

RESEARCH ARTICLE

# Substitutions for arginine at position 780 in triple helical domain of the $\alpha 1(I)$ chain alter folding of the type I procollagen molecule and cause osteogenesis imperfecta

Elena Makareeva<sup>1</sup>✉, Guoli Sun<sup>2</sup>✉, Lynn S. Mirigian<sup>1</sup>, Edward L. Mertz<sup>1</sup>, Juan C. Vera<sup>1</sup>, Nydea A. Espinoza<sup>1</sup>, Kathleen Yang<sup>2</sup>, Diana Chen<sup>2</sup>, Teri E. Klein<sup>3</sup>, Peter H. Byers<sup>2,4</sup>‡, Sergey Leikin<sup>1</sup>\*✉

**1** Section on Physical Biochemistry, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Department of Pathology, University of Washington, Seattle, Washington, United States of America, **3** Department of Genetics, Stanford University, Palo Alto, California, United States of America, **4** Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington, United States of America

✉ These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

\* [leikins@mail.nih.gov](mailto:leikins@mail.nih.gov)



**OPEN ACCESS**

**Citation:** Makareeva E, Sun G, Mirigian LS, Mertz EL, Vera JC, Espinoza NA, et al. (2018) Substitutions for arginine at position 780 in triple helical domain of the  $\alpha 1(I)$  chain alter folding of the type I procollagen molecule and cause osteogenesis imperfecta. PLoS ONE 13(7): e0200264. <https://doi.org/10.1371/journal.pone.0200264>

**Editor:** Florence Ruggiero, Ecole normale superieure de Lyon, FRANCE

**Received:** March 22, 2018

**Accepted:** June 24, 2018

**Published:** July 10, 2018

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information file.

**Funding:** This work was supported by NIH grant R01AR051582 and by funds from the Medical Genetics Training Grant at the University of Washington and from the Department of Pathology, University of Washington (T.E.K.); USPHS grants AR051582 and GM007454 (P.H.B.);

## Abstract

OI is a clinically and genetically heterogeneous disorder characterized by bone fragility. More than 90% of patients are heterozygous for mutations in type I collagen genes, *COL1A1* and *COL1A2*, and a common mutation is substitution for an obligatory glycine in the triple helical Gly-X-Y repeats. Few non-glycine substitutions in the triple helical domain have been reported; most result in Y-position substitutions of arginine by cysteine. Here, we investigated leucine and cysteine substitutions for one Y-position arginine, p.Arg958 (Arg780 in the triple helical domain) of pro $\alpha 1(I)$  chains that cause mild OI. We compared their effects with two substitutions for glycine located in close proximity. Like substitutions for glycine, those for arginine reduced the denaturation temperature of the whole molecule and caused asymmetric posttranslational overmodification of the chains. Circular dichroism and increased susceptibility to cleavage by MMP1, MMP2 and catalytic domain of MMP1 revealed significant destabilization of the triple helix near the collagenase cleavage site. On a cellular level, we observed slower triple helix folding and intracellular collagen retention, which disturbed the Endoplasmic Reticulum function and affected matrix deposition. Molecular dynamic modeling suggested that Arg780 substitutions disrupt the triple helix structure and folding by eliminating hydrogen bonds of arginine side chains, in addition to preventing HSP47 binding. The pathogenic effects of these non-glycine substitutions in bone are probably caused mostly by procollagen misfolding and its downstream effects.

and Intramural Research Program, National Institute of Child Health and Human Development, National Institutes of Health (S.L.).

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Osteogenesis imperfecta (OI) is a clinically and genetically heterogeneous disorder characterized by susceptibility to fracture [1]. Besides bone fragility, individuals with OI may present with short stature, bone deformities, dentinogenesis imperfecta, blue sclera, scoliosis, hearing loss, and other connective tissue disorders [2]. Although discoveries of causative mutations in over a dozen different genes have generated considerable excitement and identified new biochemical pathways that contribute to OI [3–9], about 90% of individuals with OI have a mutation in one of the two genes, *COL1A1* and *COL1A2*, that encode the pro $\alpha$ 1(I) and pro $\alpha$ 2(I) chains of type I collagen, the major protein of bone [7,10].

*COL1A1* and *COL1A2* mutations lead to OI by altering synthesis, assembly, folding, post-translational modification, secretion, and structure of collagen. The molecular events in the cell and extracellular matrix (ECM) that translate these effects into bone pathology are incompletely understood. In addition to the events in the ECM, disruption of procollagen folding in the osteoblast endoplasmic reticulum (ER) appears to cause bone pathology by affecting cell differentiation and/or function [6,11–13]. At the same time, deficient secretion and/or incorporation of abnormal collagen molecules into ECM can alter bone matrix assembly, mechanical properties, mineralization, and interactions with various growth factors and cells [14,15].

The majority of mutations identified in individuals with severe or lethal forms of OI result in substitutions for glycine residues in the 338 canonical Gly-X-Y repeats of collagen triple helix [1] of the pro $\alpha$ 1(I) and pro $\alpha$ 2(I) chains of type I procollagen. Substitutions of glycine (Gly) by amino acids with charged or branched hydrophobic side chains are particularly disruptive for triple helix structure and folding [16], and generally result in the most severe OI outcomes [15]. While it has become clear that such disruptions might cause malfunction of osteoblasts as well as malfunction of collagen molecules in bone matrix, the underlying biochemical pathways and their roles in bone pathology are less clear.

Less disruptive substitutions at the X and Y positions might provide important insights into pathophysiology of missense mutations. In the last few years, X- and Y-position substitutions have been identified in individuals with OI [17–21], osteopenia [22], arterial rupture [23,24], classical Ehlers-Danlos syndrome (EDS) [24,25], and Caffey disease [26]. Often these mutations result in substitutions of cysteine for arginine, but other alterations also occur, and the molecular pathophysiology of non-glycine substitutions in the triple helix remains uncertain. Since cysteine for arginine substitutions are quite common, a particularly puzzling question is whether it is the loss of the charged arginine side chain or the acquisition of the sulfhydryl cysteine side chain (or both) that cause these phenotypic changes.

In the current work we describe an individual with a mild form of OI, who has a unique mutation in one *COL1A1* allele (c.2873G>T, p.Arg958Leu). To understand some of the mechanisms by which non-glycine substitutions could result in OI phenotypes, we compared the effects of this substitution with cysteine substitution for the same arginine (c.2873G>T, p.Arg958Leu) and with two substitutions for glycine residues located in the same region of the triple helix (c.2821G>A, p.Gly941Ser and c.2830G>T, p.Gly944Cys). We refer to these substitutions mostly as Arg780Leu (R780L), Arg780Cys (R780C), Gly763Ser (G763S), and Gly766Cys (G766C), respectively. This “legacy” designation based on the triple helical “address” facilitates direct comparison with substitutions in *COL1A2* and in the triple helical domains of other fibrillar collagens.

In fibroblasts we found that both non-glycine substitutions mimicked the effects of substitutions for glycine in that triple helix folding was delayed, procollagen was retained in the ER and resulted in ER dilation and deficient matrix deposition. These observations suggest that substitutions of at least some Y-position arginine residues can destabilize the triple helix and

disrupt procollagen folding in osteoblast ER like glycine substitutions, thereby causing bone pathology.

## Materials and methods

### Human subjects

Human subject sample collection and analysis at University of Washington (UW) was approved by the Human Subjects Institutional Review Board (IRB), UW. Clinical samples were transferred to a UW research repository with a waiver of consent for participation in research studies by the UW IRB. Analysis of anonymized samples from the UW repository at NIH was approved by the NIH Office of Human Subject Research.

### Collagen purification and analysis

Dermal fibroblasts were grown from punch biopsies obtained with appropriate consent. Normal control cells from multiple individuals of different ages were used to identify the normal range of variation in different experimental observations [27]. Normal controls shown in the figures were mostly based on CRL-2127 cells purchased from American Type Culture Collection (Manassas, VA). Fibroblasts were grown to confluence at 37 °C in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 2 mM Glutamax (Invitrogen) in the presence of 5% CO<sub>2</sub>. The cells were incubated for 24 hours in DMEM/Glutamax with 0.1% fetal bovine serum and 250 μM of 2-Phospho-L-ascorbic acid trisodium salt (Fluka). The harvested medium was buffered with 100 mmol/L of Tris-HCl (pH 7.4), protected with protease inhibitors, and used for purification of collagen [27]. Pepsin-digested secreted, intracellular and matrix collagens were prepared [27].

For analysis of overmodification, collagen and procollagen were metabolically labeled with [<sup>3</sup>H]proline (GE Healthcare) and analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) [28]. For analysis of CNBr peptides, the chains of type I collagen were separated under non-reducing condition in 5% SDS-PAGE gels, the regions of the bands were excised from the gel, treated with CNBr mixture and then set atop a second 12.5% SDS-PAGE gel and analyzed [28].

Amino acid analysis to quantify lysine and hydroxylysine was contracted with Alphasys protein analysis service (Palo Alto, CA), which performed the analysis by high pressure liquid chromatography.

To demonstrate collagen posttranslational overmodification, pepsin-purified type I collagen from the medium of either control cells or cells from the patient was fluorescently labeled either by AlexaFluor 488 (AF488, Invitrogen) or Cy5 (GE Healthcare) [29]. Then the mixtures of NC\_AF488/NC\_Cy5, NC\_AF488/mutant\_Cy5 and NC\_Cy5/mutant\_AF488 were analyzed after separation of chains by SDS-PAGE.

Kinetics of collagen triple helix folding and clearance from fibroblasts were measured by pulse metabolic labeling with azidohomoalanine in methionine-free media followed by chase in media with 10 mM of methionine [30].

Collagen thermal stability was determined by sensitivity to proteases [31] and by differential scanning calorimetry (DSC) at 0.125 °C/min heating rate. [32]. Procollagen was examined in phosphate buffered saline (PBS), pH 7.4. Pepsin-purified collagen was examined in 0.5 M glycerol, 0.2 M Na-phosphate, pH 7.4 (PGB) to inhibit collagen aggregation and fibrillogenesis [32]. DSC thermograms of collagen in PBS were corrected for the contribution of type III collagen and corrected by -1.7 °C to represent collagen stability in PBS as described previously [27,32]. Denaturation thermogram of pepsin-purified intracellular α1(I)-G763S collagen in PGB was measured by differential scanning circular dichroism [27], demonstrating similar

content of mutant and normal molecules in intracellular and secreted collagen. Local reversible unfolding was studied by circular dichroism [27].

### Electron microscopy (EM) and Raman micro-spectroscopy

Cells were cultured for 3 weeks on an ACLAR<sup>®</sup> fluoropolymer film in DMEM that contained 10% FBS and 250  $\mu$ M 2-Phospho-L-ascorbic acid trisodium salt, media was changed every 2–3 days. For EM, matrix was fixed with 2.5% glutaraldehyde in 130 mM sodium cacodylate buffer, postfixed in 2% OsO<sub>4</sub>, processed into Spurr's epoxy, sectioned, stained with UO<sub>2</sub>-acetate and Pb-citrate, and examined in a JEOL 1400 electron microscope at NICHD Imaging core. For Raman spectroscopy, cell matrix was fixed with 0.7% formaldehyde solution in 0.7x PBS. Relative ratios of matrix collagen to cell organics were determined from Raman spectra within at least 10 different matrix regions as described in [33].

### Collagen cleavage with soluble MMPs

We studied cleavage kinetics of the mutant and normal control collagens by recombinant human MMP1 (gift of Prof. H. Nagase, Imperial College, London), the catalytic domain of MMP1 (Enzo Life Sciences), and MMP2 (EMD Biosciences). MMP1 was activated with 1 mM 4-aminophenylmercuric acetate [34]. For these experiments, purified type I collagen from the medium of either control cells or cells from the patient was fluorescently labeled either by AlexaFluor 488 (AF488, Invitrogen) or Cy5 (GE Healthcare) [29]. Binary mixtures of mutant-AF488/control-Cy5 and mutant-Cy5/control-AF488 in 50 mM Tris, 150 mM NaCl, 10 mM CaCl<sub>2</sub>, 0.05% Brij were prepared and co-processed with the enzyme at different temperatures. The cleavage products and kinetics were examined on precast 3–8% Tris-Acetate gels (Invitrogen) using the FLA5000 scanner (Fuji Medical Systems) for fluorescence detection and ScienceLab software supplied with the scanner for quantitative analysis.

