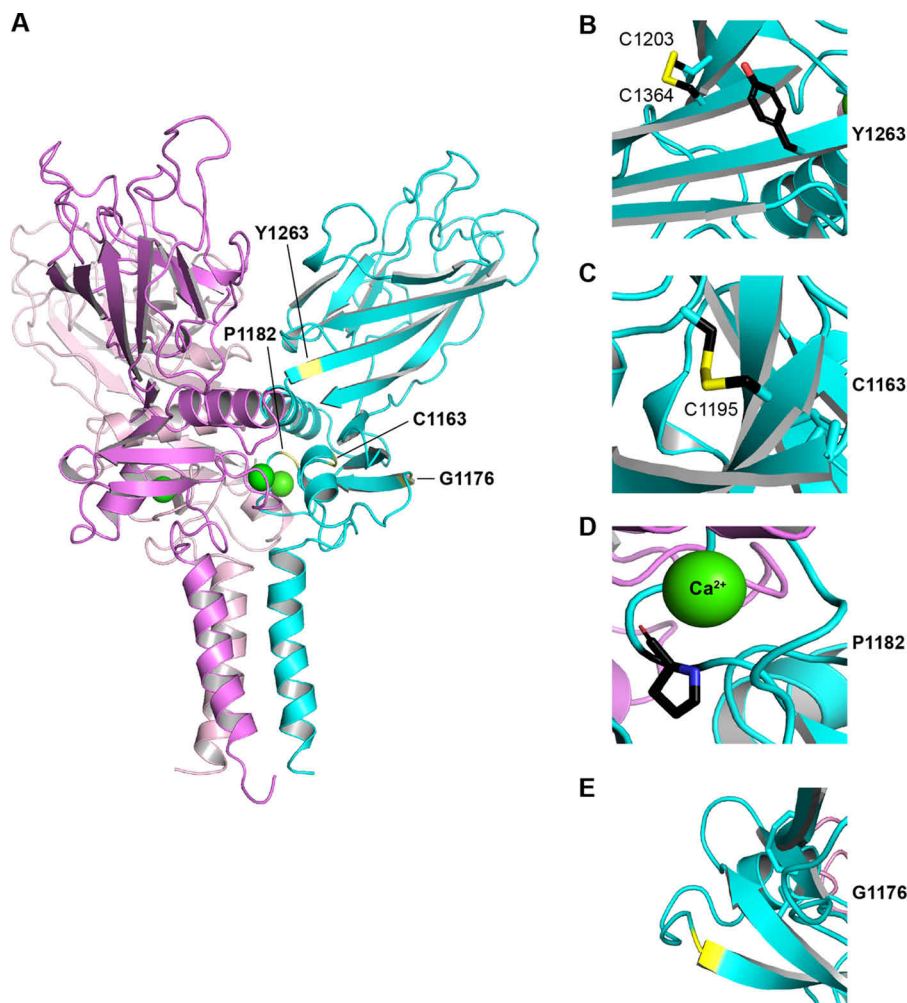
Figure 1. Structural model of the collagen-I C-Pro domain heterotrimer and locations of disease-causing amino acid substitutions. *A*, model of the 2:1 C-Proα1(I):C-Proα2(I) heterotrimer obtained by replacing a single C-Proα1(I) monomer with C-Proα2(I) using homology modeling (coordinates generously provided by Hulmes and coworkers [18]). The two α1(I) chains are colored *light* and *dark pink*, the α2(I) chain is colored *cyan*, and the Ca^{2+} ion is colored *green*. Locations of exemplary osteogenesis imperfecta (OI)-causing amino acid substitutions in C-Proα2(I) studied here are highlighted in *yellow*. *B–E*, close-up views of the locations of OI-causing mutations with the WT residue shown. Y1263C with the proximal disulfide bond (*B*), C1163R and its disulfide bond with residue C1195 (*C*), P1182R and the neighboring Ca^{2+} ion (*D*), and G1176V (*E*).

disulfide bond that covalently links the assembled monomers to each other (9, 21). The cysteine residues are important for stability and were also recently shown to be critical for controlling proper heterotrimer formation (10).

These biochemical and structural insights provide a framework to understand how C-Pro mutations can induce collagenopathies. Currently, at least thirty missense mutations in either C-Proα1(I) or C-Proα2(I) are known to cause OI (22). Some of these mutations disrupt the key cysteine residues, whereas others cause defects for less obvious reasons. In early studies, it was observed that several variants delay secretion or cause collagen-I overmodification [*e.g.* C1299W, G1272V, and T1431I in C-Proα1(I) and D1315V and G1176V in C-Proα2(I)] (19, 22, 23). Others appeared to greatly reduce procollagen secretion and cause extensive formation of Colα1(I) homotrimers [*e.g.* C1163R in C-Proα2(I)] (19). More recently, studies of C-Proα(I) variants in primary patient cells confirmed that the resulting misfolding can cause aberrant collagen-I trafficking (18).

Despite this progress, there is much that remains unknown regarding C-Pro proteostasis. Well-appreciated practical chal-

lenges associated with biochemical characterization of collagen folding, quality control, and chaperone interactions (24) present many difficulties for high-resolution characterization of defective proteostasis in the context of full-length collagen-I constructs containing disease-causing C-Pro mutations. Many variants have never been subjected to biochemical characterization. Moreover, it is unclear whether or not quality control mechanisms directly recognize misfolded C-Pro domains or if they instead require the presence of a full-length, triple-helix domain-containing collagen molecule stalled at the assembly or folding stage. Along these lines, there is considerable data now available regarding how the ER proteostasis network engages full-length collagen-I (24, 25), but which components of that network are specifically involved in C-Pro domain proteostasis still remains unclear.

We sought to obtain a more comprehensive understanding of how collagen-I folding and assembly is disrupted by disease-causing C-Pro mutations and of how the cell responds to C-Pro misfolding. We focused on C-Proα2(I) variants because, unlike C-Proα1(I), only a single copy of C-Proα2(I) is included in any
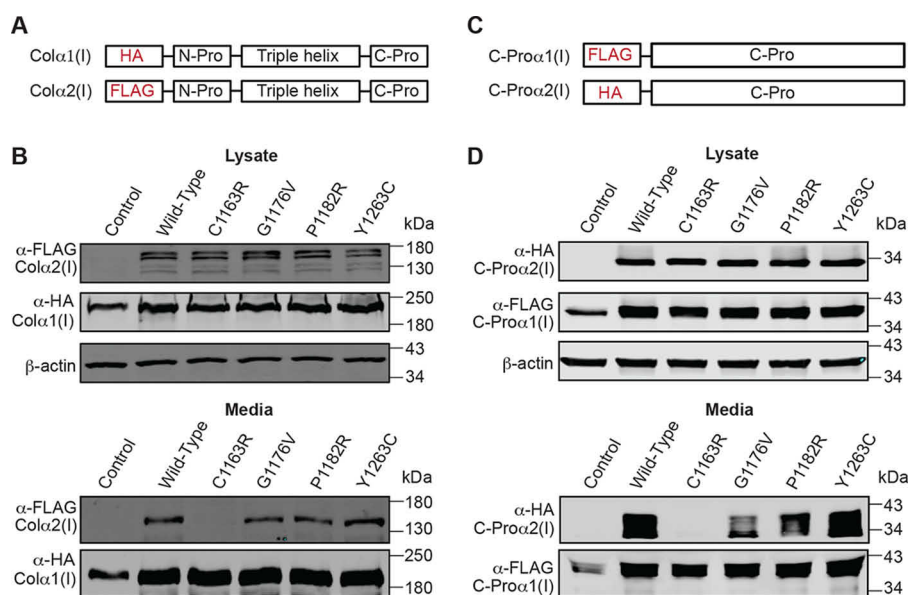
Figure 2. Effects of OI-causing mutations on full-length Colα2(I) secretion are recapitulated by expressing the biochemically amenable Colα2(I) C-Pro domain in the absence of the triple-helical domain. *A*, diagram showing full-length Colα1(I) and Colα2(I) expression constructs. *B*, immunoblots showing intracellular expression and secretion of WT Colα1(I) and WT or OI-causing mutants of Colα2(I) from the full-length collagen-I constructs. Full-length collagen is known to display anomalous mobility on SDS-PAGE gels (48). *C*, diagram showing Colα1(I) and Colα2(I) C-Pro-only constructs. *D*, immunoblots showing intracellular expression and secretion of WT C-Proα1(I) and WT or OI-causing mutants of C-Proα2(I) from the C-Pro-only constructs. HEK293T cells were transiently transfected with WT Colα1(I) and WT or mutant Colα2(I) full-length constructs or WT C-Proα1(I) and WT or mutant C-Proα2(I) constructs, as indicated. Cotransfection of Colα1(I) or C-Proα1(I) and RFP instead of Colα2(I) or C-Proα2(I) was used as a control.

given assembled heterotrimer. This stoichiometry simplifies analysis, because collagen-I heterotrimers in autosomal dominant patients must then contain only either one or no mutant C-Proα2(I) domains; there can be no cases of normal assemblies that contain both a mutant and a WT monomer. We created a set of full-length and C-Pro-only Colα2(I) constructs spanning a range of OI-causing C-Pro mutations (Figs. 1*B–E*), including mutations that add or delete Cys residues (*e.g.* Y1263C [22] and C1163R [19]) and mutations that alter charge or could otherwise be consequential for protein conformation/folding (*e.g.* P1182R [22] and G1176V [19]).

We began by showing that these C-Pro-only constructs recapitulate the behavior of the full-length collagen-I constructs, indicating that the presence *versus* the absence of unassembled or slowly assembling triple-helical domains has little impact on the trafficking of these proteins, which is instead driven mainly by the misbehaving C-Pro domain itself. Arguably, the most severe assembly defect was associated with C1163R C-Proα2(I), which was also the only variant completely retained in cells. We next applied a suite of biochemical and imaging approaches to compare the folding, trafficking, and quality control behavior of WT *versus* C1163R C-Proα2(I). Finally, we used MS-based proteomics to illuminate how the ER proteostasis network differentially engages normal *versus* misfolding collagen-I C-Pro domains. Our results provide fresh insights into the molecular determinants of collagen-I proteostasis and will inform continued efforts to resolve dysregulated collagen proteostasis in disease.

