ABSTRACT

Title of Dissertation: RECESSIVE OSTEOGENESIS IMPERFECTA: PREVALENCE AND PATHOPHYSIOLOGY OF COLLAGEN PROLYL 3-HYDROXYLATION COMPLEX DEFECTS

Wayne Anthony Cabral, Doctor of Philosophy, 2014

Dissertation directed by: Dr. Joan C. Marini, MD, PhD
Bone and Extracellular Matrix Branch, NICHD
National Institutes of Health

Professor Stephen M. Mount, PhD
Department of Cell Biology and Molecular Genetics

Osteogenesis Imperfecta (OI) is a clinically and genetically heterogeneous heritable bone dysplasia occurring in 1/15,000-20,000 births. OI is a collagen-related disorder, with the more prevalent dominant forms caused by defects in the genes encoding the α1 and α2 chains of type I collagen (*COL1A1* and *COL1A2)*. Rare recessive forms of OI are caused by deficiency of proteins required for collagen post-translational modifications or folding. We identified deficiency of components of the ER-resident collagen prolyl 3-hydroxylase (P3H) complex as a cause of recessive OI. The P3H complex, consisting of prolyl 3-hydroxylase 1 (P3H1), cartilage-associated protein (CRTAP) and cyclophilin B/peptidyl-prolyl isomerase B (CyPB/PPIB), modifies the α1(I) P986 and α2(I) P707 residues of type I collagen. The most common P3H complex defects occur in *LEPRE1*,

the gene encoding P3H1, and over one-third of these cases are due to a founder mutation we identified among individuals of West African and African American descent. Our screening of contemporary cohorts revealed that 0.4% of African Americans and nearly 1.5% of West Africans are carriers for this mutation, predicting a West African frequency of recessive OI due to homozygosity for this mutation at 1/18,260 births, equal to *de novo* dominant OI. Furthermore, haplotype analysis of affected families was consistent with a single founder for this mutation, occurring 650-900 years ago (1100-1350 C.E.). Patients deficient in P3H1 and CRTAP have consistent bone phenotypes and collagen biochemistry. However, the rare cases of CyPB deficiency have variable findings distinct from P3H1/CRTAP. To clarify the OI mechanism of CyPB deficiency, we generated a *Ppib* knock-out mouse. In the absence of CyPB, only residual collagen prolyl 3-hydroxylation is detectable in KO cells and tissues. The delay in collagen folding in KO cells is further increased upon cyclophilin inhibition, supporting CyPB's role as an isomerase and the presence of redundancy for collagen ER PPIases. Site-specific alterations of collagen post-translational modification, particularly at residues involved in helical crosslinking, suggest that CyPB is critical to the function of collagen hydroxylases, especially LH1. Thus our studies indicate novel roles for CyPB, separate from the P3H complex, which directly and indirectly regulate collagen biosynthesis and bone development.

Recessive Osteogenesis Imperfecta:

Prevalence and Pathophysiology of Collagen Prolyl 3-Hydroxylation Complex Defects

by

Wayne Anthony Cabral

Dissertation submitted to the Faculty of the Graduate School of the

University of Maryland, College Park in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

2014

Advisory Committee:

Dr. Joan C. Marini, MD, PhD, Co-Chair
Professor Stephen M. Mount, PhD, Co-Chair
Professor Ian H. Mather, PhD
Professor John Moult, PhD
Professor Adam H. Hsieh, PhD

ACKNOWLEDGEMENTS

There are a number of people that have had an undeniable, positive impact on my professional development, and to whom I wish to express my sincerest gratitude. Several people have sacrificed their time and efforts to encourage, teach and challenge me, including Dr. Denise Pettit, Dr. Krista Fischer-Stenger and Dr. Antonella Forlino. I am extremely grateful for their generosity during the early years of my career. I would also like to acknowledge my appreciation to Dr. Elisabeth Gantt, Professor Emerita, for supporting my admission to the MOCB Program. I thank members of my thesis committee, Professors Ian Mather, Stephen Mount, John Moult and Adam Hsieh, for offering their time and constructive guidance. In addition, I would like to thank Dr. Stephen Mount and Dr. Ian Mather, Professor Emeritus, for their long-term commitment and support of my non-standard route through graduate school. Most importantly, I wish to express my sincere thanks for the efforts of my advisor, mentor and coach, Joan C. Marini, M.D., Ph.D., for her unwavering support of my personal and professional development, pushing me to reach my potential, and for teaching me to never give up. To my parents, Dr. Guy Cabral and Dr. Francine Marciano-Cabral, the greatest Mom, Dad, and role models that a son could have, I strive every day to make you proud. Finally, I am indebted to my fiancée, Adriana Zingone, M.D., Ph.D., for her patience, love and support.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3Hyp	3-hydroxyproline
4Hyp	4-hydroxyproline
AA cohort	African American cohort
aBMD	areal bone mineral density
αMEM	minimum essential medium
ADAMTS	a disintegrin and metalloproteinase with thrombospondin motifs
BFR	bone formation rate
BMDD	bone mineral density distribution
BMP	bone morphogenetic protein
Brtl	Brittle mouse model of dominant osteogenesis imperfecta
BS	bone surface
BV	bone volume
Ca	calcium
CNBr	cyanogen bromide
COL1A1	gene encoding alpha1 chain of type I collagen
COL1A2	gene encoding alpha2 chain of type I collagen
COP	coat protein complex
CRTAP	cartilage-associated protein
Cs	cross-section
CsA	cross-sectional area
Ct	cortical (bone)
cTAGE	cutaneous T-cell lymphoma-associated antigen
CyP	cyclophilin
DHLNL	dihydroxylysinonorleucine crosslink
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
DXA	dual Xray absorptiometry
EDS	Ehlers-Danlos Syndrome
EDTA	ethylenediaminetetraacetic acid
ERES	endoplasmic reticulum exit sites
ERGIC	endoplasmic reticulum-golgi intermediate complex
FKBP	FK506 binding protein
GAP	GTPase activating protein
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GEF	guanine nucleotide exchange factor
GGT	glucosylgalactosylhydroxylysine
GLT25D1	galactosyltransferase 25 KDa, D1
GLT25D2	galactosyltransferase 25 KDa, D2
GPC	Golgi to plasma membrane carriers
Gros1	GROwth Suppressor located at chromosome 1p31
GT	galactosyltransferase
HERDA	hereditary equine regional dermal asthenia

HLNL	hydroxylysinonorleucine crosslink
HP	hydroxylysylpyridinoline crosslink
HSP47	heat shock protein 47KDa
Hyl	hydroxylysine
KO	gene knock-out
LC/MS	liquid chromatography / mass spectrometry
LEPRE1	gene encoding leprecan / P3H1
LH1-3	lysyl hydroxylases 1-3
LOX	lysyl oxidase
LP	lysylpyridinoline crosslink
MAR	mineral apposition rate
MGB probe	minor groove binding DNA probe
MLBR	major ligand binding region
Mlt	mineralization lag time
MMP	matrix metalloprotease
MOI	bending moment of inertia
MS	mineralizing surface
MSC	mesenchymal stem cell
NCP	non-collagenous protein
NMD	nonsense-mediated decay
Ob	osteoblast
Oc	osteoclast
OI	osteogenesis imperfecta
Oim	osteogenesis imperfect murine model
OMIM	On-line Mendelian Inheritance in Man
OOA	Old Order Amish
OPG	osteoprotegerin
OV	osteoid volume
P3H1	prolyl 3-hydroxylase 1
P4Hα	prolyl 4-hydroxylase, alpha subunit
P4Hβ	prolyl 4-hydroxylase, beta subunit (aka PDI)
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PLOD	procollagen-lysine, 2-oxoglutarate 5-dioxygenase
PPIase	peptidyl-prolyl *cis-trans* isomerase
PPIB	peptidyl-prolyl *cis-trans* isomerase B
PTC	premature termination codon
PTH	parathyroid hormone
qBEI	quantitative backscattered electron imaging
RANK	receptor activator of nuclear factor kappa
RANKL	receptor activator of nuclear factor kappa ligand
RER	rough endoplasmic reticulum
RNA	ribonucleic acid
RT-PCR	reverse transcription-polymerase chain reaction
SDS	sodium dodecyl sulfate

SERPINH1	serine protease inhibitor, clade H
SH3	SRC homology 3 domian
SIBLING	Small Integrin-Binding LIgand, N-linked Glycoprotein
SLRPs	small leucine-rich proteoglycans
SMAD	human homolog of *C. elegans* SMA and *Drosophila* MAD protein
SNP	single nucleotide polymorphism
STR	short tandem repeat
TANGO	Transport ANd Golgi Organization
Tb	trabecular (bone)
TE	Tris-EDTA
TEM	transmission electron microscopy
TGFβ	transforming growth factor beta
Tm	melting temperature
TRAPP	TRAnsport Protein Particle
WA cohort	West African cohort
WGA	whole genome amplification
WNT	human homolog of the int/Wingless genes
μQCT	micro-quantitative computed tomography

TABLE OF CONTENTS

Chapter I

Introduction

Collagens are the most abundant proteins in vertebrates, accounting for approximately one-third of total protein mass in mammals. They not only provide extracellular scaffolds for organ development and tissue support, they also play crucial roles in regulating cell growth, differentiation and migration. The discovery of genes encoding collagens in invertebrates, such as sponges, Hydra and even hydrothermal vent worms, demonstrates the importance of these proteins in the animal kingdom. In humans the collagen family of proteins now comprises at least 28 proteins encoded by 46 genes. It is believed that this diverse array of collagens arose through a series of gene duplications and divergence, a process that played a key role in vertebrate evolution.

The structural organization that defines collagenous protein is the combination of three left-handed alpha chains with the characteristic Gly-X-Y amino acid triplet repeat, where Gly is glycine and the X and Y residues are often proline and hydroxyproline. Upon synthesis in the ER, the polypeptide chains assemble into right-handed homotrimeric or heterotrimeric molecules depending on the collagen type. This helical domain of collagen functions to provide mechanical strength to tissues, while non-collagenous domains assist in subunit assembly and stability, and possess biological signaling activities. The molecular diversity of the collagen family is enhanced by the existence of isoforms that use alternative alpha chains within a given collagen type, including alpha chains of differing collagen types, alternative promoters, variant splicing and post-translational modifications.

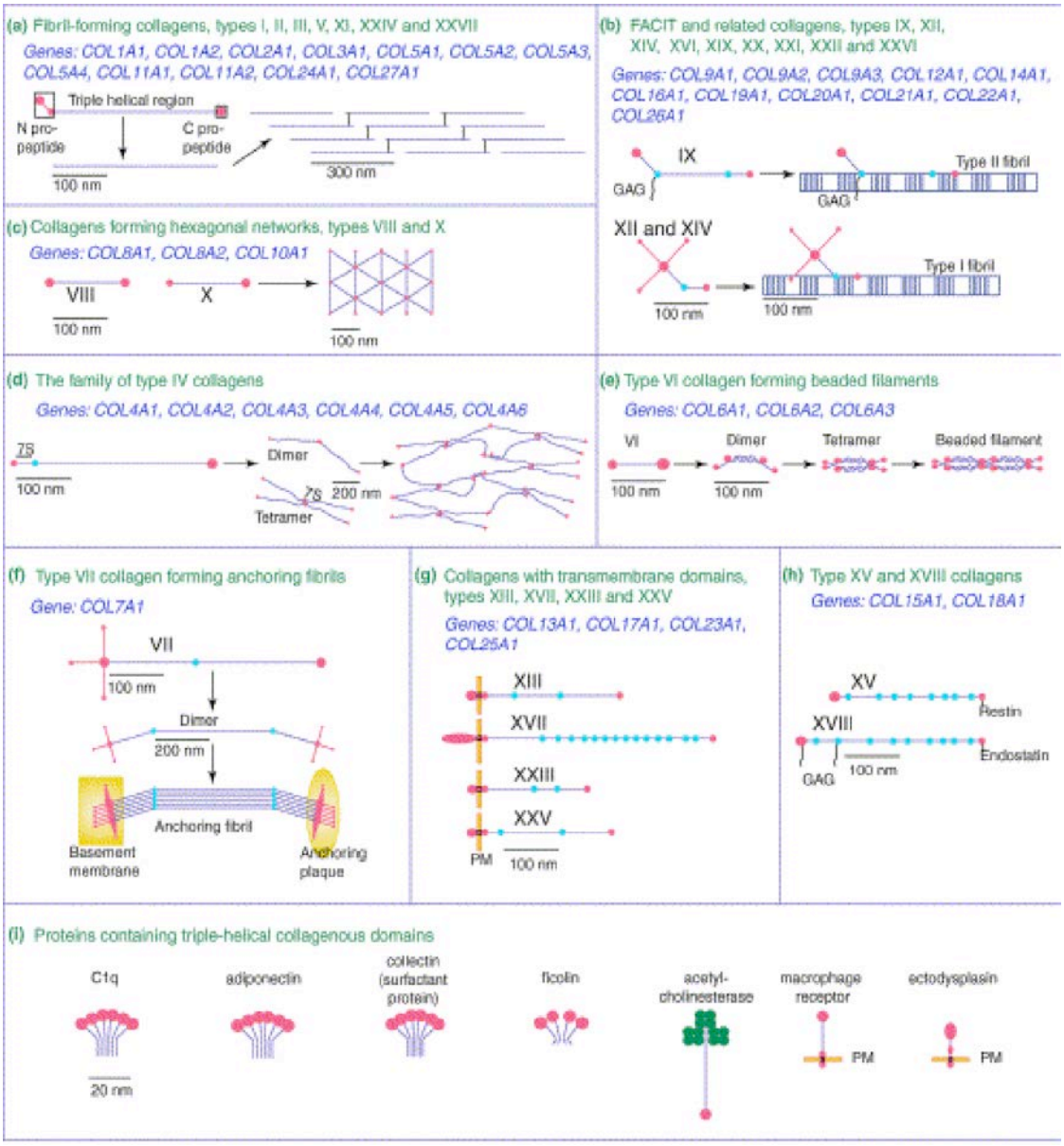

Figure 1. The Collagen Family of Proteins. Collagens are classified into three groups according to structure, function and localization. The collagens include at least 28 proteins, encoded by 46 genes, which provide structural support for the extracellular matrix of tissues. Type I collagen, the most abundant protein in vertebrates, is a fibril-forming collagen. Adapted from Myllyharju and Kivirikko [9].

Based on structural and functional characteristics, three major classes of collagens have been proposed (Figure 1). Group 1 collagens, the fibril-forming collagens, each have an uninterrupted helical domain approximately 300nm in length. These collagens, including types I, II, III, V and XI, are secreted as procollagens, processed by specific

pericellular propeptidases and assemble into higher order structures, termed fibrils, in the extracellular matrix. Group 2 collagens (types IV and VII), commonly referred to as basement membrane collagens, contain interrupted extended triple helices greater than 300nm in length. The short chain Group 3 collagens include those with uninterrupted helical regions (types VI, VIII and X), fibril-associated collagens with interrupted triple helices (FACIT collagens) such as types IX, XII and XIV, the transmembrane collagen types XIII and XVII, and collagens with multiple helices and interruptions (MULTIPLEXIN types XV and XVIII). Several additional proteins containing domains with collagenous triple helical conformations have been identified, including complement component C1q, adiponectin, ficolins, tumor necrosis factor and acetylcholinesterase. However, the absence of a structural role in extracellular matrix excludes their inclusion within the collagen family of proteins.

Type I Collagen

Type I collagen, the most abundant protein in humans, is a heterotrimer consisting of two α1(I) and one α2(I) chains encoded by the *COL1A1* and *COL1A2* genes, respectively. *COL1A1* is an 18kb gene localized to human chromosome 17q21.3-q22, while *COL1A2* is a 38kb gene at chromosome 7q21.3-q22. The structures of both genes consist of 52 exons, 42 of which encode the major triple helical domain. The majority of the exons are 54bp, with the remaining exons being multiples of 54bp or combinations of 54 and 45bp. Each exon begins with an intact glycine codon and encodes a specific

multiple of Gly-Xaa-Yaa triplets, reflecting the evolutionary origins of the separate collagen types by exon duplication [12].

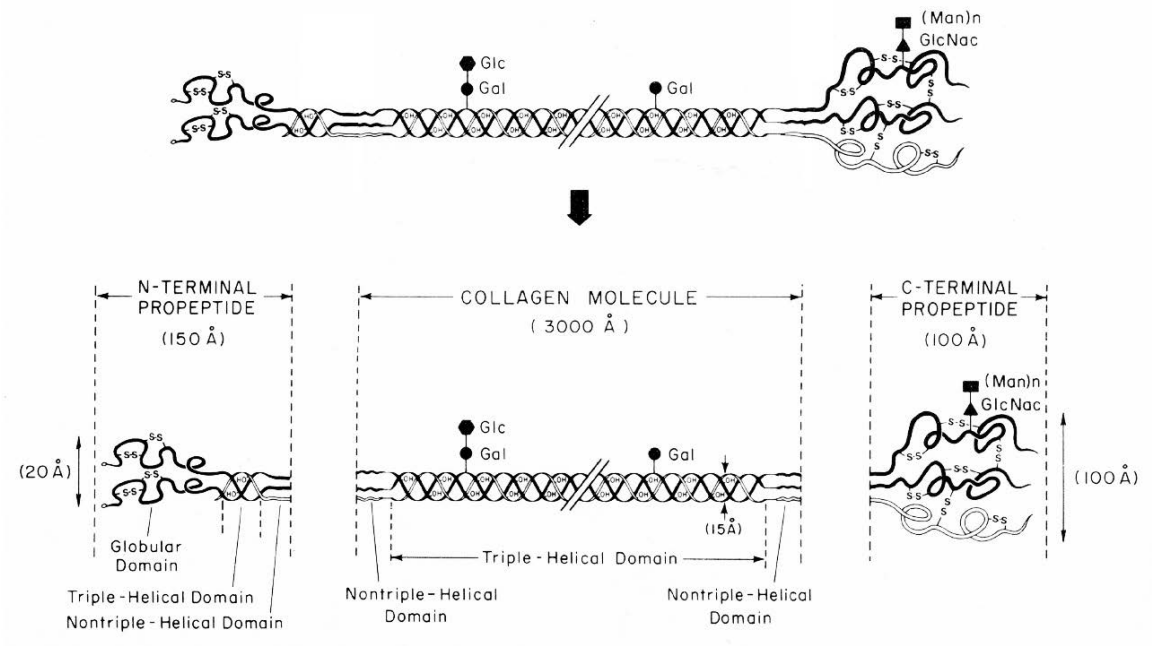

Figure 2. Type I Procollagen Structure. The three pro-alpha chains consist of a central helical region with 338 repeats of the Gly-Xaa-Yaa triplet flanked by two globular propeptide domains. The amino and carboxyl globular domains are processed upon secretion by specific pericellular propeptidases. Figure adapted from Prockop *et al* [10].

Type I collagen is mainly synthesized and secreted by fibroblasts and osteoblasts, cells whose primary function is *de novo* generation or replacement of the scaffolding of the extracellular matrix of skin, tendon and bone. The pro-alpha chains of type I collagen consist of 338 uninterrupted repeats of the Gly-Xaa-Yaa triplet comprising the helical region, flanked by globular domains at each terminus that are processed by pericellular propeptidases upon secretion from the cell (Figure 2). The amino and carboxyl propeptides of the pro-α1(I) chain consist of 139 and 246 residues, respectively, but are 57 and 247 residues in length in the pro-α2(I) chain. At each end of the helix there are short stretches of 11-26 residues, referred to as telopeptides, which connect the helix with the globular domains and also contain the propeptidase cleavage sites. Within the helical

region there is a requirement for glycine at every third position due to interior size constraints of the triple helix. Glycine is also involved in the major stabilizing force for the triple helix, which is provided by interchain hydrogen bonding between the NH-group in glycine of one alpha chain and the O=C group of the X position residue in a neighboring alpha chain. Further stabilization of the triple helical structure is provided by the X and Y-position residues, which are often proline and 4*R*-hydroxyproline. Additional stabilization of the collagen helix has been attributed to the interchain salt bridges that form between positively charged Y-position lysyl residues and negatively charged X-position glutamic/aspartic acid residues [13,14]. The combination of hydrogen bonds, electrostatic interactions and hydrophobic forces create a rigid, yet flexible, rod-like structure perfectly suited to provide support and organization to the extracellular matrix.

The organization of pro-α1(I) and pro-α2(I) peptide chains into the type I procollagen molecule begins with the coordinated expression of their mRNAs. Polysome-associated *COL1A1* and *COL1A2* transcripts have been demonstrated to occur in a 2:1 ratio in fibroblast and osteoblast extracts, and this ratio of transcripts reflects the total cellular pool of translatable transcripts [15]. Translation products from polysome-bound transcripts are spatially organized on the cytoplasmic face of the ER membrane by an incompletely understood mechanism. Investigations of early stages of procollagen assembly identified nascent polysome-bound collagen chains that were able to assemble into pepsin-resistant molecules with a heterotrimeric structure. While folding of these partially translated alpha chains was probably induced by the isolation procedure, the existence of ribosome-associated alpha chains that could trimerize provided indirect

evidence that translation products translocate into the ER lumen in a coordinated manner [16,17]. Coordinated translation is directed by a conserved stem-loop structure adjacent to the translation initiation codon in the 5'UTR of α1(I) and α2(I) mRNAs [18]. In murine fibroblasts lacking type I collagen due to the absence of endogenous *Col1a1* transcripts, pepsin-resistant heterotrimeric collagen has been generated by expression of full-length α1(I) cDNA that includes the 5' stem-loop. However, in cells expressing α1(I) cDNA with a conformation-defective mutant stem-loop or missing this noncoding structure, properly folded collagen molecules were not synthesized [19]. These findings were extended to define the 5' stem-loop as a binding motif. Incubation of cell extracts with RNA probes corresponding to the stem-loop sequence resulted in detection of an RNA-protein complex by gel-shift assay. Subsequent analyses of cDNA library clonal lines with increased 5' stem-loop binding led to the identification of La ribonucleoprotein domain family member 6 (LARP6) as the binding partner. RT-PCR amplification of α1(I) and α2(I), but not other transcripts, from immunoprecipitates of LARP6 from fibroblast extracts confirmed the *in vivo* binding of LARP6 to the stem-loop. Importantly, although mRNA levels remain unchanged, synthesis and secretion of type I collagen protein is decreased upon targeting of *LARP6* transcripts by siRNA. Furthermore, the absence of LARP6 is accompanied by loss of collagen immunostaining at discrete foci within the ER. These data are consistent with a role for LARP6 in coordinated translation of collagen alpha chains [19].

Upon completion of translation, one of the initial steps in collagen trimeric assembly is the association of the three pro-alpha chains in a conformation that is conducive to

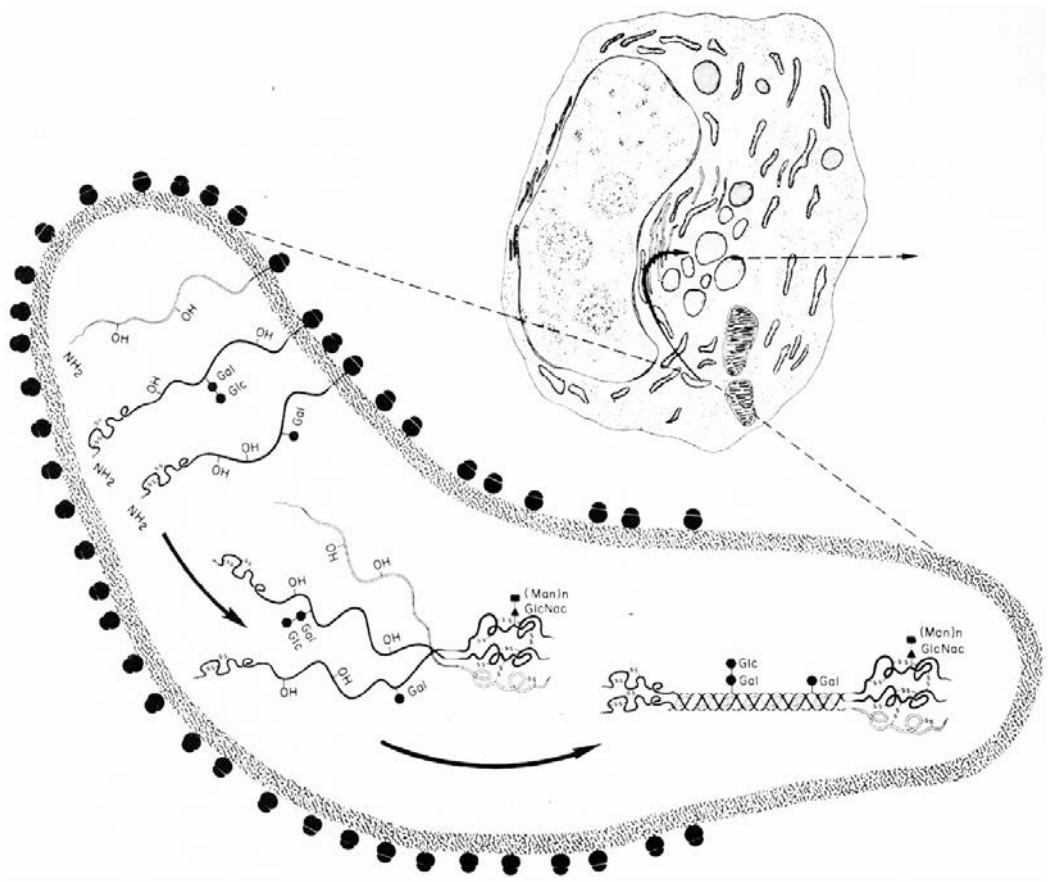

Figure 3. Procollagen Synthesis and Modification. The three pro-alpha chains assemble at the carboxyl propeptide alignment regions and are co- and post-translationally modified by lysyl hydroxylase, prolyl 3-hydroxylase and prolyl 4-hydroxylase. Some hydroxylysyl residues are subsequently glycosylated. The collagen helical region is no longer accessible to the modifying enzymes upon completion of folding. Figure from Prockop *et al* [10].

folding of the trimeric molecule from the carboxyl to amino terminus (Figure 3). In addition to conferring solubility under physiological conditions and preventing aggregation, the globular propeptide domains on procollagen contain specific residues that are required for procollagen type-specific association. By exchanging regions of the C-propeptide of pro-α1(III) chains, which only form homotrimers, with regions of the pro-α2(I), which can only assemble as heterotrimers with pro-α1(I) chains, Bulleid and colleagues determined that procollagen chain association is driven by a 23-residue molecular recognition sequence residing within the C-propeptide [20]. Although pro-α2(I)

chains are normally degraded in the absence of pro-α1(I) chains, in these experiments pro-α2(I) chains containing fragments of the pro-α1(III) propeptide were able to form homotrimers.

Another factor determining type-specific chain association resides in the highly conserved cysteine residues contained within the propeptide sequence of procollagen alpha chains. Fibrillar collagens that form homotrimers (pro-α1(I), pro-α1(II), pro-α1(III), pro-α1(V)) contain 8 cysteines, while chains requiring heterotrimer assembly (pro-α2(I), pro-α2(V) and pro-α1(XI)) contain only 7 cysteines in their C-propeptides. In type I procollagen, the last four cysteine residues in each chain form intramolecular disulfide bonds important for folding into the comformation required for association, while the remaining residues participate in intermolecular crosslinks [21,22]. It has been proposed that triple helix formation occurs before formation of interchain disulfides in the C-terminus of collagen molecules [20]. However, this is unlikely since mutations of the cysteine residues and loss of disulfide bonds in the C-propeptide disrupt chain association, possibly by altering the folded conformation of the C-propeptide [23]. In fact the importance of disulfide bonds to collagen assembly was first demonstrated by inhibition of procollagen folding in cells treated with a reducing agent [24]. Formation of these disulfide bonds can proceed absent any enzyme catalysis.

Post-translational Modifications of Collagen

Biosynthesis of procollagen is a complex process that requires several co- and post-translational modifications within the ER, including formation of inter- and intra-chain

disulfide bonds, isomerization of peptidyl-prolyl bonds, hydroxylation of prolyl and lysyl residues, and glycosylation of hydroxylysine residues. The specific post-translational modifications occur before, and to a major extent enable, collagen helical folding. Once folding is complete the collagen alpha chains are inaccessible to the modifying enzymes which require a non-triple helical conformation [25]. Given the number of post-translational modifications that are required, it is not surprising to find a diverse array of intracellular enzymes that interact with alpha chains during synthesis and maturation of the procollagen molecule. These crucial post-translational modifications are catalyzed by thioredoxins, 2-oxoglutarate dioxygenases, immunophilins and glycosyltransferases.

Protein Disulfide Isomerase

Disulfide bonds, predicted to occur in at least one-third of all secretory proteins, serve to aid in folding, stabilization and potentiation of assembly of multimeric peptides. The ER is an environment conducive to oxidation of proteins, thus a means to ensure that correct disulfide bonds are formed is essential for proper folding of cargo and ER quality control. This activity was first demonstrated for Protein Disulfide Isomerase (PDI), the founding member of the PDI family of thioredoxin proteins. In Homo sapiens, there are at least 21 members of the PDI family, although not all possess isomerase activity. PDI, encoded by the *P4HB* gene, is an ER-resident multifunctional enzyme that catalyzes the oxidation of thiol groups, as well as reduction and rearrangement of disulfide bonds in various cell surface and secreted proteins [26,27]. PDI is a 55 kDa protein with four thioredoxin domains (notated a-b-b'-a'). The internal "b" domains do not have redox cysteines, but contain hydrophobic motifs involved in substrate interactions. The two

external thioredoxin "a" domains of PDI contain the C-G-H-C active site residues that catalyze the formation, breakage and rearrangement of disulfide bonds [28,29]. Oxidized PDI first interacts with the substrate via the hydrophobic motifs. Upon binding, PDI accepts an electron from the substrate nucleophilic cysteine, creating a mixed disulfide between the two proteins. A second electron transfer completes oxidation of the substrate, resulting in an enzyme that maintains reducing potential toward the substrate. Isomerization of aberrant disulfide bonds is achieved through repeated cycles of reduction and oxidation of the substrate disulfides [30]. Release of the substrate allows PDI to be replenished to the oxidized state by a separate ER resident redox protein, Ero1. However, only PDI in the presence of glutathione has been demonstrated to accelerate formation of triple helices from reduced and denatured procollagen chains *in vitro*, connecting the ER redox system to collagen synthesis [31].

Several "non-catalytic" interactions with other proteins raise the possibility that PDI is involved in ER retention and quality control. For instance, PDI forms covalent complexes with the C-propeptides of unassembled pro-α2(I) chains in the absence of pro-α1(I) chains within the ER, but not with correctly assembled heterotrimers [32]. This interaction has been shown to assist in preventing aggregation of unassembled pro-α chains that are retained in the ER [33]. However, unassembled pro-α2(I) chains complexed with PDI lacking an ER retrieval signal can still be secreted [32]. The KDEL-containing β subunit (PDI) of the prolyl 4-hydroxylase complex also serves to retain the α subunit (P4Hα) within the ER since it does not itself contain a COOH-terminal ER retrieval signal. Furthermore, α subunits form insoluble aggregates within microsomes in the

absence of PDI in cell-free translation experiments, demonstrating the ability of PDI to act as a chaperone that inhibits aggregation of misfolded proteins [34].

Prolyl 4-hydroxylase

Proline represents 23% and 20% of the residues within the helical regions of the human α1(I) and α2(I) chains, respectively, while the Pro–4*R*Hyp–Gly sequence occurs in approximately 10% of collagen amino acid triplets [35]. The presence of 4*R*-hydroxyproline (4Hyp), which occurs exclusively in the Y-position of X-Y-Gly collagen triplets, endows collagen with its stability at physiological temperatures. This function of Hyp is easily demonstrated by the increased melting temperature of fully hydroxylated type I collagen (T_m = 39°C) compared to the unhydroxylated form (T_m = 24°C) [36]. Short synthetic peptides containing variable $(X\text{-}Y\text{-}Gly)_n$ triplets have been used extensively to investigate stabilization of the triple helix by 4-hydroxyproline. These studies led some investigators to hypothesize that hydrogen bonding between water molecules and the hydroxyl group of Hyp provides the major stabilizing force to the collagen helix [37,38]. Others have proposed that hydroxylation at the 4-position introduces an electronegative element that induces a C'-exo pucker to the pyrrolidine ring, which favors the *trans* conformation of the peptidyl-prolyl bond [39-41]. This mechanism for stabilization of the helical conformation is supported by experiments in which the hydroxyl group was replaced with an electronegative, non-hydrogen bonding flourine at the 4-position to create $(Pro\text{-}R\text{-}Flp\text{-}Gly)_{10}$ peptides, resulting in a higher thermal stability than any known collagenous decapeptide [42]. Hence, hydroxylation of prolyl residues favors the *trans* conformation of the peptidyl-prolyl bond that is required for helical folding [43].