### DNA sequencing

Genomic DNA and RNA and cDNA were prepared by standard methods. Primers and conditions for amplification of genomic and cDNA and for automated sequencing are available upon request.

### Quantitative real time PCR (qPCR)

RNA was isolated from cells using Purelink RNA kit (Life Technologies), reverse transcribed with SuperScript III First Strand Synthesis Supermix and random hexamers as primers (Invitrogen), and analyzed in a 7500 Fast Real Time PCR system (Applied Biosystems) with Taqman gene expression assays. The same C<sub>T</sub> threshold value was used for all samples. Relative mRNA quantity was calculated from  $\Delta\Delta C_T$  values assuming 100% PCR efficiency [35].

### Western blotting

Cells were grown to confluence and incubated with fresh medium that contained 250  $\mu$ M ascorbate one day before collection. At collection, cells were rinsed with PBS and lysed in 2% sodium dodecyl sulfate, lithium dodecyl sulfate, 50mM dithiothreitol, 50mM Tris, 150mM NaCl, 5mM EDTA, 1mM phenylmethylsulfonyl fluoride, 5mM benzamidine, 10mM *N*-ethylmaleimide, and PhosStop (Roche). Samples were denatured at 95°C for 10 minutes, loaded onto 3–8% Tris Acetate or 12% Bis-Tris gel and transferred onto a 0.45  $\mu$ m nitrocellulose membrane. The blots (Figures A-C in [S1 File](#)) were labeled with primary antibodies against pro $\alpha$ 1(I) C-propeptide (LF-42, a gift of Dr. Larry Fisher, NIDCR, NIH), BIP (Cell Signaling,

3177), PDI (Enzo Life Sciences, ADI-SPA-891-D), Calnexin (Abcam, ab10286), Calreticulin (Cell Signaling, 12238), EIF2 $\alpha$  (Cell Signaling, 2103), phosphorylated EIF2 $\alpha$  (pEIF2 $\alpha$ , Cell Signaling, 3398 and 9721), and b-actin (Abcam, ab8224). After staining with secondary antibodies conjugated to AlexaFluor dyes (Life Sciences), images were captured in an FLA5000 fluorescence scanner (Fuji Medical Systems, Stamford, CT) and analyzed with Multigauge 3.0 software supplied with the scanner.

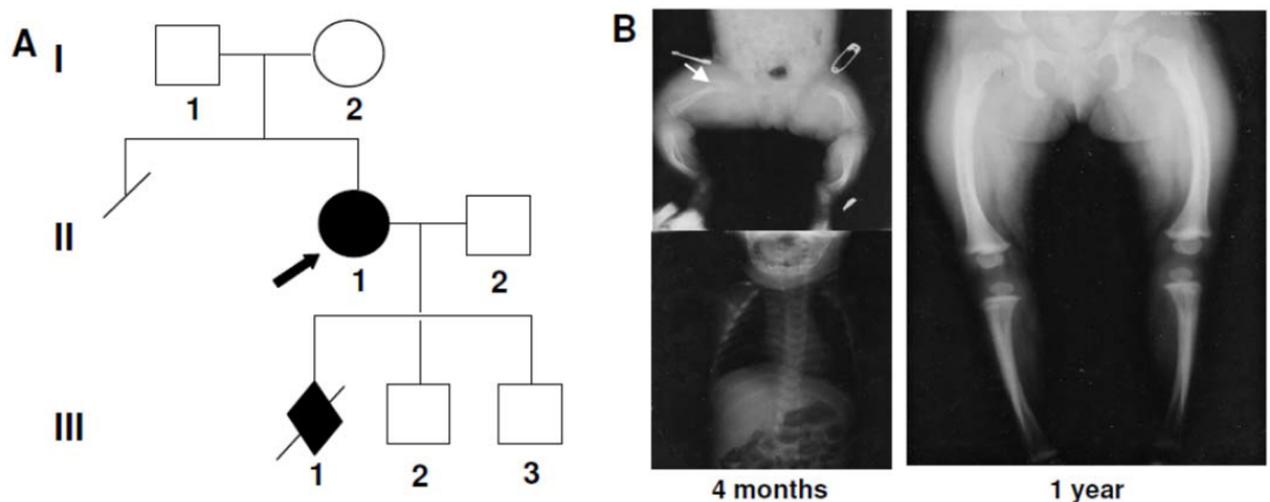
### Molecular modeling

The structural models were based on Protein Data Bank entry 1BKV [36] and molecular modeling and image generation used UCSF Chimera [37]. Leucine substitution was modeled based on the Dunbrack backbone-dependent library [38].

## Results

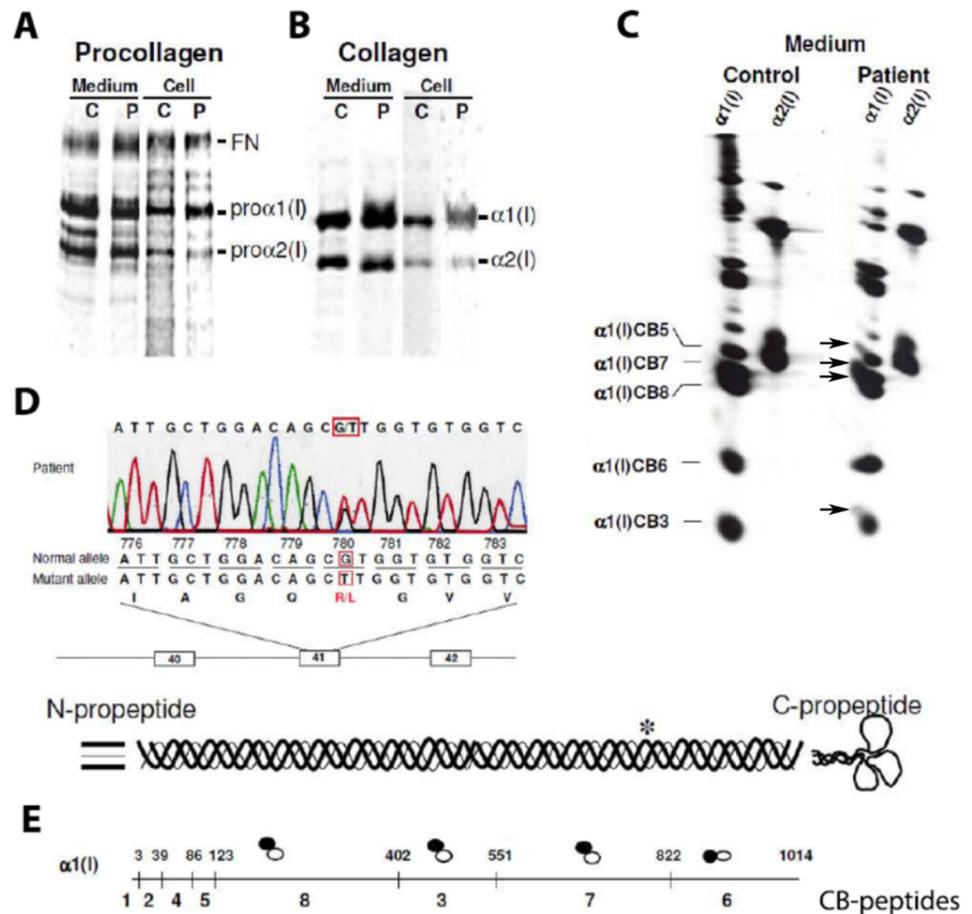
### Clinical summary and mutation identification

The index patient (Fig 1A, II-1) was recognized at birth to have a mild form of OI. Neither of her parents was affected with OI (Fig 1A, I-1, I-2). Parental DNA was not available for mutation analysis. At 4 months of age, the proband had an acute fracture of the right femur (Fig 1B, arrow). Radiographic studies at the time demonstrated that the femurs and the tibias were short and bowed; the ribs had near normal appearance although the chest was slightly small. Within a year, the femurs had grown, the bowing was diminished, and the cortex was well mineralized (Fig 1B). As an adult, she had blue sclerae, modest number of fractures, and some bowing of the tibias and is at the 25th percentile for height. Her first pregnancy (Fig 1A, III-1) was terminated at 19 weeks gestation because of the identification of an encephalocele and bowed bones by ultrasound examination. This fetus was heterozygous for the maternal mutation. The encephalocele was considered to have a different basis. Later, her second and third pregnancies were normal by ultrasound examination. Cultured CVS cells from each fetus synthesized only normal type I procollagen and analysis of DNA confirmed that neither carried the mutant allele (Fig 1A, III-2 and III-3). Healthy infants were born subsequently.



**Fig 1. Clinical summary for  $\alpha 1(I)$ -R780L.** A, Family pedigree; the proband (II-1) is marked with arrow. B, radiographs of the proband at 4 months (left) and 1 year (right) of age.

<https://doi.org/10.1371/journal.pone.0200264.g001>



**Fig 2. Analysis of procollagen (A) and collagen (B) molecules synthesized by the patient (P) and control (C) cells and mutation identification (C-E).** A,B, SDS-PAGE run with 2 M urea to enhance the difference in migration between overmodified and normal chains. Compared to control cells, pro $\alpha$ 1(I) and  $\alpha$ 1(I) chains from patient were more heterogeneous and some of the chains had delayed electrophoretic migration due to posttranslational overmodification. C, The shapes of electrophoretic bands of CNBr (CB) peptides from normal (large solid bands) and overmodified (upward streaks marked by arrows) collagen chains. The main bands might be slanted (e.g. control  $\alpha$ 1(I)) due to imperfect shape/orientation of the band cut out from the original collagen gel for CNBr digestion. The streaks of overmodified chains in the top left corner of the main bands are distinguished by their different orientation. The presence of overmodified  $\alpha$ 1(I)CB3,  $\alpha$ 1(I)CB7 and  $\alpha$ 1(I)CB8, but not  $\alpha$ 1(I)CB6, is consistent with a defect in folding within residues 551–821 of the triple helical domain (see panel E), that is, amino-terminal to the beginning of  $\alpha$ 1(I)CB6. D, The heterozygous mutation, c.2873G>T in exon 41 of one *COL1A1* allele, resulted in substitution of the arginine by leucine at position 958 of the protein (p.Arg958Leu), which is position 780 of the triple helical domain (Arg780Leu). E, The figure of CB-peptides shows the location of the methionyl residues in the  $\alpha$ 1(I) chain (vertical bars). The open circles (○) represent the electrophoretic mobility of the peptides derived from the normal chains relative to the streaks of overmodified peptides from the mutant chains (closed circles, ●).

<https://doi.org/10.1371/journal.pone.0200264.g002>

Cultured dermal fibroblasts from the proband (II-1) and the first pregnancy (III-1) but not the second pregnancy (III-2) produced both normal and abnormal type I collagen molecules with chains that had slow mobilities (Fig 2A and 2B). The shift in electrophoretic mobility was blocked by addition of  $\alpha,\alpha$ -dipyridyl (data not shown), which interferes with the post-translational hydroxylation of lysyl and prolyl residues in the triple helical domain. Analysis of the cyanogen bromide peptides derived from the chains showed that in the  $\alpha$ 1(I) chains overmodification began in peptide  $\alpha$ 1(I)CB7, which comprises position 551–821 of the triple helical domain, indicating that an alteration in that domain delays helix propagation (Fig 2C and 2E).

Sequence of the full length of the genomic DNA from *COL1A1* and *COL1A2* revealed a unique mutation in one allele of *COL1A1*: c.2873G>T, p.Arg958Leu, Arg780Leu in the triple helical domain (Fig 2D). The same mutation was identified in cells from the affected fetus (III-1) but was absent in III-2 and III-3 (data not shown). This variant has not been reported in any major population databases ([evs.gs.washington.edu/EVS](http://evs.gs.washington.edu/EVS); [gnomad.broadinstitute.org/](http://gnomad.broadinstitute.org/)) or ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>). No other coding sequencing or splice site mutations were found in either type I collagen gene and no alterations in mRNA sequence or structure of those genes was identified. The asymmetric overmodification is consistent with an altered sequence in the collagens and not in the modifying enzymes. We have used the triple helical or legacy “address” (e.g., residue 780) to identify mutation positions throughout the paper.