Results

Expression and analysis of OI-causing C-Proα2(I) variants

We began by developing a panel of constructs encoding missense variants of Colα2(I) within the C-Pro domain that are

known to cause autosomal-dominant OI (Figs. 1*B–E*). We chose Y1263C and C1163R C-Proα2(I) (Figs. 1*B* and *C*), because they either introduce an extra Cys residue or remove one of the Cys residues present in the WT protein (19, 22). The intricate network of cysteine residues in the C-Pro domain assists both monomer folding and proper assembly of 2:1 Colα1(I):Colα2(I) heterotrimers (10, 11, 26). Disruption of this network via adding or removing a Cys residue might be expected to be particularly deleterious. We selected P1182R (Fig. 1*D*), because this mutation not only removes a conformationally constrained Pro residue but also introduces an extra positive charge within the Ca^{2+}-binding region of the protein (22). Ca^{2+} binding was recently shown to be critical for dynamic, noncovalent assembly of C-Proα(I) heterotrimers. Finally, we employed the G1176V variant (Fig. 1*E*), a relatively more conservative mutation that was previously proposed to disrupt triple-helix formation (19). Overmodified Colα2(I) chains were observed in both the medium and cell layer of G1176V Colα2(I)-expressing primary patient fibroblasts (19).

We designed constructs in which full-length Colα2(I) variants were FLAG tagged at their N termini, whereas the corresponding WT Colα1(I) constructs were HA tagged (Fig. 2*A*). Given the lack of high-quality collagen-I antibodies, this approach enables robust differential detection of the various constructs. We previously showed that such N-terminal tags do not disrupt normal collagen-I assembly (24). We cotransfected WT or mutant full-length Colα2(I) plasmids along with the WT, full-length Colα1(I) plasmid in HEK293T cells. We selected HEK293T cells as an expression platform, because this cell line does not produce endogenous collagen-I, which could otherwise convolute our analyses, and because it was particularly convenient for our downstream studies focusing entirely

SASBMB

J. Biol. Chem. (2020) 295(29) 9959–9973 **9961**
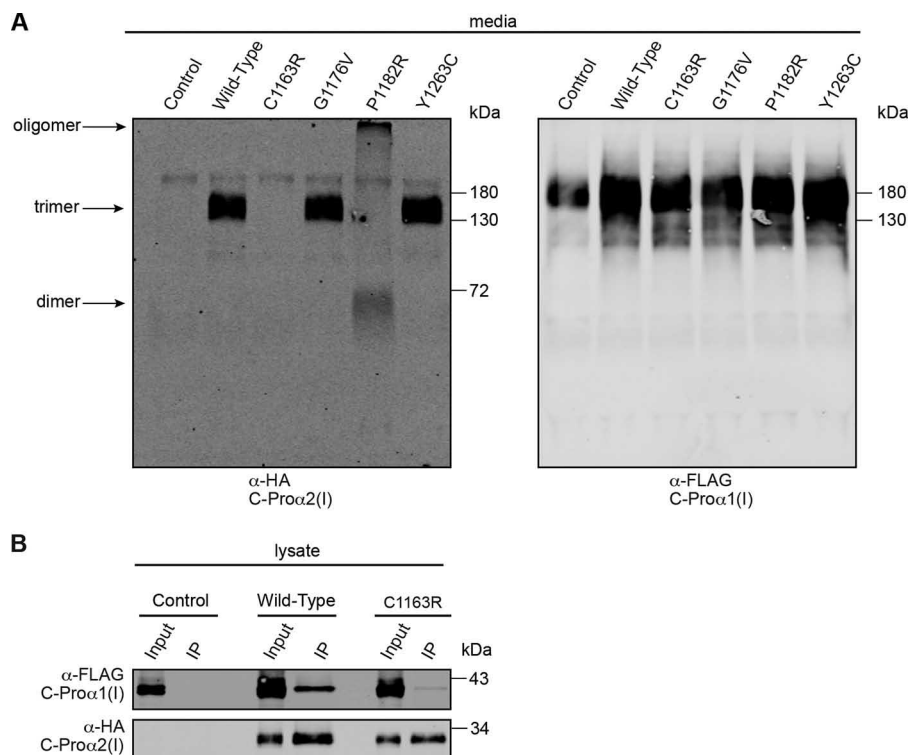
Figure 3. Influence of OI-causing mutations on C-Pro domain assembly. *A*, immunoblots of secreted fractions run under nonreducing conditions showing the formation of disulfide-linked assemblies (or the absence thereof) of C-Proα1(I) and the indicated WT and mutant C-Proα2(I) variants. Immunoblots under reducing conditions are provided in Fig. S1. G1176V and Y1263C formed disulfide-linked heterotrimers with WT C-Proα1(I). P1182R C-Proα2(I) failed to assemble into disulfide-linked heterotrimers, instead forming homooligomers mediated by incorrect disulfide bonds. As was also observed in Fig. 2, C1163R C-Proα2(I) was not secreted. *B*, immunoblot assessing intracellular interactions between WT C-Proα1(I) and WT or C1163R C-Proα2(I). WT or C1163R C-Proα2(I) was immunoprecipitated from cell lysates using anti-HA beads. The blot was then probed for coimmunoprecipitation of WT C-Proα1(I). By this assay, WT C-Proα2(I) robustly and stably assembled with WT C-Proα1(I), whereas the C1163R C-Proα2(I) variant was unable to assemble with WT C-Proα1(I) to any significant extent. HEK293T cells were transiently transfected with WT C-Proα1(I) and WT or mutant C-Proα2(I) constructs, as indicated. Cotransfection of C-Proα1(I) and RFP instead of C-Proα2(I) was used as a control.

on the C-Pro domain itself. As a control, cells cotransfected with Colα1(I) and RFP instead of Colα1(I) and Colα2(I) were included in each sample set.

We analyzed the resulting lysate and medium samples by immunoblotting of reducing SDS-PAGE gels. We observed that, just like full-length WT Colα1(I) and Colα2(I), all the full-length Colα2(I) OI variants were robustly expressed in cell lysates (Fig. 2*B*, *top*). In contrast, analysis of the corresponding medium samples revealed that, whereas all the other full-length Colα2(I) OI variants were secreted at high levels, full-length C1163R Colα2(I) was intracellularly retained (Fig. 2*B*, *bottom*). We note that the full-length WT Colα1(I) that was cotransfected with full-length C1163R Colα2(I) secreted normally into the media despite C1163R Colα2(I) retention, consistent with the notion that this mutation may prevent collagen-I heterotrimer assembly without interfering with Colα1(I) homotrimer production (19). Altogether, these observations are consistent with prior efforts to characterize collagen-I production by patient fibroblasts expressing the same variants.

We next asked whether these behaviors would be recapitulated by C-Pro-only constructs, which are much more amenable to detailed biochemical characterization. In particular, we questioned whether the presence or absence of a triple-helical domain was critical for the observed results. To test this possibility, we prepared C-Pro-only constructs lacking collagen-I triple-helical

domains such that HA epitope-tagged C-Proα2(I) WT and OI variants could be cotransfected alongside WT FLAG epitope-tagged C-Proα1(I) (Fig. 2*C*). We included a 2×-XTEN linker (27) between the FLAG epitope tag and the WT C-Proα1(I) domain to improve separation on SDS-PAGE gels and thereby facilitate differential detection of C-Proα2(I) *versus* C-Proα1(I), which have very similar molecular weights. Upon cotransfection of these plasmids (and/or an RFP control), we observed strong similarity to the behavior of the full-length constructs. In particular, C1163R C-Proα2(I) was still intracellularly retained, whereas all the other variants were robustly secreted (Fig. 2*D*).

Characterization of folding and assembly defects induced by OI-causing C-Proα2(I) variants

A key advantage of C-Pro-only constructs is that they can be much more easily subjected to detailed biochemical analysis than full-length constructs. We next made use of this feature to understand, at higher resolution, the consequences of these OI-causing C-Proα2(I) mutations for collagen-I folding and assembly. We used immunoblotting of nonreducing SDS-PAGE gels to analyze C-Pro heterotrimer assemblies in media obtained from cells cotransfected with WT C-Proα1(I) and either WT or OI-causing C-Proα2(I) variants. We found that WT C-Proα2(I) and two of the secreted OI-causing variants (G1176V and Y1263C) were able to form disulfide-linked

heterotrimers with C-Proα1(I) (see Fig. 3A for nonreducing gels and Fig. S1 for reducing gels). In contrast, the other secreted variant, P1182R, did not form disulfide-linked heterotrimers with C-Proα1(I). Interestingly, instead of forming heterotrimers with C-Proα1(I), P1182R C-Proα2(I) formed disulfide-linked homodimers and homooligomers that could be dissociated under reducing conditions and appeared to be well-defined rather than general protein aggregates. This observation is consistent with the notion that the P1182R amino acid substitution derails collagen-I proteostasis by interfering with Ca^{2+} binding (22), which we recently showed is a key early noncovalent assembly step on the pathway to formation of heterotrimers with C-Proα1(I) (10).

C1163R C-Proα2(I) was not secreted into the media, so the analysis in Fig. 3A could not reveal whether this OI-causing variant assembles with C-Proα1(I). However, immunoprecipitation of intracellular WT *versus* C1163R C-Proα2(I) from cells also expressing WT C-Proα1(I) revealed that C1163R C-Proα2(I) was indeed assembly defective (Fig. 3B), failing to trimerize with C-Proα1(I) to any detectable extent.

OI-Causing C-Proα2(I) variants do not cause acute ER stress

We next asked whether these OI-causing Colα2(I) collagen variants cause ER stress, as would be indicated by activation of the unfolded protein response (UPR). qPCR analysis revealed that transcript levels for the UPR targets (28–30) BiP, ERdj4, Grp94, CHOP, HYOU1, Sec24D, and Gadd34 were not significantly increased by expression of any of the full-length WT and OI variants of Colα2(I) studied, including the ER-retained C1163R variant and the aggregation-prone, misassembling P1182R variant (Fig. S2; thapsigargin treatment was used as a positive control for UPR activation). Similar results were also observed in cells transfected with the C-Pro-only versions of WT or OI-causing Colα2(I) (Fig. S3).