The co- and post-translational hydroxylation of proline residues in X-Pro-Gly sequences of collagens is catalyzed in the ER by prolyl 4-hydroxylase (P4H) [44,45]. The enzyme consists of two distinct α and β polypeptides, the β subunits being identical to the multifunctional enzyme protein disulfide isomerase (PDI/(P4Hβ) and the 59-kDa α subunits (P4Hα) being a unique gene product [46,47]. Three isoforms of the P4Hα subunit, α(I), α(II), and α(III), have been identified in mice and humans. In most tissues, particularly in skeletal muscle, liver, heart, lung and skin, α(I) is the most highly expressed subunit [48-50]. The α(II) subunit is the main form in chondrocytes and endothelial cells. Both α(I) and α(II) forms are highly expressed in fibroblasts and osteoblasts [51]. Although the α(III) mRNA is expressed at very low levels in many fetal and adult tissues, the highest levels of transcripts have been identified in placenta, liver and skin. Co-expression studies in insect cells have demonstrated that the isoforms associate in an $\alpha_2\beta_2$ tetramer with PDI, but do not form mixed complexes [49]. Considering that sequence analysis has revealed higher identity between the α(I) and α(II) isoforms (65%) than with the α(III) subunit (35-37%), this arrangement potentially allows for substrate-specific interactions with P4H variants [48]. Furthermore, 20% residual P4H activity towards an unhydroxylated type I procollagen substrate was observed in cell lysates from fibroblasts lacking P4Hα1 subunits. Although the P4Hα(I) null mouse is embryonic lethal, the phenotype is attributed to disruption of type IV collagen-containing basement membranes rather than defects in type I collagen thermal stability or fibril structure. Lehmann *et al* demonstrated that, in type I collagen synthesized by 13 normal control fibroblast cell lines, 46-49% of prolyl residues were hydroxylated, correlating with the fraction of prolyl residues that occur in the Y-position [52]. However, analyses of

identical type I collagen peptide fragments derived from bone, skin and tendon has revealed that the extent of prolyl hydroxylation can vary between tissues and at different ages [53,54]. These data suggest tissue-specific and substrate-specific interactions between collagen types and P4H isoenzymes [55].

The catalytic activity of P4H was first demonstrated in microsomal fractions [56], and the enzyme was subsequently purified from chick embryos [57,58]. Biochemical and sequence analyses have shown that P4H is a member of the 2-oxoglutarate dioxygenase family of enzymes that also includes the prolyl 3-hydroxylases and the lysyl hydroxylases. Accordingly, most of the conserved amino acid residues occur in the C-terminal region of the alpha subunits, which contain the catalytic active site including two histidine residues and an aspartate residue [50]. The active site is believed to consist of portions of the α and β subunits, since antibodies to the β subunit inhibit the hydroxylase activity of the tetramer, although it is unknown if these findings reflect antibody interference with subunit assembly [59]. However, it should be noted that alpha subunits from simpler life forms such as green algae are expressed and function as monomers [60]. Interestingly, the substrate-binding domain of prolyl 4-hydroxylase alpha subunits has been demonstrated to be distinct from the catalytic domain. Sequencing of trypsin-protected fragments of poly-proline peptides bound to tetrameric P4H *in vitro* identified an approximately 100-residue region separate from the C-terminal catalytic site responsible for interaction between alpha subunits and proline-containing substrate [61].

As with lysyl hydroxylase and prolyl 3-hydroxylase, activity of P4H requires a nonhelical substrate, Fe^{++}, 2-oxoglutarate, O_2 and ascorbate [25,44]. During this reaction the 2-oxoglutarate is decarboxylated, with one oxygen atom incorporated into the succinate

product and the other oxygen into the hydroxyl group of proline [62]. The role of ascorbate in this reaction is not entirely clear, since the enzyme can complete several catalytic cycles without it [63]. However, impaired 4-hydroxylation of collagen proline residues and the resulting extracellular matrix deficiency in scurvy is well characterized [64]. The minimum substrate requirement for hydroxylation is a single X-Pro-Gly tripeptide, and free proline is not hydroxylated by P4H [65]. While proline is the preferred residue in the X-position, hydroxylation can also occur with other residues, such as alanine, at this position [66]. In $(Pro\text{-}Pro\text{-}Gly)_5$ synthetic peptides, the proline residues in the Y-position of the third and fourth triplets are preferentially hydroxylated [67].

Lysyl hydroxylases

The lysyl hydroxylases (LH), members of the 2-oxoglutarate dioxygenase protein family, catalyze the formation of 5-hydroxylysine residues in collagens and other proteins with collagenous sequences by the addition of a hydroxyl group to lysine residues in X-Lys-Gly, X-Lys-Ala or X-Lys-Ser triplets [68]. Hydroxylysine residues serve as substrates for attachment of carbohydrate groups to pro-alpha chains and for the formation of intermolecular crosslinks in higher order structures of collagen. The human lysyl hydroxylase family consists of LH1-3, encoded by *PLOD1* (procollagen-lysine 2-oxoglutarate 5-dioxygenase 1), *PLOD2* and *PLOD3*, respectively. These hydroxylases have approximately 60% amino acid identity, due to their formation by gene duplication and intron lengthening during evolution [69]. The highest degree of homology occurs within the carboxyl-terminal region that contains the LH active site, structurally similar to the active site domain found in prolyl hydroxylases, which are also classified as 2-

oxoglutarate dioxygenases. Despite lacking a C-terminal KDEL ER-retrieval tetrapeptide, the three LH isoenzymes are luminally-oriented membrane-associated enzymes [70]. Although the three lysyl hydroxylases have similar substrate specificity for short synthetic peptides *in vitro* [71], recent discoveries have provided new insights into the unique functions of the individual isoenzymes *in vivo*.

In a pair of articles representing the first biochemical characterization of an inborn error of human collagen metabolism, Krane and colleagues demonstrated deficiency of procollagen lysine hydroxylation in a subset of patients with Ehlers-Danlos syndrome (EDS), a group of connective tissue disorders that share the common features of joint laxity, scoliosis and skin hyperextensibility. Amino acid analysis of dermal collagen from two sisters revealed a decrease in hydroxylysine content to 5% of normal, with variable reductions in bone, but normal levels of hydroxyproline and other amino acids [72]. These studies demonstrated that crude cell lysates from cultured patient fibroblasts showed an 85-90% reduction in enzymatic hydroxylation of fibroblast-derived unmodified procollagens. The absence of collagen lysyl hydroxylation in two affected siblings within a clinically normal extended pedigree fit a recessive inheritance pattern, suggestive of a specific enzymatic deficiency [73]. It would be twenty years before cloning of the full-length *PLOD1* cDNA led to the identification of homozygous nonsense mutations in a cohort of EDS type VIA patients, whose dermal tissue exhibited severely reduced lysyl hydroxylase activity [74].

While subsequent studies of patients with LH1 deficiency have confirmed insufficiency of helical hydroxylysine on analysis of hydrolyzed dermis or reduced enzymatic activity in cultured skin fibroblasts, amino acid quantitation of EDS VIA

patient collagens has also shown normal levels of hydroxylated lysine residues within the telopeptides [75,76]. Royce *et al* demonstrated that highly purified LH1 from chick fibroblasts was able to hydroxylate lysyl residues of CNBr-derived peptides of unmodified procollagen corresponding to the helical region, but was unable to hydroxylate lysyl residues within peptides corresponding to the non-helical regions of collagen [77]. Conversely, biochemical analysis of bone-derived collagen from Bruck Syndrome 2 patients revealed underhydroxylation of telopeptide lysyl residues, but normal hydroxylation of helical residues [78]. These findings provided strong circumstantial evidence for the existence of a unique lysyl hydroxylase specific for the collagen telopeptides. Indeed, Valtavaara *et al* reported the isolation and characterization of cDNA clones encoding a novel isoform of lysyl hydroxylase, which they called lysyl hydroxylase-2 (PLOD2) [79]. In contrast to the other isoenzymes, LH2 contains an extra exon, designated 13A, which results in 2 forms of mRNA referred to as LH2a (short form) and LH2b (long form), respectively [80]. Several lines of evidence have confirmed LH2 as the specific collagen telopeptidyl lysyl hydroxylase. First, the level of expression of *Plod2* increases drastically during the mineralization phase of differentiation in BMSC cultures, the timing of which corresponds to increased telopeptide-derived crosslink formation [81]. Second, overexpression of LH2b, but not LH2a, LH1 or LH3, increases collagen telopeptide lysyl hydroxylation and crosslink formation in MC3T3 cells in culture [82]. Third, homozygosity mapping in two pedigrees with multiple siblings affected by Bruck Syndrome localized a candidate gene defect to chromosome 3q23-q24. Sequencing identified homozygosity for *PLOD2* missense mutations in affected individuals.

The third human lysyl hydroxylase, LH3, is evolutionarily the oldest of the LH isoforms [69]. Similar to its single ortholog in *C. elegans*, human LH3 is a multifunctional enzyme with lysyl hydroxylase (LH), galactosyltransferase (GT) and glucosyltransferase (GGT) activities (see Glycosyltransferases, below). Investigations of LH3 activity using truncated or mutagenized recombinant enzyme against collagen substrates *in vitro* have established that GT and GGT activity is attributed to the N-terminal portion of LH3 [83,84]. Inactivation of LH3 lysyl hydroxylase activity in a knock-in mouse reduces hydroxylation of types IV and V collagens, but not types I, II and III collagens, and causes structural changes in tissues, particularly in basement membranes [85]. Since types IV and V collagens are more extensively hydroxylated than collagens I, II and III, the decreased helical lysine hydroxylase activity indicates a lack of compensation by LH1 and LH2 *in vivo*. Still, the moderate phenotype of this mouse model, compared to the embryonic lethality of the LH3 knockout, suggests that the most significant role for this enzyme does not reside in its LH activity [86].

Several features of collagen lysine post-translational modification remain to be clarified. Although it has been demonstrated that LH1 activity is specifically directed to collagen helical lysine residues, overexpression of rat *Plod1* cDNA in CHO cells has been shown to increase telopeptide pyridinium crosslink production, which is known to require *Plod2* activity [87]. As mentioned above, most of the collagen biochemical data on EDS VI has been limited to patient dermal samples and demonstrate almost complete loss of lysyl hydroxylation. However, in the few cases of LH1 deficiency in which other tissues have been analyzed, lysyl hydroxylation of collagen from cartilage and bone was only reduced 10-50% versus normal control collagen [72]. It has also been noted that

cultured skin fibroblasts with severely reduced lysyl hydroxylase activity (12-16% of control values) synthesize hydroxylysine-deficient collagen (8-9% of control values), while direct quantitation of dermal samples from the same patient demonstrate only moderate reductions in hydroxylysyl residues (60-80% of control values) [88]. Furthermore, quantitation of collagen hydroxylation in several tissues isolated from *Plod1*$^{-/-}$ mice have revealed marked variation in the magnitude of hydroxylysine deficiency. Supporting the ability of the two remaining isoenzymes to modify helical lysine residues, investigators found a 70-80% decrease in Hyl content in collagen isolated from skin, cornea and tail tendon, although lung, femur and aorta-derived collagens were underhydroxylated by only 15-35% compared to wild-type murine collagens [89]. These findings are not entirely surprising since each LH isoform can be found at variable levels in collagen-expressing tissues, with unique patterns of expression arising at later stages of development [90]. For instance, northern analyses show that at early embryonic stages of development (E7-15) murine *Plod1* is expressed at levels higher than *Plod2* and *Plod3*. However, transcripts for *Plod 3* increase dramatically at E15 almost to the levels seen for *Plod1*. Yamauchi and colleagues have observed dramatic increases in expression of *Plod1*, *Plod2* and *Plod3* in differentiating murine BMSC cultures, including at the proliferative preosteoblastic phase (*Plod1*), matrix producing osteoblast transition (*Plod3*), and in matrix mineralizing osteocytes (*Plod2*) [81]. Finally, there is no current explanation for expression of two separate forms of LH2; the long form (LH2b) is believed to hydroxylate telopeptides, but no function has been proposed for the short form (LH2a). These studies suggest that plasticity in substrate-specific modifications of collagens by

LH isoforms is a complex process involving developmental-, tissue- and cell -specific regulation.

FK506 Binding Protein 65 KDa

Three classes of peptidyl-prolyl isomerases (PPIases) are encoded in the human genome and include the cyclophilins (CyPs), the FK506-binding proteins (FKBPs) and the parvulins. PPIases are responsible for the catalysis of *cis-trans* isomerization of peptidyl-prolyl bonds, important for folding of proteins rich in proline. Cyclophilins and FKBPs, considered "immunophilins", are so named based on their inhibition by the immunosuppressive drugs cyclosporine A and FK506, respectively. Immunophilins have been found in every cellular compartment and tissue as membrane-bound, soluble or secreted proteins [91]. They have been implicated in various cellular functions such as the regulation of heat shock response, steroid receptor and ion channel activity, cell-cell and cell-matrix interactions, transcription and translation [92]. More importantly for proteins within the secretory pathway in humans, at least two cyclophilins and six FKBPs have been identified as ER resident proline isomerases [93]. The search for specific ligands to these isomerases as a means to understand their physiological functions has provided insights into the roles of cyclophilin B and FKBP65 in collagen biosynthesis.

Fkbp10, which encodes FKBP65, was first cloned and sequenced from a cDNA library during a search for differentially expressed genes in murine fibroblasts undergoing neoplastic progression [94]. Sequence analysis predicted the presence of four PPIase domains, two calcium-binding EF hand motifs and a C-terminal ER retrieval tetrapeptide. Subsequent studies of recombinant FKBP65 verified that the putative

isomerase was able to accelerate the isomerization of the peptidyl-prolyl bond within a succinyl-Ala-Ala-Pro-Phe-nitroanilide substrate *in vitro* with a catalytic efficiency similar to other FKBP family members. This activity was inhibited by FK506 and rapamycin [95]. The first evidence for an *in vivo* role of FKBP65 came from a series of investigations designed to identify specific chaperones for tropoelastin in chondrocytes [96]. Coimmunoprecipitation of FKBP65 with tropoelastin from cell lysates provided the first identification of a proline-rich ligand for an FKBP within the secretory pathway. The investigators speculated that FKBP65 facilitated folding and trafficking of tropoelastin through post-ER compartments, and might facilitate crosslinking of elastin into the extracellular matrix.

Since approximately one-sixth of the primary sequence of collagen alpha chains consists of proline residues, peptidyl-prolyl isomerization has long been considered the rate-limiting step in collagen folding, as only *trans* peptide bonds can be incorporated into the collagen triple helix [97]. The possibility that FKBP65 is involved in this process was investigated by Bachinger and colleagues. They isolated FKBP65, along with several other collagen-interacting proteins, as eluates of chick embryo extracts applied to a gelatin-Sepharose column [98]. Analysis of purified FKBP65 activity *in vitro* revealed that, in contrast to other FKBPs, its highest activity was towards leucine- and alanine-containing peptides that more closely resemble collagen telopeptide compared to helical sequences. In contrast to the other FKBPs, the isomerase activity of FKBP65 toward a succinyl-Ala-Ala-Pro-Phe-nitroanilide substrate is inhibited 25% *in vitro*, but only marginally against procollagen *in vivo*, in the presence of either FK506 or CsA [98,99]. Another feature of FKBP65 that distinguishes it from other FKBP family members is that

it is retrotranslocated from the ER and degraded by the proteosome under conditions of ER stress, arguing against a specific role as a folding chaperone [100]. In addition, the rate of refolding of denatured collagen *in vitro* was only minimally increased in the presence of purified FKBP65 [99]. These studies prompted investigators to consider other roles for FKBP65 in collagen biosynthesis. A putative chaperone function for FKBP65 is suggested by its ability to inhibit thermal aggregation of citrate synthase and rhodanese in a concentration-dependent manner. This activity is not affected by the presence of FK506 or CsA. Perhaps the most significant information generated on FKBP65 from the *in vitro* characterization studies was the observation that it interacted with fully folded collagen and inhibited fibril formation [101].

A fuller picture of FKBP65 function *in vivo* has now emerged with the identification and biochemical characterization of patients with bone dysplasia caused by recessive mutations in the gene encoding FKBP65. In the absence of FKBP65, *in vivo* folding is normal as suggested by normal electrophoretic migration on SDS-PAGE and amino acid analysis of type I collagen synthesized by patient fibroblasts [102]. Normal folding was demonstrated in a direct intracellular folding assay showing identical rates of collagen folding in patient and normal control fibroblasts [103]. Significantly, Barnes *et al* recently reported an almost complete loss of collagen telopeptide lysyl hydroxylation in patients lacking FKBP65 protein, resulting in a 90% decrease in type I collagen crosslinked into extracellular matrix in deposition assays in culture [103]. These biochemical findings have also been identified in patients with Bruck syndrome, caused by PLOD2 deficiency. Furthermore, clinical observations of patients with FKBP10 mutations suggest that the phenotypic spectrum of mutations in *FKBP10* overlaps with Bruck syndrome. It remains

to be determined if FKBP65 interacts directly with PLOD2 to affect its activity. Alternatively, given Bachinger's findings that FKBP65 directly interacts with type I collagen and inhibits fibril formation *in vitro* [101], it is possible that isomerization of the telopeptide substrate by FKBP65 is required for PLOD2 activity [103]. However, this is less likely since there are no prolyl residues adjacent to PLOD2-modified lysines in the telopeptide primary sequence.

Prolyl 3-Hydroxylation Complex

Trygvasson and colleagues first demonstrated that 3-hydroxyproline and 4-hydroxyproline were synthesized by unique enzymes using crude rat kidney extracts [104]. Renal cortex was used because it is rich in type IV collagen, which contains up to fifteen 3-hydroxyproline residues per alpha chain, in contrast to type I collagen which contains only two 3-hydroxyprolines per trimer. Unhydroxylated type IV procollagen incubated with kidney extracts in the presence of antiserum to prolyl 4-hydroxylase demonstrated both decreased P4H and P3H activity. In the absence of P4H antibody, purified prolyl 4-hydroxylase was unable to catalyze the 3-hydroxylation of prolines on the unmodified collagen substrate. A partially purified protein fraction with prolyl 3-hydroxylase activity was subsequently isolated from chick embryo extracts and was found to have a molecular weight of 160kDa by gel filtration [105]. Since the Gly-Pro-4Hyp sequence is found at numerous triplets and 3Hyp is only found in Gly-3Hyp-4Hyp triplets in collagen, the data suggested that the substrate sequence for 3-hydroxylation is Gly-Pro-4Hyp.

The mechanism of collagen prolyl 3-hydroxylation has, until recently, been poorly understood. In types I and II collagen, this modification system results in 3-hydroxylation

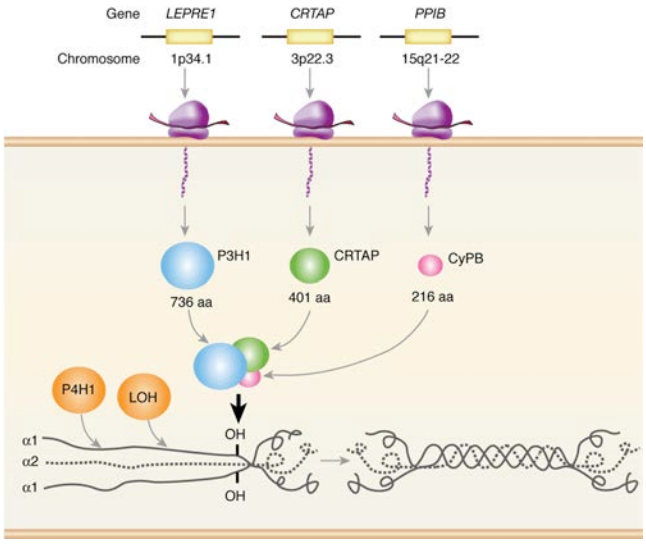

Figure 4. The Collagen Prolyl 3-Hydroxylation Complex. Prolyl 4-hydroxylase (P4H) modifies approximately 90 proline residues in the Y-positions of the Gly-Xaa-Yaa triplets in each alpha chain of type I collagen. Prolyl 3-hydroxylation, a minor post-translational modification, occurs at the X-position residues α1(I) Pro986 and to a lesser extent at α2(I) Pro707 in type I collagen. Prolyl 3-hydroxylation also occurs at α1(II) Pro986 in type II collagen, and at several residues in type V collagen. This modification is performed by the ER-resident collagen prolyl 3-hydroxylation complex, consisting of P3H1, CRTAP and CyPB. Figure from Marini *et al* [2].

of a single X-position residue (3Hyp-4*R*Hyp-Gly) in each of the α1(I) chains at Pro986. By comparison, 3-hydroxylation of several X-position proline residues occurs in types IV and V collagens. Bachinger first isolated the enzyme catalyzing the 3-hydroxyl modification of type I collagen from chick rER extracts and demonstrated that it was part of a complex consisting of prolyl 3-hydroxylase 1 (P3H1), cartilage-associated protein (CRTAP) and peptidyl-prolyl isomerase B/cyclophilin B (PPIB/CyPB) (Figure 4) [106]. Further evidence in support of a complex containing P3H1, CRTAP and cyclophilin B was provided with cosedimentation of these three proteins from affinity column-purified chick rER extracts in a 1:1:1 ratio on a sucrose gradient [107]. Confirmation of the components and function of the prolyl 3-hydroxylase complex was provided by Morello and colleagues, who demonstrated that deficiency of a single component abrogated complex activity [107]. Since P3H1 is able to independently 3-hydroxylate collagen

substrate *in vitro* [106], it is now believed that, *in vivo*, P3H1 serves as the enzymatic component while CRTAP functions as a "helper" protein for the complex.

Cyclophilin B, also known as peptidyl-prolyl cis-trans isomerase B (CyPB/PPIB), is thought to be the major, and possibly unique, peptidyl-prolyl *cis-trans* isomerase involved in collagen folding. Inhibition of CyPB by cyclosporine A (CsA) treatment slowed the rate of collagen folding, resulting in overmodification of lysyl and prolyl residues as determined by delayed electrophoretic migration of collagen alpha chains on SDS Urea-PAGE [108]. It should be noted, however, that CsA is a notably non-specific inhibitor and molecules other than CyPB would have been inhibited. As discussed in detail below, the characterization of the *Crtap* and *P3h1* knock-out mouse models, and null mutations of the *CRTAP*, *LEPRE1* (P3H1) and *PPIB* genes in patients with recessive osteogenesis imperfecta have demonstrated the importance of the proteins in this complex for normal collagen formation and bone development.

Despite the identification 50 years ago of 3-hydroxyproline residues as a minor component of fibrillar collagens, very little interest was shown in the possible function(s) of this modification until the recent discoveries that deficiency of components of the modifying complex caused severe bone disease (see below, "*Recessive OI*"). Type I collagen synthesized in culture by cells lacking components of the complex is overmodified by lysyl hydroxylase, a hallmark feature of delayed folding of the collagen helix [109,110]. Furthermore, the P3H complex has been shown to accelerate the rate of folding of denatured rhodanese and type III collagen, neither of which contains any 3-hydroxyproline residues, due to its PPIase activity towards these substrates [111]. Since cyclophilin B PPIase activity is negligible in the absence of P3H1 and CRTAP, and the

PPIase activity of the complex is inhibited by cyclosporine A *in vitro*, the P3H complex has also been characterized as an isomerase. The complex is believed to function as a collagen chaperone, since *in vitro* it is able to inhibit thermal aggregation of citrate synthase and fibril formation of type I collagen [111], suggesting an additional role in preventing premature aggregation of procollagen within the secretory pathway. Absence of P3H1 or CRTAP greatly reduces the extent to which CyPB can bind denatured collagen *in vitro*, indicating that CyPB-collagen binding depends upon the other components of the complex. These data suggest that an additional role of the complex is to shuttle cyclophilin B to its collagen substrate [112]. Given the distinct intracellular activities attributed to the P3H complex, Bachinger has described it as a post-translational modification complex, isomerase and molecular chaperone [111].

Eyre and colleagues have focused on the potential extracellular role of the 3Hyp modification. They have noted differences in 3Hyp site occupancy of fibrillar collagens from mammals and lower vertebrates as evidence of a functional role for the modification other than as a coincidental mark of intracellular complex chaperone activity [113]. Specifically, the 234-residue interval of 3Hyp sites in type I/V and II/XI heterotypic fibrils is closely aligned to the approximately 237-residue D-periodic spacing of collagen fibrils. In addition, human type III collagen lacks the α1(I) 3Hyp986 residue and *in vitro* can only form microfibrils, while *in vivo* it is found as a minor component of type I/V heterotypic fibrils. These findings suggest a putative extracellular role of 3Hyp residues in fibrillar assembly through the formation of hydrogen bonds between adjacent collagen triple helices [113]. This hypothesis is supported by the crystal structure of a synthetic peptide containing 3Hyp, showing that the hydroxyl group in the Xaa position of the

Xaa-Yaa-Gly collagen triplet points away from the helical axis [114]. In addition, the often-charged residue Xaa position of the collagen helix is the most solvent accessible, potentially allowing for its involvement in a variety of protein-protein interactions. Furthermore, in studies testing self-association of collagen peptides, absence of the hydroxyl group at the α1(I) P986 residue abrogated self-binding [115]. It is interesting that the α1(I) P986 residue overlaps a major ligand binding region (MLBR3) in type I collagen, and substitutions in this region of the α1(I) collagen chain in humans result in lethal bone disease [116]. However, *in vivo* confirmation of the extracellular role of the 3Hyp residues awaits the generation of an animal model that synthesizes collagen void of the modification.

Glycosyltransferases

Following hydroxylation of helical residues by LH1 (PLOD1), some hydroxylysine (Hyl) residues are further modified by the addition of carbohydrates, forming the disaccharide Glc(α1-2)Gal(β1-O)Hyl. Glycosylation of collagen was first described by Grassman and Schleich over 70 years ago [117]. However, it was not until the late 60's that the nature of collagen glycosylation was characterized as either a galactosyl or glucosylgalactosyl group attached via O-glycosidic linkage to hydroxylysine residues [118]. The extent and heterogeneity of hydroxylysine glycosylation in collagens varies to a large degree between collagen types and species. Furthermore, the ratio of monosaccharides and disaccharides in a given collagen type varies with age, as well as within and between tissues [119]. The factors that determine the differences in collagen glycosylation is an area

of active investigation, and cannot be solely explained by the levels of expression of the genes that encode the modifying enzymes.

As with other modifying enzymes that interact with the collagen helical region, the glycosyltransferases require an unfolded molecule since the triple helix conformation inhibits the addition of carbohydrate subunits. Two of these enzymes, which are themselves glycoproteins, have only recently been identified as ER-resident galactosylhydroxylysyl transferases, GLT25D1 and GLT25D2 [120,121]. Both enzymes, which transfer the sugar unit of UDP-galactose to hydroxylysine residues, are unable to glycosylate free hydroxylysine and require a collagen alpha chain peptide fragment length of at least 500-600Da as a substrate [122,123]. The second glycosylation reaction that occurs during collagen synthesis and folding is the attachment of UDP-glucose to galactosylhydroxylysyl residues. The multifunctional Lysyl Hydroxylase 3 (LH3/PLOD3) enzyme is believed to be the galactosyl glucosyltransferase that performs this modification *in vivo*. Although lysyl hydroxylase 3 (LH3) is a multifunctional enzyme possessing lysyl hydroxylase (LH), hydroxylysyl galactosyltransferase (GT) and galactosylhydroxylysyl glucosyltransferase (GGT) activities *in vitro*, knock-down of *Plod3* expression in MC3T3 cells has demonstrated decreased glucosylation of type I collagen, accompanied by increased galactosylhydroxylysine with normal hydroxylysine levels [124]. Further support for LH3 possessing GGT activity *in vivo* comes from studies specifically suppressing this activity in human skin fibroblasts using a polyclonal antibody directed toward the amino terminal GGT domain [83]. Knocking out the *Plod3* (LH3) gene also proved that functional redundancy for LH and GT activity exists *in vivo*, and that it is the loss of GGT activity due to the absence of LH3 that results in a

connective tissue phenotype in mice [85,125]. Recently, Salo and colleagues have identified a connective tissue disorder in a single proband caused by compound heterozygosity for mutations in the *PLOD3* gene [126]. Findings included absence of glucosyl-galactosyl-pyridinium crosslinks in urinary collagen degradation products, as well as decreased GGT activity of proband serum and lymphoblast extracts against a collagen substrate *in vitro*. However, the authors also determined that all three enzymatic activities (i.e., LH, GT, GGT) of recombinant proteins generated from each mutant allele were decreased compared to normal recombinant LH3 [126].

The biological significance of collagen glycosylation remains unclear, although several functions have been proposed. Amino acid analysis of hydrothermal vent worm cuticle collagen suggests that despite its low proline content, a high content (>18%) of threonine residues glycosylated with di- and tri-saccharides of galactose impart a high thermal stability (37°C), compared to the 28-36°C melting temperature of human type I collagen from various tissues [127,128]. Of note, as the extent of glycosylation within a specific collagen type increases, so does the temperature at which the protein denatures. This suggests a role for glycosylation in stabilizing the triple helix by restricting the conformational space of the alpha chains and "locking" them into the folded state. Another possibility is that the saccharides prevent access of water to the peptide backbone, thereby creating a hydrophobic effect preventing unfolding of collagen [129]. It has also been suggested that additional roles of glycosylation of collagen include increased resistance to enzymatic degradation, mediation of interactions with endocytic receptors involved in matrix remodeling, regulation of binding interactions with

noncollagenous extracellular matrix proteins, promotion of angiogenesis, regulation of collagen fibril growth and tissue mineralization (see below).

Heat Shock Protein 47 KDa

Serine proteinase inhibitor clade H (SERPINH1) is an ER resident collagen-specific binding protein. It is commonly referred to as Heat Shock Protein 47 kDa (HSP47) due to its initial characterization as a heat-shock response protein, rather than as a component of the intracellular ER stress response [130]. Expression of Hsp47 occurs in all cells that produce collagens at levels that correspond to the amount of collagens being synthesized. Loss of HSP47 expression in mice is incompatible with survival of developing embryos due to defective maturation of several collagen types [131]. The mechanism by which HSP47 facilitates normal collagen synthesis, secretion and processing has been a controversial topic since its initial characterization in chick embryos, as colligin in mice, and as gp46 in rats [132].

In contrast to the manner in which other chaperones interact with collagen in the ER, HSP47 preferentially binds fully folded triple helical collagen molecules. Using collagen model peptides, Koide and coworkers demonstrated that, *in vitro*, the motif recognized by Hsp47 is X-(Pro/Thr)-Gly-X-Arg-Gly, but only within the context of a folded collagen helical structure [133-136]. Recently Widmer *et al* have presented the crystal structure of HSP47 complexed with a homotrimeric collagen synthetic peptide, characterized as two molecules attached on opposite sides of the collagen helix in a clamp-like fashion [137]. As there exist 27 and 32 putative binding motifs within the helical regions of the α1(I) and α2(I) chains, respectively [134], it is possible that as many as 25-30 evenly spaced HSP47

binding sites exist on the surface of normally folded type I collagen molecule. Binding to collagen occurs in a pH-dependent manner; Hsp47 transiently interacts with folded monomers at neutral pH, which is found in the ER, and dissociates at a pH below 6.3 as is found in the Golgi. *In cellulo*, Hsp47 remains bound to collagen when ER to Golgi trafficking is inhibited with Brefeldin A treatment. However, in cells treated with monensin to inhibit vesicular transport from the trans-Golgi no association of Hsp47 and collagen was detected. These findings demonstrate binding of Hsp47 to collagen in the ER, but dissociation from collagen molecules in the ERGIC and/or *cis*-Golgi compartment [138]. Upon dissociation from collagen, HSP47 recycles back to the ER via KDEL receptors [139].

Procollagen is thermally unstable and denatures at physiologic conditions in the absence of HSP47 [140], and loss of functional HSP47 results in collagen that is susceptible to proteolytic cleavage [131,141]. Thus it would appear that HSP47 is essential for stabilization of the folded collagen helical conformation. In cells lacking expression of Hsp47, Type I collagen shows normal helical modification and slower secretion kinetics, resulting in a dilated ER due to accumulation of collagen. Other cellular findings in the Hsp47 knockout mouse include defective N-propeptide processing of secreted procollagens [142]. The requirement of both a conserved recognition sequence and proper substrate conformation for N-propeptidase cleavage suggests that, in the absence of HSP47, collagen helical folding cannot be stabilized. These findings are in addition to observations that Hsp47 inhibits fibril formation *in vitro* [143]. Since collagen is thought to be transported as (aggregate structures) through the Golgi stack by a process of cisternal maturation [144], a role for Hsp47 in preventing intracellular collagen aggregation seems

plausible. Therefore, the cumulative data supports a function for Hsp47 in assisting collagen molecules by inhibition of helical denaturation and preventing premature aggregation during transport through the secretory pathway.