The *COL1A1* c.2872G>T, p.Arg958Cys (Arg780Cys) mutation was identified in 1994/95 in a 7 wk old patient with bilateral subdural hematomas and multiple fractures, including vertebral compression fractures. The patient had osteopenia, normal sclerae, and no evidence of Wormian bones or other bone malformations. SDS-PAGE did not show significant posttranslational overmodification of type I collagen; but it revealed evidence of a disulfide bond between  $\alpha 1(I)$  chains, which was mapped to  $\alpha 1(I)$ CB7 CNBr peptide. Because of child abuse concern (mother admitted to some abusive behavior) and lack of other individuals with the same mutation, the cause of the fractures remained uncertain. There was no patient follow up. Parental DNA was not analyzed.

### Effects of $\alpha 1(I)$ -R780 substitutions on procollagen and collagen thermal stability

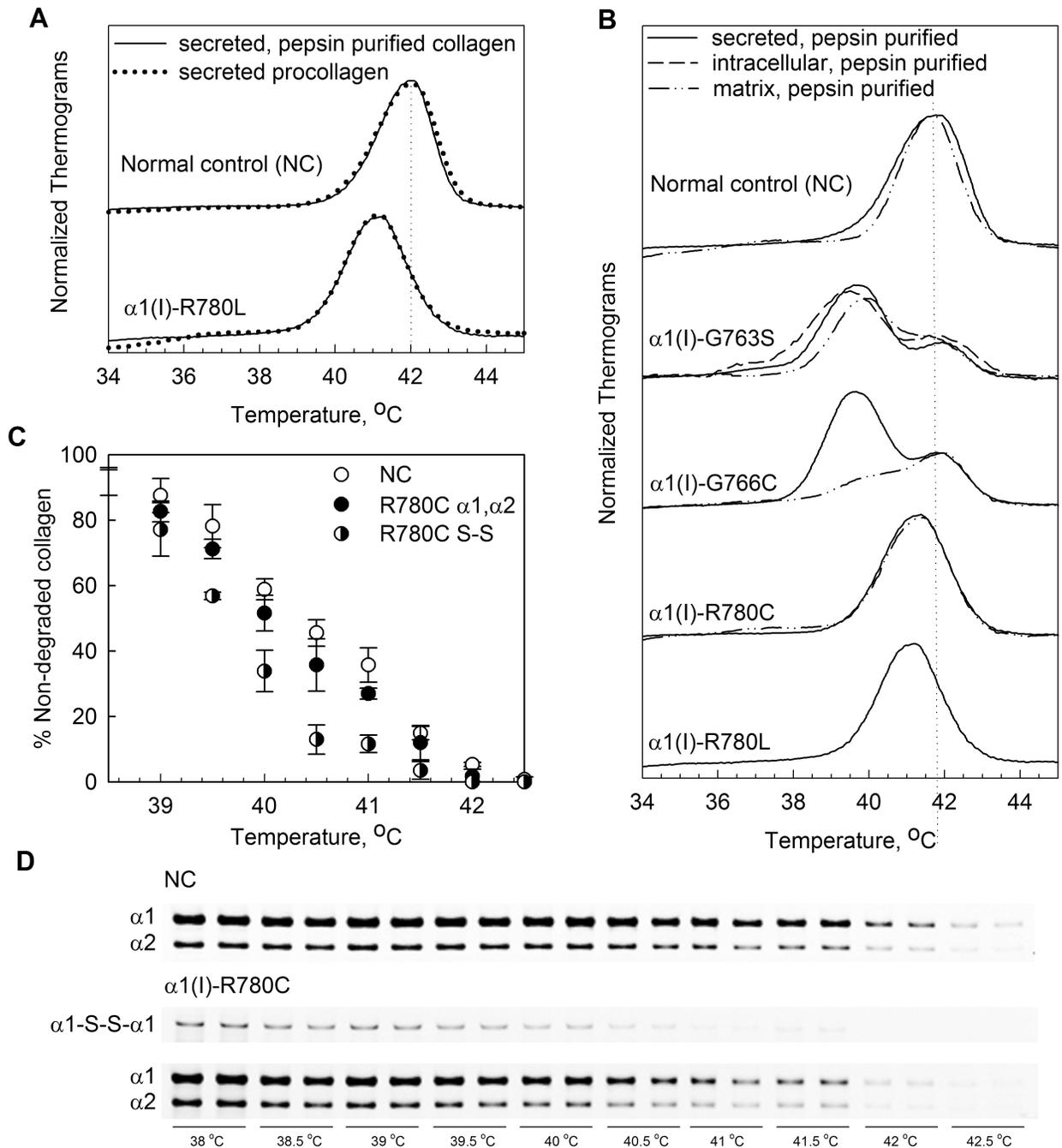
Substitutions of obligatory glycine residues decrease the thermal stability of the whole molecule and lower its apparent denaturation temperature ( $T_m$ ) from  $\sim 0.5^\circ\text{C}$  up to  $5^\circ\text{C}$  [27]. Except for the immediate vicinity of the N- and C-terminal ends of the triple helix, those substitutions reduce the thermal stability of procollagen to exactly the same extent as collagen. Several examined substitutions for Y-position Arg residues were also reported to decrease collagen  $T_m$ , all by approximately  $1^\circ\text{C}$  [17,23].

Analysis of secreted collagen and procollagen by differential scanning calorimetry (DSC) and by sensitivity to trypsin and chymotrypsin revealed  $0.5\text{--}1.0^\circ\text{C}$  lower  $T_m$  in  $\alpha 1(I)$ -R780L than in the control (Fig 3). DSC denaturation thermograms showed that the effects of  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C substitutions on the thermal stability of collagen were identical. The molecules with a Cys780-S-S-Cys780 disulfide bond appeared to have  $\sim 0.5^\circ\text{C}$  lower stability than the molecules without this bond (Fig 3D).

Substitutions for glycine residues  $\alpha 1(I)$ -G763S and  $\alpha 1(I)$ -G766C lowered collagen  $T_m$  by  $2\text{--}2.5^\circ\text{C}$ . The effect of substitutions for glycine residues on the collagen  $T_m$  is determined by their location along the triple helix rather than the identity of the substituting residue [27]. Comparison of the  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C collagens suggested that the same might be true for substitutions that do not alter triple helical glycine residues.

### Local destabilization of the triple helix at the MMP cleavage site

Measurement of collagen  $T_m$  by either DSC or chymotrypsin-trypsin digestion shows only mutation effects on the stability of the whole molecule. To investigate the extent of local destabilization of the triple helix near the mutation site, we exploited the proximity of the mutation to the collagenase MMP1 site (residues 775–776 of the triple helix) and the requirement of triple helix unwinding prior to cleavage by the enzyme.



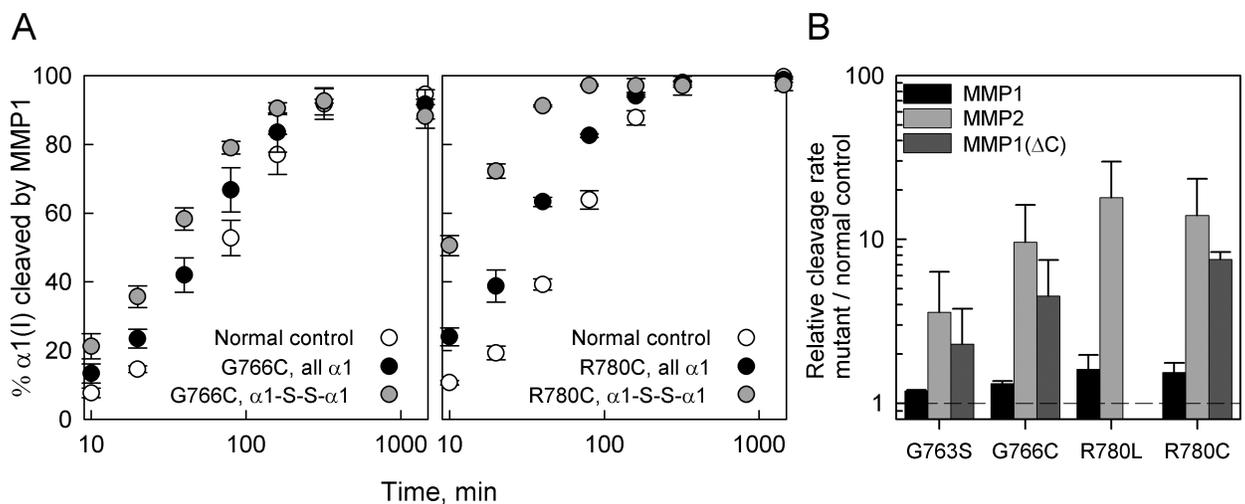
**Fig 3. Type I collagen triple helix destabilization caused by glycine and arginine mutations.** A, DSC denaturation thermograms of pepsin-purified collagen (solid lines) and procollagen (dotted lines) secreted into media by normal control (NC) and patient cells. B, Denaturation thermograms of pepsin-purified collagen from media (secreted, solid lines), cell layer (intracellular, dashed line), and extracellular matrix deposited by cells (matrix, dash-dotted lines). C, D, Analysis of  $\alpha 1(I)$ -R780C collagen thermal stability by enzymatic digestion and gel electrophoresis. In A and B, each thermogram peak represents denaturation of molecules with or without mutant chains. The peak maximum is the corresponding apparent denaturation temperature  $T_m$ , which is the same in collagen and procollagen molecules (for NC as well as most mutants) [27]. A heterozygous Gly substitution in the  $\alpha 1(I)$  chain might result in up to 3 peaks on the denaturation thermogram, representing molecules with 2, 1, and 0 mutant chains. In  $\alpha 1(I)$ -G766C collagen, the molecules with 1 and 2 mutant chains have the same  $T_m$ , producing one peak at  $\approx 39.5$  °C; the molecules without mutant chains produce the second peak at  $\approx 42$  °C. The same is true for  $\alpha 1(I)$ -G763S collagen. In  $\alpha 1(I)$ -R780L or  $\alpha 1(I)$ -R780C, molecules with 1 and 2 mutant chains produce a peak at  $\approx 41$  °C; the 42 °C peak of molecules without the mutant chains is too close to 41 °C, so that it is not resolved. The area under each peak is proportional to the fraction of the corresponding molecules. A change in this fraction upon secretion from cells or incorporation into matrix alters the shape of the denaturation thermogram [27]. A reduced intensity of the 39.5 °C peak in matrix vs. secreted collagen indicated that only 25% of  $\alpha 1(I)$ -G766C molecules in extracellular matrix contained mutant chains,

but no effects of other mutations on collagen secretion or matrix incorporation were detected. In C and D, thermal stability of secreted collagen was measured by 2 min equilibration at different temperatures followed by 5 min at room temperature, 1 min digestion with trypsin/chymotrypsin mixture, and separation of chains by gel electrophoresis at non-reducing conditions as described in [31]. Panel C shows quantitative analysis of gel electrophoresis (D), each point representing an average of 4 experiments.

<https://doi.org/10.1371/journal.pone.0200264.g003>

We found an increase in the cleavage rate of  $\alpha 1(I)$ -G763S collagen by MMP1 (Fig 4B, black bars). The increase in the cleavage rate was more pronounced for  $\alpha 1(I)$ -R780L,  $\alpha 1(I)$ -R780C and  $\alpha 1(I)$ -G766C collagens. Overall,  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C substitutions had similar effects, indicating that the global and the local changes in the triple helix stability were similar. In  $\alpha 1(I)$ -G766C and  $\alpha 1(I)$ -R780C collagens, the cleavage rate of molecules with the two  $\alpha 1(I)$  chains linked by a disulfide bond was faster than molecules with either one or no altered chains, consistent with more significant triple helix opening in these molecules (Fig 4A).

We observed similar effects of the mutations on collagen cleavage by gelatinase MMP2 (Fig 4B, light grey bars) and by the catalytic domain of MMP1 (MMP1( $\Delta C$ )) (Fig 4B, dark grey bars). In short, the collagens that contained  $\alpha 1(I)$ -G766C,  $\alpha 1(I)$ -R780L, or  $\alpha 1(I)$ -R780C chains were much more susceptible to cleavage than normal control;  $\alpha 1(I)$ -G763S collagen was less susceptible than other mutations. The quantitatively larger effects of the mutations on collagen cleavage by MMP2 and MMP1( $\Delta C$ ) confirmed that there was local triple helix destabilization at the cleavage site, since these enzymes are less able to cut the intact triple helix than MMP1 [34]. The susceptibilities of  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C collagens to cleavage were similar, consistent with similar local destabilization of the triple helix in addition to similar change in  $T_m$  for the whole molecule.