These observations are consistent with prior work indicating that collagen-I mutations, even severe disease-causing mutations, rarely induce prototypical ER stress (31–35). These general failures to induce a prototypical ER stress response raise the question of whether the ER proteostasis network is capable of differentially recognizing even very disruptive collagen-I variants.

C1163R C-Proα2(I) traffics abnormally and accumulates in the ER

We next sought to gain greater insight into the failed folding and assembly of the most severely defective of these C-Proα2(I) mutations, the C1163R variant. We first employed confocal microscopy to assess the trafficking of C1163R C-Proα2(I). We used antibodies against the ER marker protein-disulfide isomerase (PDI), the *cis*-Golgi marker GM130, and the lysosomal membrane marker LAMP1 to evaluate the localization of WT *versus* C1163R C-Proα2(I) upon cotransfection with WT C-Proα1(I) in HEK293T cells. A mAb against C-Proα1(I) was also used to track the subcellular location of that strand.

Consistent with its robust secretion, we observed that WT C-Proα2(I) trafficked extensively to the Golgi apparatus and was only weakly detected in the ER at steady state, as shown by its minimal colocalization with the ER marker PDI *versus* its

extensive colocalization with the *cis*-Golgi marker GM130 (Fig. 4A). In addition, the majority of WT C-Proα2(I) strands were found to colocalize with WT C-Proα1(I), consistent with successful heterotrimer assembly. This intracellular trafficking pattern is commonly observed for endogenous full-length, WT collagen-I (36–39).

In contrast, we observed that C1163R C-Proα2(I) was mostly retained in the ER, with only a very small proportion of this variant observed in the Golgi (Fig. 4B). These data are consistent with recognition of misfolded C1163R C-Proα2(I) by the ER quality control machinery, leading to strong intracellular retention. Notably, only a small proportion of the C1163R C-Proα2(I) colocalized with WT C-Proα1(I), likely reflecting the inability of this OI-causing Colα2(I) variant to form a heterotrimer with its WT Colα1(I) counterpart. This observation also coincides with our observation (Fig. 2D) that expression of the C1163R C-Proα2(I) variant did not prevent normal trafficking of WT C-Proα1(I).

We next used the lysosomal membrane marker LAMP1 to assess whether C1163R C-Proα2(I) was targeted for degradation through the lysosomal pathway. We observed that only a small (and similar) proportion of either WT or C1163R C-Proα2(I) colocalized with LAMP1-positive vesicles, suggesting that autophagy is not extensively involved in the degradation of accumulated, misfolded C-Proα2(I) polypeptides (Fig. S4).

C1163R C-Pro variant fails to assemble with WT C-Proα1(I) and is instead targeted for ER-associated degradation

We turned to pulse-chase analysis to better understand the fate of intracellularly retained C1163R C-Proα2(I). We began by cotransfecting WT C-Proα1(I) along with WT C-Proα2(I). Transfected cells were then metabolically labeled with a brief pulse of ^{35}S-Cys/Met-containing media, followed by a chase period in unlabeled media (Fig. 5A). RFP-transfected samples were included with each sample set to control for background signal during the subsequent analysis. We harvested cell lysate and medium samples for analysis every 30 min. To enrich the HA epitope-tagged C-Proα2(I), each sample was then immunoprecipitated using HA antibody-conjugated agarose beads. Eluted medium samples were digested with a combination of enzymes (PNGase-F and *O*-glycosidase) to remove posttranslational modifications on the C-Pro domains that would otherwise complicate SDS-PAGE separation and downstream analysis.

Analysis of the resulting autoradiographs revealed that the WT C-Proα2(I) signal steadily decreased in the lysate over time, whereas the corresponding medium signal increased, consistent with a normal secretion time course (Fig. 5B). WT C-Proα1(I) that was coimmunoprecipitated with C-Proα2(I) followed a similar trend in both lysate and medium over time. Treatment with the proteasome inhibitor MG-132 indicated little to no proteasomal degradation of WT C-Proα(I) during the experiment.

In contrast, pulse-chase analysis of C1163R C-Proα2(I) cotransfected with WT C-Proα1(I) led to detection of protein only in the cell lysates (Fig. 5C). To ensure that the lack of signal in medium was not simply because of a secretion delay, we
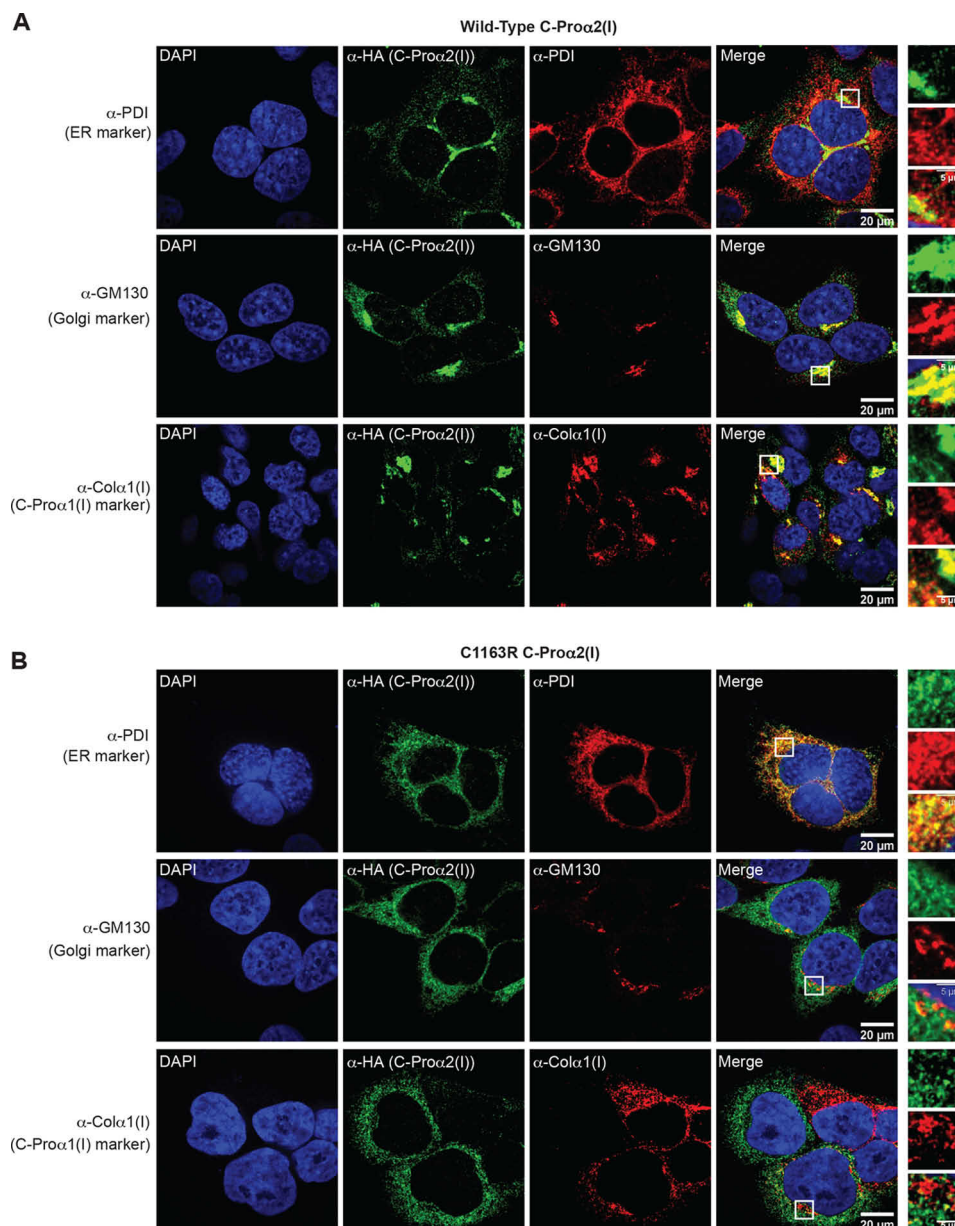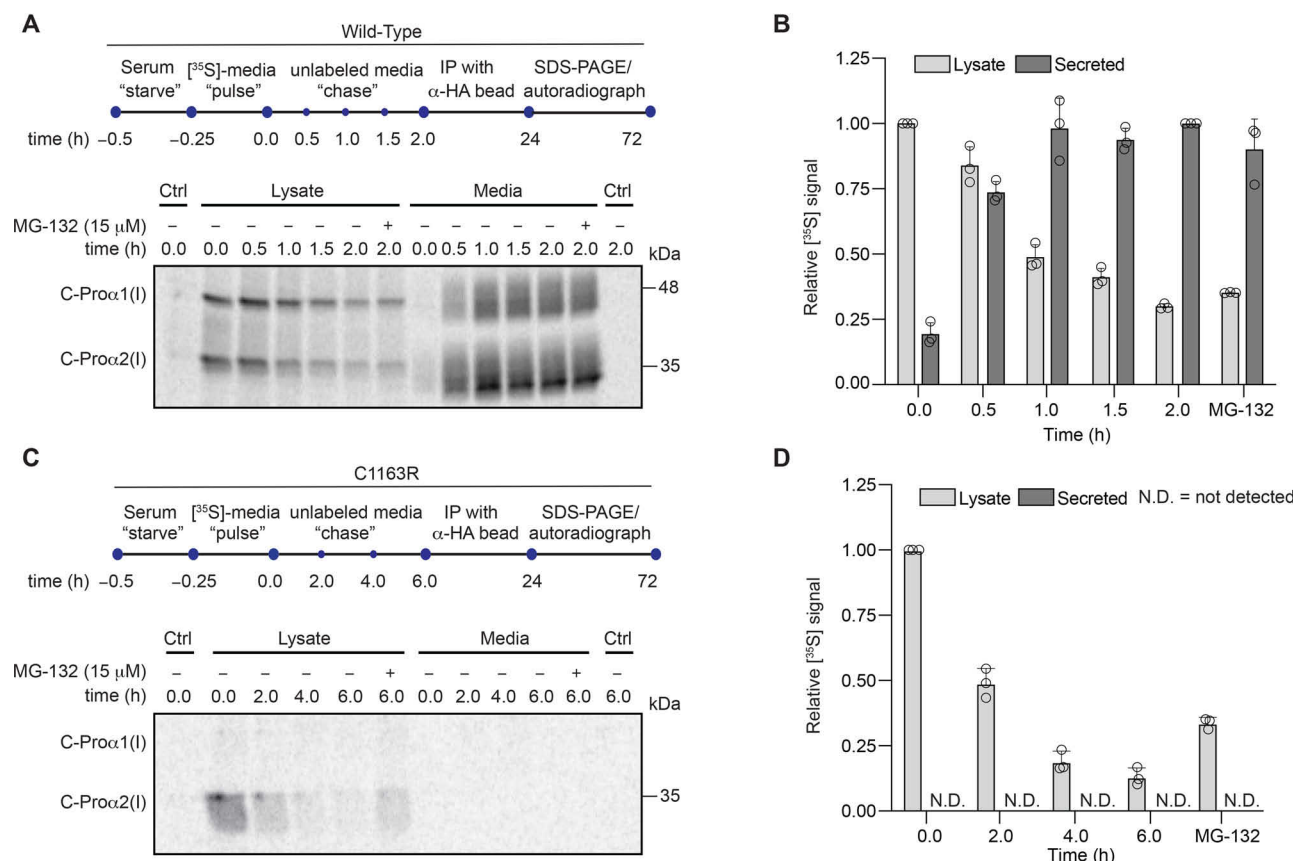
Figure 4. Trafficking of WT *versus* C1163R C-Proα2(I) in cells. *A*, confocal immunofluorescence microscopy showed minimal colocalization of WT C-Proα2 (1) with the ER luminal chaperone PDI but strong colocalization with the *cis*-Golgi matrix protein GM-130 and WT C-Proα1(I). *B*, confocal immunofluorescence microscopy showed extensive colocalization of the C1163R C-Proα2(I) variant with the PDI ER marker but limited colocalization with WT C-Proα1(I) or the *cis*-Golgi marker GM-130. HEK293T cells were transiently transfected with WT C-Proα1(I) and WT or mutant C-Proα2(I) constructs, as indicated, prior to fixing, staining with the indicated antibodies, and preparation of slides for confocal imaging. In all images, insets represent selected fields magnified 11.01× as well as their overlays. *Scale bar*, 20 μm.