Collagen Trafficking, Secretion and Processing

ER Cargo Selection and Vesicular Export

Once post-translational modification and folding is complete, the procollagen molecules must be transported through the secretory pathway for deposition into the pericellular space. In the conventional secretory pathway model, proteins destined for secretion are sorted at ER exit sites (ERES) and loaded into COPII-coated vesicles for transport to the Golgi [145]. In general, the process of ER-derived vesicle formation begins with insertion of Sec12(GEF)-activated Sar1-GTP into the outer leaflet of the ER membrane, recruiting dimerized Sec23/Sec24, which has GAP activity. Binding and concentration of cargo is thought to be mediated by Sec24. This complex forms the first layer of coat proteins on the membrane, referred to as the prebudding complex. Recruitment of Sec13/Sec31 dimers follows, with Sec31 interacting directly with Sec23, completing assembly of typical COPII-coated vesicles that are 60-90nm in diameter. Fibrillar procollagens, however, are greater than 300nm in length, suggesting that transport proceeds by a COPII-independent mechanism, or alternatively, that COPII carrier size can be specifically regulated under special circumstances to accommodate larger cargo. This debate has led to the recent identification of novel roles of previously

known proteins involved in trafficking, as well as identification of collagen-specific cargo adaptors that enable formation of "megacarriers" (Figure 5) [146].

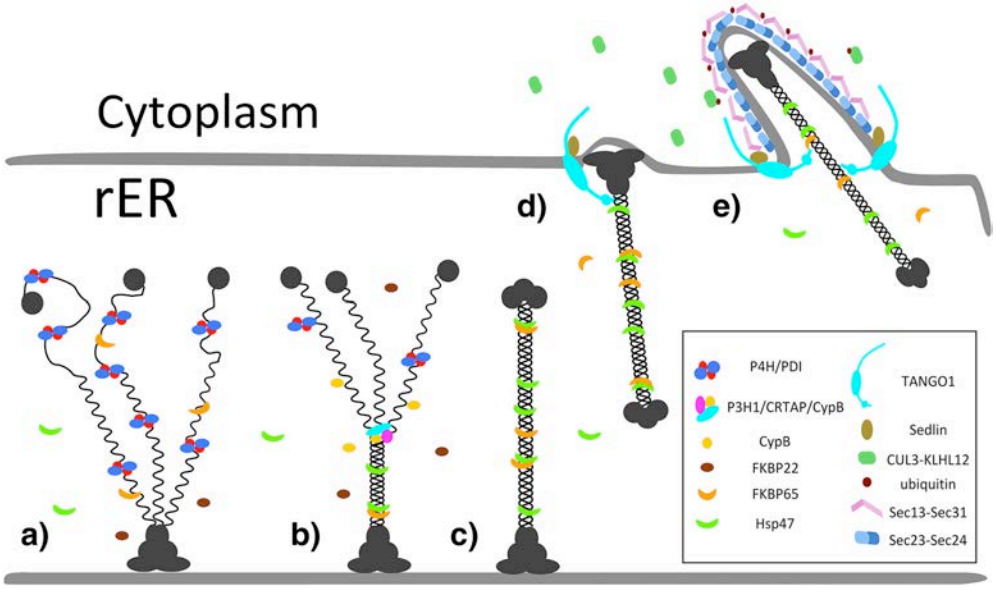

Figure 5. Procollagen Trafficking. Procollagen molecules concentrate at ER exit sites in preparation for transport to the Golgi complex. The formation of COPII-coated "megacarriers" is mediated by the interaction of collagen-specific cargo receptors (TANGO-like proteins), COPII component Sec23, SEDL and Sar1. Transport vesicle growth is regulated by an incompletely understood mechanism involving ubiquitylation of SEC13/31 by the CUL3/KLHL12 complex. HSP47 and CyPB accompany the procollagen molecule from the ER into the *cis*-Golgi compartment. Figure from Ishikawa and Bachinger [8].

The first clues to how collagens are packaged and transported evolved from experiments that took advantage of the fact that ER export of collagens is inhibited under scorbutic conditions [147]. In the absence of ascorbic acid, a cofactor for prolyl hydroxylases, collagen is underhydroxylated and retained in the ER. Upon addition of ascorbic acid, collagen is transcriptionally and translationally upregulated, while folding and secretion efficiency is increased. This "pulse" technique has allowed several investigators to follow collagen molecules through the secretory pathway. Using TEM analysis, Pacifici and Iozzo were the first to observe remodeling of the RER upon upregulation of collagen synthesis [148]. These investigators noted morphologic changes in the size and shape of specific regions of RER; rather than uniform, oval-shaped cisternae,

some portions of the organelle developed into large, saccular structures containing concentrated collagen. Interestingly, using immunofluorescence microscopy Miranov *et al* noted that concentration of procollagens occurs in regions of the ER that are distinct from the sites where vesicular stomatitis virus glycoprotein (VSVG) concentrates [149]. Since VSVG binds directly to the COPII subunit Sec23/24 [150], the authors hypothesized the existence of a collagen-specific cargo receptor for recruitment to the ERES.

Recent work in *Drosophila* has uncovered novel components of the cellular secretory pathway invloved in collagen trafficking [151]. In this study, investigators utilized a genome-wide library of 22,000 dsRNAs to screen for genes required for protein secretion in cells stably-transfected with expression constructs encoding horseradish peroxidase (HRP) driven by an inducible promoter. Knock-down of expression of genes required for secretion was performed by screening for decreased peroxidase activity by chemiluminescence. Genes with roles known to indirectly affect secretion were then filtered out, including those involved in apoptosis, mitosis, transcription, translation, and metabolism. Intracellular localization studies by immunofluorescence validated at least 10 genes whose knockdown resulted in Golgi disruption. In humans, these proteins correspond to TANGO1 (Transport ANd Golgi Organization) and cTAGE1-8 (cutaneous T-cell lymphoma-associated antigen).

TANGO1 (*MIA3*), a 1907-residue protein, colocalizes to ER structures containing Sec16L, a component of endoplasmic reticulum exit sites (ERES) in mammalian cells [152]. It has an N-terminal lumenal portion that potentially interacts with collagen through a Src homology 3 (SH3) domain that recognizes proline-rich sequences, followed by a transmembrane domain, and a cytoplasmic coiled-coil and C-terminal proline-rich

domain (PRD) [153]. The PRD of TANGO1 was demonstrated to directly interact with COPII components Sec23/24 but not Sec31 by yeast 2-hybrid, providing further evidence of its ERES localization. However, the connection to collagen was made by identification of type VII collagen as a binding partner by mass spectrometric analysis of TANGO1 immunoprecipitates from HeLa extracts. Furthermore, given the ERES localization and the observation that knockdown of TANGO1 expression by siRNA in human fibroblasts and keratinocytes led to a specific reduction in type VII collagen secretion, it was concluded that TANGO1 was a type VII collagen-specific cargo receptor [152]. Although a direct *in vivo* interaction between the SH3 domain of TANGO1 and collagens other than type VII has not been reported, *Mia3/Tango1*-null mice are perinatal lethal due to global defects in collagen secretion [154]. Increased intracellular accumulation in culture and reduction of extracellular matrix deposition in tissues of several collagen types was observed, including interstitial fibrillar collagens (I, II, III) in murine fibroblasts, and basement membrane collagens (IV) in endothelial cells.

Follow-up analyses of TANGO-like proteins have revealed additional insights into collagen export from the ER [155]. Costaining with Sec31A, but not with ERGIC53 or the COPI component β-COP, determined that cTAGE5 also localized to ERES. Although cTAGE5 is structurally similar to TANGO1, it lacks an SH3 collagen binding domain. Immunoprecipitation of either cTAGE5 or TANGO1 pulls down the other protein, suggesting the two form a complex at the ERES; binding *in vitro* occurs stoichiometrically in a 1:1 ratio. This interaction occurs between the corresponding coiled-coil motifs in each protein, but does not involve the PRD domain of TANGO1, since constructs lacking this domain are still able to from a complex. Yeast two hybrid

assays have shown that, similar to TANGO1, the PRD domain of cTAGE5 directly interacts with Sec23/24. Furthermore, knockdown of cTAGE5 expression results in decreased secretion and increased intracellular accumulation of collagens. As neither TANGO1 nor cTAGE5 have been observed at locations other than ERES, a model has been proposed suggesting that the primary role of these components is to block recruitment of Sec13/31 long enough to allow continued growth of collagen-containing COPII carriers [156]. One can also hypothesize that further specificity of large cargo is obtained by TANGO1 interactions with different cTAGE partners, however no evidence for this yet exists.

Another protein that has recently been shown to interact with TANGO1 is Sedlin (TRAPPC2), encoded by *SEDL* [146]. Mutations in *SEDL* result in Spondyloepiphyseal dysplasia tarda (SEDT), which phenotypically overlaps the osteochondrodysplasias caused by structural defects in type II collagen, encoded by *COL2A1* [157-159]. The pathophysiology of both disorders has been linked to defects in secretion of components of the extracellular matrix [160]. In chondrocytes, *SEDL* knockdown blocks ER export of type II collagen, but not VSVG. Selective ER retention of type I collagen by *SEDL* RNAi is also achieved in fibroblasts, while export of noncollagenous proteins such as CD8α, albumin and α1-antitrypsin remains unaffected. Co-immunoprecipitation confirmed a direct interaction between TANGO1 and Sedlin in these cells, explaining the selective effect on collagen secretion. Overexpression of TANGO1 increases Sedlin recruitment to the ERES and increased collagen export from the ER. Conversely, TANGO1 depletion decreases recruitment of Sedlin to the ERES. These findings prompted investigators to further characterize the role of Sedlin in collagen export from the ER [146]. Western

analysis demonstrated increased levels of Sar1-GTP in Sedlin-depleted cells, findings that were duplicated in cells depleted of the TRAPP-I complex component Bet3 [161]. Given these findings and since immunofluorescence detects overlapping signals for SEDL and Bet3 at ERES, investigators have hypothesized that, as a TRAPP-I component, Sedlin directly regulates activity of Sar1. This hypothesis is plausible given that *in vitro* Sedlin is the only TRAPP-I subunit that binds to Sar1, and Sedlin immunoprecipitates from cell lysates with Sec23 only in the presence of Sar1. In addition, fibroblasts from patients with SEDT reveal budding collagen-containing vesicles at the ERES that are reduced in size, suggesting TANGO1 and TRAPP-I coordinate growth of collagen megacarriers at ERES by regulation of Sar1 activity. Further investigation should aim to characterize the nature of the interaction between TANGO1 and the SEDL-Sec23-Sar1 complex.

While investigating the role of ubiquitylation enzymes in stem cell division, Jin and colleagues observed that knockdown of the ubiquitin ligase CUL3 caused mouse ES cells to become compacted and nonproliferative. This phenotype was reversed by growth on exogenously derived extracellular matrix [162]. Absence of cell surface integrins was noted. The preliminary findings suggesting a role for ubiquitylation in integrin-mediated cell-matrix interactions, since in the absence of surrounding extracellular matrix, cells endocytose plasma membrane-bound integrin receptors [163]. Mass spectrometry of affinity purified proteins capable of binding CUL3 revealed KLHL12 as the partner of CUL3 in ubiquitin ligase. Additional affinity purification procedures to isolate substrates of CUL3-KLHL12 identified the COPII components SEC13/31 as binding partners. Immunoprecipitation further refined this interaction as occurring between KLHL12 and

SEC31. These surprising observations correlated with studies colocalizing SEC13/31 and KLHL12 at ERES, as well as COPII coats of forming large vesicles. More importantly, transfection of cells with mutant forms of KLHL12 resulted in loss of monoubiquitylation of SEC31, and absence of collagen-containing large vesicle formation at ERES. Loss of ubiquitylation of COPII components results in accumulation of type IV collagen, with no effects on secretion of other cargoes such as fibronectin or EGF receptor. How ubiquitylation of SEC31 regulates vesicle size is unknown. Since ubiquitylation is known to target substrates to the proteasome, one possible explanation is that degradation of outer components of the COPII coat allows for continued growth of vesicles. Alternatively, modification of SEC31 may recruit an unidentified effector that delays COPII budding or promotes polymerization of vesicle coats [162].

Post-ER Trafficking of Procollagens

Little is known about intra-Golgi transport of procollagens. Based on serial sectioning and immunolabeling of procollagens released from scorbutic block, Miranov *et al* have described flattened and elongated tubular saccules, large enough to contain 300nm long procollagen trimers, protruding from the ER [149]. Upon separation from the ER, these procollagen-labeled structures were found to consist of multiple saccules that were partially stacked and devoid of ribosomes. These investigators proposed that multiple procollagen-containing saccular structures from the ER fuse to form new *cis*-Golgi cisternae. By following a synchronized wave of procollagen with antibodies recognizing helical collagen, Bonfanti *et al* have observed procollagen trafficking from the ER through the Golgi apparatus. In this study collagens were released from the ER of

scorbutic fibroblasts in a synchronized wave of migration into and through the Golgi and followed by immunoelectron microscopy. As expected, COP-I and clathrin coated buds and vesicles were observed in the vicinity of Golgi cisternae. However, no collagen-labeled vesicles were observed to separate from the Golgi stack during maturation of collagen-containing cisternae [144]. Thus the small amount of evidence so far obtained on trafficking of collagen through the Golgi apparatus is consistent with a model for cisternal maturation, in which cargo progresses in an anterograde direction without exiting the cisternae in which it is contained.

In both corneal epithileal cells and tendon fibroblasts, packaged collagen molecules have been observed to be transported to the cell surface and released into the pericellular space via "collagen condensation vacuoles" [164]. These collagen-filled vesicles were also found to be fused with the cell membrane and open to the cell exterior with portions of their contents protruding into the extracellular space. The vesicular membranes contained on their cytoplasmic surface an unknown coat material, similar to that described on coated vesicles by Roth and Porter, and later described as clathrin [165,166]. There is some evidence to suggest that the late secretory pathway is involved in the initial organization of collagen fibrils [167]. Using a combination of three-dimensional electron microscopy (serial section reconstruction) and biochemical approaches, Kadler's group has shown that during embryonic development processing can occur within Golgi to PM carriers (GPCs) co-transporting procollagens with propeptidases. Furthermore, these GPCs were found to extend to the plasma membrane, where they protruded from the cell surface. These plasma membrane extensions, termed "fibripositors", were observed to deposit pre-assembled collagen fibrils in alignment with the existing matrix [168]. However, these

findings have only been observed in murine embryonic tendon fibroblasts and likely represent a developmental and tissue-specific response to the requirement for maximal matrix deposition in the growing organism. It was not noted whether clathrin associated with the cytoplasmic surface of fibropositor membranes. At later stages of development procollagen is secreted into the pericellular environment and processed for lateral accretion into growing fibrils.

Procollagen Processing

Upon secretion into the extracellular matrix, the N- and C-terminal globular domains of procollagen are cleaved by specific propeptidases. Since procollagen molecules are secreted as highly concentrated bundles into the pericellular space, the propeptides provide solubility to the molecule while ensuring steric constraints that prevent premature fibril formation. In fact, persistence of type I collagen N- and C-propeptides in tissues results in disorganized collagen fibrils with abnormally decreased diameters [169,170]. However, immunoelectron microscopy studies have provided evidence that suggests a normal role for the presence of propeptides of types III and V collagens on heterotypic fibril surfaces in maintaining fibril diameters [171,172].

The first evidence for the existence of processed fragments of type I procollagen came from observations that fragile skin in cattle with dermatosparaxis was caused by a defect in the cleavage of N-terminal portions of type I procollagen [173]. Since cell extracts from unaffected cattle were able to remove the retained N-terminal portions of procollagen from the affected cattle, this defect was attributed to absence of a specific N-propeptidase. Since this initial study, it has been found that three procollagen N-

proteinases from the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) family of metalloproteinases are involved in specific cleavage of type I procollagen [174]. The prevalent type I procollagen N-propeptidase, also known as ADAMTS-2, cleaves the Pro-Glu and Ala-Glu bonds in pro-α1(I) and pro-α2(I) N-telopeptide regions, respectively [175]. Procollagen N-propeptidase requires a substrate with a correctly folded hairpin conformation, as demonstrated by the loss of cleavage of procollagens that are either heat-denatured or misfolded due to structural defects [176,177]. Misfolding of the cleavage site occurs due to splicing defects leading to absence of the cleavage site or substitutions in the first 90 residues of the helical region of type I collagen. In these cases missense and splicing mutations in α1(I) or α2(I) result in retention of N-propeptides and subsequent incorporation of pN-collagen into fibrils, leading to thinner diameter fibrils and causing Ehlers-Danlos types VIIA or VIIB, respectively, or Osteogenesis imperfecta/Ehlers-Danlos (OI/ED) phenotype [177,178]. Although ADAMTS-3 and -14 can also process the amino propeptide of type I procollagen, their expression is insufficient to compensate for the absence of ADAMTS-2 in cases of type VIIC Ehlers-Danlos syndrome, in which transverse images of patient dermal fibrils take on the appearance of "heiroglyphs" [169,179,180].

The procollagen C-propeptidase was termed BMP1 (bone morphogenetic protein 1) because it was initially copurified from extracts of demineralized bone with other proteins that could induce ectopic bone formation when implanted into soft tissues of rodents [181]. In addition to BMP1, procollagen C-proteinase activity is attributed to the alternatively spliced BMP1 mammalian tolloid protein (mTld) and two related homologs known as mammalian tolloid-like 1 and 2 (mTll1 and mTll2) [182-185]. The activity of these

enzymes has been shown to be increased up to ten-fold by the procollagen C-proteinase enhancer (PCOLCE1 and 2) proteins, which bind to and induce conformational changes in the procollagen substrate [186-188]. There is no dependence on substrate conformation for BMP1/type I procollagen C-propeptidase, which processes the C-propeptide by cleavage at the Ala-Asp bond in the C-telopeptides of the pro-α1(I) and pro-α2(I) chains [186,189]. However, mutations that alter the peptide sequence of the substrate cleavage site in either of the alpha chains result in retention of C-propeptides. In kinetic experiments using purified C-propeptidase against type I procollagen substrate *in vitro*, Njieh *et al* demonstrated that propeptide cleavage occurs sequentially, with both α1(I) propeptides being processed before the α2(I) propeptide [190]. Sequential cleavage is supported by *in vivo* data showing that patient fibroblasts with mutations of the α1(I) cleavage site produce very little mature collagen. However, patient fibroblasts with α2(I) cleavage site mutations produce collagen with mature α1(I) chains, but not mature α2(I) chains [191]. In these patients, incorporation of pC-collagen into extracellular matrix causes smaller diameter fibrils with irregular cross-sections, leading to an osteogenesis imperfecta phenotype with unusually high bone mass [191]. A similar phenotype, characterized by high bone density with recurrent fractures, was found in a consanguineous pedigree with a recessive missense mutation in *BMP1* [192]. Defective C-propeptide processing in dermal tissue of *Bmp1*-knockout mice results in collagen fibrils that are thinner and have a "barbed wire" appearance, presumably due to extrusion of retained C-propeptides from the fibril surfaces [193]. It is hypothesized that incorporation of partially processed procollagens with retained C-propeptides affects the normal mineralization process by increasing intrafibrillar spacing or acting as nucleators of mineralization [191]. These

studies also demonstrate that, despite the expression of three unique genes encoding four tolloids, there is an absence of redundancy in the C-propeptidase activity of these enzymes.

Beyond the well-described roles of propeptides in chain association, molecular solubility and prevention of premature fibrillogenesis, there is some experimental evidence of non-triple helical fragments of procollagen in regulating collagen expression by an inhibitory feedback mechanism. Weistner *et al* first described an inhibitory effect of type I and III procollagen N-propeptides on translation of type I procollagen mRNA in a cell-free system [194]. Bornstein's lab demonstrated that transfection of cDNA expressing type I procollagen N-propeptide reduces intracellular and secreted levels of type I procollagen, but not fibronectin, thrombospondin or SPARC in fibroblast cultures [195]. Cell fractionation studies have also shown that decreased expression of collagen transcripts coincides with the appearance of radiolabeled C-propeptides in the nucleus of fibroblasts in culture [196]. These findings are supported by the identification of propeptides bound to the plasma membranes of fibroblasts and endothelial cells, with subsequent internalization by receptor-mediated endocytosis [197,198]. Furthermore, a mechanism for this process has been proposed following the identification of type I procollagen propeptide-integrin interactions in binding assays, in addition to the use of antibodies to $\alpha 1\beta 1$ and $\alpha 2\beta 1$ integrin subunits in blocking fibroblast attachment to C-propeptide-coated surfaces in culture [199,200]. However, it is not yet clear how internalized propeptides can gain access to the cytoplasm or the nucleus to perform their inhibitory roles.

Collagen Fibrillogenesis

Fibril Structure

Collagen molecules that are secreted into the pericellular space self-assemble into higher order structures known as fibrils, which provide mechanical support, confer resistance to compressive forces within tissues, and mediate cell attachment to extracellular matrix. In bone, skin and tendon, collagen fibrils are comprised of a combination of unique collagen types, including type I, III, V and XI collagens, and are therefore considered heteropolymers [171,201]. As viewed by electron microscopy, heterotypic type I collagen-containing fibrils appear cylindrical with a cross-striated periodic banding pattern (Figure 6). This appearance is a consequence of lateral assembly of collagen molecules (monomers) with a longitudinal displacement of approximately

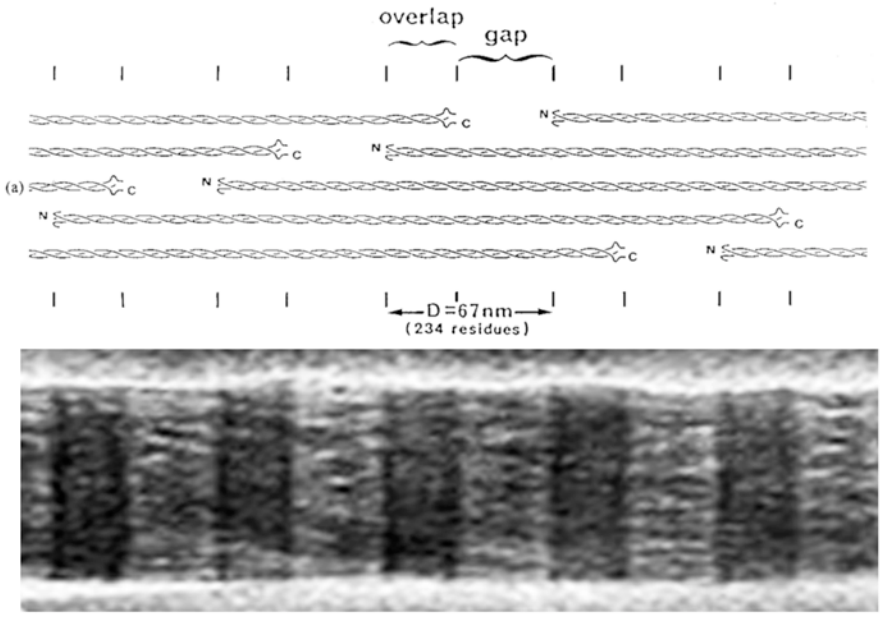

Figure 6. Collagen Fibril Structure. Upon secretion into the extracellular matrix, mature collagen molecules self-assemble into a quarter-stagger array. Visualization by electron microscopy reveals the cross-striated staining pattern of fibrils, referred to as the collagen D-period. Adapted from Kadler *et al* [6].

234 amino acids, commonly referred to as the "quarter-staggered array"[202]. For every five laterally aligned collagen molecules, this displacement creates alternating segments where all five monomers overlap and gap regions where only four molecules overlap. These repeating gap and overlap regions are approximately 67nm apart and comprise the collagen D-period. Although fibril formation is a self-assembly process driven by electrostatic and hydrophobic forces [203], regulation of longitudinal and lateral growth are separately achieved by several factors, resulting in a diverse array of fibrils that vary within and between tissues with respect to diameter, length and spatial organization.

Fibril Nucleation

Since the 1950s, investigators have sought to reproduce the process of collagen fibril formation *in vitro* to elucidate the mechanism of fibrillogenesis. Early *in vitro* experiments relied on the light scattering properties of soluble collagen monomers versus those of insoluble, aggregated collagens in solution [204,205]. For instance, Gross and colleagues first observed that the opacity of a collagen solution increased with heating but decreased with subsequent cooling in a time-dependent manner [206]. Increased opacity correlated with aggregated collagen that formed striated fibrillar structures similar to those observed in collagen-rich tissues by microscopy. These findings were quickly followed by those of Bensusan and Hoyt, who noticed that while the rate of opacity (i.e., fibril formation) could be influenced by collagen concentration, ionic strength and pH of the solution, there was always a preceding lag period for the reaction [204]. The "lag phase" was believed to entail a nucleation event whose precise nature was unknown due to limitations of the turbidity method [207]. However, the presence of a lag phase and the

requirement of a critical concentration for *in vitro* fibril formation suggested the existence of a template "microfibril" [208]. Shortening of the lag phase and increasing the rate of fibril growth could be achieved in successive rounds of fibril formation and dissolution, presumably due to the presence of trace amounts of covalently bonded microfibrils. Indeed, these unbanded template microfibrils with diameters less than 15nm were isolated from lag phase reactions and identified by electron microscopy [209,210]. Thus the nucleation event in fibrillogenesis was not driven by the formation of covalent bonds, but instead involved lateral alignment of monomers guided by electrostatic, hydrophobic forces and hydrogen bonding [211]. It is during the ensuing exponential growth, termed the propagation phase, that fibrils with characteristic D-periods are formed [212].

Regulation of Fibril Architecture

Comper and Veis first noted an effect of the presence of specific residues within collagen monomers on fibril formation [213]. Pepsin-resistant collagen helices, which are shorter than naturally processed collagens due to proteolysis of the entire telopeptide sequence, have a longer lag phase and higher critical concentration required for nucleation of fibril growth. Using synthetic peptides with sequences corresponding to the telopeptide regions, Prockop and Fertala were able to completely inhibit self-assembly of collagen I into fibrils *in vitro*, as detected by dark field light microscopy and PAGE analysis [214]. Furthermore, far western analysis demonstrated that peptides with the α2(I) C-telopeptide sequence were found to bind directly to collagen I monomers in the region corresponding to residues 781-794. These findings were later confirmed by *in vitro* fibrillogenesis studies using collagens synthesized by fibroblasts from a patient whose

mutation resulted in deletion of these same residues. Fibrils formed from type I collagen lacking residues 766-801 of the α1(I) chain were longer than fibrils comprised of only normal control type I collagen [215]. It was first noted by Kadler and colleagues that fibrils formed *in vitro* contained three distinct domains with two major sites for collagen binding: blunted ends that lacked growth, a central shaft with slow lateral growth, and a rapidly-growing tapered end [216]. These data support a model for collagen fibril growth as a two-step process that determines the rates of lateral and longitudinal accretion of collagen monomers [217]. In this model, disruption of either parameter affects the other; hence, delayed lateral accretion due to structural defects in collagen monomers results in increased length of fibrils [215]. There is additional evidence for a role of lateral interactions between collagen monomers in determining fibril diameters. Type I collagen, which forms 50-100nm diameter fibrils, has a slower rate of fibril growth than types II and III collagens, which form 10-30nm diameter fibrils both *in vitro* and *in vivo* [218]. These differences in fibril size are likely mediated by unique lateral interactions between heterotrimeric type I collagen monomers compared to homotrimer types II and III collagens.

Other proteins, including small leucine-rich proteoglycans (SLRPs) and minor collagens are known to regulate formation of type I collagen heterotypic fibrils. For instance knockout of the *Col5a1* gene in a lethal murine model of Ehlers-Danlos Syndrome results in absence of type I collagen-containing fibrils in the embryonic mesenchyme despite normal secretion of type I collagen, hinting to a possible role in fibril nucleation *in vivo* [219]. Haploinsufficiency of either *Col5a1* or *Col11a1* results in type I collagen-containing fibrils with abnormally large diameters and disorganized

surfaces, supporting a role in the regulation of lateral growth of fibrils [220]. The collagen-binding proteoglycans decorin and biglycan interfere with type I collagen incorporation into fibrils, resulting in reduction of fibril diameters *in vitro* [221]. These findings have been confirmed *in vivo*, in that mice lacking decorin or biglycan expression have larger diameter, disorganized type I collagen-containing fibrils in tendon, skin and bone [222]. Space-filling representations of several proteoglycans model horseshoe-shaped SLRPs binding collagen molecules via their concave surface, thereby blocking lateral association of additional collagen molecules [223]. Conversely, lumican- and fibromodulin-deficient mice exhibit fibrils with smaller diameters [222]. Furthermore, a study of *in vitro* reconstitution of fibrils using differentially glycosylated collagens extracted from various tissues suggested that increased extent of modification results in longer lag phase, slower growth rates and thinner diameters of fibrils, possibly due to lateral steric hindrance or interference of hydrophobic or electrostatic interactions [224]. These data demonstrate that fibril formation is regulated by specific interactions of type I collagen monomers with each other, as well as with other components of the extracellular matrix.

Crosslinking

The formation of covalent intermolecular crosslinks between collagen monomers is the final step of collagen biosynthesis. Crosslink formation not only stabilizes collagen fibrils but also provides tensile strength to the extracellular matrix (Figure 7). This reaction is now known to be catalyzed by the extracellular metalloenzyme lysyl oxidase (LOX), a monoamine oxidase that converts the side-chain amine group of specific lysine

and hydroxylysine residues to aldehydes [225]. LOX is unable to oxidize free lysyl residues or lysyl residues within unhydroxylated collagens, which cannot form fibrils. Instead,

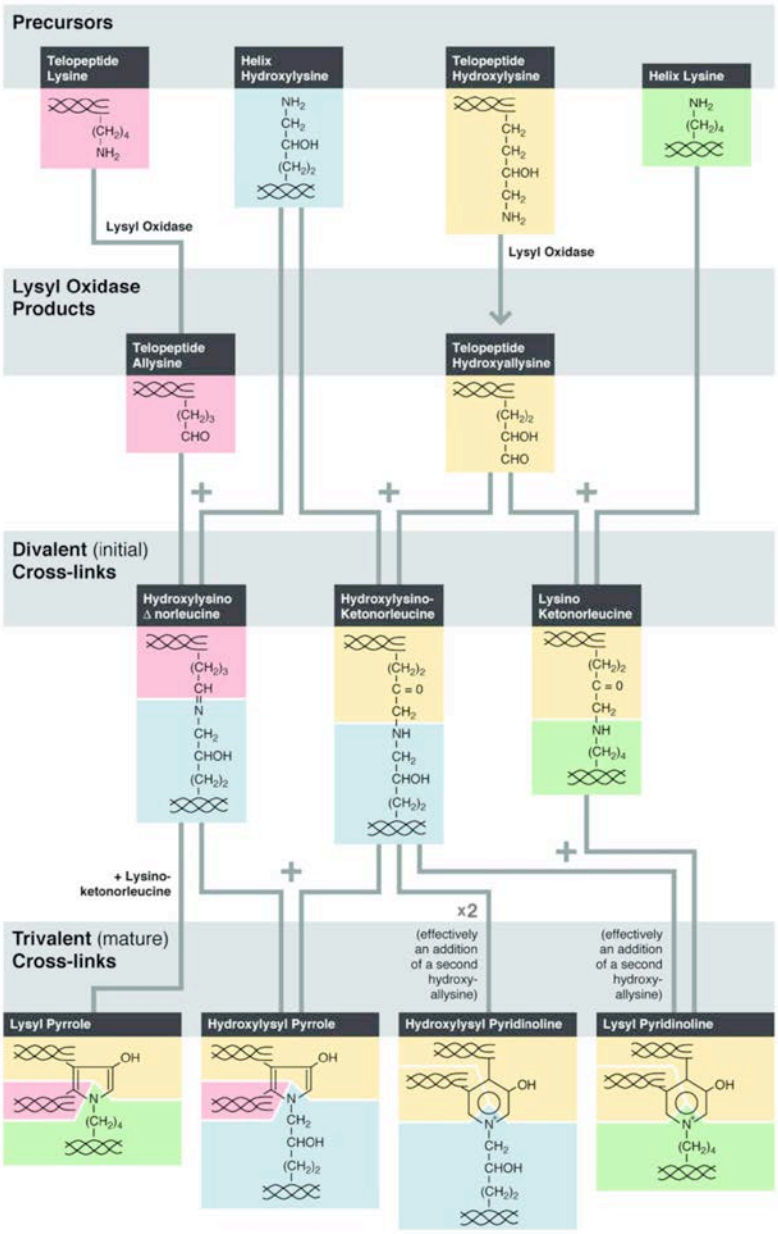

Figure 7. Lysyl Oxidase-Mediated Collagen Crosslinking. Oxidative deamination of lysyl residues by lysyl oxidase is followed by a condensation reaction between specific telopeptidyl and helical lysyl residues. The structure of crosslinks is tissue-specific and is also determined by hydroxylation status of the lysyl residues. In bone, some divalent crosslinks (HLNL and DHLNL) can mature to trivalent crosslinks (LP and HP). Figure from Eyre and Weis [11].

oxidation of these residues occurs on collagen monomers at the surface of forming fibrils, or those contained within loosely packed aggregates, but not on mature native fibrils isolated from tissues [226]. The aldehyde groups within the resulting allysine and hydroxyallysine residues then condense with corresponding aldehydes on adjacent collagen monomers [227-230]. The majority of intermolecular covalent bonds have been demonstrated by mass spectrometry of CNBr and tryptic peptides to occur between lysyl (or hydroxylysyl) residues located at distinct positions in the telopeptide and helical regions of mature type I collagen: α1(I)Ntelo9$^{LYS/HYL}$ x α1(I)930^{HYL}, α1(I)Ctelo16$^{LYS/HYL}$ x α1(I)/α2(I)87^{HYL} and α2(I)Ntelo5 $^{LYS/HYL}$ x α2(I)933^{HYL}. These intermediate divalent aldol condensation products form dehydrolysinonorleucine (Δ-LNL), dehydrohydroxylysinonorleucine (Δ-HLNL) and dehydrodihydroxy-lysinonorleucine (Δ-DHLNL) reducible crosslinks that contain Schiff base double bonds. Further maturation of crosslinks is not enzyme-catalyzed, but is instead spontaneous and reflects the local environment surrounding the crosslink site. In general, mature crosslink formation follows two pathways, depending on the tissue and types of complementary lysyl residues involved [231]. In skin, crosslink formation predominantly involves allysine, nonglycosylated hydroxylysine and/or histidine residues. Along with LNL, HLNL and DHLNL, dermal tissue contains an additional reducible crosslink, tetravalent histidinohydroxy-merodesmosine (HHMD), as well as a nonreducible trivalent histidinohydroxylysinonorleucine (HHL), which forms at α1(I)Ctelo16$^{LYS/HYL}$ x α1(I)87^{HYL} x α2(I)92^{HIS} [232]. The hydroxyallysine route for crosslink formation is found preferentially in load-bearing tissues such as bone, tendon (type I collagen), and cartilage (type II collagen). For the hydroxyallysine route, the formation of one trivalent

nonreducible 3-hydroxypyridinium (lysyl or hydroxylysyl pyrodinolines) crosslink from two pre-existing reducible divalent crosslinks occurs in most tissues except skin. The absence of this type of mature crosslink in dermal tissue may be due to its susceptibility to UV-induced bond alterations [233]. The majority of hydroxypyridinium crosslinks occur at α1(I)Ctelo16$^{LYS/HYL}$ x α1(I)Ctelo16$^{LYS/HYL}$ x α1(I)/α2(I)87^{HYL}, and less frequently at α1(I)Ntelo9$^{LYS/HYL}$ x α1(I)Ntelo9$^{LYS/HYL}$ x α1(I)930^{HYL} [234].