**Fig 4. Cleavage of mutant collagens by MMP1, MMP2, and catalytic domain of MMP1 (MMP1( $\Delta C$ )).** A, cleavage kinetics of collagens by MMP1 in molecules with Cys for  $\alpha 1(I)$  Gly or Cys for  $\alpha 1(I)$  Arg substitutions. Each time point represents an average of two experiments. In one, the mutant labeled with Alexa Fluor 488 (AF488) was mixed with normal control labeled with Cy5 and treated with MMP1 within the same sample tube. In the other, the fluorescent labels were switched. The error bar represents the corresponding standard deviation. Gray circles show the kinetics of cleavage for disulfide bonded chains formed in molecules containing two  $\alpha 1(I)$  chains with Cys for Gly or Cys for Arg substitutions. B, relative cleavage rates for all studied mutants determined as the ratio of the initial cleavage rate of all  $\alpha 1(I)$  chains in the mutant to the initial cleavage rate of  $\alpha 1(I)$  chains in the normal control within the same sample tube. The initial cleavage rates were determined by linear regression of the first 3–4 data points in the kinetic experiments shown in A and similar experiments for the other mutants and enzymes. The cleavage rates of  $\alpha 1(I)$ -R780L with MMP1( $\Delta C$ ) were not measured. The cleavage rates of all mutants were larger than in normal control with high statistical significance ( $p < 0.001$ ). For instance, the cleavage rate by MMP2 was  $0.18 \pm 0.01$  in  $\alpha 1(I)$ -R780L vs.  $0.01 \pm 0.006$  in NC (in the same units), yielding high statistical significance for  $(0.18 \pm 0.01)/(0.01 \pm 0.006) > 1$  despite noticeable uncertainty in the absolute value of this ratio.

<https://doi.org/10.1371/journal.pone.0200264.g004>

To further understand the extent of the local helix destabilization, we measured circular dichroism (CD) spectra of purified collagen in response to trains of heating pulses, during which the temperature was clamped at 37, 37.5, and 38 °C (Fig 5). At these temperatures there is rapid unfolding of unstable triple helical regions, which is reversible upon cooling, provided that the pulse is short enough to prevent denaturation of the entire molecule. The relative change in the spectra at 223.8 nm during the heating pulse is proportional to the unfolded fraction of the triple helix [27,29].

Reversible unfolding of up to 10% of the triple helix (~100 amino acids) was observed for  $\alpha 1(I)$ -G763S and  $\alpha 1(I)$ -G766C collagens (Fig 5). In contrast, CD spectra of collagens with  $\alpha 1(I)$ -R780 substitutions were similar to the normal control, placing an upper limit of ~ 3% (~ 30 amino acids) on the length of the region destabilized by the mutations. Compared to  $\alpha 1(I)$ -G763S and  $\alpha 1(I)$ -G766C, the triple helix disruptions produced by  $\alpha 1(I)$ -R780 substitutions were just as large or larger at the adjacent MMP cleavage site. However, these disruptions did not propagate as far along the triple helix, probably because of highly heterogeneous structural stability of different regions within collagen molecule [27].

### Procollagen folding delay

Local destabilization also means that formation of the triple helix within this region during procollagen folding becomes unfavorable. The triple helix folds in a zipper-like fashion from the C- to N- terminal end, and local instability causes a pause around the mutation site until the helix is renucleated beyond this region [39]. This folding delay leads to asymmetric post-translational overmodification, N-terminal to the mutation. The overmodification of CNBr peptides from  $\alpha 1(I)$  chain regions N-terminal to the mutation was consistent with such a folding delay in  $\alpha 1(I)$ -R780L procollagen (Fig 2).

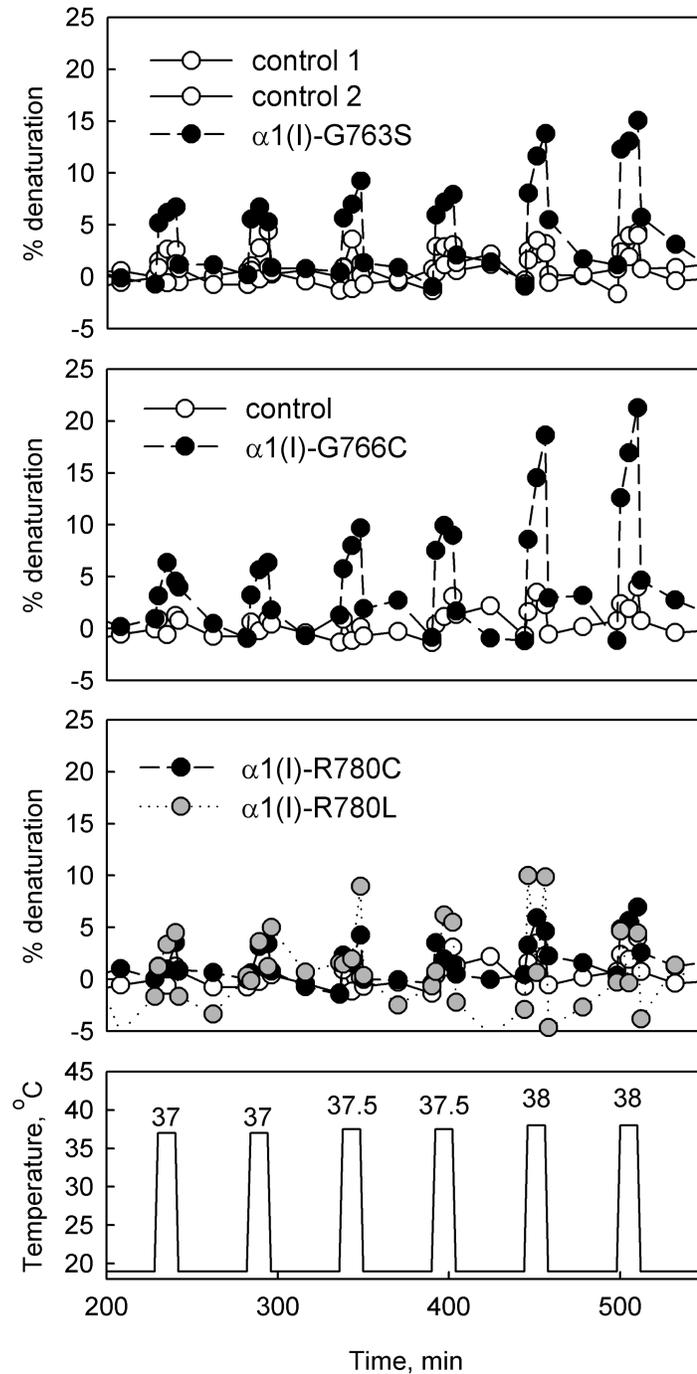
Using pulse-chase methionine substitution with a synthetic azidohomoalanine (Aha) analog followed by fluorescent labeling of Aha [30], we directly measured procollagen folding rate in human fibroblasts (Fig 6A). We confirmed that in these instances substitutions for arginine and glycine delayed triple helix folding. The most pronounced delay was observed for  $\alpha 1(I)$ -G766C. Interestingly,  $\alpha 1(I)$ -G766C molecules with the Cys766-S-S-Cys766 disulfide bond had shorter folding delay than  $\alpha 1(I)$ -G766C molecules without the bond;  $\alpha 1(I)$ -R780C molecules with the Cys780-S-S-Cys780 bond had longer folding delay than  $\alpha 1(I)$ -R780C molecules without the bond (Fig 6A).

The folding data perfectly correlated with posttranslational overmodification that we detected by slower electrophoretic migration on gel using mixtures of mutant and normal control collagens labeled with different fluorophores (Fig 6C) and by increased fraction of triple helical hydroxylysines (Fig 6D). All three parameters increased in the following order NC < R780L < G763S and R780C < G766C.

### Procollagen retention by mutant cells

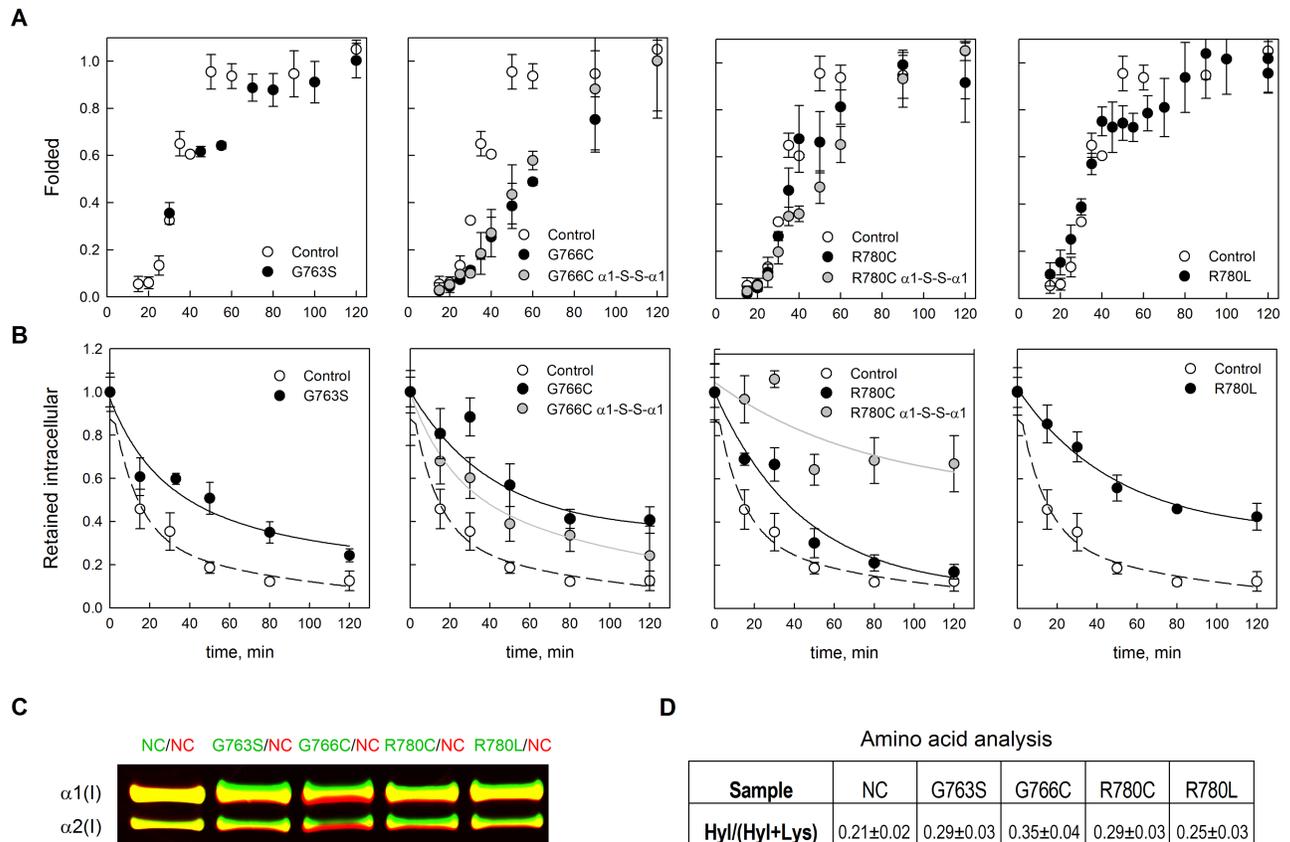
We then used pulse-chase labeling with Aha to examine the kinetics of procollagen clearance from cells [30]. The half-lifetime of Aha-labeled procollagen inside the cell increased from ~15min in NC to ~45 min in  $\alpha 1(I)$ -G763S, ~90 min in  $\alpha 1(I)$ -G766C, ~30 min in  $\alpha 1(I)$ -R780C and ~75min in  $\alpha 1(I)$ -R780L (Fig 6B). Consistent with the folding kinetics, the Cys766-S-S-Cys766 bond decreased while the Cys780-S-S-Cys780 bond increased the procollagen clearance time.