extended the time points further and still observed no medium signal even after six hours. Interestingly, the corresponding lysate samples indicated that most of the C1163R C-Proα2(I) was cleared from cells within six hours despite not being secreted. We hypothesized that the intracellularly retained C1163R C-Proα2(I) variant might be directed to ER-associated degradation (ERAD). Indeed, treatment with 15 μM of MG-132 at the latest time point (6 h) rescued a substantial fraction of the intracellular signal, consistent with clearance of C1163R C-Proα2(I) by ERAD (Fig. 5*D*). Notably, C-Proα1(I) was never detectably coimmunoprecipitated with C1163R C-Proα2(I) during this experiment, further confirming that this variant of C-Proα2(I) failed to assemble with C-Proα1(I) throughout its lifetime in the cell.

MS-based interactomics reveal how the ER proteostasis network differentially engages WT versus C1163R C-Proα2(I)

We next sought to understand how the ER proteostasis network engages WT *versus* C1163R C-Proα2(I). We began by defining the interactome of the WT C-Pro, which has not previously been established. Cells were cotransfected with RFP and WT C-Proα1(I) to create a negative control or cotransfected with WT C-Proα1(I) and WT C-Proα2(I) for the experimental sample. Following our previously developed protocol for coimmunoprecipitation of full-length collagen-I and its interactors (24, 40), we then briefly treated intact cells with the cell-permeable, lysine-reactive, reversible cross-linker dithiobis-(succinimidyl propionate) to immortalize

Figure 5. Pulse-chase experiments quantify altered trafficking kinetics for WT *versus* C1163R C-Proα2(I). *A* and *C*, representative gel images for WT (*A*) or C1163R (*C*) C-Proα2(I) kinetics of expression and secretion. HEK293T cells were transfected with WT C-Proα1(I) and WT or mutant C-Proα2(I) constructs, pulsed with ^{35}S-labeled Met- and Cys-containing media, and then chased with unlabeled media. Cell layers and media were harvested at the indicated time points, C-Proα2(I) was immunoprecipitated, and samples were separated using reducing SDS-PAGE gels and then imaged using autoradiography. *B* and *D*, quantification of relative WT (*B*) or C1163R (*D*) C-Proα2(I) ^{35}S signal in the lysate and the media, normalized to the maximum detected signal, at the indicated time points. Experiments were performed in biological triplicate, and standard deviations as well as individual data points are shown.

transient interactions between C-Proα2(I) and the proteostasis network. Samples were then immunoprecipitated using anti-HA agarose beads, followed by repeated washes and elution using a denaturing, nonreducing buffer. Eluted protein samples were then further processed for LC-MS/MS analysis (see Experimental procedures; two biological replicates of both the negative control and experimental sample were analyzed, with an additional technical replicate of the experimental sample). The resulting spectral count data for all identified proteins in these samples are provided in Table S1.

In any given experimental replicate, 15–27% of the tryptic peptides identified belonged to the 247-residue WT C-Proα2(I) protein, indicating that the bait was strongly enriched. A total of 110 proteins were identified across all negative-control and experimental samples. In Table 1, we present the results for all the high-confidence WT C-Proα2(I) interactors, as defined by meeting the following criteria: 1) absent from one or both of the negative-control replicates; 2) at least two spectral counts in at least one experimental sample; and 3) presence in at least two independent biological replicates of the experimental sample.

A number of features indicate the reliability of the resulting WT C-Proα2(I) interactome. First, based on semiquantitative spectral counting, the most abundant C-Proα2(I) inter-

actor is C-Proα1(I). Second, 75% of the identified interactors are known to be localized to the secretory pathway. Third, other than C-Proα1(I), virtually all the secretory pathway interactors are components of the ER proteostasis network and, based on their known functions, are likely to interact with and assist the folding of this protein. Fourth, most of the putative interactors were previously shown to interact with full-length collagen-I (24). Further, only a subset of the full-length collagen-I interactors were identified. Notably missing are the peptidyl prolyl isomerases, the triple-helix-specific chaperone HSP47, and a vast array of triple-helix-modifying enzymes. The absence of these proteins from the C-Pro-only interactome should be expected, as those interactors specifically support triple-helix maturation. Instead, for C-Proα2(I) we observe engagement by PDIs, Hsp40/70/90 chaperones that typically assist folding of globular proteins, and components of the ER's lectin-based proteostasis network. Because the C-Pro domain is a globular, disulfide-rich, *N*-glycosylated protein, its folding would reasonably depend specifically on assistance from these components of the ER proteostasis network.

With the first analysis of how the ER proteostasis network engages the collagen C-Pro domain in hand, we turned our attention to the misfolding, OI-causing C1163R C-Proα2(I)

Table 1
Mass spectrometry-based mapping of the wild-type C-Proα2(I) interactome[a]

Protein (common name)	Gene	Negative control (spectral counts)		Wild-type C-Proα2(I) (spectral counts)			Full-length collagen-I interactor?[b]
		Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 3	
Collagen-I							
Collagen-α2(I)	*COL1A2*	0	0	134	98	177	
Collagen-α1(I)	*COL1A1*	0	0	63	75	106	
ER HSP40/70/90 and Related							
BiP (Grp78)	*HSPA5*	0	0	27	14	23	✓
Endoplasmin (Grp94)	*HSP90B1*	0	0	3	2	9	✓
ERdj3	*DNAJB11*	0	0	1	0	2	✓
Disulfide bond formation/shuffling and ER redox chemistry							
Peroxiredoxin 4	*PRDX4*	1	0	2	3	7	✓
PDIA1	*P4HB*	0	0	5	2	5	✓
PDIA3	*PDIA3*	0	0	1	0	3	✓
PDIA4	*PDIA4*	0	0	5	3	8	✓
PDIA6	*PDIA6*	0	0	7	4	5	✓
ER lectin-based proteostasis network and related							
Glucosidase 2 β subunit	*PRKCSH*	0	0	1	1	2	✓
Neutral α-glucosidase AB	*GANAB*	0	0	4	4	7	✓
Other ER proteostasis network components							
Transitional endoplasmic reticulum ATPase	*VCP*	1	0	1	0	3	✗
Secretory pathway localized/diverse or unknown function							
Cytoskeleton-associated protein 4	*CKAP4*	0	0	3	1	9	✓
ERGIC-53	*LMAN1*	0	0	1	1	2	✗
Golgi apparatus protein 4	*GLG1*	0	0	3	0	1	✗
Golgi internal membrane protein 1	*GOLIM4*	0	0	1	0	2	✓
Multiple myeloma tumor-associated protein 2	*MMTAG2*	1	0	1	1	4	✗
Protein canopy homolog 2	*CNPY2*	0	0	9	2	4	✓
Not known to be localized to secretory pathway							
Filamin A	*FLNA*	0	3	23	3	7	✗
Heat shock cognate protein 71	*HSPA8*	1	0	10	6	14	✗
Heat shock protein 70-1B	*HSPA1B*	0	0	9	7	17	✗
Rabankyrin 5	*ANKFY1*	2	0	4	0	1	✗
Stress-70 protein (GRP75)	*HSPA9*	0	0	2	0	8	✗
Thioredoxin-dependent peroxide reductase 3	*PRDX3*	0	0	1	2	3	✗

[a]See Table S1 for completely unfiltered, raw data set. The common contaminants ribonucleoprotein/ribosomal proteins, histones, tubulin, and keratin are not included in this table.
[b]Previously identified as a component of the full-length collagen-I proteostasis network (23).

variant. Given that we had observed intracellular retention of C1163R C-Proα2(I) and targeting to ERAD, we hypothesized that a comparative analysis of the WT *versus* C1163R C-Proα2(I) interactomes would reveal differential engagement of the disease-causing variant by the ER proteostasis network.