Bone

The collagen crosslinked into the extracellular matrix of bone tissue functions as a scaffold for the development of a multifunctional mineralized tissue. Together, the protein and mineral phases of bone tissue interact to form a composite material with mechanical, physical and biological properties that exceed its constituent components. In addition to giving structural support for the body and protection to organs, the bones of the skeleton also permit movement by acting as levers for muscles, maintain mineral homeostasis, serve as reservoirs for growth factors and cytokines, and provide an environment for hematopoiesis in marrow spaces. Furthermore, these functions are dependent on the material properties and strength of the tissue, determined by its architectural organization, mineral extent and structure, organic content and the rate of tissue turnover [235].

The Process of Bone Formation

The bones of the skeletal system are formed through two distinct mechanisms, both of which involve initial condensation of mesenchymal osteochondral progenitor cells. The first type of bone formation process, intramembraneous ossification, occurs in the flat bones of the craniofacial skeleton and ribs. This process involves direct differentiation of mesenchyme-derived precursor cells into osteoblasts. In contrast, formation of the long bones, including humeri, tibiae and femurs, proceeds by a two-step process known as endochondral bone formation, where development of a cartilagenous anlage by chondrocytes precedes the development of ossified tissue (Figure 8).

During embryonic development mesenchymal precursors aggregate at sites of future long bone development. These bipotential osteochondral progenitors differentiate into chondrocytes, which organize into columnar layers and secrete components of the cartilage extracellular matrix, including type II collagen and aggrecan. By the time formation of a cartilage "model" is complete, chondrocytes with unique proliferative capacity become organized along the longitudinal axis into distinct zones within the matrix. At the center, less proliferative cells become hypertrophic, express type X collagen and mineralize the surrounding cartilaginous matrix. Formation of this central ossified region, termed the primary center of ossification, is followed by vascular invasion of the mineralized matrix, allowing for the introduction of osteoclastic and osteoblastic progenitor cells. Osteoclasts secrete proteases that degrade the type II collagen-containing cartilaginous matrix, while osteoblast precursors differentiate and replace the degraded cartilage with a type I collagen-rich matrix. The primary center of

ossification then expands towards the ends of the cartilage model as cartilage is replaced by bone matrix.

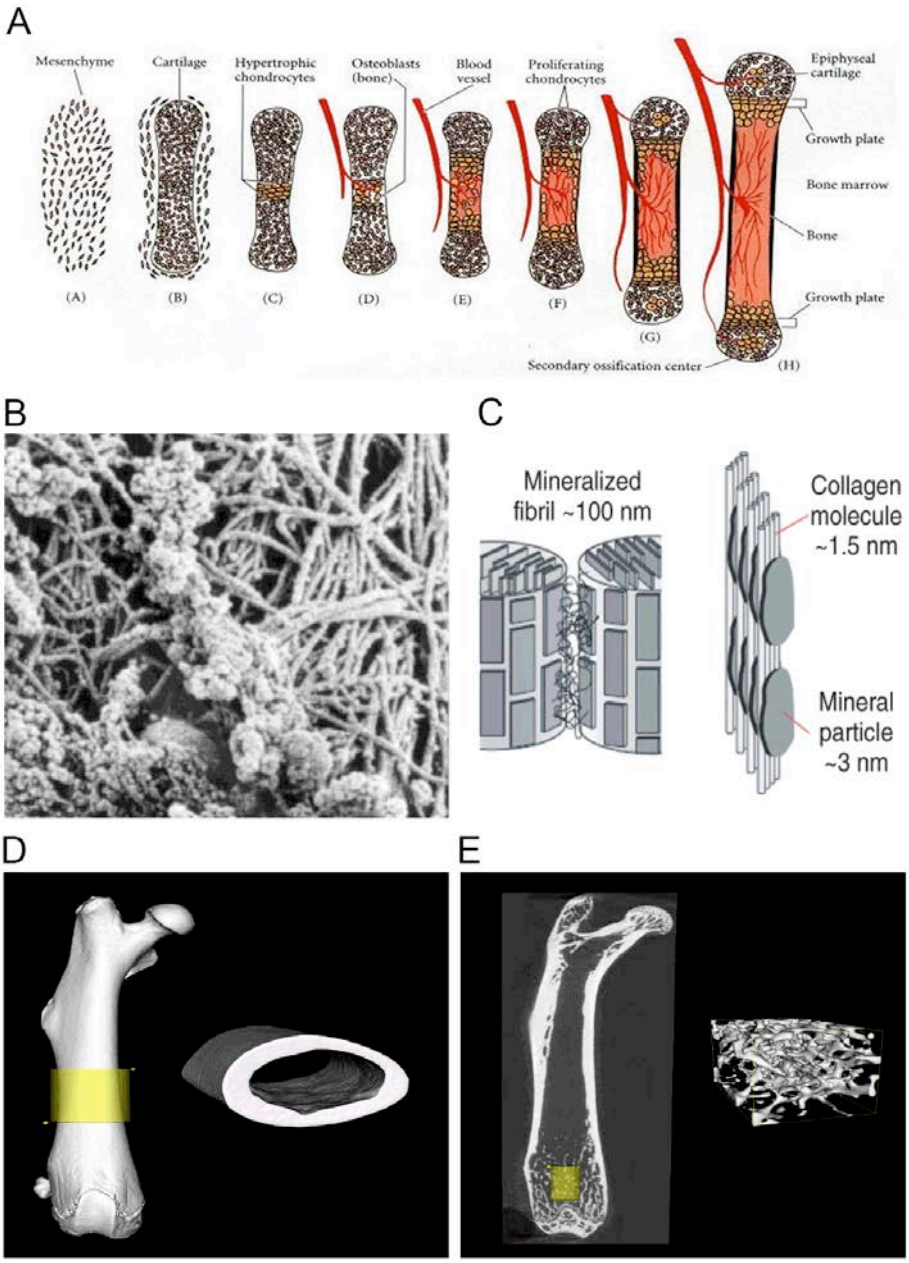

Figure 8. Endochondral Bone Formation and Mineralization. A, Endochondral bone formation begins with development of a cartilage model, followed by vascularization of the tissue and replacement with a mineralized matrix by osteoblasts. B and C, primary mineral deposition occurs on the surface and between collagen fibrils. D, cortical bone of the diaphysis. E, trabecular bone is located in the metaphyseal region of long bones. Bone formation diagram adapted from Gilbert [3]. Mineral images adapted from Weismann *et al*, Beniash *et al* [4,5]. Micro-CT reconstructions provided courtesy of Ken Kozloff, PhD.

In long bones, secondary centers of ossification form at each end of the cartilage model so that a cartilaginous growth plate remains between the primary and secondary centers of ossification. Within the growth plate, chondrocytes are organized into zones that reflect their states of differentiation. Proliferative chondrocytes are found within the cartilage matrix farthest from the primary center of ossification. Closer to the mineralizing front, chondrocytes with high secretory activity are found in multicellular clusters. Adjacent to the primary center of ossification, hypertrophic chondrocytes are arranged into columns. These cells model and mineralize their surrounding matrix prior to cell death. Therefore, bone elongation occurs as a result of growth plate migration away from the primary center of ossification, accompanied by replacement of mineralized cartilage for bone.

Bone formation without prior resorption on periosteal surfaces increases the size and modifies the shape of bone according to a genetic blueprint and in response to the prevailing mechanical loads applied [236]. This process of initial bone formation and shaping, or modeling, occurs mainly during the growth phase of the skeleton but also during fracture healing. Altering the shape of a skeletal segment also requires resorption of some surfaces without replacement of the tissue, which is also considered modeling. The replacement of the cartilage anlage by mineralized tissue and development of a marrow cavity during early endochondral bone formation is a primary example of the process of bone modeling. Modeling is most active during embryonic development and infancy, and subsides upon skeletal maturity [237].

Types of Bone

The bone collar that forms between the growth plates eventually becomes a hollow, marrow-filled shaft termed the diaphysis. This tissue is composed primarily of compact, nonporous cortical bone. The outer periosteal surface of the diaphysis is an area of active bone deposition. The inner endosteal surface faces the marrow cavity, where bone resorption exceeds formation. Thus, the diaphysis and marrow cavity tend to expand with increasing age. Below the growth plates, bone is composed of a network of mineralized matrix with a sponge-like appearance, termed trabecular bone, that comprises the metaphysis. The secondary ossification center at each end of long bones, enclosed within the epiphysis, is also composed of trabeculae. Both trabecular and cortical bone are composed of osteons, which consist of concentric layers of mineralized tissue. These concentric layers, termed lamellae, acquire their extraordinary strength as a result of the alternating orientations of collagen fibrils within the rings. In cortical bone each osteon surrounds a Haversian canal, which contains the bone's nerve tissue and blood vessels.

Bone Mineralization

The organic phase of bone is composed mostly of collagen fibrils, which function as a scaffold upon which calcification occurs. In addition to collagen, unmineralized matrix contains non-collagenous proteins believed to regulate mineralization, including proteoglycans, glycoproteins, and γ-carboxyglutamic acid-containing molecules, and phosphoproteins [4]. The most important of these mineral-regulating proteins are the SIBLINGs (Small Integrin-Binding LIgand, N-linked Glycoproteins), which include osteopontin (OPN), bone sialoprotein (BSP), dentin matrix protein 1 (DMP1), dentin

sialophosphoprotein (DSPP), and matrix extracellular phosphoglycoprotein (MEPE) [238]. Some of these proteins have been studied for their role in promoting or inhibiting mineralization *in vivo* and *in vitro*. DMP1, BSP and DSPP appear to promote mineralization, since they are highly acidic proteins capable of binding Ca^{++}, while OPN has been reported to inhibit mineralization [239-242]. The NCPs, collagen scaffold, and mineral components are deposited within the extracellular matrix by osteoblasts, demonstrating the central role that these cells have in bone modeling.

The inorganic phase of bone tissue consists of hydroxyapatite, which forms a hexagonal crystal lattice made of calcium, phosphate, and hydroxide ($Ca_5(PO_4)_3(OH)$). However, approximately 4-7% of bone mineral is apatite in which a phosphate group or a hydroxyl ion is substituted by carbonate [243]. The mechanism for mineral nucleation and growth remains a controversial topic. An early hypothesis proposed direct nucleation of crystals by precipitation of apatite upon the surface of matrix components, which include NCPs as well as collagen. More recently, evidence has emerged for matrix mineralization as a cell-mediated process of vesicular delivery of mineral components. Specifically, vesicles have been observed by microscopy to bud off from the plasma membrane of hypertrophic chondrocytes and mineralizing osteoblasts, and associate with sites corresponding to initial mineral formation prior to matrix mineralization [244,245]. Originally discovered by TEM analysis of growth plate cartilage and bone, these "matrix vesicles" have been described as 20-200nm spherical bodies bounded by a lipid bilayer, and often found associated with small crystals of apatite mineral [246-248].

Proteomic and mass spectrometry analysis of components isolated from cartilage and osteoblast cultures has demonstrated that matrix vesicles are enriched in tissue non-

specific alkaline phosphatase (TNAP), nucleotide pyrophosphatase phosphodiesterase (NPP1/PC-1), PHOSPHO1, annexins (ANX), integrins and phosphatidylserine (PS) [249]. Identification of these components suggests a mechanism for mineralization. For instance, annexins are Ca^{++} and phospholipid binding proteins which form calcium channels through membranes [250]. NPP1 suppresses mineralization by increasing the extracellular concentration of the calcification inhibitor, pyrophosphate (PPi), whereas TNAP promotes mineralization by hydrolyzing PPi and increasing the concentration of the mineralization promoter, inorganic phosphate (Pi) [251]. Wu and coworkers have provided evidence for matrix vesicles binding to native, but not denatured Type I collagen via interaction with integrins and other proteins [252]. Furthermore, Golub *et al* have described the appearance of hydroxyapatite crystals precipitating inside vesicles [251]. Thus, it seems likely that mineralization proceeds by initial formation of apatite within vesicles and integrin-mediated binding of matrix vesicles to fibrils, followed by a mineral growth phase on the fibril (Figure 8).

Bone matrix is not uniformly mineralized. Primary mineralization begins, on average, 5-10 days after deposition of the organic matrix, and proceeds until approximately 50-60% of the mineralizing region is filled [253,254]. In early bone, mineralized tissue has a composition of approximately 25% organic, 50% inorganic, and 25% water. Following primary mineralization, the rate of mineralization slows down and a secondary phase begins. Propagation of the mineral from the surface to the interior of the fibril continues through the gap spaces, which accounts for the differences in the ratios of organic, inorganic and water components of bone tissue during development. At later stages of development, mature bone tissue has a composition of approximately 30% organic

material, 60% inorganic material, 7% water and less than 3% lipids [255]. As bone tissue matures, mineral replaces water in the longitudinal gap spaces separating collagen monomers and between fibrils, a process that can take years to complete [254].

The Cellular Components of Bone

The cell type responsible for bone formation is the osteoblast (Figure 9). Although the events that regulate the development of osteoblasts from their precursor cells *in vivo* remain to be fully characterized, studies have demonstrated that these cells are derived from bone marrow mesenchymal stem cells (MSCs) that differentiate in response to local and systemic factors, including various cytokines, growth factors and hormones. In

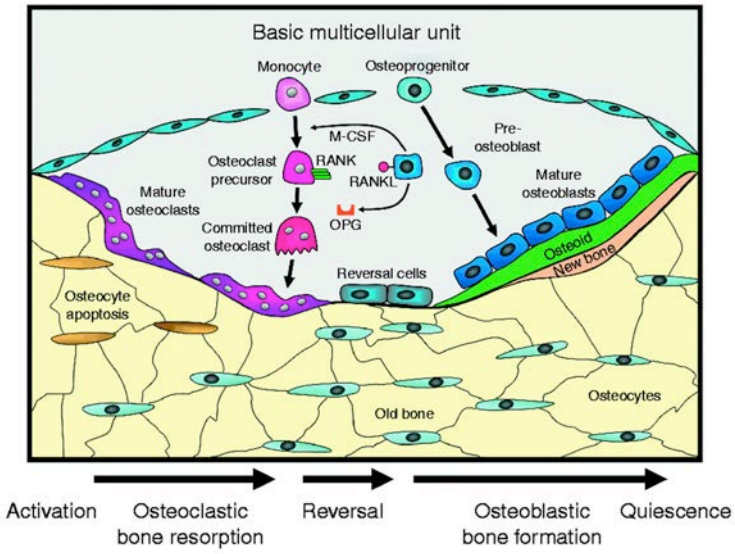

Figure 9. The Cellular Components of Bone. Old bone is removed (resorbed) by osteoclasts, which are myeloid cells that develop from the hematopoietic lineage. New bone is generated from osteoblasts that develop from mesenchymal progenitors in the bone marrow. Once they are completely encased within the mineralized extracellular matrix osteoblasts terminally differentiate to osteocytes. Bone health is regulated by a balance between osteoclasts and osteoblasts, which are able to influence each other's activities through cell-cell interactions. Osteocytes function to monitor the status of the surrounding tissue. Figure from Nicholls *et al* [1].

addition to members of the TGFβ and WNT superfamilies, several interleukins and non-collagenous proteins (NCPs) of the extracellular matrix affect osteoblast activity. Vitamin D_3, parathyroid hormone (PTH), estrogen and leptin are among the systemic factors that influence recruitment and maturation of bone forming cells [256]. However, MSCs in culture require only serum, ascorbate, the glucocorticoid dexamethasone, and β-glycerophosphate to differentiate into committed osteoblasts [257]. The initial signaling events associated with osteoblast differentiation occurs via Indian hedgehog (*Ihh*), which is required for endochondral bone formation [258]. IHH induces expression of the transcription factor RUNX2 (*Cbfa1*), the master regulator in osteoblast differentiation. RUNX2 regulates the expression of osteoblastic markers such as type I collagen (*Col1a1*), bone sialoprotein (*Ibsp*), osteocalcin (*Bglap2*), and osteopontin (*Spp1*), all key components of mineralizing bone matrix. It is maximally expressed in osteochondral progenitors, as well as in early stages of osteoblastic differentiation. *Runx2* KO mice display absence of intramembranous and endochondral ossification [259,260]. An additional target gene of RUNX2 is Osterix (Sp7), which is essential for commitment of preosteoblastic cells to differentiate into mature matrix-producing osteoblasts [261].

Osteocytes comprise over 90% of the cell population in adult bone. Formation of these cells was historically thought to be a passive process of osteoblasts becoming encased in osteoid that mineralizes, followed by terminal differentiation into osteocytes (Figure 9). However, osteocytogenesis involves changes in morphology and function that differ from matrix-depositing osteoblasts. As the osteoblast transitions to an osteocyte, the expression of genes required for matrix deposition shifts to those required for mineralization and tissue maintenance, including upregulation of osteocalcin (BGLAP2),

phosphate-regulating gene with homologies to endopeptidases on the X chromosome (PHEX), matrix extracellular phosphoglycoprotein (MEPE), dentin matrix protein 1 (DMP-1), fibroblast growth factor 23, (FGF-23) and sclerostin (SOST) [262]. These proteins are directly involved in the regulation of bone mineralization, since inactivating mutations in the genes that encode them result in a number of metabolic bone disorders. For example, autosomal recessive hypophosphatemic rickets in patients is due to mutations in *DMP1* [263]. *Phex* mutations result in osteomalacia and rickets, owing to elevated levels of FGF-23 in osteocytes [264]. Since both *Dmp1* and *Phex* downregulate FGF-23 expression, mutations in these genes cause downstream defects in the role of FGF23 in maintaining normal phosphate levels, which in turn maintains normal bone mineral content. This occurs because, in the absence of either *Dmp1* or *Phex*, FGF-23 is elevated in osteocytes and in the circulation, leading to phosphate excretion by the kidney, thereby reducing circulating phosphate, which is in turn responsible for osteomalacia and rickets [265].

The primary morphological feature of the osteocyte is the formation of multiple branched filipodial processes, similar to dendrites, that extend towards the mineralizing front and vascular spaces in bone. This is an invasive process requiring cleavage of collagen and other matrix molecules by MT1-MMP, a membrane-anchored proteinase, the absence of which results in reduced dendritic processes [266]. This unique morphological feature may be controlled by E11/gp38, also known as podoplanin, which is a specific marker of osteocytes [267]. One of the proposed roles of these dendritic processes is to allow these cells to sense and communicate responses involved in mineral homeostasis and bone repair. This is made possible by virtue of a network of

interconnecting canaliculi within bone tissue in which osteocytic processes make intercellular connections through gap junctions consisting of connexins. Very little is known about the bone fluid that flows through the osteocyte canalicular system, except that a molecular weight/mass cutoff of 70 kDa, the size of bovine serum albumin (BSA), exists [268]. However, this size cut-off would not exclude small signaling molecules.

Mature osteoclasts are multinucleated cells of hematopoietic origin, derived from bone marrow mononuclear precursors of the monocyte-macrophage lineage [269], whose primary role in bone is to resorb existing bone tissue (Figure 9). Osteoclastogenesis is a multistage process that includes commitment, differentiation, multinucleation, and activation of osteoclast precursors. Analagous to the role of Runx2 in osteoblast maturation, the transcription factor PU.1 is responsible for commitment of precursor to the osteoclastogenic phenotype. PU.1-null mice are unable to form osteoclasts and macrophages, but are still able to form monocytes [270]. Expression of the c-FMS gene, the receptor for M-CSF (macrophage colony stimulating factor), is dependent on both PU.1 and MITF transcription factors. However, M-CSF can upregulate expression of its own receptor by stimulating PU.1 expression [271]. Activation of c-FMS by M-CSF has been shown to be necessary for the proliferation and survival of osteoclast progenitor cells, and loss of function mutation in the *Mcsf* gene (*op/op* mice) results in an osteopetrotic phenotype due to the lack of osteoclasts [272].

Differentiation and multinucleation of osteoclasts occurs upon interaction of RANK+ mononuclear precursors and a cell expressing RANKL. Osteoblastic cells in the bone marrow express two cytokines required for osteoclast precursor (OCP) differentiation into mature osteoclasts: macrophage-colony stimulating factor (M-CSF) and receptor

activator of NF-κB ligand (RANKL). M-CSF binds to its receptor, c-fms, on OCPs and activates signaling through MAPK and ERKs during the early phase of OCP differentiation [273]. RANKL binds to its receptor, RANK, on the surface of OCPs, activating signaling through NF-κB, c-Fos, phospholipase Cγ (PLCγ) and nuclear factor of activated T cells c1 (NFATc1) to induce differentiation of OCPs into osteoclasts [274]. T and B lymphocytes also express RANKL and promote osteoclast formation in pathologic states such as rheumatoid arthritis [275,276]. Osteoclast formation and activity is also balanced by a number of cytokines, most notably by osteoprotegerin (OPG), which binds to RANKL as a decoy receptor and prevents its interaction with RANK [277]. Thus the RANK/RANKL/OPG axis is considered a major determinant of bone resorption and bone mass.

Activation of immature multinucleated osteoclasts to the mature bone-resorbing multinucleated osteoclast begins with the generation of a ruffled membrane in contact with the bone surface via αvβ3 integrin. Simultaneously, an intracellular actin ring forms the sealing zone, which isolates the resorptive microenvironment between the ruffled membrane and bone from the surrounding extracellular space. Osteoclasts resorb bone by secreting H^+ via Atp6v0d2 and Cl^- through the osteoclast chloride channel, ClC-7, in addition to proteases, such as cathepsin K and matrix metalloproteinases (MMPs), into the extracellular compartment beneath the ruffled part of their basal cell membrane to dissolve the mineral and matrix components of bone simultaneously [278]. In addition to their role in bone resorption, osteoclasts also regulate differentiation of osteoblast precursors through RANK-RANKL interaction, participate in immune responses, and

secrete cytokines that can affect their own functions and those of other cells in inflammatory and neoplastic conditions affecting bone [279].

Cellular Crosstalk

Wnts are a family of 19 secreted glycoproteins involved in multiple signaling cascades essential for embryonic development and tissue regeneration. Binding of Wnts to Frizzled (Fzd) and its co-receptor LRP5/6 recruits Dishevelled (Dsh) to stabilize intracellular β-catenin, a signaling protein which enters the nucleus and complexes with T-cell factor/lymphoid enhancer factor (TCF/LEF) as a transcription factor [280]. Several *in vivo* studies have demonstrated a role for Wnt signaling in the regulation of osteoblast survival and differentiation. For example, activating mutations in LRP5 in humans result in a high bone mass phenotype with decreased bone marrow fat content [281,282]. Conversely, LRP5 inactivating mutations are the genetic cause of osteoporosis pseudoglioma (OPPG) syndrome, which is a recessive disorder characterized by low bone mass and increased marrow fat [283]. *Lrp6* KO mice are not viable and show defective limb development, while haploinsufficiency reduces total and trabecular bone mineral density [284,285]. Conditional knockout of β-catenin in skeletal mesenchymal cells promotes chondrogenesis but inhibits osteoblasts differentiation [286]. Knockdown in osteoblasts has also shown a role for β-catenin in postnatal bone remodeling in mice, since its absence leads to osteopenia and increased osteoclast recruitment [287]. In addition, overexpression of *Frzb* (frizzled-related protein), a Wnt antagonist, blocks ectopic bone formation by human periosteal cells in mice [288]. These studies have demonstrated that Wnts directly promote bone formation by activating target genes required for osteoblastic

differentiation, including Osterix, Runx2 and alkaline phosphatase (Alp), while simultaneously suppressing transcription of adipogenic transcription factor peroxisome proliferator-activated receptor-γ (PPAR-γ) [289,290].

Another antagonist of Wnt which has received recent attention as a potential target for anabolic treatment of bone disease is Sclerostin (Scl), encoded by *SOST*. The role of Sclerostin in bone development first came to attention when the progressive skeletal overgrowth disorders, sclerosteosis and van Buchem disease, were found to be caused by inactivating mutations in the *SOST* gene [291,292]. Loss of sclerostin activity in human patients is associated with high bone mass, abnormally increased osteoblast activity and elevated bone formation markers [293]. These findings have been recapitulated in *Sost* KO mice, which have elevated bone density and bone mechanical strength [294]. Overexpression of *Sost* in transgenic mice induces osteopenia by blocking the association of coreceptors Lrp5/6 with Fzd, which inhibits differentiation and stimulates apoptosis of osteoblasts [295,296]. Thus by secreting sclerostin into the canilicular network of bone, osteocytes are able to regulate bone mass by inhibition of Wnt signaling [297].

The transforming growth factor-beta (TGFβ) superfamily is a large family of growth factors that includes TGFβs, Nodal, activins, and bone morphogenetic proteins (BMPs). TGFβ is assembled as a biologically inactive propeptide homodimer, termed the small latent complex (SLC). The SLC is secreted into the extracellular matrix as a noncovalent complex with latent TGFβ binding proteins (LTBPs), which together are referred to as the large latent complex (LLC) [298]. Since LTBPs bind to fibrillins, which are major structural components of connective tissue microfibrils and elastic fibers, TGFβ can be retained in its latent form until it is released by osteoclast-mediated bone turnover.

Hence, many of the manifestations of Marfan Syndrome (OMIM #154700) in patients with mutations of fibrillin-1 (*FBN1*) reflect excessive TGFβ signaling, including aortic aneurysm and lung pathology [299].

TGFβ transmits signals across the plasma membrane through formation of heteromeric complexes of type I and II serine/threonine kinase receptors (TGFβRs). The type I receptor is phosphorylated following activation of the type II receptor and initiates intracellular signaling through phosphorylation of Smad2 and -3 [300]. Activated Smads form a complex with co-Smad and Smad4 and translocate into the nucleus to direct transcriptional response involved in osteoprogenitor commitment, proliferation and differentiation [301]. TGFβ1 has been shown to activate β-catenin signaling and modulate osteoblastogenesis in hMSCs [302]. TGFβ1-deficient mice display reduced bone growth and mineralization [303]. In humans, heterozygous missense mutations of TGFβ1 have been associated with Camurati-Engelmann disease (OMIM #131300), a rare autosomal dominant bone dysplasia with cortical thickening of the diaphyses of the long bones [304]. TGFβ2 and TGFβ3 double knockout mice are embryonic lethal and lack the distal portions of the ribs [305]. Since Tgfbr2 directly phosphorylates the PTH type I receptor (PTH1R) cytoplasmic domain, mice with Tgfbr2 deleted in osteoblasts have increased bone mass due to constitutive activation of PTH1R [306,307]. PTH couples the processes of bone resorption and formation by enforcing simultaneous internalization of Tgfbr2 and PTH1R [307].

The bone morphogenetic proteins (BMPs) were first identified as constituents of demineralized bone extracts capable of inducing endochondral bone formation [308]. BMPs were subsequently purified from bovine bone and identified as members of the TGF-β

superfamily [309]. Similar to the TGFβRs, there are two receptors subtypes, BMPR-I and BMPR-II, that dimerize upon ligand binding and transduce their signal through activated Smads 1, 5 and 8. BMP-activated Smads interact with Runx2 to induce osteoblast-specific gene expression and differentiation [310]. In human and rat marrow stromal cells, BMP2, 4, 5, 6, and 7 all have strong osteogenic capacity [311,312]. Addition of BMP2 to these cells substantially increases *ALP* and *BGLAP2* (osteocalcin) expression, while short-term exposure irreversibly induces bone formation [313,314].

Since loss of either BMP2 or BMP4 results in embryonic lethality with severe impairment of osteogenesis, most studies utilizing mouse models for BMPs have relied upon conditional KO alleles to delineate their function in skeletal morphogenesis. Mice lacking BMP2 in long bones develop spontaneous fractures that fail to resolve [315]. Combined loss of Smads 1, 5 and 8 results in severe chondrodysplasia [316], while deletion of either Smad1 or -4 in osteoblasts leads to lower bone mineral density, decreased bone volume, decreased bone formation rate, and reduced numbers of osteoblasts [317,318]. However, not all BMP activity induces bone formation. BMP-3 limits skeletal progenitor cell differentiation to mature osteoblasts, Trabecular bone volume of BMP-3-deficient mice is two-fold greater than that of wild-type mice [319]. Mice with conditional deletion of BMPR-IA in collagen-expressing tissues have increased bone volume and mass, with shortened limbs and growth deficiency, due to decreased osteoclastogenesis through the RANKL-OPG pathway, as well as inhibition of Wnt signaling through increased expression of sclerostin [320-323].

It should be noted that signaling between bone cells is multidirectional and regulates both osteoblast and osteoclast recruitment and differentiation. For example, conditioned

medium from osteocyte-like cells is known to support osteoblast differentiation, as well as mesenchymal stem cell differentiation, supporting the theory that osteocytes orchestrate bone remodeling through increased secretion of growth factor M-CSF [324,325]. More importantly, immunohistochemical staining for RANKL enhances the visibility of dendritic processes of osteocytes, showing RANKL expression on these processes and suggesting that osteocytes support osteoclast formation. It has also been reported that apoptotic osteocytes are capable of inducing bone remodeling through an unidentified, RANKL-independent mechanism, by recruiting and evoking a phagocytic response from osteoclasts [326].

Cellular crosstalk between osteoblast and osteoclast lineage cells, as described above, is the key mechanism by which bone tissue is replaced in the adult skeleton. This process of tissue turnover, or remodeling, is the only normal physiologic process for altering the organic and inorganic components of bone after puberty. The process of remodeling does not change the size or shape of a bone, but is instead a process of maintaining the health of the existing tissue. Resorption of existing tissue occurs at focally discreet points on endocortical and trabecular surfaces, followed by replacement with collagenous osteoid. During remodeling, osteoclasts and osteoblasts colocalize at the bone surface in bone-remodeling compartments (BRC), which are separated from the surrounding environment by a monolayer of flat cells that show osteoblast-like cell markers [327]. These bone-lining cells deposit a thin layer of collagen matrix, generating an enclosed "canopy" that surrounds the site of active bone remodeling [328]. Canopy cells have also been reported to degrade matrix left over by osteoclasts by secreting high levels of MMP-13 at locations adjacent to the osteoclast-bone interface [329]. Although little is known about canopy/bone-

lining cells, it has been proposed that these cells represent early OB lineage cells that later differentiate into bone-forming osteoblasts [330]. BRC canopies are physically associated with capillaries, which allows for recruitment of both osteoclast and osteoblast progenitors from the bone marrow, and pre-osteoclasts associate with canopy-capillary connection points [330].