Larger delays in the clearance time compared to the folding delays indicated retention of misfolded procollagen in the cell. This conclusion was supported by electron microscopy observation of dilated ER cisternae in all mutant but not NC cells (Fig 7). Yet, we found no



**Fig 5. Reversible unfolding of mutant collagens.** The fraction (% denaturation) of the triple helix length unfolded in response to a train of heating pulses (bottom figure). This fraction was measured by monitoring % change in ellipticity of collagen solution in 2 mM HCl (pH 2.7) at 223.8 nm compared with the ellipticity change upon complete denaturation. Complete denaturation of collagen triple helix under the same conditions (e.g. after a 55 °C heating pulse) was irreversible. Reversible changes in the ellipticity were therefore associated with unfolding of a region(s) within the helix rather than the whole molecule. Assuming ellipticity of the unfolded region being the same as of a fully unfolded collagen chain and the ellipticity of the folded region being the same as of a fully folded triple helix, the measured % change in ellipticity represents the % of the triple helix length unfolded in response to the heating pulse. Thus, ~10% reversible change in the ellipticity observed in  $\alpha 1(I)$ -G763S and  $\alpha 1(I)$ -G766C collagen solutions indicated unfolding of 10% of the triple helix length.

<https://doi.org/10.1371/journal.pone.0200264.g005>



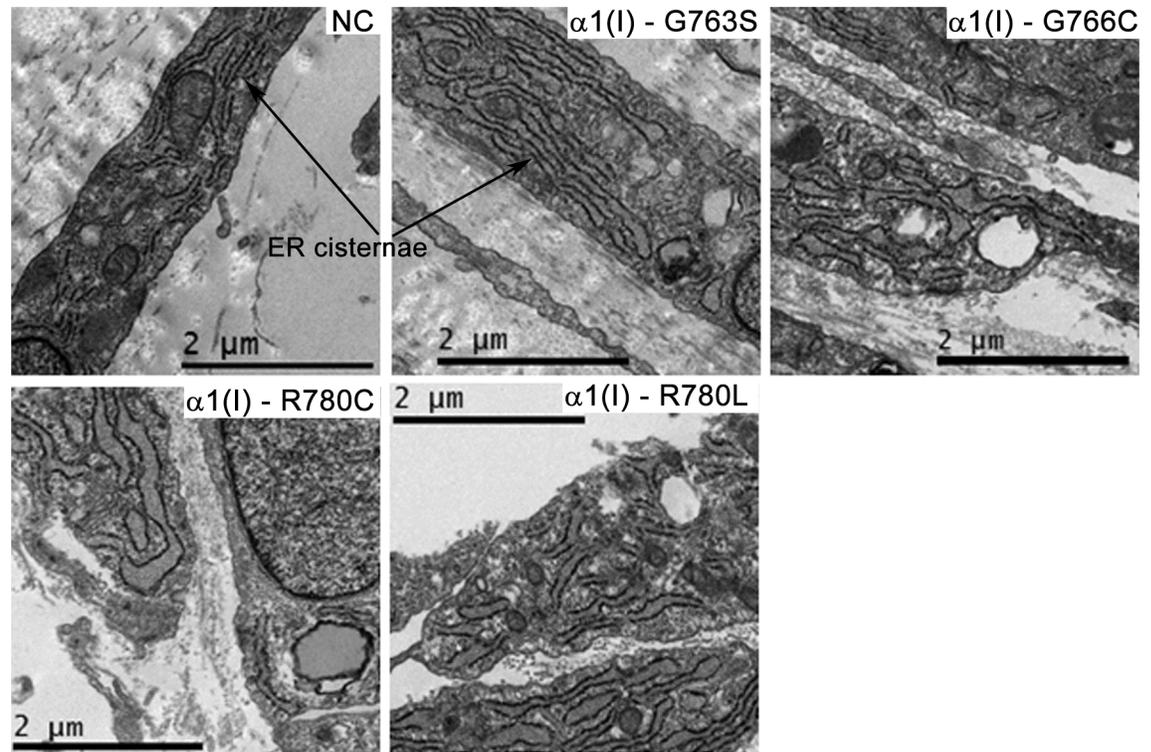
**Fig 6. Delay in type I collagen folding (A) and intracellular retention (B) caused by substitutions for glycine and arginine in the triple helical domain of pro $\alpha$ 1(I) chains.** A, Collagen triple helix folding measured by incorporation of azidohomoalanine (Aha) into newly synthesized collagen chains after 10 min Aha pulse, followed by chymotrypsin-trypsin digestion, labeling with DIBO dye, and quantifying  $\alpha$ 1(I) chains resistant to chymotrypsin-trypsin digestions by SDS-PAGE. The fraction of folded (chymotrypsin-trypsin resistant) molecules was measured from the intensity of Aha-DIBO-labeled  $\alpha$ 1(I) bands on SDS-PAGE relative to the intensity of these bands after 2 h chase. B, Residence time of intracellular Aha-DIBO-labeled collagen after 2 h Aha-pulse. C, Visualization of a shift in migration of overmodified mutant molecules vs. normal (NC) molecules on SDS-PAGE by double fluorescent labeling. Binary mixtures of AlexaFluor488-labeled NC, G763S, G766C, R780L, or R780C (green) with Cy5-labeled NC (red) were used to compare the migration of mutant vs normal control in the same SDS-PAGE lane. Relative to Fig 2B, the shift is less pronounced since urea was not used to enhance the separation, but it is still clearly visualized by the separation of red and green colors. D, Amino acid analysis shows higher fraction of hydroxylysines in all mutant collagens. The standard deviation was measured in NC (5 samples) and assumed to have similar relative value in other collagens.

<https://doi.org/10.1371/journal.pone.0200264.g006>

evidence of significant selective degradation of mutant chains inside the cells. The fractions of mutant chains in secreted procollagen estimated from denaturation thermograms appeared to be consistent with the corresponding expected values and we did not observe significant differences between the thermograms of intracellular and secreted molecules (Fig 3A and 3B).

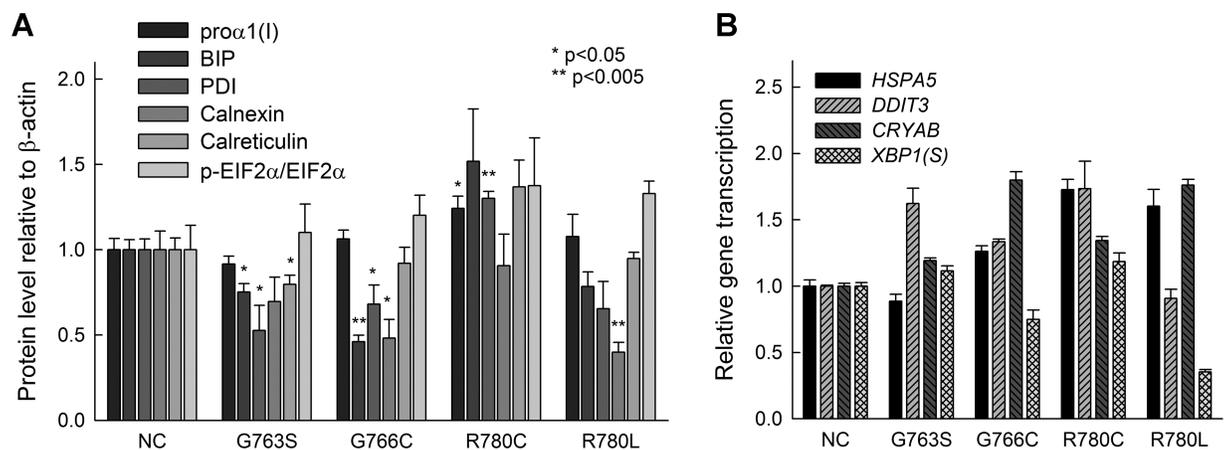
### Cell stress response

Retention of misfolded procollagen and resulting ER dilation suggested the possibility of cell stress. Western blot analysis (Fig 8A) provided evidence of increased EIF2 $\alpha$  phosphorylation expected in cell stress, but this trend did not reach statistical significance. Consistently, real-time quantitative PCR (qPCR, Fig 8B) suggested upregulation of *DDIT3/CHOP* ( $\alpha$ 1(I)-G763S, G766C, R780C) and *CRYAB/* $\alpha$ B-crystalline ( $\alpha$ 1(I)-G763C, R780L, R780C). We also observed decreased level of calnexin and PDI in  $\alpha$ 1(I)-G763S,  $\alpha$ 1(I)-G766C and  $\alpha$ 1(I)-R780L cells, indicating some disruption of ER function. Except for  $\alpha$ 1(I)-R780C cells, we found no evidence of



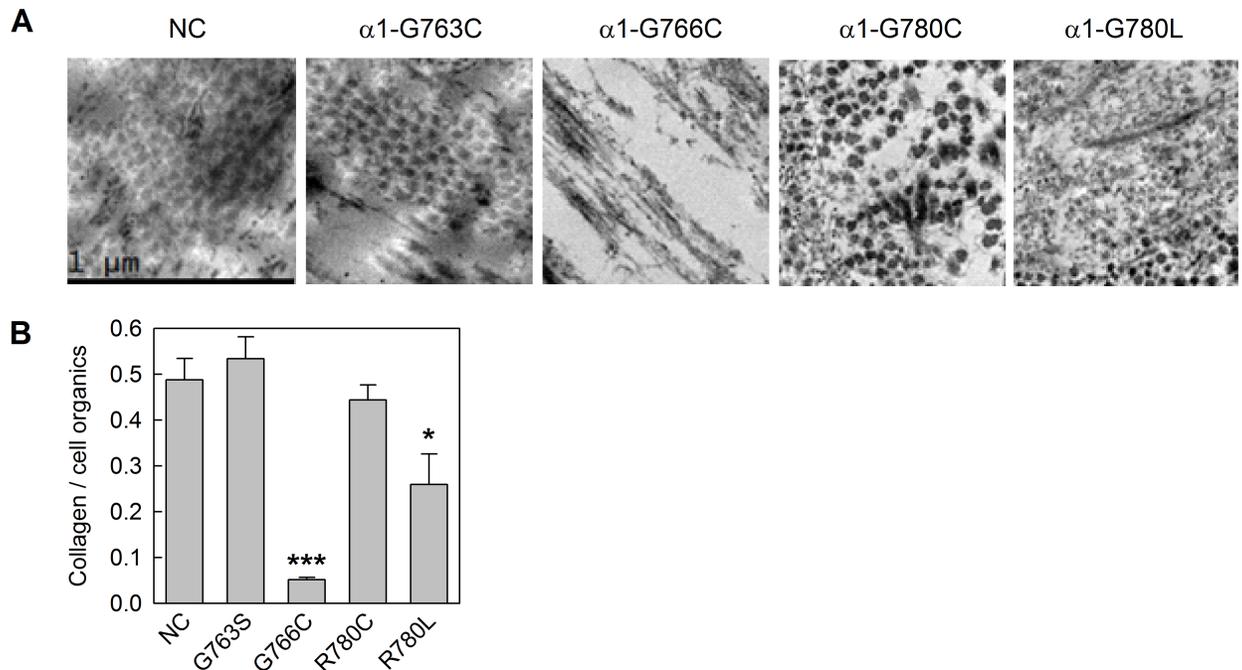
**Fig 7. Effect of type I collagen mutations on endoplasmic reticulum (ER) dilation.** Representative electron microscopy images of human fibroblasts show dilated ER cisternae in all mutant but not normal control cells. Rough ER cisternae are distinguished by ribosomes (black spots) lining the ER membrane. ER dilation was more subtle in  $\alpha 1(I)$ -G763S,G766C and more pronounced in  $\alpha 1(I)$ -R780C,R780L cells.

<https://doi.org/10.1371/journal.pone.0200264.g007>



**Fig 8. Cellular stress caused by glycine and arginine mutations in type I collagen.** A, Markers of integrated cell stress response (p-EIF2 $\alpha$ ) and unfolded protein response (BIP) as well as ER chaperones PDI, calnexin and calreticulin were measured in biological-triplicate by Western blotting relative to  $\beta$ -actin. B, Unfolded protein response (*HSPA5*/BiP and *XBP1(S)*) and integrated cell stress response (*DDIT3*/CHOP and *CRYAB*/ $\alpha$ B-crystallin) mRNA markers were measured in technical-triplicate by qPCR ( $\Delta\Delta C_T$  method) using a geometric mean of HPRT1, B2M and GAPDH as an endogenous control.