Samples from cells cotransfected with WT C-Proα1(I) and WT C-Proα2(I) or cotransfected with WT C-Proα1(I) and C1163R C-Proα2(I) (two biological replicates and one additional technical replicate of each sample) were immunoprecipitated and processed by following the same procedure as that used for establishing the WT C-Proα1(I) interactome. Following LC-MS/MS analysis, protein identification, and quantitation of spectral counts (see Table S2 for the complete data set), we set out to establish the differences in these two interactomes.

In Table 2, we present the results for proteins that interact differentially with WT *versus* C1163R C-Proα2(I) based on the following criteria: 1) absence from both the negative-control samples from Table S1; 2) >2-fold enrichment in bait-normalized spectral counts in at least two biological replicates of either the WT or C1163R C-Proα2(I) samples; and 3) at least two spectral counts present in at least one of the enriched samples.

Three key features of this comparative analysis support the reliability of the results. First, >85% of the proteins shown to interact more with C1163R C-Proα2(I) than with WT C-Proα2(I) are known to be localized to the secretory pathway. Second, as expected based on our results in Fig. 3*B*, the comparative proteomic analysis shows that C-Proα1(I) interacts much more strongly with WT than with C1163R C-Proα2(I). Third, we used immunoprecipitation and Western blotting to corroborate several of the proteins identified to preferentially engage C1163R C-Proα2(I). All three (PDIA3, PDIA6, and BiP) were shown to interact with C1163R more than with WT C-Proα2(I) (Fig. 6).

The results (Table 2) provide a compelling picture of greatly enhanced engagement of the misfolding C1163R C-Proα2(I) variant by the ER proteostasis network. Of particular note, the ER's HSP40/70/90 and lectin-based chaperone and quality control systems differentially recognize the misfolding variant. Further, the PDI machinery interacts much more extensively with C1163R C-Proα2(I). This observation is consistent with the notion that the loss of a key cysteine disrupts the C-Pro disulfide network in a particularly deleterious manner. The outcome of this enhanced proteostasis network engagement is the complete intracellular retention coupled with effective quality control of C1163R C-Proα2(I).

Table 2

MS-based comparative analysis of how the ER proteostasis network differentially engages WT *versus* C1163R C-Proα2(I)[a]

Protein (common name)	Gene	C1163R spectral counts / WT spectral counts[b]		
		Replicate 1	Replicate 2	Replicate 3
Collagen-I				
Collagen-α2(I)[c] (bait protein)	COL1A2	$\frac{55}{63}$	$\frac{82}{98}$	$\frac{203}{177}$
Collagen-α1(I)	COL1A1	$\frac{5}{23}$	$\frac{29}{75}$	$\frac{35}{106}$
ER HSP40/70/90 and related				
BiP (Grp78)	HSPA5	$\frac{82}{19}$	$\frac{166}{14}$	$\frac{170}{23}$
Endoplasmin (Grp94)	HSP90B1	$\frac{14}{7}$	$\frac{29}{2}$	$\frac{34}{9}$
ERdj3	DNAJB11	$\frac{1}{0}$	$\frac{10}{0}$	$\frac{17}{2}$
ERdj5	DNAJB10	$\frac{1}{0}$	$\frac{1}{0}$	$\frac{5}{1}$
Hypoxia-upregulated protein 1	HYOU1	$\frac{2}{0}$	$\frac{14}{0}$	$\frac{23}{2}$
ER disulfide bond formation/shuffling and ER redox chemistry				
PDIA1	P4HB	$\frac{5}{2}$	$\frac{16}{2}$	$\frac{24}{5}$
PDIA3	PDIA3	$\frac{3}{1}$	$\frac{38}{0}$	$\frac{40}{3}$
PDIA4	PDIA4	$\frac{3}{1}$	$\frac{31}{3}$	$\frac{52}{8}$
PDIA6	PDIA6	$\frac{13}{5}$	$\frac{20}{4}$	$\frac{24}{5}$
Thioredoxin domain-containing protein 5 (Erp46)	TXNDC5	$\frac{2}{1}$	$\frac{0}{0}$	$\frac{9}{0}$
ER lectin-based proteostasis network and related				
Calreticulin	CALR	$\frac{1}{0}$	$\frac{6}{0}$	$\frac{10}{0}$
Dolichol-diphosphooligosaccharide-protein glycosyltransferase subunit 1	RPN1	$\frac{3}{1}$	$\frac{4}{0}$	$\frac{4}{1}$
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	$\frac{3}{0}$	$\frac{11}{0}$	$\frac{17}{0}$
Secretory pathway localized/diverse or unknown function				
Collagen-α1(V)	COL5A1	$\frac{3}{1}$	$\frac{2}{0}$	$\frac{1}{1}$
SLIT and NTRK-like protein 4	SLITRK4	$\frac{1}{0}$	$\frac{4}{0}$	$\frac{0}{0}$
Stromal cell-derived factor 2-like protein 1	SDF2L1	$\frac{1}{0}$	$\frac{4}{0}$	$\frac{3}{0}$
Not known to be localized to secretory pathway				
Flap endonuclease 1	FEN1	$\frac{8}{1}$	$\frac{0}{0}$	$\frac{1}{0}$
Interleukin enhancer-binding factor 2	ILF2	$\frac{2}{8}$	$\frac{0}{2}$	$\frac{1}{3}$
Sam68	KHDRBS1	$\frac{5}{2}$	$\frac{2}{1}$	$\frac{2}{2}$
Single-stranded DNA-binding protein	SSBP1	$\frac{5}{2}$	$\frac{1}{1}$	$\frac{3}{1}$

[a] Proteins displaying a >2-fold change in bait-normalized spectral counts across at least two biological replicates are included in this table. See Table S2 for complete, unfiltered data set. The common contaminants ribonucleoprotein/ribosomal proteins, histones, tubulin, and keratin are not included in this table.
[b] Spectral counts shown for C1163R C-Proα2(I) interactors are normalized based on the amount of C1163R *versus* WT C-Proα2(I) spectral counts observed in a given replicate.
[c] Spectral counts shown for C-Proα2(I) are not normalized.
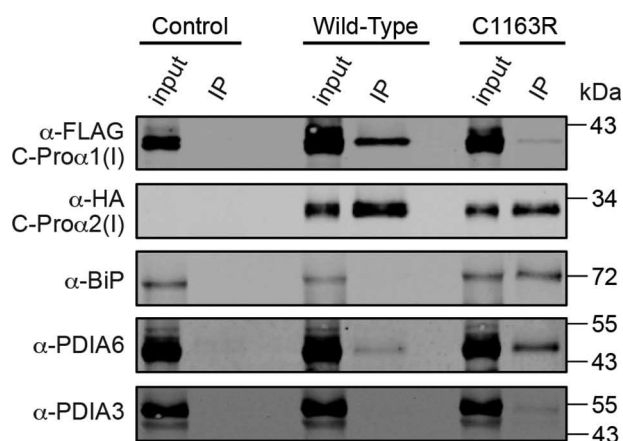
Figure 6. Confirmation of MS-based interactome data showing how the ER proteostasis network differentially engages WT *versus* misfolded C-Proα2(I). Shown is an immunoblot assessing intracellular interactions between WT or C1163R C-Proα2(I), expanded to include additional interactors beyond C-Proα1(I) that is shown in Fig. 3*B*. WT or C1163R C-Proα2(I) was immunoprecipitated from cell lysates using anti-HA beads. The blot was then probed for coimmunoprecipitation of BiP, PDIA3, and PDIA6, validating increased interaction of these ER proteostasis network components with the misfolding C1163R variant. Cotransfection of C-Proα1(I) and RFP instead of C-Proα2(I) was used as a control.

Discussion

The critical first step in the production of functional fibrillar collagens is the folding and proper assembly of C-Pro domains (41). However, the proteostasis network mechanisms that specifically engage the C-Pro domain, the molecular consequences of disease-causing C-Pro mutations, and how the cell recognizes and responds to misfolded C-Pro domains have all remained
ill-defined (28). Here, we studied the folding, misfolding, processing, and interactome of collagen-I C-Pro domains in cells using a biochemically amenable system in which C-Pro domains were expressed in the absence of their associated triple-helical domains. This C-Pro-only system provides a valuable context to elucidate how selected OI-causing C-Pro domain mutations disrupt collagen-I assembly and proteostasis.

We first showed that the behavior of WT *versus* mutant C-Pro-only constructs recapitulated the processing of the corresponding full-length collagen-I constructs. Comparison of the medium and lysate samples of mutant C-Pro constructs with the corresponding full-length constructs confirmed similar secretion and retention patterns. It is noteworthy that, whereas C-Pro domain mutations have been shown to be targeted to degradation pathways, it was previously unclear whether the quality control system directly recognizes and engages the misfolded C-Pro domain itself or rather recognizes unfolded triple-helical domains whose assembly is stalled by the presence of a misfolded C-Pro (42, 43). Our data indicate that, at least for the mutations studied here, the presence of a triple-helical domain is not an important driver of the behavior of the mutant collagens. Misfolded C1163R C-Proα2(I) is still retained in the absence of the triple-helical domain, whereas misfolded P1182R Colα2(I) is still secreted when the triple-helical domain is present. This observation suggests that cells have limited capacity to specifically recognize triple-helix domain defects.

We characterized the consequences of four OI-causing C-Proα2(I) mutations reported to result in type I OI: specifically, mutations resulting in the G1176V, Y1263C, P1182R, and C1163R amino acid substitutions. The G1176V and Y1263C C-Proα2(I) variants formed disulfide-linked heterotrimers with C-Proα1(I) and did not significantly impact collagen-I secretion. These results are consistent with prior observations that these variants only modestly delay collagen-I production in patient cells and result in a mild OI phenotype (19, 22).