Assessment of Bone Formation and Quality

Bone strength, as defined by resistance to fracture under normal loading conditions, is determined by material and structural properties [235]. The cortical femoral diaphysis is structurally similar to a tube, and its geometric properties can be calculated from measurements of cortical dimensions to provide estimates of bone strength. The predicted strength, or moment of inertia (MOI), reflects both the amount and distribution of tissue in cortical bone and correlates well with breaking strength. Independent of size and shape, the material properties of bone encompass the properties of the constitutive components of bone, including the organic and inorganic phases of the mineralized tissue [331]. In healthy bone tissue, the mineral density distribution (BMDD), as determined by quantitative backscattered electron imaging (qBEI), should reflect a heterogenous density. Heterogeneity of mineral density encompasses newer, less mineralized tissue, and older, more mineralized tissue. A higher rate of turnover, as is seen in osteoporosis and other bone pathologies, is indicated by an increased fraction of tissue with mineral density that is lower than in healthy bone since there is less time for secondary mineralization to occur [332]. In contrast, a decreased rate of bone turnover will result in a shift to greater mineral density. Although it is counterintuitive to consider mineral density

higher than normal as detrimental to bone strength, mineral increases the stiffness of bone. This property of mineral causes bone to become more brittle, requiring less energy to fracture [333]. Thus, hypermineralized bone may be just as susceptible to fracture as undermineralized bone.

Bone strength, however, is not only determined by the amount of mineral, but also by the number, size and orientation of bone mineral crystals. Crystallinity, as measured by Fourier-Transform Infrared Imaging (FT-IR) Spectroscopy, is a parameter that expresses crystal size and perfection [334]. Homogeneity of crystal, with a preponderance of either small or large crystals, is associated with increased bone brittleness as is seen in osteoporosis [335]. An additional component contributing to bone material properties is the organic (collagen) phase. Increased collagen content relative to mineral indirectly decreases the stiffness of bone by lowering the BMDD. Furthermore, collagen crosslinks provide tensile strength to bone tissue by increasing its flexibility, allowing it to absorb energy on impact. Both mineral and collagen content of bone tissue can be assessed by Raman spectroscopy, since differences in mineral or organic matrix composition affect the Raman spectrum. Alterations of vibrational frequencies of amide groups of collagen, and phosphate and carbonate groups of bone mineral can be used to determine differences in mineral-to-matrix ratios, as well as crystallinity [336].

The structural properties of bone are described by the microarchitecture and geometry of the bone tissue [235], which can be assessed by magnetic resonance imaging (MRI) and computed tomography (CT) [337]. For cortical bone, diameter and thickness (CtTh) have a dramatic impact on mechanical integrity [338]. Long bones are geometrically similar to tubes, making the periosteal radius, cortical cross-sectional area (CsAr) and length factors

that contribute to bone strength, in addition to material properties. In fact, increased bone strength is associated with increases in cortical diameter and thickness, while individuals with thinner cortices are at an elevated risk for fractures [339]. Also, increases in tissue porosity can be caused by an imbalance of bone formation and resorption during remodeling as the individual ages, contributing to decreased bone strength. In trabecular bone, the microarchitecture of the trabecular network has a substantial affect on bone strength. Trabecular bone volume (TrBV), which is a reflection of trabecular number (TbN) and thickness (TbTh), determines strength in a non-linear manner [340]. Also, the orientation of trabeculae relative to loading, as well as their connectivity, can determine the extent to which strain is resisted [341]. Bone with widely separated, thick trabeculae having few connections is less resistant to compressive forces than bone with numerous, highly connected trabeculae because increased connectivity reduces the area that is unsupported [342]. Still another parameter for bone strength is anisotropy, which is the trait of an object that possesses different values for a given parameter, such as strength, when measured in different directions. In trabecular bone, an anisotropic structure is one that would demonstrate different strength or mechanical integrity attributes dependent on the direction of loading. Accordingly, femoral trabecular tissue is better constructed to withstand applied vertical loads compared to horizontal stresses. Anisotropy of trabecular bone is not only affected by trabecular thickness, orientation, and connectivity, but also by material properties [343].

The contributions of the structural and material properties of bone can be determined by examination of its mechanical properties. Like other objects in nature, the application of force to a segment of bone will result in the generation of an internal resistance known

as stress. *Stress*, expressed as force/area, is equal in magnitude but opposite in direction to the applied force. Three types of stress exist: tension (stretching), compression and shear (parallel, but opposite forces produced by torsion). However, most forces applied to bone are a combination of these three types of stress. For example, bending of a long bone produces tensile forces on the convex side and compression on the concave side. The deformation of the bone, or *strain*, is the difference in the length of the long bone due to the stress. By knowing the size of the specimen, the force applied, and the extent of the deformation produced by that force, the mechanical properties of a bone sample can be deduced by analysis of the stress versus strain comparison [7]. By convention, the loading condition (*stress*) is plotted on the y-axis and *strain* on the x-axis (Figure 10). The resulting curve plots the load-versus-deformation relationship. Two regions of the curve are used to define the material properties of the sample being analyzed. The linear

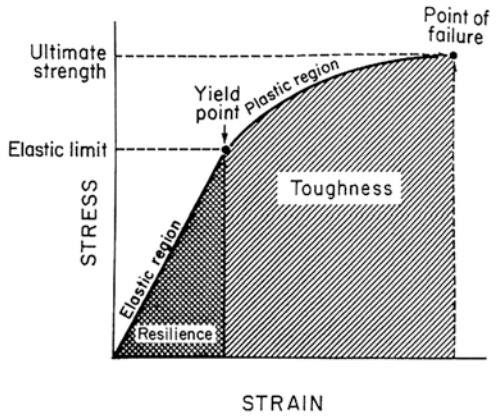

Figure 10. Stress-Strain Curve for Cortical Bone. The relationship between load applied (stress) and resulting deformation that occurs (strain) for a given bone sample can be plotted as a stress-strain curve of bone loaded in bending. The slope of the curve within the elastic region is the elastic modulus and is a measure of the sample stiffness. The area under the curve indicates the amount of energy required for failure of the sample, and is independent of the size and shape of the bone (i.e., material properties). The yield point represents the amount of load required to permanently damage the bone. The post-yield strain is inversely proportional to the brittleness of the bone sample. Figure from Einhorn [7].

portion of the stress-strain curve, the *elastic region*, represents the amount of force that can be applied without inflicting permanent damage on the sample. In the linear region, the slope of the stress-strain curve represents the *elastic modulus*, which is a measure of the stiffness or rigidity of the sample. The maximum stress that can be applied without permanent deformation of the sample is the *elastic limit*; the point on the curve corresponding to this limit is the *yield point*. Further loading beyond this point will cause permanent deformation of the material, represented on the curve as the *plastic region*, and continues to the point at which the sample being analyzed reaches the breaking point (i.e., *point of failure*). The strength of a bone specimen is determined by calculating the maximum stress applied at the point of failure, and is referred to as the *ultimate strength* of the sample. When the various stresses that are applied to bone exceed the ultimate strength of the tissue, a fracture will occur.

What information about bone material properties can be obtained by observation of its mechanical properties? Under certain pathological conditions, bone fails before undergoing permanent deformation; this lack of plasticity is characteristic of brittle material. In contrast, a flexible material will demonstrate a decreased slope within its elastic region, which will be shifted since the material can endure greater strain before reaching its yield point. However, more specific characteristics of bone material properties can be ascertained. Biomechanical testing of demineralized bone has demonstrated that the mineral component contributes the major portion of the tensile strength, while the slope of the plastic region is a function of the collagen component [344]. Furthermore, the contribution of the mineral component to bone strength is a function of

the crystallinity of the mineral [345]. In general, elasticity (stiffness) is related to the mineral phase and plasticity (ductility, toughness) to the organic matrix [344,346].

Another tool that is useful for investigating bone pathology is histomorphometry. Bone histomorphometry is a quantitative histological examination of an undecalcified bone biopsy to obtain quantitative information on bone remodeling and structure [347]. Labeling agents taken before the procedure deposit at sites of bone formation allowing a dynamic analysis. Labeling fluorochromes, such as tetracyclines, are incorporated into the bone at sites of new bone formation, binding irreversibly to hydroxyapatite at the mineralization front. In animal models, bone formation is often measured by dual-calcein and xylenol-orange labeling. Administration of these fluorochromes, with a drug-free interval of 10 to 14 days between the two courses, results in the formation of dual-labeled sites where bone formation was ongoing during the entire labeling sequence. Bone histomorphometric parameters are assessed in two dimensions by means of histology, where the structural and remodeling parameters are measured on sections. The static parameters that can be obtained with this technique provide information about osteoid volume (OV/BV), given as the percent of a specified volume of bone tissue that consists of unmineralized bone (osteoid). More importantly, dual-labeled histomorphometry can yield informative dynamic bone remodeling parameters related to bone formation rate. Mineralizing surface (MS/BS, %) is a measure of the proportion of bone surface upon which new mineralized bone was deposited during the period of labeling. Mineral apposition rate (MAR), a measurement of the linear rate of new bone deposition, is the mean distance between the double labels, divided by the time interval between their application. Bone formation rate (BFR/BS), the amount of new bone formed in unit time

per unit of bone surface, is calculated by multiplying the mineralizing surface by the mineral apposition rate. In addition, the mineralization lag time (Mlt/day) represents the average time interval between osteoid formation and its subsequent mineralization, which is calculated by dividing the osteoid width by the apposition rate. Thus, dynamic histomorphometric parameters can provide insight into defects of osteoblast function and generation of extracellular matrix of bone in certain pathological conditions.

Osteogenesis Imperfecta

Osteogenesis imperfecta (OI), also known as Brittle Bone Disease, is a heritable disorder of connective tissue with an estimated incidence of approximately 1 in 15,000-20,000 births [348-351]. The most common clinical features of OI include bone fragility and deformity, and growth deficiency (Figure 11). Additional variable symptoms associated with OI include blue sclerae, hearing loss, dentinogenesis imperfecta and pulmonary deficiency [352]. The original Sillence Classification, which preceded the identification of the molecular causes of OI, is used to describe the range of phenotypic severity in this disorder and is based on radiological and clinical criteria [348]. Type I, the mildest form of OI, results from quantitative defects of type I collagen synthesis which causes an osteoporotic phenotype due to matrix insufficiency [353]. More severe bone disease, including perinatal lethal type II, progressive deforming type III and moderate type IV OI, results from defects in type I collagen primary structure. An updated form of the original classification has been proposed to attribute the Sillence types only to cases with

collagen quantitative and structural defects, with additional OI types assigned to recently discovered novel gene defects [354].

Dominant Osteogenesis Imperfecta

OI is primarily caused by dominant mutations in the two genes encoding the type I collagen α1(I) and α2(I) chains, *COL1A1* and *COL1A2*, respectively. More than 1500 independent OI-causing mutations have now been identified in these genes [116]. A majority of these mutations (80%) are *de novo* single nucleotide changes that result in

Figure 11. Clinical Features of Classical Osteogenesis Imperfecta. The hallmark clinical symptoms of osteogenesis imperfecta (OI) include growth deficiency, fractures from minimal trauma and blue sclerae. X-rays of OI probands reveal bowing and fractures of long bones, compression of vertebrae, and a bell-shaped rib cage. Severity of symptoms, including frequency of fractures, can vary and is described by the original Sillence Classification system. Figures courtesy of Joan C. Marini, MD, PhD.

primary structural defects of type I collagen alpha chains, leading to glycine substitutions within the Gly-Xaa-Yaa amino acid triplet. The second most common mutations occur at splice sites, which lead to exon skipping or intronic retention [116]. Nearly one-third of glycine substitutions in the α1(I) chain, but only one-fifth of substitutions in the α2(I)

chain, result in a lethal phenotype. Additionally, substitutions for glycine by residues with charged or branched nonpolar side chains are associated with the most severe clinical phenotypes. Lethality also correlates with position of the substitution on the alpha chain; those that occur within major ligand-binding regions (MLBR2 and MLBR3) of the collagen helix on the α1(I) chain and within α2(I) proteoglycan binding sites along the fibril are exclusively lethal [116].

The mechanism by which collagen structural mutations cause disease is incompletely characterized. Collagen trimer formation begins with the association of pro-alpha chains through an alignment region in the carboxyl globular domains, followed by folding which proceeds toward the amino terminus in a zipper-like fashion. Since only glycine residues can fit into the sterically restricted inner aspect of the collagen helix, substitutions by amino acid residues with larger or charged side chains interfere with the folding process. The delay in helix folding caused by the glycine substitution in a collagen alpha chain exposes the portion of the chains amino-terminal to the substitution to the modifying enzymes in the endoplasmic reticulum for a longer period of time [355,356]. Hence, defects in the primary structure of an alpha chain result in overmodification of all chains in the trimer. Detection of collagen overmodification, as observed by delayed electrophoretic migration of alpha chains on polyacrylamide gel electrophoresis, is the basis for the biochemical diagnostic test for classical osteogenesis imperfecta (Figure 12). A gradient in overmodification can be demonstrated by PAGE analysis of CNBr cleavage products of collagens harboring structural defects occurring along the alpha chain; substitutions located at the carboxyl terminus result in a greater extent of overmodification than those at the amino terminus [357]. At the molecular level, both glycine substitutions and splicing

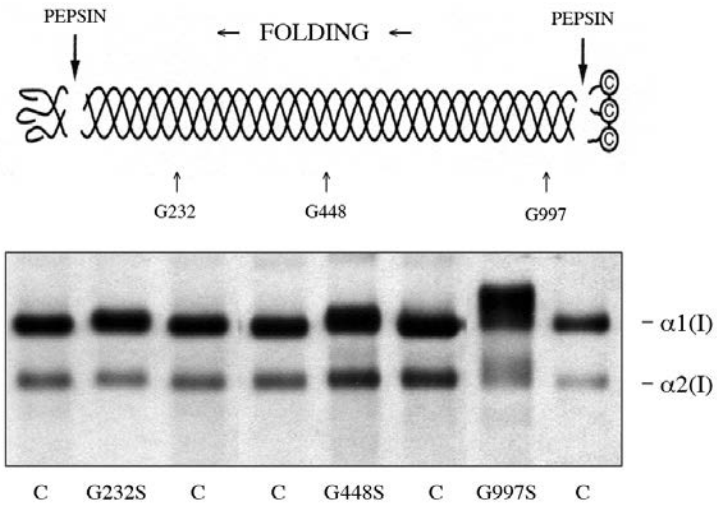

Figure 12. Standard Biochemical Determination of Type I Collagen Overmodification. Pepsin digestion of radiolabeled procollagen removes the propeptides from the molecule. The resulting collagen molecules are electrophoresed on SDS-Urea polyacrylamide gels and visualized by autoradiography. Structural defects in the collagen molecule delay folding of the helix and expose the alpha chains to the post-translational modification enzymes for a longer period of time. Due to the carboxyl to amino direction of folding, a gradient of overmodification can be detected and correlates with the position of the structural defect on the alpha chain. Figure adapted from Marini *et al* [2].

defects are believed to disrupt noncovalent inter- and intrachain bonds, leading to local unwinding of the triple helix. At the tissue level, abnormal collagen that is secreted into the extracellular matrix may disrupt cell-matrix and cell-cell interactions, collagen-noncollagenous protein interactions, fibril formation, or mineralization of bone in a dominant-negative manner [354]. However, the tissue-level pathologies in OI are ultimately due to defects in the structure of the collagen molecule itself.

Mouse Models of Dominant Osteogenesis Imperfecta

Although collagen is also the most abundant structural protein of the extracellular matrix of skin and tendon, the primary structural defects of type I collagen result in a clinical phenotype defined primarily as a bone disorder. The invasive nature of bone tissue procurement and scarcity of pediatric normal control and OI bone tissue has

limited the majority of the investigations on human samples to skin and fibroblasts. The generation of mouse models of OI has made it possible to investigate the relevant tissues and cells involved in OI pathology. Mouse models of dominant OI that express mutant collagen transgenes have been generated but their usefulness is limited since they reflect a mosaic condition, consistent with a less severe phenotype than in fully-expressing heterozygous OI patients [358-360].

Three murine models of OI are currently in common use, including two mice with classical OI knock-in substitutions and a third mouse model that synthesizes type I collagen lacking an α2(I) chain. The Brtl mouse, which was the first knock-in murine OI model developed , is heterozygous for a missense mutation in *Col1a1* that results in an α1(I) G349C substitution [361]. This mouse reproduces the molecular and phenotypic features of the child with moderate type IV OI on whose mutation it was modeled [361]. Heterozygous Brtl mice demonstrate a 30% perinatal lethality rate. Although surviving Brtl mice have normal life expectancy, they exhibit undermineralization of calvaria, long bones and vertebral bodies, flared rib cages with multiple beads of callus, and dentinogenesis imperfecta. Growth deficiency in the Brtl mouse is evident, as heterozygous mice weigh 30% (5-7g) less than their wild-type littermates. The Brtl mouse has been more thoroughly investigated than the α2(I) G610C mouse, also referred to as the OOA mouse since it models a *COL1A2* founder mutation in an Old Order Amish population in the Ohio River Valley [362]. The phenotypically less severe α2(I) G610C mouse has mild to moderate OI (types I/IV), as is seen in the OOA population [362]. Offspring are produced with normal Mendelian distribution, suggesting absence of perinatal lethality with normal life expectancy. The main features of this mouse model of

mild to moderate dominant OI include osteoporotic long bones and moderate growth deficiency; mice with the α2(I) G610C weigh approximately 10% less than wild-type littermates through their lifespan. The third mouse model with a collagen defect, the *oim*/*oim* mouse, is the most studied mouse model of OI. However, it should be noted that the specific mutation in this mouse strain is atypical of OI because it causes the occurrence of $\alpha1(I)_3$, which is not a normal component of bone or skin tissue. *Oim* harbors a naturally occurring single nucleotide deletion in the portion of *Col1a2* that encodes the interchain alignment region of the C-propeptide, so that α2(I) chains cannot be incorporated into normal heterotrimers, resulting in synthesis and matrix deposition of exclusively α1(I) homotrimers in homozygous mice [363]. A similar mutation has been described in only one human case, who harbored a 4-bp deletion in *COL1A2*, causing severe progressive deforming OI and leading to death from pulmonary complications in the 3rd decade [364]. Interestingly, most of the small number of reported human cases lacking α2(I) chain synthesis have mutations at the N-terminus of the alpha chain, resulting in Ehlers-Danlos syndrome rather than OI [365,366], so homotrimer itself is not responsible for OI. Production of α1(I) homotrimers causes a severe phenotype in the mouse, including limb deformities, osteopenia, growth deficiency, skeletal fractures and progressive kyphoscoliosis. These features are consistent with the severe progressive deforming OI in the human patient, making *oim* the most severe mouse model of OI caused by collagen structural defects. Specifically, *oim* weighs 30% (7-15g) less than wild-type littermates throughout its lifespan and, unlike Brtl and OOA mice, suffers from spontaneous fractures of long bones and ribs.

Common bone features caused by collagen structural defects are revealed by μCT scans of trabecular and cortical bone. Brtl has proven to be especially useful for investigating the contributions of the organic and inorganic phases to bone strength [367]. Brtl mice have significantly lower aBMD than their wild-type counterparts. Micro-CT analysis of 2-month Brtl bone has revealed femoral cross-sectional area (CSA) and cortical thickness (CtTh) that are decreased by 28% and 32%, respectively. Based on geometric parameters alone, the predicted bone strength of Brtl femurs is less than wild-type femurs, with MOI that is decreased by one-half. Trabecular bone parameters also reflect the decreased bone mass in Brtl; BV/TV is decreased by one-third, due to decreased trabecular number (TbN). By 14 weeks of age, however, Brtl demonstrates slight improvements in bone structural parameters. CtTh and CSA are decreased 10% and 14%, respectively, with MOI predicting a 10% decrease in bone strength compared to wild-type mice. Trabecular parameters at 14 weeks show BV/TV reduced by one-third and TbN decreased by 25%. In the α2(I) G610C defect mouse, micro-CT analysis of 2 month femurs demonstrate a modest 10% reduction in cross sectional area with normal CtTh. Trabecular BV/TV and TbN are decreased by half compared to wild-type littermates. In the *oim/oim* mouse, whole bone structural properties reflect decreased bone matrix deposition with severe reductions in bone mass, femoral length, and femoral cortical thickness (CtTh) and cross-sectional area (CSA) throughout life [368-370]. At 14 weeks of age *oim*/*oim* mice, the severest of the three models, show 20% and 26% reductions in CtTh and CSA, respectively, versus wild-type mice [369]. Trabecular bone parameters are also reduced; BV/TV is decreased 25% and TbN is slightly reduced by

6% compared to wild-type. These geometric parameters predict that *oim*/*oim* bone strength (MOI) is decreased by half.

On mechanical testing, murine OI bone with collagen defects has reduced stiffness, breaking strength and post-yield displacement. Brtl bone on mechanical testing showed significantly lower stiffness (↓25%), yield (↓30%), ultimate load (↓33%) and post-yield displacement (↓50%). At 14 weeks Brtl bone has improved mechanics, with stiffness, yield load and ultimate load reduced 20%, 15% and 23%, respectively. Mechanical testing of α2(I) G610C femurs shows about the same extent of reduced bone strength as Brtl. While the mineral content of bone contributes substantially to its stiffness, collagen tends to dominate the properties of the post-yield region of the stress-strain curve. It would appear from the post-yield displacement that once the bone starts to break, the presence of the mutated collagen offers less resistance to complete failure in G610C OI mice, suggesting that the abnormal collagen in the bone is organized in a structurally abnormal manner due to the mutation and/or there is a general reduction in the amount of collagen per unit volume of bone. In *oim/oim,* mechanical testing has demonstrated a weaker bone than Brtl or OOA. *Oim/oim* femurs withstand significantly less load to failure (~50%) compared to wild-type bone; stiffness and ultimate load are comparably decreased. In addition*, Oim*/*oim* bone has significantly less post-yield deformation than wild-type bone [368,369,371]. Interestingly, these reduced material properties do not exist on analysis of demineralized samples; the ultimate strains of wild-type bone matrix and *oim*/*oim* bone matrix are essentially equivalent [372].

Bone mineralization is altered in all three OI murine models, all of which model the paradoxical hypermineralization of OI bone. The Brtl mouse demonstrates post-pubertal

adaptations in geometric parameters and material properties of Brtl femurs. From 2 to 3.5 months of age the aBMD of Brtl femurs increased at a rate greater than wild-type, although aBMD remains decreased relative to wild-type mice at 1, 2 and 6 months of age. However, increased vBMD at these ages relative to wild-type mice, with a more dense and homogenous mineral content of cortical and trabecular tissue, illustrates the hypermineralization of bone that is found in patients with OI [367,373,374]. Raman spectral analysis of Brtl bone also suggests a change in mineral:matrix ratios in Brtl; at 2 months of age the mineral:matrix ratio is greater than wild-type, but at 6 months of age the mineral:matrix ratio is significantly lower than wild-type [367]. An additional difference between 2- and 6-month bone that was revealed by Raman analysis was that the mineral component of Brtl bone has a greater carbonate-to-phosphate ratio than wild-type. It is possible that this represents an alteration in non-collagenous components of the extracellular matrix that influence mineralization. The OOA mouse has not been studied in the same detail as Brtl. It does share the reduced femoral aBMD measured by DXA analysis, and the increased mineral:collagen ratio measured by Raman, and also found in Brtl. Altered mineralization of *oim*/*oim* bone has also been reported. Femoral tissue from these mice has a mineral phase with abnormal chemical composition, a two-fold higher mineral:matrix ratio, lower mineral carbonate content and less crystalline structure [375,376]. In both cortical and trabecular bone qBEI analysis shows that bone tissue has a greater material density (increased Ca^{++} peak), while the heterogeneity of mineral content is decreased in cortical bone, but increased in trabeculae [369]. However, the hypermineralization of *oim/oim* bone does not result in increased material strength. Thus, while the brittleness of *oim/oim* bone is largely related to disorganization of the mineral

phase, the abnormal collagen scaffold serves as a template for abnormal mineral deposition. Hypermineralization of bone tissue in the mouse models of OI is consistent with higher mean mineralization density in OI patient bone samples measured by qBEI.

Histomorphometric analysis offers insight into cellular function in the Brtl and *oim* mouse models of OI. Brtl osteoblast surface/bone surface (ObS/BS), mineral apposition rate (MAR) and bone formation rate (BFR) are equivalent to wild-type at 2 months, suggesting quantitatively normal osteoblast matrix production at this age. However, Brtl mice demonstrate a postpubertal decline in MAR and BFR that reflects reduced matrix production by osteoblasts [377]. The concurrent disproportionate increase in osteoclast number and activity in postpubertal Brtl bone by a RANKL-independent mechanism suggests a secondary effect of the matrix on these cells [377]. These observations offer a possible explanation of how post-pubertal adaptions of Brtl bone occur, in which bone stiffness and strength attain wild-type values through stronger bone material properties, despite having reduced geometry [367]. Importantly, the concurrence of improved bone parameters and development of cellular uncoupling in postpubertal bone of Brtl mice offers insight into patients with dominant osteogenesis imperfecta, in which fracture frequency is noted to decrease after puberty [378]. Histomorphometric studies of *oim*/*oim* bone conclude that the fraction of mineralizing surface per bone surface (MS/BS) is significantly increased, while mineral apposition rate (MAR) and bone formation rate (BFR) are significantly decreased [379]. Concurrent with decreased osteoblast functioning, bone remodeling in *oim/oim* mice is increased, as determined by greater levels of bone degradation products in urine [380]. While the specific mechanism underlying dysregulated bone remodeling is unknown, increased osteoclast precursors with greater resorptive

activity in *oim/oim* have been observed [381], and may be due to a higher *Rankl/Opg* ratio and higher expression of TNFα in immature osteoblasts [382].

Although the cellular and tissue data from mouse models increase our understanding of OI pathophysiology, they are limited by an imperfect correlation with findings in patients with OI. For example, the mechanism of bone dysplasia in mice with collagen structural defects is not the same in each model and also differs from OI patients. Neither Brtl nor *oim* has femoral histomorphometry that is identical to that reported from two groups of children with types III and IV OI, who have high bone turnover with coupled osteoblasts and osteoclasts [383,384]. The murine models do mirror the changes evident in bone structural parameters reported in two studies analyzing iliac crest bone biopsies bone from OI pediatric populations [383,384]. Patients with collagen structural defects (types III and IV OI) in both studies demonstrated decreased structural parameters, including reduced CtWi, BV/TV, TbN and TbTh. Indicators of bone formation and turnover, such as decreased MAR and increased OcS/BS are found in both humans and mouse models of OI. In contrast to the mouse models, children with OI have increased ObS and BFR/BS, demonstrating the high turnover state of OI bone [383,384]. However, the *oim* mouse has a higher rate of bone turnover than is seen in OI children, with relatively greater elevation of ObS, OcS, and BFR [380], compared with normal controls. OI children also have decreased MAR [383], in contrast to *oim* osteoblasts which maintain an MAR comparable to WT mice [380]. In Brtl mice, there is an uncoupling of cellular responses; OcS is elevated, although ObS remains normal, and both bone formation rate and MAR are decreased. In OI proband bone indicators of bone formation and resorption are increased [377,383,384].

Collagen-Related Osteogenesis Imperfecta

In addition to the well-described collagen mutations, a new paradigm for osteogenesis imperfecta as a collagen-related bone disease has emerged with the discovery of novel gene defects. Among OI patients whose collagen modification was found to be normal, whole exome sequencing has identified mutations in *SERPINF1* and *IFITM5*. In 2000, a subset of OI patients exhibiting a dominant mode of inheritance in the absence of *COL1A1* and *COL1A2* mutations were classified as having type V OI [385]. These patients exhibited a distinct triad of distinguishing features, including calcification of the forearm interosseous membrane, a radiodense metaphyseal band at the growth plates of long bones and a tendency towards formation of hyperplastic callus at sites of bone trauma [385,386]. Histological examination of type V OI bone described an irregular "meshlike" pattern of lamellar bone under polarized light in these patients. In 2012, two groups identified an identical heterozygous mutation in the 5' untranslated region of *IFITM5* (interferon-induced transmembrane protein 5), encoding the 15kDa osteoblast cell surface-localized bone-restricted ifitm-like protein (Bril) [387,388]. The *IFITM5* c.-14C>T mutation, which introduces a novel upstream translation initiation codon, has since been found in all patients presenting with type V OI.

SERPINF1 encodes pigment epithelium-derived factor (PEDF), a 50-kDa secreted glycoprotein that binds type I collagen and functions as an inhibitor of angiogenesis [389]. Patients with deficiency of PEDF have a recessively inherited severe skeletal phenotype due to a mineralization defect, referred to as type VI OI [390]. Alteration of the orientation of bone lamellae in type VI OI is described as a "fish-scale" pattern on polarized light microscopy [391].

The majority of the recent novel gene defects identified in collagen-related OI are recessively inherited and involve proteins responsible for post-translational modification or processing of type I collagen. For example, defects in the specific pericellular C-propeptidase for type I collagen (bone morphogenetic protein 1, BMP1) have been associated with a high bone-mass form of OI [192]. Interestingly, a similar phenotype is found in OI patients with dominant mutations in the C-propeptidase cleavage site of type I procollagen [191]. Also, a mutation in *SERPINH1*, which encodes a collagen-specific folding chaperone (HSP47), has been characterized in a single case of severe recessive OI due to homozygosity for a missense mutation (p.Leu78Pro) [141]. A missense mutation of *SERPINH1* results in a similar severe phenotype in dachshunds [392]. Homozygosity for a null *SerpinH1* allele in mice results in a lethal phenotype [131]. Deficiency of HSP47 causes delayed cellular secretion of collagen, which shows increased susceptibility to proteases *in vitro*, supporting a role for this protein in intracellular stabilization of the collagen helix. Mutations in *FKBP10*, which encodes the collagen chaperone and peptidyl-prolyl *cis-trans* isomerase FKBP65, result in a range of phenotypes that includes OI, OI with contractures (Bruck syndrome type 1), and a contracture syndrome with osteopenia (Kuskokwim disease) [102,393,394]. Defects in FKBP65 result in deficiency of collagen telopeptidyl hydroxylation, a modification known to be catalyzed by lysyl hydroxylase 2 (LH2; *PLOD2*). The overlapping phenotypes of Bruck syndrome type 2, caused by LH2 deficiency, with FKBP65 deficiency suggests that interaction of these proteins is required for LH2 activity, or alternatively, that FKBP65 isomerase activity is required for the correct collagen substrate conformation to allow LH2 recognition [393,394].

Collagen 3-Hydroxylation Defects: Recessive OI

Recessive forms of OI were first postulated in the original Sillence classification. Although additional gene defects have since been identified, it was the discovery that deficiency of proteins involved in collagen prolyl 3-hydroxylation underlie the majority of recessive cases that generated the new paradigm for OI. This expanded view of OI as a collagen-related bone dysplasia was the result of the convergence of four independent lines of investigation incorporating bone histomorphometry, collagen biochemical analyses, genetic mouse models and clinical molecular genetics approaches to unsolved cases of severe to lethal bone disease. The importance of the collagen 3-hydroxylation complex to collagen assembly has been made clear by the characterization of a *Crtap* knock-out mouse, and null mutations of the genes encoding P3H1, CRTAP and CyPB in patients with severe or lethal bone dysplasia.

Morello and colleagues demonstrated that CRTAP was important for development by characterizing its expression in several chick, murine and human tissues [395-397]. *Crtap* was first identified in a subtracted cDNA library during screening for uniquely expressed genes in hypertrophic chick chondrocytes in culture [395]. Northern analysis of mRNA from various embryonic chick tissues detected *Crtap* transcripts in tissues including skin, heart and lungs. Using an antibody raised against a short peptide sequence from the chick protein, *Crtap* expression was detected in tibial and femoral epiphyses, articular cartilage and lower hypertrophic cartilage. In human tissues, expression of CRTAP was identified in heart, lung, small intestine and skeletal tissue [397]. A more direct role for *Crtap* in skeletal development was suggested by subsequent localization of transcripts to several murine tissues by *in situ* hybridization using the full-length cDNA as a probe. Murine

tissues containing high levels of *Crtap* transcripts corresponded to tissues also containing types I and II fibrillar collagens, such as the growth plate, bone collar and calcified cartilage of the chondro-osseous junction [107]. This distribution of CRTAP was confirmed by immunohistochemical analysis using an antibody directed against CRTAP-specific peptides. Although light staining for CRTAP protein was identified in the cartilage matrix by immunohistochemistry, the majority of CRTAP was intracellular. Accordingly, co-transfection of murine chondrocytes with myc-tagged full-length *Crtap* cDNA and ER-targeted GFP showed colocalization of CRTAP with the endoplasmic reticulum. More specifically pertaining to bone, RT-PCR analysis confirmed expression of *Crtap* in both osteoblasts and osteoclasts [107].