<https://doi.org/10.1371/journal.pone.0200264.g008>



**Fig 9. Effect of type I collagen mutations on matrix deposition.** A, Electron microscopy images of collagen fibrils deposited during 3-week culture. B, quantitative analysis of collagen content in extracellular matrix relative to cells by Raman microscopy; \*,  $p < 0.05$ ; \*\*\*,  $p < 0.001$ . The collagen/cell organics is the ratio of intensities of collagen-specific Raman bands to cytoplasm-specific Raman bands of organic material, which is present in cells but not in extracellular matrix. This ratio is proportional to the amount of collagen per cell present in the matrix.

<https://doi.org/10.1371/journal.pone.0200264.g009>

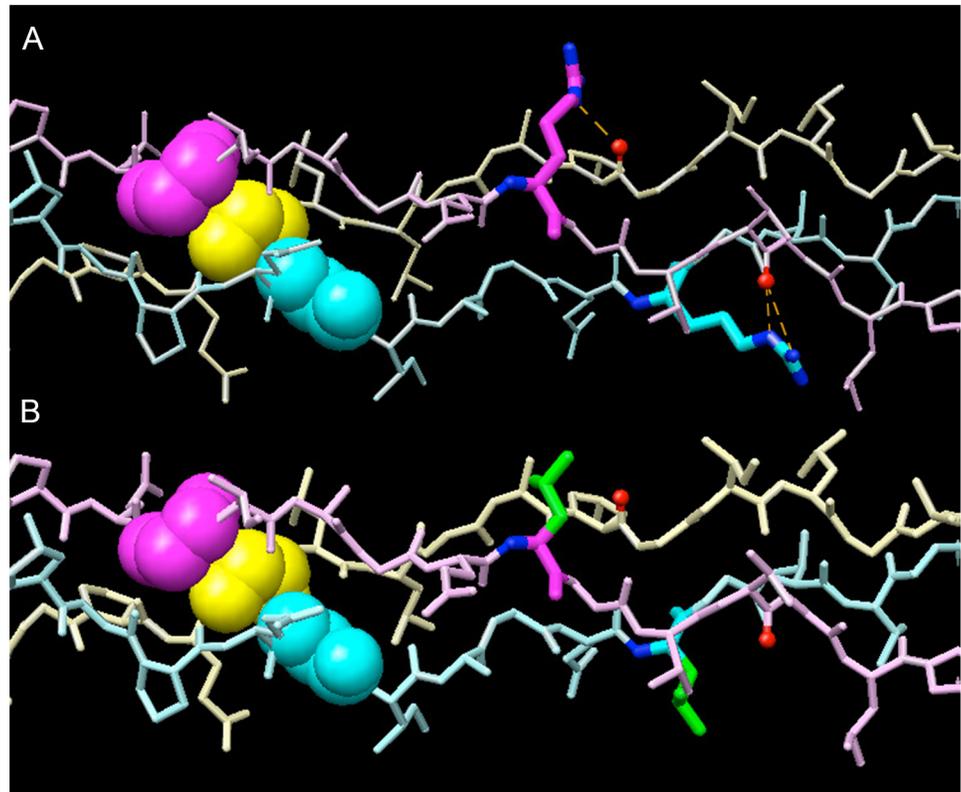
BIP upregulation either by Western blotting or qPCR. We also found no evidence of increased *XBPI* splicing.

### Collagen matrix deposition and structure

Comparison of collagen denaturation thermograms from cell culture media and matrix isolates revealed abnormal matrix incorporation only for  $\alpha 1(I)$ -G766C molecules (Fig 3B). Consistently, ~10% of the amount incorporated by control cells was deposited into matrix by the cells (Fig 9B) and the resulting matrix contained only very thin, disorganized fibers (Fig 9A). The matrix produced by  $\alpha 1(I)$ -G763S and  $\alpha 1(I)$ -R780C cells contained collagen fibers with normal size and appearance as well as normal amount of collagen, although  $\alpha 1(I)$ -R780C matrix appeared to be less well-organized. The  $\alpha 1(I)$ -R780L matrix had slightly less collagen and appeared to be more disorganized and heterogenous across the culture than the  $\alpha 1(I)$ -R780C matrix.

### Molecular model predicts that loss of arginine causes destabilization

To understand how the loss of Arg780 affects the triple helix in the context of the surrounding sequence at this particular site in type I collagen, we performed molecular modeling experiments. Previous studies of homotrimeric model peptides revealed much higher triple helix stability of sequences with Y-position arginine or hydroxyproline compared to other amino acids [40]. Computational analysis [41] and the crystal structure of a triple helical peptide incorporating arginine-containing fragment at the collagenase cleavage site in type III collagen [42] indicated that arginine side chains form hydrogen bonds with backbone carbonyl groups that result in significant stabilization of the triple helix [40]. Our computational analysis of the



**Fig 10. Model of effects of substitution of arginine by leucine at position 780 of the triple helix in the  $\alpha 1(I)$  chain.** A stretch of 14 residues are shown in three chains which in  $\alpha 1(I)$  are Pro773-Gln-Gly-Ile-Ala-Gly-Gln-Arg-Gly-Val-Val-Gly-Leu-Pro786. The glycines at position 775 are represented by the space filling models in which cyan (chain A) and magenta (chain B) are the  $\alpha 1(I)$  chains and the  $\alpha 2(I)$  chain is in yellow (chain C). The chain order is assumed to be  $\alpha 1(I)$ - $\alpha 2(I)$ - $\alpha 1(I)$  in the helical domain. A, The normal sequence in which the residue at position 780 of the triple helix is arginine shown in magenta and cyan on their respective chains. The hydrogen bonds are shown by the dotted lines. B, The mutant sequence in which both residues at position 780 are occupied by leucine. No hydrogen bonds form to stabilize the molecule.

<https://doi.org/10.1371/journal.pone.0200264.g010>

collagenase cleavage site in type I collagen is most compatible with the presence of hydrogen bonds between the  $\epsilon$ -nitrogen of Arg780 on one  $\alpha 1(I)$  chain and backbone of the  $\alpha 2(I)$  chain as well as between the  $\epsilon$ - and terminal nitrogens of Arg780 on the other  $\alpha 1(I)$  chain and backbone of the first  $\alpha 1(I)$  chain (Fig 10A, dashed lines). Leucine or cysteine substitutions for Arg780 (Fig 10B) prevent formation of these hydrogen bonds, destabilizing the helix.

## Discussion

Although *COL1A1* or *COL1A2* mutations that result in substitutions at X or Y-positions in the triple helix are much less common among patients than those that result in substitutions of obligatory glycines, they provide insights into mechanisms by which altered collagens can produce bone pathology in OI. In particular, reduced bone mineral density with or without recurring fractures was observed in most individuals reported with Y position Arg-to-Cys substitutions but not those in which the substitutions were at the X-position. The interpretation of these phenotypic differences has to take into account several considerations. Arg-to-Cys substitutions might lead to bone pathology as a result of osteoblast malfunction caused by misfolding of mutant procollagen molecules and their retention in the ER and/or due to disruption of extracellular matrix properties by incorporation of secreted mutant collagen molecules. Both

intra- and extra-cellular effects of the mutation could be related to the loss of a Y-position arginine residue, which stabilizes the collagen triple helix locally [40], serves as a binding site for HSP47 in the process of maturation of the triple helix in the ER [43], and might be important for collagen interactions in the ECM. The gain of a cysteine residue, which is not present in normal triple helices, might lead to detrimental formation of disulfide bonds in the ER, the Golgi, or in the ECM. Finally, the intra- and extracellular effects of free cysteine residues could be different for X- vs. Y-position cysteine residues perhaps because of different preferred conformations of the side chains at these positions and different bond formation efficiencies.

In the family identified here, the  $\alpha 1(I)$ -R780L substitution occurred at the same site as the  $\alpha 1(I)$ -R780C substitution in an unrelated individual with mild OI. Although we did not find other patients with the latter substitution, the present study supports its pathogenicity by demonstrating the similarity of the effects of both mutations on collagen properties. Analysis of the physical and chemical properties of collagen secreted by dermal fibroblasts from these patients provided us with a unique opportunity to distinguish the effects of Arg loss from Cys gain. It also enabled us to get additional insights into intra- vs extra-cellular mechanisms of bone pathology in OI.

### **$\alpha 1(I)$ -R780 is not essential for binding or cleavage of collagen chains by collagenases**

One possible effect of  $\alpha 1(I)$ -R780L or  $\alpha 1(I)$ -R780C molecules in the ECM could be inhibition of collagen cleavage by collagenases, because Arg780 appeared to be important for interaction of these enzymes with collagen [44]. A recent crystal structure revealed a hydrogen bond between the corresponding Arg residue in the leading chain of a triple helical peptide and MMP1, but substitution of this Arg to Ala did not have a significant effect on the enzyme-peptide interaction [45]. We found that cleavage of  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C containing molecules was faster rather than slower by MMP1 and MMP2, as a result of triple helix destabilization at the cleavage site. Increased susceptibility to collagenases might affect ECM remodeling, but it is not likely to be responsible for the OI phenotype, since collagenases do not appear to play a major role in bone [46].

### **Y-position Arg substitutions destabilize the triple helix and affect procollagen folding**

The common feature of OI-causing substitutions of Y-position arginine residues [17,23] is that all of them destabilize and affect the folding of the triple helix, similar to substitutions for glycine. Despite small changes in the collagen  $T_m$  (Fig 3), the increased susceptibilities to cleavage by MMP2 and catalytic domain of MMP1 (Fig 4) clearly show significant local destabilization of the helix already at 25 °C. This destabilization is comparable to effects of some substitutions for glycine residues described in this and our earlier studies [27]. CD measurements indicate that this destabilization is limited to less than ~ 30 adjacent amino acids (Fig 5). However, it disrupts the propagation of triple helix folding through the mutation site, as indicated by the folding delay (Fig 6A) and posttranslational overmodification of triple helical residues on the N-terminal side of the mutation (Figs 2C & 6C and 6D).

Y-Arg loss is the primary factor in decreasing the triple helix stability while the identity of the substituting amino acid appears to be secondary, as found previously for substitutions of obligatory Gly [27]. Similar effects of Leu and Cys suggest that the destabilization is not associated with detrimental interactions of their side chains. Even formation of a Cys780-S-S-Cys780 disulfide bond has only a relatively minor effect on the helix stability (Fig 3D) and folding (Fig 6C), like the formation of a Cys766-S-S-Cys766 disulfide bond. Instead, studies of triple helical

peptides [40], our previous studies [41], and our computer simulations all suggest that the triple helix destabilization is caused by the loss of hydrogen bonds between Y-position Arg side chains and triple helix backbone carbonyls.

Furthermore, the loss of Y-Arg affects procollagen folding not only because it destabilizes the triple helix but also because these Arg residues are essential for HSP47 binding [43], which is required for the final stages of triple helix folding [47–49]. Although  $\alpha 1(I)$ -R780 is a medium affinity HSP47 binding site, simultaneous triple helix destabilization and loss of HSP47 binding might explain the folding delay and posttranslational overmodification comparable to Gly-X-Y repeat disruption (Figs 2 and 6).

### Y-Arg loss causes ER accumulation of misfolded procollagen similar to Gly substitutions, likely resulting in cell stress and malfunction

Even relatively minor procollagen folding delay (like in  $\alpha 1(I)$ -G763S,  $\alpha 1(I)$ -R780C and  $\alpha 1(I)$ -R780L cells) causes major procollagen misfolding and accumulation of the misfolded molecules inside the cells. Indeed, the half-time required for clearing procollagen molecules from all these cells increases more than two-fold. Although  $\alpha 1(I)$ -G763S,  $\alpha 1(I)$ -R780C and  $\alpha 1(I)$ -R780L mutations cause more subtle changes in the folding rate than  $\alpha 1(I)$ -G766C, their effects on the clearance kinetics are much more pronounced and similar (Fig 6B). Curiously, neither Y-Arg nor Gly substitutions we studied significantly reduce the fraction of mutant chains in secreted procollagen. The fraction of the molecules retained inside the cells might not be large enough for detecting the change in the chain composition in secreted procollagen from denaturation thermograms. Moreover, mutant cells might non-selectively retain molecules with and without the mutant chains, e.g. because higher ER concentration of unfolded chains might cause nonspecific and therefore non-selective chain aggregation into gelatin-like structures [11].