On the other hand, the P1182R and C1163R C-Proα2(I) variants caused much more substantive issues from the perspective of collagen proteostasis. P1182R C-Proα2(I) was secreted at high levels but did not detectably assemble with C-Proα1(I) at all. Stable formation of collagen-I heterotrimers is a process that depends on dynamic noncovalent assembly of various trimeric species in a Ca^{2+}-mediated process, followed by covalent immortalization of appropriately assembled heterotrimers using interstrand disulfide bonds (10). Disruption of Ca^{2+} binding, which is a plausible consequence of positioning a positively charged arginine residue near the Ca^{2+}-binding site (Fig. 1*D*), therefore would be expected to prevent the assembly of disulfide-linked heterotrimers. Interestingly, we found that this mutation resulted in the exclusive formation of what clearly must be badly misassembled, disulfide-linked homodimers and homooligomers of P1182R C-Proα2(I). This defect did not prevent secretion of P1182R C-Proα2(I), consistent with observations by us here and others elsewhere that the P1182R substitution does not substantially impede the secretion of full-length Colα2(I) (22). Intriguingly, the P1182R amino acid substitution causes type IV OI with moderate phenotypes rather than causing a severe form of OI (22), despite the severe misassembly into homooligomers. A likely explanation is 2-fold: 1) P1182R oligomers are well secreted, avoiding risks such as severe ER stress or cellular dysfunction, and 2) because P1182R does not significantly assemble with Colα1(I), the simultaneous presence of WT Colα2(I) in patient cells enables the formation of normal collagen-I heterotrimers; P1182R-driven homooligomers may not be stable or extensively deposited into extracellular matrices, although it may still be disruptive.

The C1163R C-Proα2(I) amino acid substitution proved similarly problematic from a protein-folding and assembly perspective. The loss of C1163 disrupts a critical intrachain disulfide bond in C-Proα2(I). We found that the resulting protein was recognized as misfolded by the ER proteostasis network, and secretion was entirely prevented. Notably, the disease phenotype caused by the C1163R C-Proα2(I) mutation is quite mild type I OI (19), despite the severe folding defect. Our data suggest three factors that contribute. 1) Very little, or none, of the C1163R C-Proα2(I) variant escapes quality control to be secreted, avoiding extracellular matrix defects. 2) Unlike the case for many intracellularly retained triple helical variants (32, 42), the cell robustly directs retained C1163R to ERAD, avoiding extensive buildup in the ER and consequent ER dysfunction or stress. 3) Very little, or none, of the C1163R Colα2(I) assembles with Colα1(I), instead being quality controlled and degraded. Therefore, in heterozygous patients,

WT Colα2(I) can still assemble normally with Colα1(I) to form heterotrimers. In this regard, it is notable that Colα1(I) homotrimers were also produced by C1163R mutant patient cells, likely reflecting the deficiency in Colα2(I) caused by the misfolding mutation (19). Our data also help to explain why the C1163R Colα2(I) mutation is not as deleterious from a pathology perspective as the P1182R Colα2(I) mutation. For example, P1182R patients suffer from bone fractures, whereas C1163R patients typically do not. The ER proteostasis network can prevent secretion of C1163R Colα2(I) and direct it to degradation. In contrast, P1182R Colα2(I) misassembles into homooligomers, escapes quality control, and is secreted into the extracellular matrix, where it may introduce some structural deficiencies.

Intrigued by the efficient ER retention and clearance observed with the C1163R C-Proα2(I) variant, we turned to MS to establish how the ER proteostasis network engages WT C-Pro domains and to identify how the ER chaperone and quality control system specifically identifies misfolding C1163R C-Proα2(I) to retain it intracellularly and direct it to ERAD. Our proteomic studies revealed increased interaction of C1163R C-Proα2(I), compared with WT, with the PDI family (*e.g.* PDIA3, PDIA4, and PDIA6), as well as general ER chaperones (*e.g.* Hyou1, BiP, and ERdj3). The PDIs are known to promote proper disulfide bond formation and engage clients with improperly assembled disulfides. The increased interaction of C1163R C-Proα2(I) with these proteostasis network components highlights the impact of disulfide bond network disruption. The increased interaction with the HSP40/70/90 system likely reflects the mechanism of targeting to ERAD.

In conclusion, these results highlight the potential of the biochemically amenable C-Pro-only system to provide molecular-level insights into how cells handle collagen proteostasis defects. The assembly, pulse-chase, and interactome experiments are all challenging to perform with this level of resolution using the full-length protein, because of the inherent challenges associated with working with GC-rich and repetitive genes, MS-grade immunoprecipitations of collagen, and the difficulty of obtaining high-quality immunoblotting data. Thus, this approach sets the stage not just for fundamental understanding of collagen folding but also for future studies aimed at identifying whether and how we can resculpt cellular proteostasis networks to better address collagen folding defects (35). Modulation of the cellular proteostasis network is an emerging and promising strategy for targeting other protein misfolding-related diseases (44). Improving the proteostasis network's ability to identify and successfully fold or prevent the misfolding of mutant collagen-I strands could prove beneficial in therapeutic settings. A critical first step is studies such as that presented here, characterizing proteostasis defects in detail and identifying how the proteostasis network attempts to address such defects.

Experimental procedures

Plasmids

Plasmids encoding procollagen-α1(I) and procollagen-α2(I) with the preprotrypsin signal sequence upstream of an HA or FLAG epitope tag, respectively, were described previously (24). Plasmids (pcDNA3.1) encoding FLAG- and HA-tagged C-Proα1(I) and C-Proα2(I), respectively, were prepared by PCR amplifying the C-Pro domains spanning from the endogenous C-proteinase cleavage site (45) to the C terminus of procollagen using primers that incorporated the NotI and EcoRV restriction enzyme cut sites for insertion downstream of a preprotrypsin signal sequence and the indicated HA or FLAG tag. The FLAG-tagged C-Proα1(I) construct was then further modified to incorporate a 2×-XTEN linker by inserting annealed oligonucleotides into the NotI site. The additional amino acids introduced by this linker enabled separate detection of C-Proα1(I) *versus* C-Proα2(I) on autoradiography gels from pulse-chase experiments. OI variants were introduced by site-directed mutagenesis using the QuikChange XL II kit from Agilent Technologies and the primers listed in Table S3. Complete open reading frames for the WT constructs are provided in the supporting information.

Cell culture and transfections

HEK293T cells (ATCC) were cultured in Dulbecco's modification of Eagle's medium (DMEM; Corning) supplemented with 15% fetal bovine serum (FBS; Corning), 100 IU penicillin (Corning), 100 μg/ml streptomycin (Corning), and 2 mM L-glutamine (Corning) at 37 °C in a humidified 5% CO_2 atmosphere. Cells were regularly tested for mycoplasma contamination using the Agilent MycoSensor PCR assay kit. Transient transfections of full-length and C-Pro domain-encoding collagen-I and RFP-encoding plasmids were performed using Lipofectamine 3000 (Thermo Fisher Scientific). For all experiments, medium was changed to fresh DMEM supplemented with 50 μM L-ascorbate (Amresco) 24 h posttransfection. Media and lysates were harvested 24–48 h posttransfection for analysis. Cells were harvested, washed with 1× PBS, and then lysed at 4 °C in a buffer containing 50 mM Tris-HCl at pH 7.5, 150 mM sodium chloride, 1 mM EDTA, 1.5 mM magnesium chloride, 1% Triton X-100 (Integra), and protease inhibitor tablets (Sigma). Each experiment was performed in biological triplicate.

Immunoblotting

Prior to SDS-PAGE analysis, media and lysate samples were diluted in 6× gel loading buffer (300 mM Tris at pH 6.8, 15% glycerol, 6% SDS, and 10% [w/v] bromphenol blue) containing 100 mM DTT and boiled for 15 min prior to separation on homemade 12% SDS-PAGE gels. Gels were transferred using a Trans Blot Turbo (Bio-Rad) onto nitrocellulose membranes using transfer buffer at pH 9.2 containing 48 mM tris, 39 mM glycine, 0.1% SDS, and 20% ethanol. Membranes were blocked with 5% milk and then probed with the following primary antibodies diluted in 5% BSA (VWR): mouse anti-HA (1:10,000; Thermo, 26183), rat anti-DYKDDDDK (1:500; Agilent Technologies, 200474), mouse anti-β-actin (1:5000; Sigma, A1978), rabbit anti-BiP (1:1000; CST, 3177), rabbit anti-PDIA6 (1:1000), and rabbit anti-PDIA3 (1:1000; CST 2881). Blots were imaged after incubation with appropriate primary and 800CW or 680LT secondary antibodies (LI-COR) by scanning on an Odyssey IR imager (LI-COR).

Immunoprecipitations

24 h posttransfection with the indicated constructs in a 70% confluent 10-cm dish, HEK293T cells were washed with $1\times$ PBS, diluted to 10 ml with $1\times$ PBS, and treated with a final concentration of 100 μM dithiobis(succinimidyl propionate) (Thermo Fisher) for 30 min. The reaction was quenched by treatment with 1 ml of 1 M Tris buffer at pH 8.0 for 10 min. Cell pellets were collected by centrifugation and then lysed by treatment for 20 min with lysis buffer (see above) at 4 °C. Cell debris was removed by centrifugation at $10,000 \times g$ for 15 min. Protein concentration in the supernatant was quantified using a Bradford assay (Bio-Rad). A ratio of 1 mg of each sample to a 60-μl slurry of HA-antibody conjugated agarose beads (Sigma) was then diluted to 1 ml in lysis buffer. Samples were mixed end-over-end at 4 °C overnight and centrifuged at $1500 \times g$ for 5 min at 4 °C, and then the supernatant was removed and the beads were washed three times with lysis buffer. Next, the beads were boiled in 100 μl of 6% SDS in 300 mM Tris at pH 6.8 for 30 min to elute the proteins. After spinning samples at $2000 \times g$, the eluent was carefully separated from the beads for further analysis.