Prolyl 3-hydroxylase 1 (P3H1) is a multifunctional protein that has now been independently identified three times. It was first isolated from a rat yolk sac tumor cell line as a novel matrix chondroitin sulfate proteoglycan [398]. A 100kDa protein was purified from a high-density proteoglycan fraction of conditioned media and used for generation of a polyclonal antibody. This antibody was subsequently used for immunoscreening of a rat cDNA library. The authors isolated a single clone, found to contain a fragment coding for the COOH terminal region of a novel cDNA. The sequence was used to isolate a full-length transcript by a cDNA capture assay, producing a 3.1 kb transcript predicted to encode a 728 amino acid leucine-proline enriched proteoglycan. The COOH terminal fragment was translated for production of a peptide to generate a polyclonal antibody for immunodetection. Termed Leprecan, the proteoglycan was detected in the ER and Golgi of CHO cells transfected with the full-length cDNA. Furthermore, despite containing a carboxyl terminus KDEL endoplasmic reticulum

retrieval signal, Leprecan was shown to be secreted by rat L2 cells and localized to basement membrane of several rat tissues by immunostaining. Using the polyclonal antibody recognizing the C-terminus to Leprecan, strong staining was seen in basement membranes of the vasculature and smooth muscle of several organs, including glomerulus, mesangial matrix and tubules of kidney, liver, and skeletal muscle, as well as the pericellular matrix surrounding chondrocytes in cartilage.

The human homolog of Leprecan, Gros1, was isolated as a potential growth suppressor located at chromosome 1p31 [399]. In this study two alternatively spliced transcripts encoding 41- and 84-kDa proteins denoted as Gros1-S and Gros1-L, respectively, were identified. The first 360 residues of the 736 amino acid, 84 kDa isoform are identical to the predicted translated sequence of the Gros1-S variant. Gros1-S, however, utilizes a cryptic exon (exon 5b) in intron 5 while also retaining the entire intron 7 sequence, leading to a premature stop codon and, presumably, the generation of a truncated 363 amino acid peptide. This smaller variant would lack the carboxyl leucine repeat and P4H domain. Semiquantitative PCR analysis has demonstrated that Gros1-L is the dominantly expressed form in adult human tissues, such as heart, placenta, lung and kidney [400]. In human fetal tissues Gros1-S predominates in brain and thymus, while both forms of the transcript are seen in equimolar ratios in liver, skeletal muscle, spleen and pancreas. It is important to note, however, that expression of the Gros1S variant has not been confirmed at the protein level.

The isolation of the protein responsible for collagen 3-hydroxylation was accomplished when Bachinger and colleagues purified P3H1 [106]. By gelatin affinity chromatography, several proteins from rER extracts of chick embryos, including P3H1,

were shown to interact with denatured collagen alpha chains. A low pH elution fraction of the partially purified rER extract was then used to generate a monoclonal antibody to P3H1. Subsequent coelution of P3H1, CRTAP and CyPB from a P3H1 affinity column suggested the presence of a collagen-interacting complex in the ER. Immunostaining localized P3H1 to tissues expressing fibrillar collagens, such as dermis, cartilage and tendon. Additional staining was observed in kidney calyx, interlobular septum of the liver and smooth muscle layer of the aorta. In contrast to the staining pattern for Leprecan in rat tissues [398], no staining for P3H1 was seen in kidney cortex, liver parenchyma and skeletal muscle. These differences in immunostaining may be due to the use of different species (rat versus chicken) at different developmental stages (postnatal versus embryonic). More importantly, the antibody used to stain rat tissues recognizes the carboxyl half of the protein, which has a high degree of homology to the carboxyl halves of the P3H2 and P3H3 isoforms [2]. The localization issue has been addressed by Myllaharju and coworkers, who have investigated the substrate specificity of the P3H2 isoform and found that it preferentially 3-hydroxylated synthetic peptides corresponding to sequences found in type IV collagen compared to a peptide containing the type I collagen 3-hydroxylation site [400]. Since basement membranes are rich in type IV collagen, it is possible that P3H1 is the primary isoform expressed in tissues rich in type I collagen while P3H2 functions in basement membranes. Support for this is seen in the pattern of immunostaining with polyclonal antibodies generated from synthetic peptides corresponding to regions of nonidentity in the murine P3H1 and P3H2 sequences. It was shown that P3H2 was strongly expressed in mouse tissues where basement membranes,

and therefore type IV collagen, are found, while staining for P3H1 was weak in these tissues [400].

The third member of the 3-hydroxylation complex is peptidyl-prolyl isomerase B (PPIB), which does not have any structural similarities to CRTAP or the prolyl 3-hydroxylases (P3H1-3). Peptidyl-prolyl *cis-trans* isomerase B (PPIB), also known as Cyclophilin B, is an immunophilin, a family of proteins with PPIase activity. Immunophilins are responsible for the catalysis of *cis-trans* isomerization of peptidyl-prolyl bonds and are divided into three groups: cyclophilins, which are inhibited by cyclosporine A (CsA), FK506-binding proteins, and the parvulins. PPIB is known to play a central role in the *in vivo* folding of procollagens in the endoplasmic reticulum. Procollagen alpha chain association occurs at the carboxyl-terminal propeptides, which are the last portions of the polypeptide to be translated. Although the collagen triple helix can only accommodate peptide bonds that are in the *trans* conformation, there is sufficient time for the formation of an equilibrium state of unfolded alpha chains in which as much as 12% of X-Pro bonds in type I procollagen are in the *cis* conformation, as determined by nuclear magnetic resonance [401]. Hence, it is the function of peptidyl-prolyl *cis-trans* isomerase to catalyze the efficient folding of procollagen alpha chains into the triple helical structure. PPIase activity is especially important for type I collagen folding, since nearly one-sixth of its primary sequence consists of proline residues.

Evidence in support of the role of PPIB in collagen folding has been presented by Bachinger and colleagues, who used partially purified enzyme to study the *in vitro* rate of folding of type III procollagen [402]. In this study the authors observed the formation of collagen helices in the presence and absence of PPIB, monitored by optical rotary

dispersion, as well as formation of protease-resistant collagen helices determined by PAGE analysis. It was concluded that PPIB increased the rate of folding 3-fold, and that *cis-trans* isomerization of the X-Pro bonds in collagen alpha chains is the rate-limiting step of helical folding [403]. Complementary findings by Steinmann and colleagues supported an *in vivo* role of PPIB [108] in collagen folding. Treatment of fibroblasts with cyclosporine A (CsA) induced overmodification of type I collagen along the full length of the alpha chains. Since CsA binds to and inhibits the isomerase activity of PPIB by an incompletely characterized mechanism [404], it is believed that loss of *cis-trans* prolyl isomerization and subsequent delayed folding results in collagen overmodification. This was supported by amino acid analysis, which demonstrated an increase in the proportion of hydroxylysine residues from 21% of lysyl residues in normal control collagen to 28% of lysyl residues in treated cells. Furthermore, electrophoretic mobility of collagens synthesized during inhibition of modification by α,α'-dipyridyl in the presence of CsA was indistinguishable from collagens synthesized in the presence of α,α'-dipyridyl alone, showing that the delayed electrophoretic mobility of collagens synthesized in the presence of CsA was due to overmodification. These data demonstrate that direct inhibition of PPIB results in collagen overmodification and imply that helix folding is delayed. However, it has yet to be determined whether treatment with CsA disrupts the functioning of PPIB as a component of the 3-hydroxylation complex, or if CsA blocks PPIB from becoming a component of the complex. It should also be noted that CsA is a nonspecific inhibitor, and that its effect on collagen folding may not be due solely to inhibition of PPIB/CyB. Furthermore, defects in CRTAP or P3H1 have not yet been linked to defective isomerization of proline residues in collagen.

Cartilage-Associated Protein (CRTAP) Deficiency

Identification of the first human cases of CRTAP deficiency paralleled characterization of the *Crtap* knockout mouse [109]. To elucidate the *in vivo* function of CRTAP, a knock-out mouse was generated by replacing the first two exons of the *Crtap* gene with a neo cassette [107]. This mouse model of *Crtap* deficiency had recessively inherited impaired bone development characterized as an osteochondrodysplasia, which included growth deficiency, osteopenia, shortening of the femurs (rhizomelia) and progressive kyphosis. Consistent with localization of *Crtap* to the growth plate, the proliferating chondrocytes of the *Crtap*$^{-/-}$ mouse were disorganized. Absence of CRTAP resulted in loss of 3-hydroxylation of the P986 residue in α1(I) chains of skin and bone, as well as α1(II) collagen chains of cartilage, demonstrating CRTAP's *in vivo* role as a crucial component of the collagen 3-hydroxylation complex [107,405]. The electrophoretic migration of collagen synthesized by *Crtap*$^{-/-}$ calvarial osteoblasts was examined; collagen bands were broad, consistent with post-translational overmodification. In addition, collagen fibrils in the extracellular matrix of skin and bone were reported to have increased diameters [107].

Loss or decrease of type I collagen prolyl 3-hydroxylation in *Crtap*$^{-/-}$ mice has revealed multiple abnormalities of connective tissue, especially abnormal structural and material properties of bone [107,405]. At the tissue level, absence of *Crtap* results in deficient production of osteoid and excess mineral in bone tissue. In 12-week trabecular bone BV/TV, TbTh and TbN are decreased 62%, 37% and 50%, respectively, while TbSp is twice that of wild-type mouse trabecular bone. The reduced structural parameters correlate with static and dynamic histomorphometric analysis, which demonstrates

reduced bone formation parameters in *Crtap*$^{-/-}$ mice [107]. Both ObS/BS and MS/BS are normal, indicating normal osteoblast numbers. However, the mineral apposition rate (MAR) and bone formation rate (BFR/BS) are decreased 37% and 50%, respectively. Also, indices of matrix deposition *in vivo*, such as osteoid thickness (OTh), osteoid surface/bone surface (OS/BS), osteoid volume/bone volume (OV/BV) and mineralization lag time (Mlt) were significantly decreased [107]. Hence, the osteoblasts in Crtap$^{-/-}$ bone tissue are producing matrix at a slower rate, but the matrix is mineralizing faster than in wild-type bone. These findings are supported by the BMDD in femurs from *Crtap* deficient mice, which reveals increased mineral content compared to wild-type mice [406]. As expected, the heterogeneity of mineral (CaWidth) and the fraction of low mineralized bone matrix (CaLow) were highest in metaphyseal bone and lowest in the cortical midshaft, since metaphyseal bone is metabolically more active than cortical bone. However, in both cortical and trabecular bone the average (CaMean) and most frequently measured (CaPeak) calcium concentrations are increased in *Crtap*$^{-/-}$ as compared to wild-type bone, indicating either a decrease of bone turnover or an alteration in mineralization kinetics. However, since surface-based parameters of bone remodeling (i.e., osteoclast numbers and activity) are normal in the *Crtap*$^{-/-}$ mice, the increase in mean and peak calcium concentrations is consistent with an increased rate of mineralization [107]. This is presumably due to either a loss of the 3-hydroxylation modification or an altered collagen:noncollagenous protein ratio in bone tissue. Thus the bone phenotype of the *Crtap* knockout mouse generated interest among researchers investigating potential candidate genes for recessive OI.

With the development of diagnostic methods to completely sequence the type I collagen genes from patients with osteogenesis imperfecta, it became clear that 10-15% of these patients did not have collagen mutations. A subset of these patients were also shown to synthesize overmodified collagens despite the lack of a defect in collagen primary structure, suggesting that overmodification was caused by a defect in a protein(s) that interact(s) with collagen in the endoplasmic reticulum [2]. In addition to the recessive osteochondrodysplasia of the *Crtap*$^{-/-}$ mouse and the association of P3H1 and CRTAP with denatured collagen, the mapping of type VIII OI and *CRTAP* to chromosome 3p22 [397,407] suggested that defects in components of the collagen 3-hydroxylation complex might cause recessive osteogenesis imperfecta.

The first cases of recessive OI with collagen overmodification and absence of a primary structural defect were identified in three probands with severely reduced or absent *CRTAP* mRNA in fibroblasts by real-time RT-PCR [109]. Sequencing of the gene in these patients identified mutations consistent with loss of expression of *CRTAP*. Furthermore, in fibroblasts of all probands no CRTAP protein was detected by western analysis. In both the *Crtap* knock-out mouse and patients with null mutations of the *CRTAP* gene, the resulting absence of this protein abrogates 3-hydroxylation of type I collagen; only 0-21% of α1(I) Pro986 residues were 3-hydroxylated as demonstrated by tandem mass spectrometry of trypsin-digested collagen peptides [107,109].

It was demonstrated that loss of CRTAP protein resulted in delayed folding of the collagen triple helix; collagen alpha chains exhibited delayed electrophoretic migration on PAGE analysis. Amino acid analysis of type I collagen from CRTAP-null fibroblasts revealed an increase in lysyl hydroxylation compared to normal control collagen (31-35%

versus 24% respectively) [109]. The additional hydroxylysine residues provide increased substrate for glycosylation, which is consistent with the 1°C increase in thermal stability of the patient type I collagens on differential scanning calorimetry. These findings showed that absence of CRTAP protein or loss of the putative collagen chaperone function of the 3-hydroxylation complex causes severe to lethal bone dysplasia. An additional potential mechanism by which loss of the 3-hydroxylation modification causes bone dysplasia was demonstrated by Valli and coworkers, who observed by immunofluorescence of long-term cultures severe deficiency (10-15% of control) and disorganization of collagen deposited in extracellular matrix of CRTAP-deficient fibroblast cultures [408]. Interestingly, CRTAP deficiency in humans results in a more severe phenotype than in the knock-out mouse.

Prolyl 3-Hydroxylase 1 (P3H1) Deficiency

Demonstration that deficiency of CRTAP resulted in severe to lethal OI suggested that the other components of the 3-hydroxylation complex were candidates for remaining cases with suspected recessive inheritance. Shortly after reports of CRTAP deficiency as the cause of type VII OI, screening for loss of expression of *LEPRE1* (P3H1) in patient fibroblasts identified an additional cohort with 3-hydroxylation defects [110]. In all cases, severely reduced or absent mRNA could be partially rescued by treatment of the fibroblasts with emetine, suggesting that mutations in *LEPRE1* triggered nonsense-mediated decay (NMD) of transcripts due to the presence of a premature termination codon (PTC). Sequencing of the 15 exons and surrounding intronic regions of *LEPRE1* genomic DNA from each of these cell lines revealed the causative mutations, all of which

are consistent with the reduction of *LEPRE1* transcripts, as the resulting frameshifts lead to introduction of a PTC.

The first report of P3H1 deficiency as the cause of type VIII OI noted a common mutation, *LEPRE1* IVS5+1G>T (c.1080+1G>T), in seven of the 14 alleles identified [110], all from infants of African-American or contemporary West African descent. The recurrence of this mutation was noted in additional probands of African-American and Caribbean descent and has been identified in approximately one quarter (22/86) of mutant *LEPRE1* alleles reported to date [409-412]. Subsequent screening of contemporary West African and African-American cohorts using a high-throughput SNP assay yielded a 0.40% carrier frequency for the recurring mutation among Mid-Atlantic African Americans, predicting an incidence of type VIII OI at approximately 1 in 250,000 African American births [413]. In addition, nearly 1.5% of contemporary West African

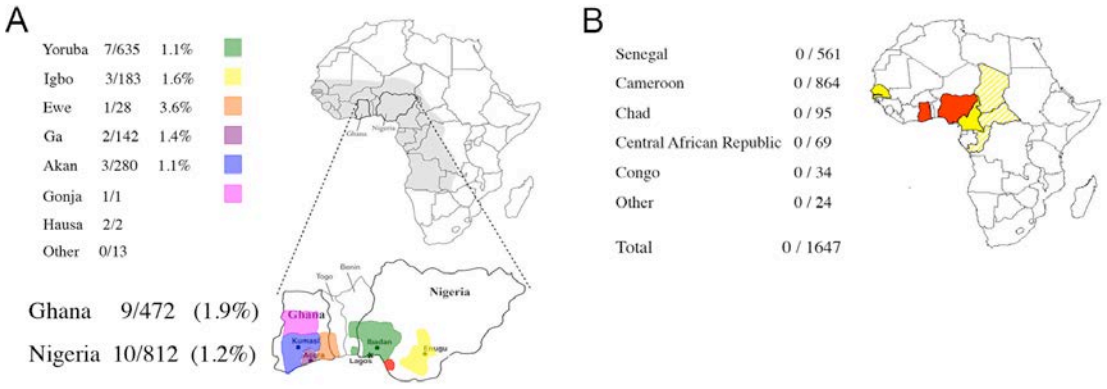

Figure 13. West African Origin of *LEPRE1* c.1080+1G>T Mutation. Homozygosity for the c.1080+1G>T mutation causes type VIII osteogenesis imperfecta. Approximately 0.4% of African-Americans from the Mid-Atlantic United States, and 1.5% of West Africans from Ghana to Nigeria are carriers for this mutation. The mutation was not identified in genomic DNA samples from any African regions outside of West Africa, suggesting that it is not a Pan-African

genomic DNA samples representing several ethnic groups from Ghana, Benin and Nigeria were identified as heterozygous for this mutation (Figure 13), predicting an incidence of type VIII OI due to homozygosity for this mutation at approximately 1 in

18,000 births [413], equivalent to the rate of dominant OI in American and European Caucasian populations. Haplotype analysis of 15 African American and West African families support the conclusion that this founder mutation arose between the 12th and 14th centuries and was transported to the Americas with the Trans-Atlantic slave trade [413].

Consistent with the mutations and resulting degradation of transcripts, P3H1 protein was absent from proband fibroblasts on Western analysis. Only one of these patients had a residual level of protein from his paternal allele, presumably translated from the small amount of in-frame transcripts that were detected. It was also confirmed that absence of P3H1 abolished 3-hydroxylation of the α1(I) Pro986 residue in type I collagen. The 3-hydroxylation modification at this residue was detected by mass spectrometry in 0-21% of α1(I) chains from the collagens synthesized by patient fibroblasts. Deficiency of P3H1 also affects other modifications of the helical region of the type I collagen chains. Type I collagen chains synthesized by fibroblasts derived from each patient migrate as broad bands on SDS-gels, consistent with overmodification by lysyl hydroxylase and the glycosylating enzymes. Amino acid analysis showed hydroxylation of more helical lysine residues in patient collagen than in a control sample. In collagen secreted from P3H1-null fibroblasts, 32-35% of lysyl residues were hydroxylated, compared to 24% hydroxylation of lysine residues in normal control collagen. This extent of lysyl hydroxylation is comparable to the level of overmodification found in collagen with a structural defect near the carboxyl end of the helix, as well as with patients with CRTAP deficiency, and show an increased melting temperature of 1°C compared to normal control, which is the expected consequence of collagen overmodification in the absence of a primary structural defect [109,414]. Since these data are consistent with longer exposure of the collagen chains

to lysyl hydroxylase it has been concluded that loss of P3H1 protein or its role in the 3-hydroxylation complex results in delayed folding of the collagen helix, which results in the overmodification. Of note, no discernable collagen modification abnormalities have been detected in the parents who are heterozygous carriers for the *LEPRE1* mutations; gel migration of parental collagen is normal, as are the levels of *LEPRE1* mRNA on real-time RT-PCR. Parental 3-hydroxylation was only slightly reduced to 85-90% of Pro986 residues versus normal control collagen.

There is an additional metabolic aspect unique to recessive OI caused by deficiency of collagen 3-hydroxylation components. In classical OI, overmodified collagen resulting from glycine substitutions is secreted slower than from normal control cells [415]. Pulse-chase experiments comparing fibroblast collagen synthesis from cases with P3H1 deficiency and normal controls to observe collagen secretion kinetics demonstrate that, similar to classical OI, there is a 15-20 minute delay in the rate of type I collagen secretion from P3H1-null fibroblasts [110]. However, in contrast to dominant OI, total collagen secretion per cell by P3H1-deficient cells is increased 20-50% compared to normal controls. Slower secretion of overmodified collagens is expected to cause ER stress with upregulation of the Unfolded Protein Response (UPR) pathway [416-418]. Since the UPR response results in downregulation of translation, the occurrence of increased collagen synthesis in P3H1-deficient fibroblasts is counterintuitive but may hint to a direct or indirect role for P3H1 in regulating collagen synthesis and secretion, as well as proline metabolism [110].

The *P3h1*$^{-/-}$ mouse, which contains a frameshifting deletion of exons 1-3 of the *Lepre1* gene, has a phenotype similar to that of the *Crtap*$^{-/-}$ mouse. As with the *Crtap*$^{-/-}$

mouse, collagen secretion and electrophoretic migration on PAGE analysis is delayed [419]. These findings are consistent with the overmodification of lysine residues of type I collagen that was found in tissues; collagens have a 32-42% increase in lysyl hydroxylation, and 49-140% increase in glucosylgalactosyl-hydroxylysine content in skin and bone, respectively. P986 3-hydroxylation was completely absent in type I collagen from tendon, skin and bone [419,420]. Collagen fibrils in dermal tissue of *P3h1*$^{-/-}$ mice have normal diameters, but are less dense and are interspersed with aggregates compared to wild-type skin, while fibrils from tendon are increased in diameter. Similar to observations in the *Crtap*$^{-/-}$ mouse, the net phenotypic effect of absence of the P3H complex components and the abnormal collagen that is deposited into extracellular matrix of *P3h1*$^{-/-}$ mice is 10-25% growth deficiency from 3 to 13 weeks of age, progressive kyphoscoliosis and rhizomelia. Long bone parameters are decreased, including a 7% reduction in CtA and 6% reduction in CtTh at 20 weeks of age. In addition to the 14% decrease in volumetric BMD, femurs from 20-week mice demonstrate 12-13% reductions in stiffness and force to failure, respectively, again demonstrating the weakness of bone tissue as a result of loss of prolyl 3-hydroxylase complex activity. However, a more detailed analysis of bone formation parameters, including mineral density and histomorphometric measurements, is lacking for this mouse model of recessive OI.

Cyclophilin B Deficiency

A total of eight human cases with *PPIB* mutations, causing type IX OI, have been reported [103,421-424]. The majority of these mutations lead to PTCs or misfolded protein and

result in a severe to lethal phenotype [422,423]. The probands with CyPB deficiency lack the overlapping phenotypic and collagen biochemical findings that are consistent with a dysfunctional 3-hydroxylation complex and observed for types VII and VIII OI, including the occurrence of rhizomelia. Of the four cases that have been investigated beyond mutation identification, two lethal cases have mutations that decrease collagen 3-hydroxylation to 30-33% of α1(I) P986 residues. The two other cases of CyPB deficiency present with a moderate phenotype and normal levels of α1(I) P986 hydroxylation. The most interesting of the moderate cases, a *PPIB* mutation (c.26T>G) in siblings with moderate OI, defines the *PPIB* start codon [109]. Although the NCBI Reference Sequence (NP_000933.1) for CyPB protein predicts a M9R substitution, it was never demonstrated which nucleotides of the cDNA sequence were utilized as the initiation codon [425]. However, M1 is not evolutionarily conserved among species and expression plasmids lacking M1 produce functional protein [426]. Furthermore, the complete absence of CyPB protein in guanidinium extracts of proband cell extracts after proteasomal inhibition demonstrated that *PPIB* c.25-27 was indeed the actual start codon [425]. P3H1 and CRTAP levels are one-third and two-thirds normal in the *PPIB*-null cells by western analysis, demonstrating a non-critical role of CyPB in enhancing complex stability. However, despite the absence of CyPB, P3H complex activity is sufficient to normally hydroxylate α1(I) Pro986. This data suggests that partial modification of α1(I) P986 in two of the reported cases of PPIB deficiency may be a result of a truncated CyPB protein product that interferes with complex function. Furthermore, normal collagen modification in the total absence of CyPB suggests a normal rate of folding, and supports the existence of additional *cis-trans* isomerases involved in collagen folding in human cells [421].

As with humans, absence of CyPB protein causes severe bone dysplasia in mice. A *Ppib* knockout mouse containing an out-of-frame exon 2/4 junction has non-lethal recessive OI [112]. Although CyPB is undetectable on western blots of thymocyte, splenocyte and embryonic fibroblast lysates, *Ppib* transcripts are only reduced to 25% versus wild-type in splenocyte cultures. In contrast to the type IX OI cases, murine collagen biochemistry differs from both human PTC and start codon mutations. Collagen migration demonstrates a delayed baseline rather than a broadened band on PAGE analysis, but helical modification was not quantitated in the mouse collagen. Pro986 3-hydroxylation was absent in collagen from murine skin, bone and tendon, although partial persistence of complex components is suggested by detection of reduced levels of P3H1 in murine fibroblasts. The differences in collagen modification between the KO mouse and human start codon mutation may reflect alternative binding partners supporting P3H1/CRTAP stability or redundancy for PPIases in humans. Furthermore, procollagen is retained in the ER and is not properly transported into the Golgi, as determined by fluorescent microscopy. At the tissue level, dermal collagen fibrils from CyPB-deficient mice have diameters that are increased by 50% compared to wild-type collagen dermal fibrils in longitudinal sections. Phenotypically, this mouse shows no size difference with wild-type littermates at weaning, but kyphosis and growth deficiency progress with age until knockout mice weigh 31% (10g) less than wild-type at 6 months of age. From 7 weeks to 29 weeks of age, these mice have reduced trabecular bone volume (BV/TV, ↓50-75%) due to increased trabecular spacing (TbSp, ↑25-40%) and reduced trabecular number (TbN, ↓10-33%) [112].

An additional animal model with a recessively inherited *PPIB* mutation exists in American Quarter Horses, which contain a missense mutation resulting in the skin condition hereditary equine regional dermal asthenia (HERDA) [427]. The mutation, a *PPIB* c.115G>A missense mutation, results in an amino terminus p.Gly6Arg substitution and occurs with a 17.3% frequency in the Quarter Horse population. Horses homozygous for the HERDA mutation are normal at birth, but by two years of age begin to present with skin ulceration and poorly resolving wounds with skin sloughing. No bone phenotype has been reported in these animals. Although the p.Gly6Arg substitution does not interfere with peptidyl-prolyl isomerase activity of recombinant mutant protein against a succinyl-Ala-Ala-Pro-Phe-methyl coumarylamide substrate *in vitro*, cell culture assays demonstrate delayed folding of type I collagen in HERDA fibroblasts. Rather than delayed electrophoretic migration of collagen due to overmodification caused by delayed folding, dermal collagen from the HERDA horse migrates faster than wild-type collagen, while tendon-derived fibrils are smaller and branched. The apparent undermodification of collagen in the presence of CyPB p.Gly6Arg protein has been hypothesized to be due to the loss of an interaction between mutant CyPB and lysyl hydroxylase 1 (LH1), as ascertained in CyPB pull-down assays [427]. It is unknown, however, if CyPB is required for LH1 activity or if isomerization of peptidyl-prolyl bonds in the collagen substrate is required for LH1 binding.

Chapter II

A Founder Mutation in *LEPRE1* Carried by 1.5% of West Africans and 0.4% of African Americans Causes Lethal Recessive Osteogenesis Imperfecta

Wayne A. Cabral[1], Aileen M. Barnes[1], Adebowale Adeyemo[2], Kelly Cushing[3], David Chitayat[4], Forbes D. Porter[5], Susan R. Panny[6], Fizza Gulamali-Majid[7], Sarah A. Tishkoff[8], Timothy R. Rebbeck[9], Serigne M. Gueye[10], Joan E. Bailey-Wilson[11], Lawrence C. Brody[3], Charles N. Rotimi[2], and Joan C. Marini[1].

[1]Bone and Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD 20892
[2]Center for Research on Genomics and Global Health, NHGRI, NIH, Bethesda MD 20892
[3]Genome Technology Branch, NHGRI, NIH, Bethesda, MD, 20892
[4]The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital; Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children; Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, M5G 1X8
[5]Program in Developmental Genetics and Endocrinology, NICHD, NIH, Bethesda, MD 20892
[6]Office for Genetics and Children with Special Health Care Needs, Maryland Department of Health and Mental Hygiene, Baltimore, MD 21201
[7]Laboratories Administration, Maryland Department of Health and Mental Hygiene, Baltimore, MD 21201
[8]Departments of Genetics and Biology, University of Pennsylvania, Philadelphia, PA 19104
[9]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine and Abramson Cancer Center, Philadelphia, PA 19104
[10]Department of Urology-Andrology, Hopital General de Grand Yoff, Dakar, Senegal
[11]Inherited Disease Research Branch, NHGRI, NIH, Baltimore, MD 21224

Communicating Author:
Joan C. Marini, MD, PhD
Chief, Bone and Extracellular Matrix Branch
NICHD, NIH
Bldg. 10; Rm 10N260
9000 Rockville Pike
Bethesda, MD 20892

ABSTRACT

Purpose: Deficiency of prolyl 3-hydroxylase 1, encoded by *LEPRE1*, causes recessive osteogenesis imperfecta. We previously identified a *LEPRE1* mutation, exclusively in African Americans and contemporary West Africans. We hypothesized that this allele originated in West Africa and was introduced to the Americas with the Atlantic slave trade. We aimed to determine the frequency of carriers for this mutation among African Americans and West Africans, and the mutation origin and age. **Methods:** Genomic DNA was screened for the mutation using PCR and restriction digestion, and a custom TaqMan genomic SNP assay. The mutation age was estimated using microsatellites and short tandem repeats spanning 4.2 Mb surrounding *LEPRE1* in probands and carriers. **Results:** Approximately 0.4% (95% C.I.: 0.22-0.68%) of Mid-Atlantic African Americans carry this mutation, estimating recessive OI in 1/260,000 births in this population. In Nigeria and Ghana, 1.48% (95% C.I.: 0.95-2.30%) of unrelated individuals are heterozygous carriers, predicting 1/18,260 births will be affected with recessive OI, equal to the incidence of *de novo* dominant OI. The mutation was not detected in Africans from surrounding countries. All carriers shared a haplotype of 63-770 Kb, consistent with a single founder for this mutation. Using linkage disequilibrium analysis, the mutation was estimated to have originated between 650 and 900 years before present (1100-1350 C.E.). **Conclusions:** We identified a West African founder mutation for recessive OI in *LEPRE1*. Nearly 1.5% of Ghanians and Nigerians are carriers. The estimated age of this allele is consistent with introduction to North America via the Atlantic slave trade (1501 – 1867 C.E).

Key Words: *LEPRE1*, osteogenesis imperfecta, founder mutation, West Africa

INTRODUCTION

Classical osteogenesis imperfecta (OI), or "brittle bone disease", is a connective tissue disorder characterized by susceptibility to bone fractures, blue sclerae and growth deficiency [352]. Most (85%) individuals with OI have autosomal dominant mutations in *COL1A1* or *COL1A2* (OMIM #166200, #166210, #166220, #259420), which alter the structure or synthesis of type I collagen [116]. Collagen mutations cause a range of OI phenotypes, from mild to perinatal lethality, described by the Sillence Classification [348]. Autosomal dominant OI has an incidence of approximately 1/15-20,000 births, with over 90% of cases resulting from *de novo* mutations [348,349,351]. Mutation "hotspots" in *COL1A1* and *COL1A2* are associated with independent recurrences of mutations, rather than founder mutations [116].

Type I collagen is a heterotrimer of two α1(I) and one α2(I) chains which is subject to post-translational modification during chain synthesis and folding by prolyl 4-hydroxylase (P4H) and lysyl hydroxylase (PLOD1), with subsequent glycosylation of some hydroxylysine residues [9]. An additional post-translational modification system exists for types I, II and V collagen, which results in 3-hydroxylation of the α1(I) and α1(II) Pro986 residues [428]. Partial modification of the α1(II) Pro944 and α2(I) Pro707 residues has also been reported [429]. These modifications occur in the endoplasmic reticulum by the collagen prolyl 3-hydroxylation complex, consisting of prolyl 3-hydroxylase 1 (P3H1, encoded by *LEPRE1*), cartilage-associated protein (CRTAP) and cyclophilin B (CyPB/*PPIB*). Deficiency of proteins comprising the collagen prolyl 3-hydroxylation complex causes autosomal recessive OI. Null mutations in *CRTAP* (type VII OI, OMIM # 610682), *LEPRE1* (type VIII OI, OMIM# 610915) and *PPIB* (type IX

OI, OMIM #259440) cause a moderate to lethal osteochondrodystrophy that overlaps phenotypically with dominant types II, III and IV OI, but has several distinctive clinical features [107,109,110,409,421,422], including white sclerae, rhizomelia and extreme bone undermineralization. Surviving children with type VIII OI also present with extreme growth deficiency and bulbous metaphyses [110,409]. Recessive OI caused by deficiency of components of the collagen 3-hydroxylation complex accounts for about 5-7% of severe cases of OI in North America [109,410].

Among the first cases of type VIII OI reported, we identified homozygosity for the same mutation within several African American and West African émigré families in North America [110]. This mutation accounts for about one-third (21/62) of all mutant alleles currently reported in type VIII OI [430,431]. The mutation (*LEPRE1* c.1080+1G>T) results in multiple alternatively spliced transcripts, each containing a premature termination codon (PTC). Homozygosity for this mutation is perinatal lethal, while compound heterozygosity with a different *LEPRE1* mutation is compatible with survival into the second decade of life, and heterozygous carriers are clinically unaffected [110]. We hypothesized that the *LEPRE1* c.1080+1G>T allele arose in West Africa prior to the African diaspora and was introduced to the Americas during the Atlantic slave trade. We screened contemporary African American and African cohorts to determine the prevalence of carriers and illuminate the history of this mutation. Haplotype analysis of mutant alleles demonstrated one shared core haplotype surrounding the *LEPRE1* gene on chromosome 1. Our findings are consistent with a founder mutation in West Africa more than 650 years ago, and was transported to the Americas during the transatlantic slave trade.