Consistent with the significantly delayed procollagen clearance from the cells, Y-Arg and Gly substitutions cause similar ER dilation, indicating misfolded procollagen accumulation inside the ER (Fig 7) and possible cell stress. In agreement with other studies [11,50,51], the ER accumulation of misfolded procollagen does not trigger the conventional unfolded protein response (UPR) that is accompanied by upregulation of BIP expression and XBP1 splicing. An apparent increase in EIF2 $\alpha$  phosphorylation, CHOP and  $\alpha$ B-crystalline transcription, and decrease in calnexin observed here for Y-Arg and Gly substitutions (Fig 8) are consistent with the observations for cell stress response in osteoblasts from G610C mice [11], yet inconclusive. Our experience gained from the latter G610C mouse study suggests that conclusive evidence of cell stress response requires analysis of primary osteoblast cultures from multiple animals at low (preferably zero) passage. Dermal fibroblasts produce less collagen than osteoblasts, experience subtler ER dilation, and therefore less pronounced cell stress. The cell stress is particularly difficult to detect in human cells, since genetic variants unique for each subject might affect its severity.

Formation of a Cys766-S-S-Cys766 disulfide bond at the Gly substitution site improves procollagen folding and clearance from cells (Fig 6A and 6B), potentially reducing the severity of bone pathology. This bond might promote triple helix renucleation beyond the mutation site by keeping the three pro $\alpha$  chains together. Formation of a Cys780-S-S-Cys780 bond at the Y-Arg substitution site has an opposite effect. This bond appears to inhibit triple helix renucleation, probably by disrupting the alignment of the chains. Reduced misfolding and accumulation of molecules with Gly-to-Cys substitution in two  $\alpha 1(I)$  chains due to disulfide bond formation might contribute to lower cell stress and therefore explain the surprising milder phenotype of homozygous vs. heterozygous mice with a Gly349-to-Cys substitution [52].

Note that Arg-to-Cys substitutions in the triple helical region of type II collagen were also reported to cause pathology [53,54]. In contrast to  $\alpha 1(I)$ -R780 mutations, conventional UPR and cell apoptosis were observed for  $\alpha 1(II)$ -R789C and  $\alpha 1(II)$ -R992C [55,56]. Similar to  $\alpha 1(I)$ -R780 mutations,  $\alpha 1(II)$ -R75C and  $\alpha 1(II)$ -R519C were found to cause some ER dilation without conventional UPR or apoptosis [56].

### Cell stress response to accumulation of misfolded procollagen might explain why Y- but not X-Arg substitutions cause bone pathology

Despite the inconclusive evidence for the cell stress response caused by Y-Arg substitutions, we believe that the idea of resulting osteoblast malfunction provides the most logical explanation for the following observations. Bone fragility and low mineral density were identified in a number of individuals with Cys for Y-Arg substitutions at positions 396, 858, 888, and 915 in the  $\alpha 1(I)$  chain of the triple helix [17,18,23]. In contrast, classical EDS [25] and Caffey disease [26] were seen with Cys for X-Arg substitutions at positions 134 and 836, respectively. Procollagen folding is expected to be significantly affected only by Arg substitution at Y-positions while ECM interactions of collagen should be affected by substitutions at both X- and Y-positions. Since decreased bone density and fractures are observed only with substitutions for Y-position Arg, this bone pathology is more likely to be related to osteoblast malfunction in response to abnormal procollagen folding than to direct effects of Arg to Cys substitutions on bone matrix and its mineralization. For the same reasons, skin and joint laxity as well as arterial abnormalities are more likely to be related to aberrant interactions of Cys in the ECM. These EDS symptoms were reported in patients with both X- and Y-position Arg substitutions to Cys, which is not present in normal type I collagen.

Note that similarly mild bone pathology in patients with  $\alpha 1(I)$ -R780L and  $\alpha 1(I)$ -R780C substitutions is also more difficult to explain by disruptions of collagen interactions in the ECM. As expected, hydrophobic Leu and sulfhydryl Cys appear to have different effects on ECM composition and structure (Fig 9) but similar effects on procollagen folding (Fig 6) and ER dilation (Fig 7).

The idea of osteoblast malfunction as a major factor in bone pathology caused by procollagen misfolding has been discussed in several recent studies of dominant OI caused by Gly substitutions as well as recessive OI caused by mutations in procollagen chaperones [3,6,11,12,51,57–59]. It might contribute to the reported TGF- $\beta$  signaling abnormalities in OI [11,60,61]. We think that the present analysis of Y-Arg substitutions provides further support for this idea, even though we have been able to study only patient fibroblasts.

### Supporting information

**S1 File. Representative Western blots. Figure A. Top panel.** BIP and p-eIF2a were fluorescently labeled with Alexa Fluor 550 goat anti-rabbit IgG (red channel); eIF2a and  $\beta$ -actin were fluorescently labeled with Alexa Fluor 488 goat anti-mouse IgG (green channel). Magenta labels mark gel lanes from normal control (C5),  $\alpha 1(I)$ -G763S (763),  $\alpha 1(I)$ -G766C,  $\alpha 1(I)$ -G780C (780C), and  $\alpha 1(I)$ -G780L (780L) cells; the 811 sample was from a different study. Yellow molecular weight labels mark positions of the corresponding molecular weight standards, visualization of which requires contrast enhancement (70 kDa standard is visible as a faint red band at the left edge of the blot). Low intensity red p-eIF2a bands are masked by much higher intensity green eIF2a bands. **Bottom panel.** Contrast-enhanced red fluorescence channel shows thin, low intensity p-eIF2a bands and residual fluorescence of wide  $\beta$ -actin bands labeled with Alexa Fluor 488 (which is not completely eliminated by the Alexa Fluor 550 filter set). **Figure B.** Calnexin and calreticulin were fluorescently labeled with Alexa Fluor 550 goat

anti-rabbit IgG (red channel); PDI and  $\beta$ -actin were fluorescently labeled with Alexa Fluor 488 goat anti-mouse IgG (green channel). ~85 kDa red bands (top panel) were identified as calnexin and ~55 kDa green bands (top panel) were identified as PDI. Magenta and yellow labels are the same as in Figure A, S1 File. **Figure C.** Procollagen  $\alpha$ 1(I) chain and  $\beta$ -actin were fluorescently labeled with Alexa Fluor 550 goat anti-rabbit IgG (red channel) and Alexa Fluor 488 goat anti-mouse IgG (green channel), respectively. Magenta and yellow labels are the same as in Figure A, S1 File.  
(PDF)

## Author Contributions

**Conceptualization:** Teri E. Klein, Peter H. Byers, Sergey Leikin.

**Formal analysis:** Peter H. Byers, Sergey Leikin.

**Funding acquisition:** Peter H. Byers, Sergey Leikin.

**Investigation:** Elena Makareeva, Guoli Sun, Lynn S. Mirigian, Edward L. Mertz, Juan C. Vera, Nydea A. Espinoza, Kathleen Yang, Diana Chen, Teri E. Klein, Peter H. Byers, Sergey Leikin.

**Supervision:** Peter H. Byers, Sergey Leikin.

**Writing – original draft:** Elena Makareeva, Guoli Sun, Teri E. Klein, Peter H. Byers, Sergey Leikin.

**Writing – review & editing:** Elena Makareeva, Guoli Sun, Teri E. Klein, Peter H. Byers, Sergey Leikin.

## References

1. Shapiro JR, Byers PH, Glorieux FH, Sponseller PD, editors (2014) *Osteogenesis Imperfecta: A Translational Approach to Brittle Bone Disease*. Amsterdam: Elsevier. 555 p.
2. Sillence DO, Senn A, Danks DM (1979) Genetic heterogeneity in osteogenesis imperfecta. *J Med Genet* 16: 101–116. PMID: [458828](#)
3. Forlino A, Marini JC (2016) Osteogenesis imperfecta. *Lancet* 387: 1657–1671. [https://doi.org/10.1016/S0140-6736\(15\)00728-X](https://doi.org/10.1016/S0140-6736(15)00728-X) PMID: [26542481](#)
4. Boudko S, Pokidysheva E, Bachinger HP (2013) Autosomal Recessive Osteogenesis Imperfecta: A Puzzle for Bone Formation, Structure and Function. *Curr Genet Med Rep* 1: 239–246.
5. Byers PH, Pyott SM (2012) Recessively inherited forms of osteogenesis imperfecta. *Annu Rev Genet* 46: 475–497. <https://doi.org/10.1146/annurev-genet-110711-155608> PMID: [23145505](#)
6. Makareeva E, Aviles NA, Leikin S (2011) Chaperoning osteogenesis: new protein-folding disease paradigms. *Trends Cell Biol* 21: 168–176. <https://doi.org/10.1016/j.tcb.2010.11.007> PMID: [21183349](#)
7. Bardai G, Moffatt P, Glorieux FH, Rauch F (2016) DNA sequence analysis in 598 individuals with a clinical diagnosis of osteogenesis imperfecta: diagnostic yield and mutation spectrum. *Osteoporosis Int* 27: 3607–3613.
8. Mendoza-Londono R, Fahiminiya S, Majewski J, Care4Rare-Canada-Consortium, Tetreault M, Nadaf J, et al. (2015) Recessive osteogenesis imperfecta caused by missense mutations in SPARC. *Am J Hum Genet* 96: 979–985. <https://doi.org/10.1016/j.ajhg.2015.04.021> PMID: [26027498](#)
9. Marini JC, Forlino A, Bachinger HP, Bishop NJ, Byers PH, Paeppe A, et al. (2017) Osteogenesis imperfecta. *Nat Rev Dis Primers* 3: 17052. <https://doi.org/10.1038/nrdp.2017.52> PMID: [28820180](#)
10. Dagleish R (2014) Osteogenesis Imperfecta Genotypes and Genotype–Phenotype Relationships. In: Shapiro JR, Byers PH, Glorieux FH, Sponseller PD, editors. *Osteogenesis Imperfecta: A Translational Approach to Brittle Bone Disease*. Amsterdam: Elsevier. pp. 103–112.
11. Mirigian LS, Makareeva E, Mertz EL, Omari S, Roberts-Pilgrim AM, Oestreich AK, et al. (2016) Osteoblast Malfunction Caused by Cell Stress Response to Procollagen Misfolding in  $\alpha$ 2(I)-G610C