Pulse-chase analyses

Following our previously published protocol (46), 6×10^6 HEK293T cells were transfected with 14 μg of DNA (2:1 ratio of 1A1:1A2 or 1A1:control plasmids) 24 h postseeding in a 10-cm dish. 16 h posttransfection, cells were split (1:6) and seeded on poly(D)lysine (Sigma)-coated 6-well plates. After 24 h, cells were starved for 15 min in DMEM containing dialyzed 10% FBS lacking Cys and Met. Cells were then metabolically labeled with pulse medium containing [^{35}S]-Cys/Met (final concentration of ~0.1 mCi/ml; MP Biomedical) for 15 min, washed three times with complete DMEM, and then incubated in prewarmed chase medium (complete DMEM). MG-132 (15 μM; Enzo) was included as a control at the longest time point for each replicate. The experiments were performed in biological triplicate. Samples were harvested at the indicated time points. Medium was collected, spun down at $1500 \times g$ to remove debris, and then added to HA-antibody agarose beads (30 μl; Sigma). Cells were washed with $1\times$ PBS and lysed for 20 min in lysis buffer (see above). Lysates were spun down at $10,000 \times g$ for 15 min to remove cell debris, and then the supernatant was added to HA-antibody agarose beads. All immunoprecipitations were allowed to incubate on an end-over-end rotator at 4 °C overnight. The following day, the supernatant was removed and the beads were washed three times with radioimmunoprecipitation assay buffer containing 50 mM Tris at pH 7.5, 150 mM sodium chloride, 1% Triton X-100, 0.5% DOC, 0.1% SDS, and 1 mM EDTA. Immunoisolates were eluted by boiling in $6\times$ Laemmli buffer lacking SDS for 15 min and then spun down to separate the eluent from the beads. WT medium samples were treated with a combination of PNGase-F, neuraminidase, and *O*-glycosidase (NEB), and digestions were performed according to the manufacturer's instructions. Samples were denatured by boiling in $6\times$ Laemmli buffer (300 mM aqueous Tris at pH 6.8, 15% glycerol, 6% SDS, and 10% [w/v] bromphenol blue) supple- mented with 167 mM 1,4-DTT. Eluted samples were separated on homemade 15% SDS-PAGE gels. Gels were dried, exposed to a phosphorimager plate, and then imaged on a Typhoon imager. Band intensities were quantified using ImageQuant TL (GE Healthcare). Experiments were performed in biological triplicate with standard deviation shown.

Quantitative RT-PCR

Transfected cells were harvested and washed with PBS at 4 °C. Total RNA was extracted using the Omega RNA purification kit. RNA concentrations were quantified and normalized to 1 μg total RNA for cDNA reverse transcription. cDNA was synthesized in a Bio-Rad Thermocycler using the Applied Biosystems reverse transcriptase cDNA kit. Kapa BioSystems Sybr Fast qPCR master mix, appropriate primers (Sigma Aldrich), and cDNA were used for amplification in a Light Cycler 480 II real-time PCR Instrument. Primer sequences are shown in Table S4. Primer integrity was assessed by thermal melt to ensure homogeneity. Transcripts were normalized to the housekeeping gene *RPLP2*. Standard deviation of $n = 3$ is shown in plots.

Confocal microscopy

Transfected HEK293T cells (4×10^4 cells) suspended in complete DMEM were plated on a 24-well plate with poly (D)lysine-coated coverslips (Chemglass Life Sciences) and incubated for 48 h at 37 °C in a humidified 5% CO$_2$ atmosphere. Culture medium was removed and the coverslips were carefully washed with PBS. Cells were fixed with 4% formaldehyde (Mallinckrodt) for 3 h at 4 °C and then permeabilized with 0.1% Triton X-100 in PBS for another 30 min at room temperature. Coverslips were then incubated for 1 h at room temperature in a blocking buffer containing 1% BSA in TBS at pH 7.5. Double labeling was performed by incubating coverslips overnight at 4 °C in TBS (5% BSA, 0.01% sodium azide) containing mouse anti-HA (1:200; Abcam, ab9110) and then mouse anti-PDI (1:200; Abcam, ab2792), anti-LAMP1 (1:200; Abcam), rabbit anti-Col1A1 (1:200; Sigma, HPA008405), or mouse anti-GM130 (1:500; BD Biosciences, 610822). Secondary antibodies (1:1000), Alexa Fluor 488-conjugated anti-mouse (Invitrogen) or Alexa Fluor 568-conjugated anti-rabbit (Invitrogen), were then applied to the coverslips for 2 h at room temperature. Coverslips were rinsed at least $3\times$ with TBS after each incubation. Nuclei were stained with DAPI (1 μM; Invitrogen) for 15 min at room temperature. After final washes ($3\times$ with PBS), the coverslips were mounted with ProLong (Thermo) to prevent photobleaching. Negative controls for nonspecific binding of the secondary antibodies obtained by omitting primary antibodies in the staining protocol were included for each experiment. Images were acquired on a Zeiss AxioVert200M microscope with a $63\times$ oil immersion objective and a Yokogawa CSU-22 spinning disk confocal head with a Borealis modification (Spectral Applied Research/Andor) and a Hamamatsu ORCA-ER charge-coupled device camera. The MetaMorph software package (Molecular Devices) was used to control the

hardware and image acquisition. The excitation lasers used to capture the images were 405 nm, 488 nm, and 561 nm. Image processing was performed using ImageJ (NIH).

MS-based interactome analyses

Eluted C-Pro samples obtained from immunoprecipitations (as described above) were precipitated by vortexing with 450 μl of MeOH. 150 μl of CHCl$_3$ was then added and the sample was vortexed again. 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 carefully removed, whereas the white precipitate at the solvent interface was preserved. The collected precipitate was then washed 3× with 0.5 ml of MeOH. The washed pellet was dried using a SpeedVac and then resuspended in an aqueous solution containing 8 M urea, 50 mM ammonium bicarbonate, and 10 mM DTT. The samples were incubated in a 56 °C water bath for 45 min, cooled for 2 min at room temperature, and then incubated with 55 mM iodoacetamide for 1 h in the dark. Samples were next incubated with 1 μg of sequencing-grade trypsin (Promega) overnight at room temperature. Proteolyzed samples were acidified to a final concentration of 5% formic acid and subjected to C$_{18}$ stage tips for desalting. After eluting tryptic peptides from C$_{18}$ stage tips, the samples were dried by SpeedVac and then resuspended in 0.1% formic acid.

Samples were injected onto an EASY-nLC 1000 nanopump system connected to a Thermo Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer and analyzed using the LC-MS/MS parameters specified in Tables S1 and S2. For database searching, the five samples used for the WT interactome analysis (Table S1) and the six samples used for the comparative interactome analysis (Table S2) were analyzed separately to obtain accurate statistics. Tandem mass spectra were extracted, and charge states were deconvoluted using ProteomeDiscoverer v2.3. Deisotoping was not performed. Samples were analyzed using Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version IseNode in Proteome Discoverer 2.3.0.523). Sequest was used to search Uniprot_Human, updated 1 December 2019 with 20,533 entries, containing common contaminant proteins and assuming fully tryptic peptides with at most 2 missed cleavages. Sequest was searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 10.0 ppm. Carbamidomethylation of cysteine was specified in Sequest as a fixed modification. Oxidation of methionine and acetylation of the N terminus were specified in Sequest as variable modifications. For criteria for protein identification, Scaffold (Scaffold_4.10.0, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at >6.0% probability to achieve a false discovery rate of <1.0% by the Scaffold Local false discovery rate algorithm. Protein identifications were accepted if they could be established at >97.0% probability to achieve a false discovery rate of <1.0% and contained at least two identified peptides in at least one sample. Protein probabilities were assigned by the Protein Prophet algorithm (47). Proteins

that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. See Table S5 for detailed results of these analyses.

Data availability

All the proteomic data are available on MassIVE with accession number MSV000085340.

Author contributions—A. S. H., A. S. D., and M. D. S. conceptualization; A. K. data curation; A. K. formal analysis; M. D. S. funding acquisition; N.-D. D., A. S. H., A. A. B., M. S. H., and A. S. D. investigation; N.-D. D., A. S. H., A. A. B., A. S. D., and L. J. P. methodology; M. D. S. project administration; M. D. S. supervision; A. A. B. and M. D. S. visualization; N.-D. D., A. S. H., A. A. B., and M. D. S. writing-original draft; N.-D. D., A. S. H., A. A. B., M. S. H., A. S. D., L. J. P., A. K., and M. D. S. writing review and editing.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: ER, endoplasmic reticulum; OI, osteogenesis imperfecta; PDI, protein-disulfide isomerase.

References

1. Shoulders, M. D., and Raines, R. T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* **78,** 929–958 CrossRef Medline
2. Eyre, D. R. (1980) Collagen: molecular diversity in the body's protein scaffold. *Science* **207,** 1315–1322 CrossRef Medline
3. Ricard-Blum, S. (2011) The collagen family. *Cold Spring Harb. Perspect. Biol.* **3,** a004978 CrossRef Medline
4. Khoshnoodi, J., Cartailler, J. P., Alvares, K., Veis, A., and Hudson, B. G. (2006) Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. *J. Biol. Chem.* **281,** 38117–38121 CrossRef Medline
5. Makareeva, E., Aviles, N. A., and Leikin, S. (2011) Chaperoning osteogenesis: new protein-folding disease paradigms. *Trends Cell Biol.* **21,** 168–176 CrossRef Medline

6. Prockop, D. J., Kivirikko, K. I., Tuderman, L., and Guzman, N. A. (1979) The biosynthesis of collagen and its disorders (first of two parts). *N. Engl. J. Med.* **301,** 13–23 CrossRef Medline

7. Olsen, B. R., Guzman, N. A., Engel, J., Condit, C., and Aase, S. (1977) Purification and characterization of a peptide from the carboxy-terminal region of chick tendon procollagen type I. *Biochemistry* **16,** 3030–3036 CrossRef Medline

8. Leung, M. K., Fessler, L. I., Greenberg, D. B., and Fessler, J. H. (1979) Separate amino and carboxyl procollagen peptidases in chick embryo tendon. *J. Biol. Chem.* **254,** 224–232 Medline