MATERIALS AND METHODS

Cohort Selection and Acquisition

African American cohorts from Pennsylvania (NeoGen)[432] and Maryland (Maryland Department of Health and Mental Hygiene) contained randomly chosen anonymized newborn metabolic screening cards with parental racial designation. Because samples were anonymized it was not possible to exclude twins or multiple samples within a family. Leukocyte genomic DNA from African Americans and contemporary Africans enrolled in the African American Diabetes Mellitus Study (AADM)[433] and Howard University Family Study of Hypertension (HUFS)[434] were obtained from the District of Columbia, and the West African countries of Ghana and Nigeria. Related individuals in AADM and HUFS were excluded by pedigree analysis. Genomic DNA from individuals originating from the Central African countries of Cameroon, Chad, Central African Republic and Congo was collected in Cameroon for an African genetic diversity study [435]. Senegalese samples were collected for a prostate cancer susceptibility study in apparently healthy individuals [436]. All samples were collected through IRB-approved protocols.

Sample Screening

DNA was extracted from newborn metabolic screening cards and screened for the *LEPRE1* mutation by PCR and restriction enzyme digestion with *Bsl*I (New England Biolabs, Ipswich, MA) and the resulting restriction products were analyzed by 10% PAGE. In brief, a 6 mm punch from each card was rehydrated in TE buffer, then treated with 5% Chelex-100 resin (BioRad, Hercules, CA) for 25 minutes at 55°C to remove

heme groups. Samples were centrifuged to pellet the resin and the supernatant was used directly for PCR, as follows: 50 μl reactions contained 15 μl of dried blood extract, 1% Perfect Match (Stratagene, La Jolla, CA), 3% DMSO, 15 pmol *LEPRE1* I4 sense (5'GGCCATCATGTTAAGTAGCAGGCAC CAGCT-3') and *LEPRE1* I5 antisense (5'-CTCCCTGTGCTCCCTTCTCCTCTGAATAAC-3') primers with 1.0 U High Fidelity Platinum Taq polymerase (Invitrogen, Carlsbad, CA). After an initial 5 minute denaturation at 94°C, reactions were incubated for 40 cycles of 94°C for 1 minute, 63°C for 1 minute and 72°C for 1 minute, followed by 7 minutes at 72°C.

Genomic DNA was isolated from leukocytes (District of Columbia, Ghana, Nigeria, Cameroon, Chad, Central African Republic and Congo) or buccal swabs (Senegal) and whole-genome amplification (WGA) was performed using the Illustra GenomiPhi DNA Amplification kit (GE Healthcare, Piscataway, NJ). Amplification products were quantitated with PicoGreen (Invitrogen, Carlsbad, CA) using a standard curve generated from normal control genomic DNA. For each sample, 20 ng of WGA DNA was screened using a custom TaqMan genomic SNP assay (Applied Biosystems, Foster City, CA) to detect the mutation with probes specific for each allele. A 91 bp PCR product corresponding to nucleotides g.9130 – 9221 of the *LEPRE1* gene was amplified under standard conditions on an ABI 7000 Sequence Detection System using forward (LEPRE1-AFR_F, 5'-TTGGCCTATTATGCAGCTATGCTT-3') and reverse (LEPRE1-AFR_R, 5'-GGCAGCTGTCATAACAGAAGGAA-3') primers at a concentration of 0.9 μM. Detection of PCR products was obtained by inclusion of 0.2 μM MGB probes specific for the normal "G" allele (LEPRE1-AFR_V, 5'-VIC-CCCGTGAGGTGAGAG A-3') mutant "T" allele (LEPRE1-AFR_M, 5'-FAM-CCCGTGAGTTGAGAGA-3'), and

Table 1. Primers Utilized in Haplotype Analysis

Marker	Primers	$MgCl_2$ (mM)	T_A (°C)	Fragment Size Range
D1S2706	5'-AGGCAGTAGTTTCAGTGAGTC-3' 5'-Fam-CCTGGTGACCTCCTAGAAG-3'	1.50	55	206-232bp
D1S2722	5'-CAAATAATGCTACCATTGC-3' 5'-Fam-TTCTGGTCATTTACCCTG-3'	1.0	55	195-221bp
D1S2645	5'-TTCCCCACATTGGTGTATAG-3' 5'-Fam-CTGGCTTGAGTGTAGACTCG-3'	0.50	50	266-284bp
D1S463	5'-AGCTACAATAACTCAGTACATGGAA-3' 5'-Fam-TTGCTTTGAACTAGGAACCC-3'	1.50	54	212-234bp
D1S1586	5'-AGAAACAAACAGTTCCATTGC-3' 5'-Hex-TCAAAGTTCTTTCTGTTTCAGTG-3'	0.50	54	83-112bp
D1S193	5'-ACTCTAGCCTGGGTGACAAG-3' 5'-Fam-AGACTGGGAAAATGCAAATG-3'	0.50	54	88-106bp
D1S2861	5'-TAGAGTCAGATGCCTGGTGC-3' 5'-Fam-CAGCCTTGAACTCCACACAT-3'	0.50	54	180-240bp
STR3	5'-AAGAGGATAAGTTAGGAGC-3' 5'-Fam-AGCGTAATCTAGCTCTTTCT-3'	0.50	54	92-104bp
STR5	5'-ATTGTTTGTAAAGCGCCTAG-3' 5'-Hex-GGGAACACACATATGCGT-3'	0.50	54	84-102bp
D1S443	5'-AGCTGAGATCATGGCACT-3' 5'-Fam-TCTTCCAGAGCAAAAGTTGA-3'	0.25	54	310-326bp
D1S447	5'-TTAGTCTGAGTTTGTGGGGG-3' 5'-Fam-GTTTTAACTTCATGGCTGCC-3'	0	57	102-126bp
D1S211	5'-AGCTACATGGCAGGATCAGA-3' 5'-Fam-GGATTCCTTGCTCTGGAAAG-3'	0.50	52	154-192bp
D1S2733	5'-TGCGGCGAGACAGACATC-3' 5'-Fam-AGGACCAGCGTGTGCG-3'	1.50	56	106-120bp
D1S421	5'-AGCTGGCATCTCACCC-3' 5'-Fam-CTGACTGGCGTTACATTCTC-3'	0.50	55	147-151bp

Assays designed and performed by W. A. Cabral

compared to normal (G/G), heterozygous (G/T) and homozygous (T/T) control samples. The TaqMan allelic discrimination assay was used for automated allele calling of the measured fluorescence. Samples designated as carriers by this assay were confirmed by independent PCR and restriction analysis, as described above.

Haplotype Analysis

To confirm the origin and determine the age of the mutation, the size of the conserved haplotype on the mutant allele was determined using microsatellites and short tandem repeats (STRs). Fourteen markers covering a 4.2 Mb region surrounding the *LEPRE1* gene on chromosome 1p were chosen using the UCSC Genome Browser (http://genome.usc.edu). PCR amplification of the microsatellite markers was performed using 200 ng genomic DNA or WGA DNA in 25 μl reactions using the HotStar PCR amplification System (Qiagen, Germantown, MD). Amplification was performed using 15 pmol fluorescently labeled primers and $MgCl_2$ concentrations as described (Table 1). After an initial 15 min denaturation at 95°C, reactions were incubated at 94°C for 30 sec, at primer-specific annealing temperatures of 52-57°C (see T_A in Table 1) for 30 sec and 72°C for 30 sec for 35 cycles, followed by 72°C for 10 min. Amplification products were separated by capillary electrophoresis on an ABI 3730 Automated Sequencer and analyzed using Genemapper software (Applied Biosystems, Foster City, CA). Allele calls were binned to confirm the expected distribution of fragment sizes.

Calculation of the mutation age in generations was performed with the method of Ranalla and Slatkin [437,438] using the formula:

$$t= (1/\Theta)(\log(1-p_0) – \log(Y_0 - n\ p_0) + \log(n))$$

where (t) is the mutation age in generations, (n) is the number of mutation bearing chromosomes sampled, (Y_0) is the number of chromosomes carrying both the mutation and the ancestral marker allele, Theta (Θ) is the recombination fraction between the marker and the mutation, and (p_0) is the frequency of the ancestral marker allele in the normal chromosomes (non-mutation carrying chromosomes). We assumed a 20-year time period per generation.

RESULTS

Identification of Recurring Mutation in African Americans

To determine the carrier frequency of the *LEPRE1* c.1080+1G>T mutation among African Americans, we screened genomic DNA from African Americans from the Mid-

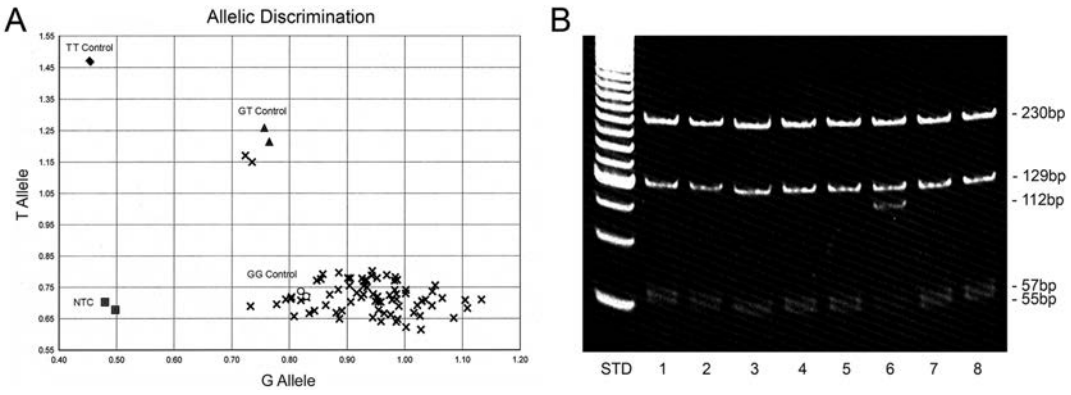

Figure 14. Screening of genomic DNA for the *LEPRE1* c.1080+1G>T mutation. (*A*) Scatter plot of the results from a representative genotyping of WGA DNA using a custom genomic SNP assay. Amplification of the normal "G" allele (VIC-labeled probe) is plotted along the X axis; amplification of the mutant "T" allele (FAM-labeled probe) is plotted along the Y axis. Individual results plotted as x's. NTC, no template control, squares; TT, homozygous mutant genomic DNA control sample, diamonds; G/T, heterozygous genomic DNA control sample, triangles; GG, homozygous normal DNA control sample, circles. (*B*) Genotyping of genomic DNA extracted from newborn metabolic screening cards by PAGE analysis of *Bsl*I-digested PCR products. The mutation eliminates a *Bsl*I site and results in the presence of a 112 bp fragment (lane 6) after restriction digestion. STD, 25 bp DNA size standard; lanes 1-8, reactions from 8 anonymized samples. Assays designed and performed by W.A. Cabral and A.M. Barnes.

Atlantic United States. DNA from 1429/1594 (89.6%), 631/757 (83.4%) and 995/1002 (99.3%) individuals from Pennsylvania, Maryland and the District of Columbia could be genotyped (Figure 14A). Twelve carriers were identified among the 3055 genotyped African Americans, and confirmed by *Bsl*I digestion of independent PCR products (Figure 14B). Heterozygous carriers (Table 2) were detected in 5 of 1429 (0.35%) Pennsylvania samples, 2 of 631 (0.32%) Maryland samples, and 5 of 995 (0.50%) samples from the District of Columbia, supporting a 0.39 ± 0.11% (12/3055) overall carrier frequency for the *LEPRE1* mutation among Mid-Atlantic African Americans. These data predict that cases of type VIII OI, due to homozygosity of the c.1080+1G>T mutation, will occur in 1 in 160-400,000 African American births.

Table 2. *LEPRE1* c.1080+1G>T Carrier Frequency in African Americans

Cohort	Tested	Carriers	Frequency
Maryland	631	2	0.32% (1/316)
Pennsylvania	1429	5	0.35% (1/286)
District of Columbia	995	5	0.50% (1/199)
Total	3055	12	0.39% (1/255)

Pennsylvania and Maryland newborn metabolic screening cards, and WGA genomic DNA samples from AA cohorts were generously provided by F.D. Porter, S.R. Panny, F. Gulamali-Majid, A. Adeyemo and C.N. Rotimi, and studied under protocols 08-CH-N050 (NICHD, Principle Investigator J.C. Marini) and 07-45 (Maryland Department of Health and Mental Hygiene, Principle Investigator J.C. Marini). Assays performed by W.A. Cabral and A.M. Barnes.

Determination of West African Origin and Frequency of *LEPRE1* c.1080+1G>T Mutation

We previously noted that carriers of the c.1080+1G>T mutation were African Americans or African immigrants of West African origin [110]. Therefore, we investigated the prevalence of carriers among contemporary West Africans. Using a genomic SNP assay, 1665/1690 (98.5%) WGA gDNA samples representing 7 groups related by language from Ghana and Nigeria were genotyped (Table 3). Within the cohort, 381 samples were eliminated because they were from related individuals according to pedigree analysis. We identified 9 carriers among 453 samples (1.99%) from Ghana, including 2/142 (1.4%) Ga, 3/280 (1.1%) Akan, and 1/28 (3.6%) Ewe, as well as among the Gonja and Hausa ethnic groups, in which frequency cannot be meaningfully calculated because of low sample number. The 818 Nigerian samples yielded 10 carriers (1.22%), including 7/635 (1.10%) Yoruba and 3/183 (1.64%) Ibo samples. Combined with 13 samples not affiliated with a specific West African ethnic group, our screening identified 19/1284 (1.48%) carriers for the *LEPRE1* mutation among contemporary West Africans. The high carrier frequency predicts that this mutation alone will cause recessive type VIII OI in approximately 1/18,260 births in West Africa, equivalent to the incidence of *de novo* dominant OI.

Pan-African Screening

To determine whether the *LEPRE1* mutation was localized specifically to the Ghana/Nigeria region of West Africa or whether it was Pan-African, we screened DNA samples for the mutation in individuals from regions east, west and south of Ghana and

Table 3. *LEPRE1* c.1080+1G>T Carrier Frequency in Contemporary Africans

Cohort		Tested	Carriers	Frequency
Ghana	Ewe	28	1	3.57%
	Ga	142	2	1.41%
	Akan	280	3	1.07%
	Gonja	1	1	-
	Hausa	2	2	-
	Total Ghana	453	9	1.99% (1/50)
Nigeria	Yoruba	635	7	1.10%
	Igbo	183	3	1.64%
	Total Nigeria	818	10	1.22% (1/82)
Other		13	0	0.0%
Total West Africa		1284	19	1.48% (1/68)
Senegal		561	0	0.0%
Cameroon		864	0	0.0%
Chad		95	0	0.0%
Central African Republic		69	0	0.0%
Congo		34	0	0.0%
Other		24	0	0.0%
Total Pan-Africa		1647	0	0.0%

WGA genomic DNA samples from African cohorts were acquired by S.A. Tishkoff, T.R. Rebbeck, S.M. Gueye, A. Adeyemo and C.N. Rotimi. Assays performed by W.A. Cabral and A.M. Barnes under protocol 08-CH-N050 (NICHD, Principle Investigator J.C. Marini).

Nigeria (Table 3). Among the 1647 genotyped African samples, we were unable to identify a single carrier for the mutation among 561 Senegal, 864 Cameroon, 95 Chad, 69 Central African Republic, 34 Congo or 24 samples collected in Cameroon but of unspecified origin.

Haplotype Analysis and Estimated Age of *LEPRE1* c.1080+1G>T Mutation

We next determined the haplotype surrounding this mutation in African American and West African alleles from 15 families (Figure 15). In the 12 unrelated West African families with 16 independent mutant alleles, we found a conserved haplotype of approximately 63 - 425 Kb surrounding *LEPRE1* (Table 4), extending from between D1S2861 and the *LEPRE1* gene to the region between markers STR#3 and STR#5. In one African family (080), the mutation chromosome could not be determined with certainty for these three markers immediately surrounding *LEPRE1*, although the genotype of the mutant allele carrier in this family is compatible with the ancestral haplotype. Because of this uncertainty, this individual was excluded from calculations. Two (#3F4 and #3G4) of 25 independent normal chromosomes in these African families carried the conserved haplotype, suggesting that the mutation occurred on an ancestral haplotype which exists with an undetermined, but possibly non-negligible, frequency among West Africans.

For the three African American families with four independent mutant alleles, the conserved region surrounding *LEPRE1* was approximately 770 Kb - 1.2 Mb (Table 5) and extended from between markers D1S2645 and D1S463 to the region between markers STR#3 and STR#5. Furthermore, the conserved region around African

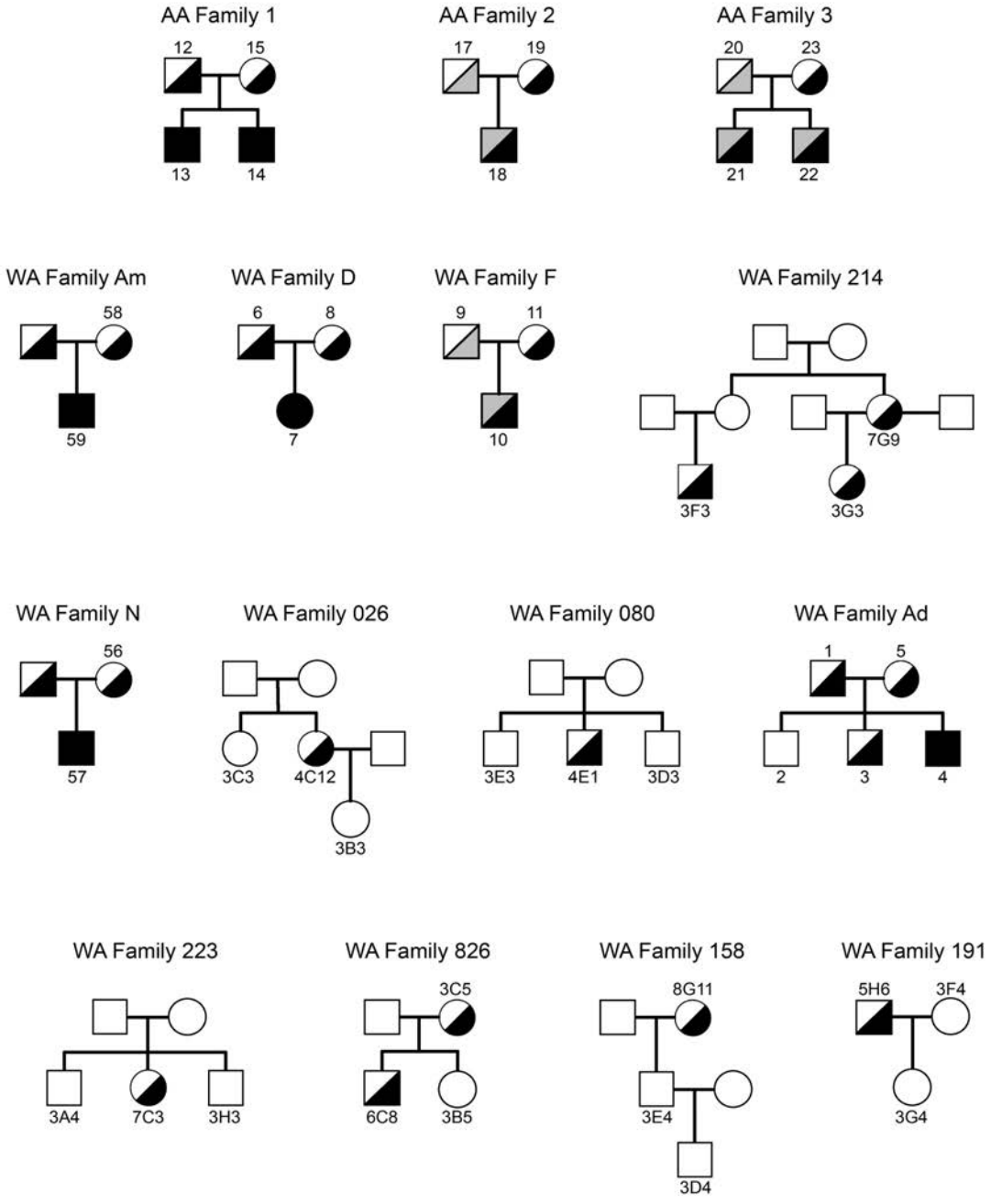

Figure 15. Pedigrees of African American (AA) and West African (WA) families included in haplotype analysis. Black shading denotes the *LEPRE1* c.1080+1G>T allele, while gray shading denotes a second mutant *LEPRE1* allele. Pedigree analysis by W.A. Cabral, D. Chitayat, C.N. Rotimi and J.C. Marini.

American mutant alleles corresponds to the conserved region from West African families 026, 080 and 223. It is noteworthy that the full haplotype was not found in nine African Americans unrelated to our study families who are not carriers of the mutation, although one of these individuals (AA cohort, #30) shared the ancestral haplotype markers around *LEPRE1* from D1S2861 to STR#5.

The age of the mutation was estimated using linked markers on chromosome 1 haplotypes from carrier families. Based on marker assignments for all individuals in the 15 combined African and African American families, we estimate the mutation age between 416 and 1066 years at markers STR#3 and D1S2861, respectively. However, calculations using independent mutant chromosomes from the combined African and African American families narrowed the mutation age range to 458 - 548 years. The number of independent African American carriers was insufficient for a separate calculation.

When analysis was limited to our 12 West African triads, calculations including all family members yielded an estimated age of 622 - 2340 years old. The mutation age range was further refined to 648 - 894 years using only independent mutant chromosomes at markers STR#3 and D1S2861. Thus, the West African haplotypes date this mutation to the 12^{th} - 14^{th} century, prior to the transatlantic slave trade, which occurred from the 16^{th} through 19^{th} centuries.

Table 4. *LEPRE1* c.1080+1G>T Allele Haplotypes in West Africans

Chrom 1p34 Physical position Genethon (cM) Marshfield (cM) deCODE (cM)	ID#	Mutation Genotype	D1S2706 40845025 72.20 71.13 65.31	D1S2722 41326116 73.70 72.59 65.77	D1S2645 42029305 74.30 73.21	D1S463 42285320 74.30 73.21 66.44	D1S1586 42549936 73.21 66.44	D1S193 42791926 73.80 73.21 66.64	D1S2861 42879555 73.70 72.59 66.64	LEPRE1 42994968	STR#3 43057447	STR#5 43304759	D1S443 43361239 74.80 73.81	D1S447 43730229 74.80 73.81	D1S211 44004031 75.30 73.81	D1S2733 44269456 74.80 73.81	D1S421 45029231 73.81 68.90
Family Am	58	+/-	216 220	**218** 218	266 **278**	223 **229**	101 **107**	**90** 96	182 **200**		93 **99**	**97** 99	327 **333**	122 **124**	151 **179**	**116** 120	146 148
	59	-/-	**216 220**	**218 222**	**278 284**	**229 233**	**107 107**	**90 98**	**200 200**		**99 99**	**85 97**	**319 333**	**124 124**	**164 179**	**110 116**	**146 148**
Family 191	5H6	+/-	**224** 228	**202** 218	**274** 280	215 **225**	89 **101**	**98** 104	198 **200**		99 **99**	**99** 103	311 **313**	118 126	169 183	120 **122**	**148** 150
	3G4	+/+	222 228	218 218	280 286	215 215	89 107	98 104	198 200		99 99	97 103	311 329	118 126	169 183	112 120	150 150
	3F4	+/+	218 222	216 218	284 286	215 233	107 107	98 98	200 200		99 99	97 97	325 329	118 126	169 183	112 122	148 150
Family N	56	+/-	214 **228**	210 222	284 **284**	229 **233**	**107** 113	**98** 100	198 **198**		**99** 103	95 **99**	**313** 321	106 **116**	**169** 181	**116** 118	148 150
	57	-/-	**224 228**	**210 222**	**266 284**	**231 233**	**104 107**	**98 98**	**198 198**		**99 99**	**93 99**	**313 329**	**116 116**	**169 185**	**110 116**	**148 150**
Family F	9	+/+	216 220	204 222	272 280	229 233	107 107	96 102	196 196		99 105	97 99	327 331	114 126	179 187	110 120	148 152
	10	+/-	**216** 220	**220** 222	272 **282**	229 233	**107** 107	96 **102**	**182** 196		**99** 105	99 **99**	331 **335**	**104** 114	187 **193**	**114** 120	**150** 152
	11	+/-	**216** 218	218 **220**	266 **282**	229 233	**107** 107	98 **102**	**182** 182		97 **99**	97 **99**	313 **335**	**104** 118	183 **193**	**114** 120	**150** 150
Family 214	7G9	+/-	**218** 228	214 **218**	266 284	233 **233**	89 **107**	100 **102**	**182** 200		95 **99**	**97** 97	325 **333**	**120** 122	**151** 179	**118** 124	148 **150**
	3G3	+/-	**218** 230	**218** 220	272 **284**	225 **233**	89 **107**	**102** 106	**182** 196		91 **99**	**97** 97	329 **333**	114 **120**	**151** 181	**118** 118	150 **150**
	3F3	+/-	**218** 218	**218** 220	266 **284**	229 **233**	107 **107**	**102** 106	**182** 198		99 **99**	**97** 99	317 **333**	**120** 122	**151** 185	**118** 118	148 **150**
Family Ad	1	+/-	214 **214**	214 **218**	280 **284**	215 **233**	**107** 113	**98** 98	**200** 200		93 **99**	**97** **97**	325 **333**	104 **124**	171 **179**	110 **120**	150 **150**
	2	+/+	214 218	210 214	280 282	215 223	98 113	98 98	200 202		93 97	95 97	313 325	104 120	171 183	110 120	150 152
	3	+/-	214 **214**	214 **218**	**266** 280	215 **229**	**104** 113	98 **98**	200 **200**		93 **99**	97 **97**	325 **333**	**104** 104	171 **187**	110 **118**	**148** 150
	4	-/-	**214 214**	**218 218**	**266 284**	**229 233**	**104 107**	**98 98**	**200 200**		**99 99**	**97 97**	**333 333**	**104 124**	**179 187**	**118 120**	**148 150**
	5	+/-	**214** 218	210 **218**	**266** 282	223 **229**	**104** 98	**98** 98	**200** 202		97 **99**	95 **97**	313 **333**	**104** 120	183 **187**	**118** 120	**148** 152
Family D	6	+/-	214 **216**	**204** 218	268 284	225 **233**	**107** 110	90 **98**	**200** 182		95 **99**	85 **97**	323 **333**	**104** 122	169 **189**	114 **118**	150 **150**
	7	-/-	**216 216**	**204 204**	**284 284**	**233 233**	**107 107**	**98 98**	**200 200**		**99 99**	**97 97**	**333 333**	**104 104**	**189 189**	**118 118**	**148 150**
	8	+/-	**216** 216	**204** 210	266 **284**	231 **233**	**107 107**	**98** 100	**200** 198		**99** 101	**97** **97**	313 **333**	**104** 118	187 **189**	**118** 118	**148** 150
Family 158	8G11	+/-	**214** 222	**216 216**	280 **280**	223 **223**	89 **107**	**98** 100	198 **200**		93 **99**	**97** 99	313 **333**	**104** 126	169 **189**	120 **124**	150 **150**
	3E4	+/+	222 226	200 216	280 280	223 227	89 113	100 106	186 198		93 97	85 99	313 321	126 126	169 179	120 122	150 150
	3D4	+/+	218 226	200 212	280 282	223 227	104 113	98 106	186 200		97 101	85 85	321 327	104 126	179 189	118 122	150 150
Family 223	3A4	+/+	214 214	218 218	266 266	223 223	101 110	98 106	194 200		93 97	99 99	327 327	118 126	179 187	118 120	144 148
	7C3	+/-	218 **220**	**214** 218	266 **284**	223 **233**	101 **107**	98 **98**	**200** 200		97 **99**	97 99	**313** 327	118 **124**	**169** 187	**116** 118	144 **150**
	3H3	+/+	214 218	218 218	266 266	223 223	101 110	98 106	194 200		93 97	99 99	327 327	118 126	179 187	118 120	144 148
Family 826	3C5	+/-	**218** 218	202 **218**	266 **274**	225 **233**	89 **107**	90 **98**	182 **200**		97 **101**	85 **97**	**311** 323	104 **104**	169 **179**	120 **122**	148 150
	6C8	+/-	**218** 220	210 **218**	**274** 282	231 **233**	89 **107**	98 **98**	198 **200**		93 **101**	95 **97**	**311** 327	104 **104**	**179** 189	118 **122**	148 152
	3B5	+/+	218 220	210 218	266 282	225 231	89 89	90 98	182 198		93 97	85 95	323 327	104 104	169 189	118 120	150 152
Family 026	3C3	+/+	206 218	216 218	280 280	215 229	89 89	96 98	184 184		93 103	99 103	313 313	114 126	169 171	110 124	148 150
	4C12	+/-	218 **222**	**218** 218	280 **284**	215 **233**	89 **107**	98 **98**	184 **200**		99 103	97 99	313 **327**	114 **122**	171 **189**	110 **116**	**148** 150
	3B3	+/+	218 222	210 218	280 280	215 223	89 104	98 102	184 200		93 103	97 99	313 317	114 126	171 175	110 118	150 150
Family 080	3E3	+/+	208 232	222 224	266 268	223 229	89 107	98 100	182 198		97 99	99 99	313 313	114 122	179 181	110 120	146 148
	4E1	+/-	**222** 232	**218** 222	266 **284**	229 **233**	107 **107**	98 **98**	182 **200**		97 **99**	97 99	**313** 313	**104** 114	**177** 181	**120** 120	**148** 150
	3D3	+/+	208 232	222 224	266 268	223 229	89 107	98 100	182 198		97 99	99 99	313 313	114 122	179 181	110 120	148 150

Haplotypes surrounding *LEPRE1* on chromosome1 in 12 West African pedigrees. Pedigrees are shown in Figure 15. Physical positions and recombination distances of 14 microsatellites and short tandem repeats (STRs) are given. Individual genotypes are noted, with presence of the *LEPRE1* c.1080+1G>T transversion denoted as "-" due to the null allele resulting from the mutation. The data represents the determined size (bp) of the PCR products amplified for each allele at a given marker, with the mutant allele in bold. Gray shading indicates the conserved haplotype within each pedigree. Sample preparation performed by W.A. Cabral. Fragment sizes were determined by K. Cushing and L.C. Brody. Data analysis performed by W.A. Cabral and L.C. Brody. Mutation age estimation calculated by J. Bailey-Wilson.