- Mouse Model of Osteogenesis Imperfecta. *J Bone Miner Res* 31: 1608–1616. <https://doi.org/10.1002/jbmr.2824> PMID: 26925839
12. Bianchi L, Gagliardi A, Maruelli S, Besio R, Landi C, Gioia R, et al. (2015) Altered cytoskeletal organization characterized lethal but not surviving Brl+/- mice: insight on phenotypic variability in osteogenesis imperfecta. *Hum Mol Genet* 24: 6118–6133. <https://doi.org/10.1093/hmg/ddv328> PMID: 26264579
  13. Bianchi L, Gagliardi A, Gioia R, Besio R, Tani C, Landi C, et al. (2012) Differential response to intracellular stress in the skin from osteogenesis imperfecta Brl mice with lethal and non lethal phenotype: a proteomic approach. *J Proteomics* 75: 4717–4733. <https://doi.org/10.1016/j.jprot.2012.01.038> PMID: 22846432
  14. Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, et al. (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J Biol Chem* 283: 21187–21197. <https://doi.org/10.1074/jbc.M709319200> PMID: 18487200
  15. Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, et al. (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat* 28: 209–221. <https://doi.org/10.1002/humu.20429> PMID: 17078022
  16. Beck K, Chan VC, Shenoy N, Kirkpatrick A, Ramshaw JA, Brodsky B (2000) Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc Natl Acad Sci U S A* 97: 4273–4278. <https://doi.org/10.1073/pnas.070050097> PMID: 10725403
  17. Cabral WA, Makareeva E, Letocha AD, Scribanu N, Fertala A, Steplewski A, et al. (2007) Y-position cysteine substitution in type I collagen (alpha1(I) R888C/p.R1066C) is associated with osteogenesis imperfecta/Ehlers-Danlos syndrome phenotype. *Hum Mutat* 28: 396–405. <https://doi.org/10.1002/humu.20456> PMID: 17206620
  18. Lund A, Joensen F, Christensen E, Duno M, Skovby F, Schwartz M (2008) A novel arginine-to-cysteine substitution in the triple helical region of the alpha1(I) collagen chain in a family with an osteogenesis imperfecta/Ehlers-Danlos phenotype. *Clin Genet* 73: 97–101. <https://doi.org/10.1111/j.1399-0004.2007.00926.x> PMID: 18028452
  19. Pollitt R, McMahon R, Nunn J, Bamford R, Affii A, Bishop N, et al. (2006) Mutation analysis of COL1A1 and COL1A2 in patients diagnosed with osteogenesis imperfecta type I-IV. *Hum Mutat* 27: 716.
  20. Fuccio A, Iorio M, Amato F, Elce A, Ingino R, Filocamo M, et al. (2011) A novel DHPLC-based procedure for the analysis of COL1A1 and COL1A2 mutations in osteogenesis imperfecta. *J Mol Diagn* 13: 648–656. <https://doi.org/10.1016/j.jmoldx.2011.06.006> PMID: 21884818
  21. Lindahl K, Astrom E, Rubin CJ, Grigelioniene G, Malmgren B, Ljunggren O, et al. (2015) Genetic epidemiology, prevalence, and genotype-phenotype correlations in the Swedish population with osteogenesis imperfecta. *Eur J Hum Genet* 23: 1042–1050. <https://doi.org/10.1038/ejhg.2015.81> PMID: 25944380
  22. Spotila LD, Colige A, Sereda L, Constantinou-Deltas CD, Whyte MP, Riggs BL, et al. (1994) Mutation analysis of coding sequences for type I procollagen in individuals with low bone density. *J Bone Miner Res* 9: 923–932. <https://doi.org/10.1002/jbmr.5650090618> PMID: 8079666
  23. Malfait F, Symoens S, De Backer J, Hermanns-Le T, Sakalihasan N, Lapiere CM, et al. (2007) Three arginine to cysteine substitutions in the pro-alpha (I)-collagen chain cause Ehlers-Danlos syndrome with a propensity to arterial rupture in early adulthood. *Hum Mutat* 28: 387–395. <https://doi.org/10.1002/humu.20455> PMID: 17211858
  24. Gaines R, Tinkle BT, Halandras PM, Al-Nouri O, Crisostomo P, Cho JS (2015) Spontaneous ruptured dissection of the right common iliac artery in a patient with classic Ehlers-Danlos syndrome phenotype. *Ann Vasc Surg* 29: 595 e511–594.
  25. Nuytinck L, Freund M, Lagae L, Pierard GE, Hermanns-Le T, De Paepe A (2000) Classical Ehlers-Danlos syndrome caused by a mutation in type I collagen. *Am J Hum Genet* 66: 1398–1402. <https://doi.org/10.1086/302859> PMID: 10739762
  26. Gensure RC, Makitie O, Barclay C, Chan C, Depalma SR, Bastepe M, et al. (2005) A novel COL1A1 mutation in infantile cortical hyperostosis (Caffey disease) expands the spectrum of collagen-related disorders. *J Clin Invest* 115: 1250–1257. <https://doi.org/10.1172/JCI22760> PMID: 15864348
  27. Makareeva E, Mertz EL, Kuznetsova NV, Sutter MB, DeRidder AM, Cabral WA, et al. (2008) Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J Biol Chem* 283: 4787–4798. <https://doi.org/10.1074/jbc.M705773200> PMID: 18073209
  28. Bonadio J, Holbrook KA, Gelinis RE, Jacob J, Byers PH (1985) Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. *J Biol Chem* 260: 1734–1742. PMID: 2981871
  29. Makareeva E, Cabral WA, Marini JC, Leikin S (2006) Molecular mechanism of alpha 1(I)-osteogenesis imperfecta/Ehlers-Danlos syndrome: unfolding of an N-anchor domain at the N-terminal end of the type

- I collagen triple helix. *J Biol Chem* 281: 6463–6470. <https://doi.org/10.1074/jbc.M511830200> PMID: 16407265
30. Mirigian LS, Makareeva E, Leikin S (2014) Pulse-chase analysis of procollagen biosynthesis by azido-homoalanine labeling. *Connect Tissue Res* 55: 403–410. <https://doi.org/10.3109/03008207.2014.959120> PMID: 25159826
  31. Bulleid NJ, Wilson R, Lees JF (1996) Type-III procollagen assembly in semi-intact cells: chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation. *Biochem J* 317 (Pt 1): 195–202.
  32. Leikina E, Merts MV, Kuznetsova N, Leikin S (2002) Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* 99: 1314–1318. <https://doi.org/10.1073/pnas.032307099> PMID: 11805290
  33. Barnes AM, Cabral WA, Weis M, Makareeva E, Mertz EL, Leikin S, et al. (2012) Absence of FKBP10 in recessive type XI osteogenesis imperfecta leads to diminished collagen cross-linking and reduced collagen deposition in extracellular matrix. *Hum Mutat* 33: 1589–1598. <https://doi.org/10.1002/humu.22139> PMID: 22718341
  34. Chung L, Dinakarandian D, Yoshida N, Lauer-Fields JL, Fields GB, Visse R, et al. (2004) Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *Embo J* 23: 3020–3030. <https://doi.org/10.1038/sj.emboj.7600318> PMID: 15257288
  35. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C-T method. *Nat Protoc* 3: 1101–1108. PMID: 18546601
  36. Kramer RZ, Bella J, Brodsky B, Berman HM (2001) The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J Mol Biol* 311: 131–147. <https://doi.org/10.1006/jmbi.2001.4849> PMID: 11469863
  37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
  38. Dunbrack RL Jr. (2002) Rotamer libraries in the 21st century. *Curr Opin Struct Biol* 12: 431–440. PMID: 12163064
  39. Engel J, Prockop DJ (1991) The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Biophys Chem* 20: 137–152. <https://doi.org/10.1146/annurev.bb.20.060191.001033> PMID: 1867713
  40. Persikov AV, Ramshaw JA, Kirkpatrick A, Brodsky B (2000) Amino acid propensities for the collagen triple-helix. *Biochemistry* 39: 14960–14967. PMID: 11101312
  41. Vitagliano L, Nemethy G, Zagari A, Scheraga HA (1993) Stabilization of the triple-helical structure of natural collagen by side-chain interactions. *Biochemistry* 32: 7354–7359. PMID: 8338832
  42. Kramer RZ, Bella J, Mayville P, Brodsky B, Berman HM (1999) Sequence dependent conformational variations of collagen triple-helical structure. *Nat Struct Biol* 6: 454–457. <https://doi.org/10.1038/8259> PMID: 10331873
  43. Koide T, Nishikawa Y, Asada S, Yamazaki CM, Takahara Y, Homma DL, et al. (2006) Specific recognition of the collagen triple helix by chaperone HSP47. II. The HSP47-binding structural motif in collagens and related proteins. *J Biol Chem* 281: 11177–11185. <https://doi.org/10.1074/jbc.M601369200> PMID: 16484215
  44. Fields GB (1991) A model for interstitial collagen catabolism by mammalian collagenases. *J Theor Biol* 153: 585–602. PMID: 1666905
  45. Manka SW, Carafoli F, Visse R, Bihan D, Raynal N, Farndale RW, et al. (2012) Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc Natl Acad Sci U S A* 109: 12461–12466. <https://doi.org/10.1073/pnas.12049911109> PMID: 22761315
  46. Krane SM, Inada M (2008) Matrix metalloproteinases and bone. *Bone* 43: 7–18. <https://doi.org/10.1016/j.bone.2008.03.020> PMID: 18486584
  47. Makareeva E, Leikin S (2007) Procollagen triple helix assembly: an unconventional chaperone-assisted folding paradigm. *PLoS ONE* 2: e1029. <https://doi.org/10.1371/journal.pone.0001029> PMID: 17925877
  48. Nagai N, Hosokawa M, Itohara S, Adachi E, Matsushita T, Hosokawa N, et al. (2000) Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J Cell Biol* 150: 1499–1506. PMID: 10995453
  49. Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Drogemuller C, et al. (2012) Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc Natl Acad Sci U S A* 109: 13243–13247. <https://doi.org/10.1073/pnas.1208072109> PMID: 22847422

50. Chessler SD, Byers PH (1993) BiP binds type I procollagen pro alpha chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. *J Biol Chem* 268: 18226–18233. PMID: [8349698](https://pubmed.ncbi.nlm.nih.gov/8349698/)
51. Ishida Y, Yamamoto A, Kitamura A, Lamande SR, Yoshimori T, Bateman JF, et al. (2009) Autophagic elimination of misfolded procollagen aggregates in the endoplasmic reticulum as a means of cell protection. *Mol Biol Cell* 20: 2744–2754. <https://doi.org/10.1091/mbc.E08-11-1092> PMID: [19357194](https://pubmed.ncbi.nlm.nih.gov/19357194/)
52. Reich A, Cabral WA, Marini JC (2012) Homogeneous mutant collagen in homozygous *Brtl* or *Brtl/mov* compound mice leads to improved bone phenotype through altered osteoblast differentiation. *Bone* 50: S166.
53. Barat-Houari M, Sarrabay G, Gatinois V, Fabre A, Dumont B, Genevieve D, et al. (2016) Mutation Update for COL2A1 Gene Variants Associated with Type II Collagenopathies. *Hum Mutat* 37: 7–15. <https://doi.org/10.1002/humu.22915> PMID: [26443184](https://pubmed.ncbi.nlm.nih.gov/26443184/)
54. Arnold WV, Fertala A (2013) Skeletal diseases caused by mutations that affect collagen structure and function. *Int J Biochem Cell Biol* 45: 1556–1567. <https://doi.org/10.1016/j.biocel.2013.05.017> PMID: [23707199](https://pubmed.ncbi.nlm.nih.gov/23707199/)
55. Arita M, Fertala J, Hou C, Steplewski A, Fertala A (2015) Mechanisms of aberrant organization of growth plates in conditional transgenic mouse model of spondyloepiphyseal dysplasia associated with the R992C substitution in collagen II. *Am J Pathol* 185: 214–229. <https://doi.org/10.1016/j.ajpath.2014.09.003> PMID: [25451152](https://pubmed.ncbi.nlm.nih.gov/25451152/)
56. Hintze V, Steplewski A, Ito H, Jensen DA, Rodeck U, Fertala A (2008) Cells expressing partially unfolded R789C/p.R989C type II procollagen mutant associated with spondyloepiphyseal dysplasia undergo apoptosis. *Hum Mutat* 29: 841–851. <https://doi.org/10.1002/humu.20736> PMID: [18383211](https://pubmed.ncbi.nlm.nih.gov/18383211/)
57. Forlino A, Kuznetsova NV, Marini JC, Leikin S (2007) Selective retention and degradation of molecules with a single mutant alpha1(I) chain in the *Brtl* IV mouse model of OI. *Matrix Biol* 26: 604–614. <https://doi.org/10.1016/j.matbio.2007.06.005> PMID: [17662583](https://pubmed.ncbi.nlm.nih.gov/17662583/)
58. Lisse TS, Thiele F, Fuchs H, Hans W, Przemecck GK, Abe K, et al. (2008) ER stress-mediated apoptosis in a new mouse model of osteogenesis imperfecta. *PLoS Genet* 4: e7. <https://doi.org/10.1371/journal.pgen.0040007> PMID: [18248096](https://pubmed.ncbi.nlm.nih.gov/18248096/)
59. Mertz EL, Makareeva E, Mirigian LS, Koon KY, Perosky JE, Kozloff KM, et al. (2016) Makings of a brittle bone: Unexpected lessons from a low protein diet study of a mouse OI model. *Matrix Biol* 52–54: 29–42. <https://doi.org/10.1016/j.matbio.2016.03.005> PMID: [27039252](https://pubmed.ncbi.nlm.nih.gov/27039252/)
60. Bi X, Grafe I, Ding H, Flores R, Munivez E, Jiang MM, et al. (2017) Correlations Between Bone Mechanical Properties and Bone Composition Parameters in Mouse Models of Dominant and Recessive Osteogenesis Imperfecta and the Response to Anti-TGF-beta Treatment. *J Bone Miner Res* 32: 347–359. <https://doi.org/10.1002/jbmr.2997> PMID: [27649409](https://pubmed.ncbi.nlm.nih.gov/27649409/)
61. Grafe I, Yang T, Alexander S, Homan EP, Lietman C, Jiang MM, et al. (2014) Excessive transforming growth factor-beta signaling is a common mechanism in osteogenesis imperfecta. *Nat Med* 20: 670–675. <https://doi.org/10.1038/nm.3544> PMID: [24793237](https://pubmed.ncbi.nlm.nih.gov/24793237/)