9. Sharma, U., Carrique, L., Vadon-Le Goff, S., Mariano, N., Georges, R. N., Delolme, F., Koivunen, P., Myllyharju, J., Moali, C., Aghajari, N., and Hulmes, D. J. (2017) Structural basis of homo- and heterotrimerization of collagen I. *Nat. Commun.* **8,** e14671 CrossRef

10. DiChiara, A. S., Li, R. C., Suen, P. H., Hosseini, A. S., Taylor, R. J., Weickhardt, A. F., Malhotra, D., McCaslin, D. R., and Shoulders, M. D. (2018) A cysteine-based molecular code informs collagen C-propeptide assembly. *Nat. Commun.* **9,** e4206 CrossRef

11. Boudko, S. P., Engel, J., and Bächinger, H. P. (2012) The crucial role of trimerization domains in collagen folding. *Int. J. Biochem. Cell B* **44,** 21–32 CrossRef Medline

12. Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. *Eur. J. Biochem.* **106,** 619–632 CrossRef Medline

13. Doege, K. J., and Fessler, J. H. (1986) Folding of carboxyl domain and assembly of procollagen I. *J. Biol. Chem.* **261,** 8924–8935 Medline

14. Buehler, M. J. (2006) Nature designs tough collagen: explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. U S A* **103,** 12285–12290 CrossRef Medline

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

16. Forlino, A., Cabral, W. A., Barnes, A. M., and Marini, J. C. (2011) New perspectives on osteogenesis imperfecta. *Nat. Rev. Endocrinol.* **7,** 540–557 CrossRef Medline

17. Lamandé, S. R., Chessler, S. D., Golub, S. B., Byers, P. H., Chan, D., Cole, W. G., Sillence, D. O., and Bateman, J. F. (1995) Endoplasmic reticulum-mediated quality-control of type-I collagen production by cells from osteogenesis imperfecta patients with mutations in the proα1(I) chain carboxyl-terminal propeptide which impair subunit assembly. *J. Biol. Chem.* **270,** 8642–8649 CrossRef Medline

18. Barnes, A. M., Ashok, A., Makareeva, E. N., Brusel, M., Cabral, W. A., Weis, M., Moali, C., Bettler, E., Eyre, D. R., Cassella, J. P., Leikin, S., Hulmes, D. J. S., Kessler, E., and Marini, J. C. (2019) COL1A1 C-propeptide mutations cause ER mislocalization of procollagen and impair C-terminal procollagen processing. *Biochim. Biophys. Acta Mol. Basis Dis.* **1865,** 2210–2223 CrossRef Medline

19. Pace, J. M., Wiese, M., Drenguis, A. S., Kuznetsova, N., Leikin, S., Schwarze, U., Chen, D., Mooney, S. H., Unger, S., and Byers, P. H. (2008) Defective C-propeptides of the proα2(I) chain of type I procollagen impede molecular assembly and result in osteogenesis imperfecta. *J. Biol. Chem.* **283,** 16061–16067 CrossRef

20. Oliver, J. E., Thompson, E. M., Pope, F. M., and Nicholls, A. C. (1996) Mutation in the carboxy-terminal propeptide of the proα1(I) chain of type I collagen in a child with severe osteogenesis imperfecta (OI type III): possible implications for protein folding. *Hum. Mutat.* **7,** 318–326 CrossRef

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

22. Symoens, S., Hulmes, D. J., Bourhis, J. M., Coucke, P. J., De Paepe, A., and Malfait, F. (2014) Type I procollagen C-propeptide defects: study of genotype-phenotype correlation and predictive role of crystal structure. *Hum. Mutat.* **35,** 1330–1341 CrossRef Medline

23. Pace, J. M., Kuslich, C. D., Willing, M. C., and Byers, P. H. (2001) Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the proα1(I) chain of type I procollagen permits slow assembly and

24. DiChiara, A. S., Taylor, R. J., Wong, M. Y., Doan, N.-D., Rosario, A. M., and Shoulders, M. D. (2016) Mapping and exploring the collagen-I proteostasis network. *ACS Chem. Biol.* **11,** 1408–1421 CrossRef Medline

25. Ishikawa, Y., and Bächinger, H. P. (2013) A molecular ensemble in the rER for procollagen maturation. *Biochim. Biophys. Acta* **1833,** 2479–2491 CrossRef Medline

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

27. Schellenberger, V., Wang, C. W., Geething, N. C., Spink, B. J., Campbell, A., To, W., Scholle, M. D., Yin, Y., Yao, Y., Bogin, O., Cleland, J. L., Silverman, J., and Stemmer, W. P. (2009) A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27,** 1186–1190 CrossRef Medline

28. Wong, M. Y., DiChiara, A. S., Suen, P. H., Chen, K., Doan, N.-D., and Shoulders, M. D. (2018) Adapting secretory proteostasis and function through the unfolded protein response. *Curr. Top. Microbiol. Immunol* **414,** 1–25 CrossRef Medline

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

30. Wong, M. Y., Chen, K., Antonopoulos, A., Kasper, B. T., Dewal, M. B., Taylor, R. J., Whittaker, C. A., Hein, P. P., Dell, A., Genereux, J. C., Haslam, S. M., Mahal, L. K., and Shoulders, M. D. (2018) XBP1s activation can globally remodel N-glycan structure distribution patterns. *Proc. Natl. Acad. Sci. U S A* **115,** E10089–E10098 CrossRef Medline

31. Lisse, T. S., Thiele, F., Fuchs, H., Hans, W., Przemeck, G. K., Abe, K., Rathkolb, B., Quintanilla-Martinez, L., Hoelzlwimmer, G., Helfrich, M., Wolf, E., Ralston, S. H., and Hrabé de Angelis, M. (2008) ER stress-mediated apoptosis in a new mouse model of osteogenesis imperfecta. *PLoS Genet.* **4,** e7 CrossRef Medline

32. Bateman, J. F., Sampurno, L., Maurizi, A., Lamandé S. R., Sims, N. A., Cheng, T. L., Schindeler, A., and Little, D. G. (2019) Effect of rapamycin on bone mass and strength in the α2(I)-G610C mouse model of osteogenesis imperfecta. *J. Cell. Mol. Med.* **23,** 1735–1745 CrossRef

33. Mirigian, L. S., Makareeva, E., Mertz, E. L., Omari, S., Roberts-Pilgrim, A. M., Oestreich, A. K., Phillips, C. L., and Leikin, S. (2016) Osteoblast malfunction caused by cell stress response to procollagen misfolding in α2(I)-G610C mouse model of osteogenesis imperfecta. *J Bone Miner. Res.* **31,** 1608–1616 CrossRef

34. Chessler, S. D., Byers, P. H. (1993) Bip binds type-I procollagen proα chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. *J. Biol. Chem.* **268,** 18226–18233 Medline

35. Wong, M. Y., and Shoulders, M. D. (2019) Targeting defective proteostasis in the collagenopathies. *Curr. Opin. Chem. Biol.* **50,** 80–88 CrossRef Medline

36. Canty, E. G., and Kadler, K. E. (2005) Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* **118,** 1341–1353 CrossRef Medline

37. Nabavi, N., Pustylnik, S., and Harrison, R. E. (2012) Rab GTPase mediated procollagen trafficking in ascorbic acid stimulated osteoblasts. *PLoS ONE* **7,** e46265 CrossRef Medline

38. Bonfanti, L., Mironov, A. A., Jr., Martínez-Menárguez, J. A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H. J., Mironov, A. A., and Luini, A. (1998) Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* **95,** 993–1003 CrossRef Medline

39. Van Duyn Graham, L., Sweetwyne, M. T., Pallero, M. A., and Murphy-Ullrich, J. E. (2010) Intracellular calreticulin regulates multiple steps in fibrillar collagen expression, trafficking, and processing into the extracellular matrix. *J. Biol. Chem.* **285,** 7067–7078 CrossRef Medline

40. Doan, N.-D., DiChiara, A. S., Del Rosario, A. M., Schiavoni, R. P., and Shoulders, M. D. (2019) Mass spectrometry-based proteomics to define

ASBMB

intracellular collagen interactomes. *Methods Mol. Biol.* (1944) 95–114 CrossRef

41. Hulmes, D. J. (2002) Building collagen molecules, fibrils, and suprafibrillar structures. *J. Struct. Biol.* **137,** 2–10 CrossRef Medline

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

43. Fitzgerald, J., Lamandé, S. R., and Bateman, J. F. (1999) Proteasomal degradation of unassembled mutant type I collagen proα1(I) chains. *J. Biol. Chem.* **274,** 27392–27398 CrossRef Medline

44. Sebastian, R. M., and Shoulders, M. D. (2020) Chemical biology framework to illuminate proteostasis. *Annu. Rev. Biochem.* **89,** in press CrossRef Medline

45. Lindahl, K., Barnes, A. M., Fratzl-Zelman, N., Whyte, M. P., Hefferan, T. E., Makareeva, E., Brusel, M., Yaszemski, M. J., Rubin, C. J., Kind-mark, A., Roschger, P., Klaushofer, K., McAlister, W. H., Mumm, S., Leikin, S., *et al.* (2011) *COL1* C-propeptide cleavage site mutations cause high bone mass osteogenesis imperfecta. *Hum. Mutat.* **32,** 598–609 CrossRef Medline

46. Dewal, M. B., DiChiara, A. S., Antonopoulos, A., Taylor, R. J., Harmon, C. J., Haslam, S. M., Dell, A., and Shoulders, M. D. (2015) XBP1s links the unfolded protein response to the molecular architecture of mature N-glycans. *Chem. Biol.* **22,** 1301–1312 CrossRef Medline

47. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75,** 4646–4658 CrossRef Medline

48. Hayashi, T., and Nagai, Y. (1980) The anomalous behavior of collagen peptides on sodium dodecyl sulfate-polyacrylamide gel electrophoresis is due to the low content of hydrophobic amino acid residues. *J. Biochem.* **87,** 803–808 CrossRef Medline

ASBMB

J. Biol. Chem. (2020) 295(29) 9959–9973   **9973**