Table 5. *LEPRE1* c.1080+1G>T Allele Haplotypes in African Americans

Chrom 1p34 Physical position Genethon (cM) Marshfield (cM) deCODE (cM)	ID#	Mutation Genotype	D1S2706 40845025 72.20 71.13 65.31	D1S2722 41326116 73.70 72.59 65.77	D1S2645 42029305 74.30 73.21	D1S463 42285320 74.30 73.21 66.44	D1S1586 42549936 73.21 66.44	D1S193 42791926 73.80 73.21 66.64	D1S2861 42879555 73.70 72.59 66.64	LEPRE1 42994968	STR#3 43057447	STR#5 43304759	D1S443 43361239 74.80 73.81	D1S447 43730229 74.80 73.81	D1S211 44004031 75.30 73.81	D1S2733 44269456 74.80 73.81	D1S421 45029231 73.81 68.90
African Americans																	
Family 2	17	+/+	214 216	214 214	284 284	223 229	101 101	96 96	196 198		93 105	97 97	315 315	118 118	191 191	108 118	146 150
	18	+/-	216 **218**	214 **216**	**274** 284	223 **233**	101 **107**	96 **98**	196 **200**		**99** 105	97 **97**	315 **327**	118 **122**	**169** 191	108 **110**	150 **152**
	19	+/-	**218** 228	216 **216**	266 **274**	229 **233**	107 **107**	**98** 100	200 **200**		93 **99**	97 **97**	315 **327**	122 **122**	**169** 191	**110** 116	146 **152**
Family 1	12	+/-	214 **216**	218 **220**	282 **284**	229 **233**	**107** 110	98 **98**	198 **200**		93 **99**	97 97	325 **331**	114 **116**	**181** 181	**116** 116	148 **150**
	13	-/-	**216 216**	**218 220**	**284 284**	**233 233**	**107 107**	**98 98**	**200 200**		**99 99**	**97 97**	**331 333**	**104 116**	**181 185**	**116 118**	**148 150**
	14	-/-	**216 216**	**218 220**	**284 284**	**233 233**	**107 107**	**98 98**	**200 200**		**99 99**	**97 97**	**331 333**	**104 116**	**181 185**	**116 118**	**148 150**
	15	+/-	**216** 226	214 **218**	274 **284**	229 **233**	101 **107**	**98** 100	198 **200**		**99** 101	**97** 99	313 **333**	**104** 126	179 **185**	118 **118**	**148** 148
Family 3	20	+/+	216 216	214 218	272 272	229 229	113 113	104 106	192 198		93 97	97 99	313 313	104 116	157 189	116 120	146 150
	21	+/-	216 **216**	214 **218**	272 **284**	229 **233**	**107** 113	**98** 106	192 **200**		97 **99**	97 **99**	**313** 313	116 **118**	157 **183**	120 **120**	146 **148**
	22	+/-	216 **216**	214 **218**	272 **284**	229 **233**	**107** 113	**98** 106	192 **200**		97 **99**	97 **99**	**313** 313	116 **118**	157 **183**	120 **120**	146 **148**
	23	+/-	216 **216**	218 **218**	266 **284**	229 **233**	107 **107**	98 **98**	200 **200**		99 **99**	99 **99**	**313** 313	114 **118**	183 **183**	116 **120**	**148** 150
African American Controls	24	+/+	216 216	196 222	266 278	223 223	89 110	96 100	186 200		93 93	97 97	317 327	118 124	173 187	116 120	148 150
	25	+/+	216 218	216 218	272 274	221 231	101 107	96 98	184 184		93 99	99 99	313 313	122 126	179 179	116 120	148 150
	26	+/+	216 216	214 220	272 282	223 231	89 101	104 104	196 202		95 99	97 99	313 315	118 124	169 191	110 116	146 150
	27	+/+	216 216	198 216	274 276	229 231	101 107	96 106	184 196		97 103	97 97	311 327	118 120	169 191	118 120	148 150
	28	+/+	216 218	212 222	278 280	223 229	107 110	100 106	198 202		91 103	97 103	313 315	110 122	185 191	114 118	146 150
	29	+/+	216 216	214 218	274 284	225 233	89 107	100 100	180 182		93 97	97 99	321 325	122 124	169 179	120 120	150 150
	30	+/+	206 226	204 212	276 280	215 231	89 101	104 104	194 200		95 99	97 103	315 325	120 124	183 191	110 118	150 150
	31	+/+	216 230	218 222	266 282	213 221	89 107	104 106	194 196		99 101	93 97	323 329	104 122	179 189	118 118	152 152
	32	+/+	216 232	212 218	268 278	221 223	86 116	104 104	198 198		97 101	85 97	315 319	120 124	169 171	116 120	148 148

Haplotypes surrounding *LEPRE1* on chromosome 1 in 3 African American pedigrees and 9 African American normal controls. Pedigrees are shown in Figure 15. Physical positions and recombination distances of 14 microsatellites and short tandem repeats (STRs) are given. Individual genotypes are noted, with presence of the *LEPRE1* c.1080+1G>T transversion denoted as "-" due to the null allele resulting from the mutation. The data represents the determined size (bp) of the PCR products amplified for each allele at a given marker, with the mutant allele in bold. Gray shading indicates the conserved haplotype within each pedigree. Sample preparation performed by W.A. Cabral. Fragment sizes were determined by K. Cushing and L.C. Brody. Data analysis performed by W.A. Cabral and L.C. Brody. Mutation age estimation calculated by J. Bailey-Wilson.

DISCUSSION

We have identified a founder mutation in the gene encoding P3H1 that causes autosomal recessive type VIII osteogenesis imperfecta, a severe to lethal bone dysplasia. The *LEPRE1* c.1080+1G>T mutation originated in West Africa, where nearly 1.5% of contemporary Ghanians and Nigerians are carriers. The mutant allele was presumptively brought to North America via the Atlantic slave trade, resulting in a carrier frequency among African Americans of 0.32-0.50%. Type VIII OI joins sickle cell disease, and end-stage kidney disease (ESKD) resulting from *APOL1* variants as examples of single gene defects transported from Africa to the Americas [439]. Previously reported founder mutations for recessive OI have occurred in genetically isolated populations, including a First Nations Tribe of Quebec with type VII OI [107,440], and the Irish Traveler population with type VIII OI [409,441].

The high carrier frequency of this *LEPRE1* mutation among West Africans results in distinctive OI inheritance patterns in West Africa compared to North America and Europe. While severe to lethal OI has been reported in West Africa [442,443], this study is the first to address the prevalence of OI in West African populations. Our data predicts that the incidence of recessive OI in West Africa due to this mutation is about 1/20,000 births, equivalent to the incidence of *de novo* dominant OI in North America and Europe. In comparison, recessive OI accounts for approximately 5 - 7% of severe OI cases in North America and Europe and type VIII OI is calculated to occur in approximately 1/250,000 African American births.

The occurrence of *LEPRE1* c.1080+1G>T within a conserved haplotype confirms a founder effect, rather than the presence of a mutation "hotspot". Furthermore, no

individuals from outside this region of West Africa were identified as carriers, suggesting that *LEPRE1* c.1080+1G>T is not a Pan-African mutation. The carrier incidence is higher in Ghana than Nigeria, consistent with a mutation that began in Ghana and was carried eastward. The mutation expanded among groups with similar languages; all carriers except one belong to ethnic groups that speak languages of the Kwa or Volta-Niger branches of the Niger-Congo family. Haplotype analysis also dates this mutation to approximately 648 - 894 years ago, placing its origin within the 12^{th} - 14^{th} centuries. This date is prior to the Atlantic slave trade, which peaked from the late 17^{th} to mid-19^{th} centuries [444]. Interestingly, combining the West African and African American haplotype data decreases the estimated age of the mutation, compared to West African haplotypes alone (458 - 548 years versus 648 - 894 years), consistent with descent of the African American alleles from a limited number of West African founders who did not represent the full African allelic variation. In fact, the conserved region around *LEPRE1* in African American carriers corresponds to a subset of contemporary West African haplotypes. However, we cannot rule out incomplete ascertainment because only 4 independent African American mutant alleles were analyzed. It is also possible that the West African haplotypes represent continued recombination events since the end of the slave trade.

During the nearly 400 years of the Atlantic slave trade, more than 12 million slaves were transported, about 10.7 million of whom survived to disembark in the New World [444,445]. More than a third (38.1%) of these Africans originated from the region of West Africa then known as the Gold Coast, the Bight of Benin and the Bight of Biafra [445,446]. These three areas included the territories of the Akan, Ewe, Ga, Yoruba, and Ibo, whose contemporary members are carriers of the *LEPRE1* mutation. Thus, it is not surprising to

identify mutation carriers among African Americans of the Mid-Atlantic United States. Assuming a 1.48% carrier incidence among the approximately 130,000 slaves from West Africa who disembarked in the American colonies [445], it is possible that as many as 1900 carriers were transported to North America. This number of carriers is more than sufficient to make a second founder effect unlikely, as would have occurred if only a few carriers were transported. The incidence of carriers among African Americans would not have been decreased by immigration from the Caribbean after World War II [447], or by direct immigration from Africa to the United States during the last two decades, which have both included a high proportion of West Africans [448].

Only about 3.6% of all slaves transported in the Atlantic slave trade were sent to the future United States. The majority of slaves were transported to the Caribbean and South and Central America, with their origins and destinations reflecting the European colonial regions of influence [444,445]. Since the majority of Africans sent to Brazil by the Portuguese originated in West Central and Southeast Africa (Angola, Congo and Mozambique), we do not expect a significant number of carriers for this *LEPRE1* mutation among Brazilians of African descent. In contrast, the British, French, and Spanish traders obtained a large proportion of their slaves from the current Ghana/Nigeria area. Over 1.5 million West Africans, 39% of all West Africans transported during the slave trade, were sent to the British Caribbean, including Jamaica, Barbados, Antigua, Trinidad and the Grenadines, where they constituted 68% of disembarking slaves. The French and Spanish areas of the Caribbean each received over 10% of all West African slaves, with nearly half of the slaves in Martinique, Cuba and Spanish Central America originating in West Africa [445]. Therefore, the carrier frequency of this mutation on

multiple Caribbean islands may be higher than among Mid-Atlantic African Americans. The *LEPRE1* mutation may be useful for African ancestry studies among Caribbean populations because of its specific region of origin and simple, rapid analysis.

A high carrier frequency for a lethal mutation is rare within a population that is not geographically isolated. Disease founder alleles have expanded within non-geographically isolated ethnic groups such as Eastern European Jews, with Tay-Sachs and Gaucher disease carrier incidences of 1:31 and 1:20 [449,450], and among Finnish isolates, who have carrier frequencies for multiple recessive diseases as high as 1:45 or 1:60 [451]. There are several possible causes for the high West African carrier frequency of the *LEPRE1* allele. First, despite its lethality in a homozygous state, the *LEPRE1* mutation may act as a neutral variant in the heterozygous state. In fact, the heterozygous parents of our patients are apparently healthy. A second possibility is that heterozygosity for this mutation may provide some protection against one of the infectious diseases endemic to this region, for example, by maintaining the integrity of connective tissues. The best-known disease mutation which confers selective advantage is sickle-cell trait (HbS), which provides resistance to malarial infection and is found in 8% of African Americans, and as much as a quarter of the West African population [452,453]. In addition, variants of apolipoprotein L-1 (*APOL1*) which are associated with lysis of *T. brucei rhodesiense* are common in Africa and are now linked to susceptibility to end-stage kidney disease (ESKD) in African Americans [439]. The *APOL1* gene cluster occurs on a region of chromosome 22 which shows strong evidence of evolutionary selection. However, the *LEPRE1* carrier frequency is lower than expected for a mutation conferring selective advantage, and a protective role against West African endemic diseases is not

apparent. A third possibility is the *LEPRE1* lethal allele may be linked to a neighboring gene on chromosome 1 that is undergoing selection, as a so-called "hitchhiker", a possibility requiring further sequencing to investigate.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Ostrander Lab (NHGRI) and Peter Chines (NHGRI) for assistance with microsatellite genotyping and haplotype data analysis. For assistance with African sample preparation, we thank William Beggs, Amy Walker, Teo Tran, Charnita Zeigler-Johnson and Elaine Spangler at the University of Pennsylvania. For providing access to the Maryland Department of Mental Health and Hygeine newborn metabolic screening cards, we thank John M. DeBoy, Dr.P.H. We also thank Professor David Eltis, coeditor of the Transatlantic Slave Trade database, for critical reading of the manuscript. This work was supported by the intramural research programs of NICHD (JCM, FDP) and NHGRI (CNR, LCB, JEB-W), as well as grant support to SAT (R01-GM076637-05 and DP1-OD-006445-01) and TRR (R01-CA085074 and P50-CA105641).

DISCLOSURES

The authors declare no conflict of interest.

Chapter III

Abnormal Type I Collagen Post-translational Modification and Crosslinking in a Cyclophilin B KO Mouse Model of Recessive Osteogenesis Imperfecta

Wayne A. Cabral[1], Irina Perdivara[2], MaryAnn Weis[3], Masahiko Terajima[4], Angela R. Blissett[1], Weizhong Chang[1,5], Elena N. Makareeva[6], Edward Mertz[6], Sergey Leikin[6], Kenneth B. Tomer[2],
David R. Eyre[3], Mitsuo Yamauchi[4], and Joan C. Marini[1]

[1]Bone and Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD, USA
[2]Laboratory of Structural Biology, NIEHS, NIH, Research Triangle Park, NC, USA
[3]Orthopaedic Research Laboratories, University of Washington, Seattle, WA, USA
[4]North Carolina Oral Health Institute, University of North Carolina, Chapel Hill, NC, USA
[5]Current Address: Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[6]Section on Physical Biochemistry, NICHD, NIH, Bethesda, MD, USA

Communicating Author:

Joan C. Marini, MD, PhD
Chief, Bone and Extracellular Matrix Branch
NICHD, NIH
Building 10; Room 10D39
9000 Rockville Pike
Bethesda, MD 20892
(T) 301-594-3418
(F) 301-480-3188
e-mail: oidoc@helix.nih.gov

KEY WORDS: Osteogenesis imperfecta, Cyclophilin B, PPIB, crosslinking, type I collagen

ABSTRACT

Cyclophilin B (CyPB), encoded by *PPIB*, is an ER-resident PPIase that occurs both independently and as a component of the collagen prolyl 3-hydroxylation complex. CyPB is proposed to be the major PPIase catalyzing the rate-limiting step in collagen folding. Mutations in *PPIB* cause recessively inherited osteogenesis imperfecta type IX, a moderately severe to lethal bone dysplasia. To investigate the role of CyPB in collagen folding and post-translational modifications, we generated *Ppib*$^{-/-}$ mice that recapitulate the OI phenotype. KO mice are small, with reduced femoral aBMD, BV/TV, and MOI. *Ppib* transcripts are absent in skin, fibroblasts, femora and calvarial osteoblasts, and CyPB is absent from KO osteoblasts and fibroblasts on western blots. Only residual (2-11%) collagen prolyl 3-hydroxylation is detectable in KO cells and tissues. Collagen folds more slowly in the absence of CyPB, supporting its rate-limiting role in folding. However, treatment of KO cells with cyclosporine A causes further delay in folding, supporting the potential existence of another collagen PPIase. *Ppib*$^{-/-}$ fibroblast and osteoblast collagen has normal total lysyl hydroxylation, while increased collagen diglycosylation is observed. LC/MS analysis of bone and osteoblast type I collagen revealed site-specific alterations of helical lysine hydroxylation, in particular, significantly reduced hydroxylation of helical crosslinking residue K87. Consequently, underhydroxylated forms (HLNL and LP) of di- and trivalent crosslinks are strikingly increased in KO bone, leading to increased total crosslinks and decreased helical hydroxylysine- to lysine-derived crosslink ratios. The altered crosslink pattern was associated with decreased collagen deposition into matrix in culture and altered fibril structure in tissue, and presumably reduced bone strength. These findings suggest a novel

role for CyPB in supporting the activity of specific collagen lysyl hydroxylase(s), especially LH1, by an unknown mechanism. Thus, CyPB not only facilitates collagen folding directly, but also indirectly regulates collagen hydroxylation, glycosylation, crosslinking and fibrillogenesis.

AUTHOR SUMMARY

Osteogenesis imperfecta (OI), or brittle bone disease, is characterized by susceptibility to fractures from minimal trauma and growth deficiency. Deficiency of components of the collagen prolyl 3-hydroxylation complex, CRTAP, P3H1 and CyPB, cause recessive types VII, VIII and IX OI, respectively. We have previously shown that mutual protection within the endoplasmic reticulum accounts for the overlapping severe phenotype of patients with CRTAP and P3H1 mutations. However, the bone dysplasia in patients with CyPB deficiency is distinct in terms of phenotype and type I collagen biochemistry. Using a knock-out mouse model of type IX OI, we have demonstrated that CyPB is the major, although not unique, peptidyl prolyl *cis-trans* isomerase that catalyzes the rate-limiting step in collagen folding. CyPB is also required for activity of the collagen prolyl 3-hydroxylation complex; collagen α1(I) P986 modification is lost in the absence of CyPB. Unexpectedly, CyPB was found to also influence collagen helical lysyl hydroxylation in a tissue-, cell- and residue-specific manner. Thus CyPB facilitates collagen folding directly, but also indirectly regulates collagen hydroxylation, glycosylation, crosslinking and fibrillogenesis through its interactions with other collagen modifying enzymes in the endoplasmic reticulum.

INTRODUCTION

Type I collagen, the most abundant protein component of the extracellular matrix of skin, tendon and bone, is a heterotrimer consisting of two α1(I) and one α2(I) chains encoded by the *COL1A1* and *COL1A2* genes, respectively. The pro-alpha chains of type I collagen contain an uninterrupted helical region consisting of 338 repeats of the Gly-Xaa-Yaa triplet. Biosynthesis of procollagen is a complex process that requires several co- and post-translational modifications within the endoplasmic reticulum, including formation of disulfide bonds within the propeptide extensions, isomerization of peptidyl-prolyl bonds, hydroxylation of Yaa lysyl and prolyl residues, and glycosylation of hydroxylysines [9]. The post-translational modifications occur before, and to a major extent stabilize, collagen helical folding. After secretion into the pericellular space and processing of propeptide extensions, the mature collagen heterotrimer is incorporated into heterotypic collagen fibrils. The fibrils are then stabilized by intermolecular aldehyde-derived crosslinks formed from specific collagen lysyl and hydroxylysyl residues by lysyl oxidases (LOX) [454,455].

Dominant mutations in *COL1A1* or *COL1A2* cause classical osteogenesis imperfecta (OI), with susceptibility to fractures from minimal trauma and growth deficiency [116]. Glycine substitutions in the collagen alpha chains delay folding and increase exposure to modifying enzymes, resulting in collagen overmodification. Some OI cases (≈10%) have recessive inheritance, caused by deficiency of proteins that interact with collagen for folding or post-translational modification [354]. Most commonly, recessive OI involves the collagen prolyl 3-hydroxylation complex, consisting of prolyl 3-hydroxylase 1 (P3H1/LEPRE1, leucine- and proline-enriched proteoglycan 1), cartilage-associated

protein (CRTAP) and cyclophilin B (CyPB/PPIB, peptidyl-prolyl *cis-trans* isomerase B). This complex is responsible for 3-hydroxylation of the Xaa position α1(I) P986 residue in types I and II collagen [2,106,107]. Loss of individual components abrogates collagen P986 3-hydroxylation [107,110,422,456]. CRTAP and P3H1 are mutually supportive in the complex; deficiency of either component causes severe to lethal OI with rhizomelia, classified as OI types VII and VIII, respectively [457]. However, loss of the CRTAP/P3H1 complex does not decrease the level of CyPB, and, conversely, the CRTAP/P3H1 complex is only partially decreased in the absence of CyPB.

CyPB/PPIB, is an ER-localized member of the immunophilin family of proteins with PPIase activity [93,458]. *Cis-trans* isomerization of peptidyl-prolyl bonds is especially important for type I collagen folding because prolines constitute nearly one-sixth of its primary sequence. Studies published over twenty years ago demonstrated that exposure of cells to the cyclophilin inhibitor cyclosporine A (CsA) slows the rate of collagen folding and results in overmodification of lysyl residues [108]. Thus, CyPB is thought to be the major, and possibly unique, collagen peptidyl-prolyl *cis-trans* isomerase [403]. However, collagen biochemical data from the few reported patients with CyPB deficiency is inconsistent. Two patients with *PPIB* mutations causing moderate OI have normal levels of collagen prolyl 3-hydroxylation [421,423], while in two lethal OI cases of CyPB deficiency, α1(I) P986 3-hydroxylation decreases to 30% of normal [422,423]. Overall collagen overmodification was detected in both lethal cases, but in only one moderately severe patient.

In an attempt to clarify the inconsistencies among the human cases of PPIB/CyPB deficiency, we generated a *Ppib* knockout mouse model. Collagen biochemical studies

demonstrate cell- and tissue-specific dysregulation of collagen helical lysyl hydroxylation and glycosylation in the absence of CyPB. In addition, reduced hydroxylation of specific collagen helical lysine residues involved in intermolecular crosslinks led to a shift in the pattern of crosslinks in bone tissue, and reduced collagen deposition into matrix. These studies imply that CyPB has a role in lysyl hydroxylation that is independent of collagen folding, and establish a novel function for CyPB in collagen biosynthesis.

MATERIALS AND METHODS

Generation of Ppib-Null Mice

ES cell gene trap lines with a β-geo reporter construct inserted in intron 1 of the *Ppib* gene were produced by BayGenomics (UCSF, CA) and obtained from the Mutant Mouse Regional Resource Center (Davis, CA) [459]. The gene trap construct contains intron 1 and a portion of exon 2 to include splice acceptor sequence from the mouse *En2* gene, followed by a β-galactosidase /neomycin (β-geo) reporter-selection cassette and SV40 polyadenylation signal. Cells were expanded for isolation of mRNA and quantitation of *Ppib* expression. A heterozygous clone from cell line RST139 was shown to have half-normal *Ppib* expression by real-time RT-PCR, and was injected into C57BL/6 blastocysts. Founders were generated by mating chimeric males with 129/P2/OlaHsd females to retain the 129 background of the ES cell line. A second line was generated for experiments by backcrossing F1 mice into C57BL/6 for 5 generations.

Genomic DNA for genotyping was isolated using the Red Extract-n-Amp tissue PCR kit (Sigma) and amplified by hemi-nested PCR using a sense primer located in *Ppib* exon

1 (5'-TGCCCGGACCCTCCGTGGCCAACGATAAGA-3'), an antisense primer corresponding to intron 1 of *En2* (5'-GGCATCTCCCCTTCAGTCTTCCTGTCCAGG-3'), and an antisense primer downstream of the inserted construct internal to *Ppib* intron 1 (5'-GGGGGGCTGGGGGAGTCTGGGTTATTCTCT-3'). Complete absence of *Ppib* transcripts was verified by real-time RT-PCR of mRNA isolated from femoral tissue and skin of F2 *Ppib* homozygous knock-out mice. Animal care and experiments were performed in accordance with a protocol approved by the NICHD ACUC committee.

Skeletal Staining, Growth curves, X-ray, DXA analysis, micro-CT

For skeletal staining, skin and viscera were removed from dead P1 newborn pups. Pups were fixed in 95% ethanol for 7 days and stained with 0.3% Alcian Blue 8GS and 0.1% Alizarin Red S [460]. Excess stain was removed with 1% KOH and increasing concentrations of glycerol. For growth curves, mice were fed regular rodent chow and weighed weekly from 3 to 24 weeks of age. Skeletal characteristics of F6 wild-type, heterozygous and homozygous *Ppib*-null littermates were analyzed at age 8 weeks. Radiographs were performed by Faxitron (30 kV for 1 min). Areal bone mineral density (aBMD) scans of mouse femurs were acquired using a GE Lunar PIXImus2 (GE Healthcare) and internal calibration standards. To determine femoral length, femora were dissected and cleaned of soft tissue, leaving the epiphyses intact. Femurs were measured from the proximal head to the distal end of the medial and lateral condyles with a digital caliper.

Left femora of 8-week old mice were analyzed by μQCT for both structural and mineral parameters using a SkyScan1174 compact micro-CT scanner (MicroPhotonics)

operating at 50kV with an X-ray source current of 800μA, according to manufacturer directions. BMD was calibrated with hydroxyapatite phantoms. Morphometric analyses of trabecular and cortical regions was performed using CTAn software (v.1.13) and 3D images were generated using CTvol software (v.2.2). The trabecular region of interest (ROI) was located just proximal to the distal femoral growth plate and extended 10% total femoral length. The diaphyseal cortical ROI centered on the femur midpoint, spanning 15% total femoral length.

Cell culture

Primary fibroblast (FB) and calvarial osteoblast (OB) cultures were derived from 3-day old pups by standard procedures [461]. Cells from digestions 3-5 were plated at a density of 5,000 cells/cm^2 and cultured in αMEM with 10% FBS, 2mM glutamine, 1% pen-strep and 8% CO_2. FB cultures were derived from dermal tissue dissected from the abdomens of newborn pups. FB were allowed to grow out from dermal samples for 2 weeks, released by trypsin digestion and cultured in DMEM containing 10% fetal bovine serum, 2mM glutamine, 1% pen-strep and 5% CO_2.

Analysis of Gene Expression

Total RNA was extracted from cell cultures or tissues dissected from F6 mice using TriReagent (Molecular Research Center) according to the manufacturer's protocol. Total RNA was treated with DNA-free (Life Technologies), then reverse-transcribed using a High Capacity cDNA Archive Kit (Life Technologies). Real-time RT-PCR was performed using Taqman Assays on Demand (Life Technologies, *Ppib*,

Mm00478295_m1; *Plod1*, Mm01255760_m1; *Plod3*, Mm00478798_m1; *Glt25d1*, Mm00600638_m1; *Glt25d2*, Mm01290012_m1; *Gapdh*, Mm99999915_g1). Relative expression of genes of interest was measured in triplicate, normalized to *Gapdh* transcripts, and quantified relative to wild-type calvarial OB.

Western Blot Analysis

FB and OB cultures were lysed in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris, pH 7.4) supplemented with protease inhibitor cocktail (Sigma). Protein concentration was determined using the BCA Protein Assay Kit (Thermo Scientific). Samples (15 μg protein) were subjected to SDS-PAGE on 10% acrylamide gels under denaturing conditions and electroblotted onto nitrocellulose membranes. The membranes were blocked overnight in 5% non-fat milk in TBST. After washing in TBST, membranes were incubated overnight at 4°C in TBST containing 2.5% non-fat milk and primary antibody (diluted 1:1000). After washing in 1X TBST, membranes were incubated with the corresponding IRDye infrared secondary antibody (diluted 1:10,000) (LI-COR Biosciences). Proteins were visualized using an Odyssey Infrared Imaging System (LI-COR Biosciences). The primary antibodies used include anti-P3H1 (Novus Biologicals), rabbit anti-actin (Santa Cruz Biotechnology), rabbit anti-CyPB (Abnova), rabbit anti-PLOD1 (Santa Cruz Biotechnology), rabbit anti-PLOD3 (Proteintech), and goat anti-GLT25D1 (Santa Cruz Biotechnology). Anti-mouse CRTAP antibody was a generous gift from Dr Brendan Lee, Baylor College of Medicine.

Biochemical Analysis of Type I Collagen

Steady-state collagen analysis was performed as previously described [355]. Collagens were prepared by pepsin digestion (50 μg/ml) of procollagen samples, separated on 6% SDS-urea- polyacrylamide gels and visualized by autoradiography.

For analysis of collagen modification in cell culture, confluent FB and OB were stimulated for collagen synthesis in DMEM or αMEM containing 0.1% FBS and 100 μg/ml ascorbate for three days, with daily collection. Collected medium was buffered with 100 mM Tris-HCl, pH 7.4, and cooled to 4°C. Protease inhibitors were added to the following final concentrations: 25 mM EDTA, 0.02% NaN_3, 1 mM phenylmethylsulfonylfluoride, 5 mM benzamidine, and 10 mM N-ethylmaleimide. Procollagens were precipitated from media with ammonium sulfate overnight at 4°C. Procollagen was collected by centrifugation, resuspended in 0.5 M acetic acid and digested with 0.1 mg/ml pepsin at 4°C overnight. Selective salt precipitation of collagen with 0.9 M NaCl in 0.5 M acetic acid was performed twice. Purified collagen samples were resuspended in 0.5 M acetic acid, dialyzed against 5mM acetic acid overnight and lyophilized before further analyses.

Differential Scanning Calorimetry (DSC) scans were performed as previously described [462]. Thermograms were recorded in 0.2 M sodium phosphate, 0.5 M glycerol, pH 7.4, from 10 to 50°C at 0.125 and 1°C/min heating rates in a Nano II DSC instrument (Calorimetry Sciences Corporation).

Analysis of tissue-derived collagens was performed using skin, femora and humeri from 2 month old mice (n=5). After marrow was flushed with cold PBS, bone samples were pulverized in liquid N_2, demineralized with EDTA at 4^{o}C for 2 weeks, and

lyophilized. Skin dissected from the backs of mice was minced and lyophilized after removal of hair, fat and muscle. Two mg of the dried samples were then reduced with standardized NaB^3H_4, hydrolyzed with 6N HCl and subjected to amino acid and cross-link analyses as previously described [463-465]. The extent of Lys hydroxylation in collagen was calculated as hydroxylysine (Hyl) / hydroxyproline (Hyp) X300 (i.e. ~300 residues of Hyp/collagen). The cross-link precursor aldehydes and reducible cross-links were measured as reduced forms and cross-links were quantified as moles/mole of collagen.

Characterization of collagen post-translational modifications

Collagen 3-hydroxylation was analyzed by in-gel tryptic digestion of SDS-PAGE-purified type I collagen alpha chains. Electrospray mass spectrometry was performed on the tryptic peptides using an LCQ Deca XP ion-trap mass spectrometer equipped with in-line liquid chromatography (Thermo Finnigan) using a C8 capillary column (300 μm x 150 mm; Grace Vydac 208MS5.315) eluted at 4.5 μl per min.

Site-specific modification of lysyl residues was determined by LC/MS/MS on a Waters Q-Tof Premier mass spectrometer coupled to a nanoACQUITY UPLC system (Waters Corporation) as reported [124]. Tryptic peptides containing Lys residues, their hydroxylated and/or glycosylated forms were identified from the LC/MS/MS analyses using manual interpretation of the MS/MS spectra. Relative quantitation of lysine (Lys), hydroxylysine (Hyl), galactosyl-Hyl (G-Hyl) and glucosylgalactosyl-Hyl (GG-Hyl) at a specific glycosylation site was performed by dividing the total ion abundance determined for each species by the sum of the ion abundances of all observed species containing that particular site.

Collagen Folding Assays

An intracellular collagen folding assay was performed as described [108]. Confluent cells were stimulated overnight in media with 10% FBS and 100 μg/ml ascorbic acid, and then incubated in serum free media containing 100μg/ml ascorbic acid for 2 hr. Cells were pulsed with 1.4 μCi/ml 14C-proline for 15 min to label procollagen chains, followed by collection of the cell layer every 5 min. Each sample was digested for 2 min at 20^{o}C with 0.2% Triton X-100, 100 μg/ml trypsin, and 250 μg/ml chymotrypsin in PBS (Sigma). Digestions were stopped by addition of 1 mg/ml soybean trypsin inhibitor (Sigma). Samples were precipitated overnight, collected by centrifugation, electrophoresed on 3-8% Tris-acetate gels (Life Technologies) and quantitated by densitometry of autoradiograms.

Collagen Synthesis and Secretion Kinetics

For pulse-chase assays, performed as described [466], wild-type and homozygous *Ppib*-null FB were grown to confluence. For each cell line, two wells were used for cell counts. Procollagens were harvested at the indicated times, digested with pepsin and precipitated. Samples were loaded for equivalent cell number, on 3-8% Tris-acetate gels and quantitated by densitometry.

Collagen Matrix Deposition

Wild-type and homozygous *Ppib*-null FB and OB were grown to confluence and stimulated every other day for 14 days with fresh DMEM (fibroblasts) or αMEM (osteoblasts) containing 10% FBS and 100 μg/ml ascorbic acid, as described [467]. Matrix

collagens were sequentially extracted at 4°C, with neutral salt for newly incorporated collagen, then acetic acid for collagens with acid-labile cross-links, and, finally, by pepsin digestion for collagens with mature cross-links [468]. All fractions were electrophoresed on 6% polyacrylamide-urea-SDS gels. Samples were loaded for equivalent densitometry signal; the total signal for each fraction was calculated by adjusting the gel signal by the total volume of that fraction. In separate experiments, quantitation of matrix deposited in culture was performed using Raman microspectroscopy. Cultures were fixed in 1% paraformaldehyde and analyzed as previously described [103]. Matrix collagen:cell organics ratios were evaluated from decomposition of corrected spectra of collagen-free cytoplasm and purified collagen in the amide III spectral region.

TEM Analysis of Murine Dermal Collagen Fibrils

A dermal biopsy was obtained from abdominal skin of wild-type and homozygous null mice, then processed as described [177]. Representative areas of the stained grids were photographed in a Zeiss EM10 CA transmission electron microscope (JFE Enterprises).

RESULTS

Generation and phenotype of Ppib-Null Mice

Ppib-null mice were produced from an ES cell line carrying a gene trap insertion in intron 1 of *Ppib*. Two ESC lines, RST059 and RST139, were screened by real-time RT-PCR. Expression of *Ppib* in RST059 and RST139 was decreased to 76 ± 5% and 54 ± 3%, respectively, of wild-type levels. Because RST139 ES cells were more likely to have one null *Ppib* allele, we proceeded with generation of a mouse line in the C57BL/6 background using these cells.

Genomic DNA from homozygous mutant mice was amplified by heminested PCR (Figure 16A) and sequenced to demonstrate insertion of the gene trap vector 123 bp from the 5'-end of the1066 bp *Ppib* intron 1. The insertional mutation interrupts transcription of *Ppib* in cells and tissues isolated from $Ppib^{-/-}$ mice. Expression of *Ppib* in primary fibroblast (FB) and osteoblast (OB) cultures from newborn mice, as well as in dermal and femoral tissues of 8 week-old mice, was reduced in cells and tissues heterozygous for the gene-trapped allele, and completely absent from homozygous cells and tissues (Figure 16B). Thus, *Ppib* expression is completely blocked in the gene trap mutant allele, versus a previously reported *Ppib*-null mouse, which retained 30% of wild-type transcripts [112].

Heterozygous mice were bred into the C57BL/6 line. The genotype distribution among offspring of F4 and F5 het x het crosses deviated from the Mendelian ratio, with 30 and 50% lethality of homozygous pups from F4 and F5 matings, respectively. In surviving homozygous offspring of F5 matings, growth deficiency became apparent soon after weaning. Knockout mice weighed about 25% less than wild-type and heterozygous littermates from 3 to 24 weeks of age (Figure 16D). At 8 weeks of age, $Ppib^{-/-}$ femoral

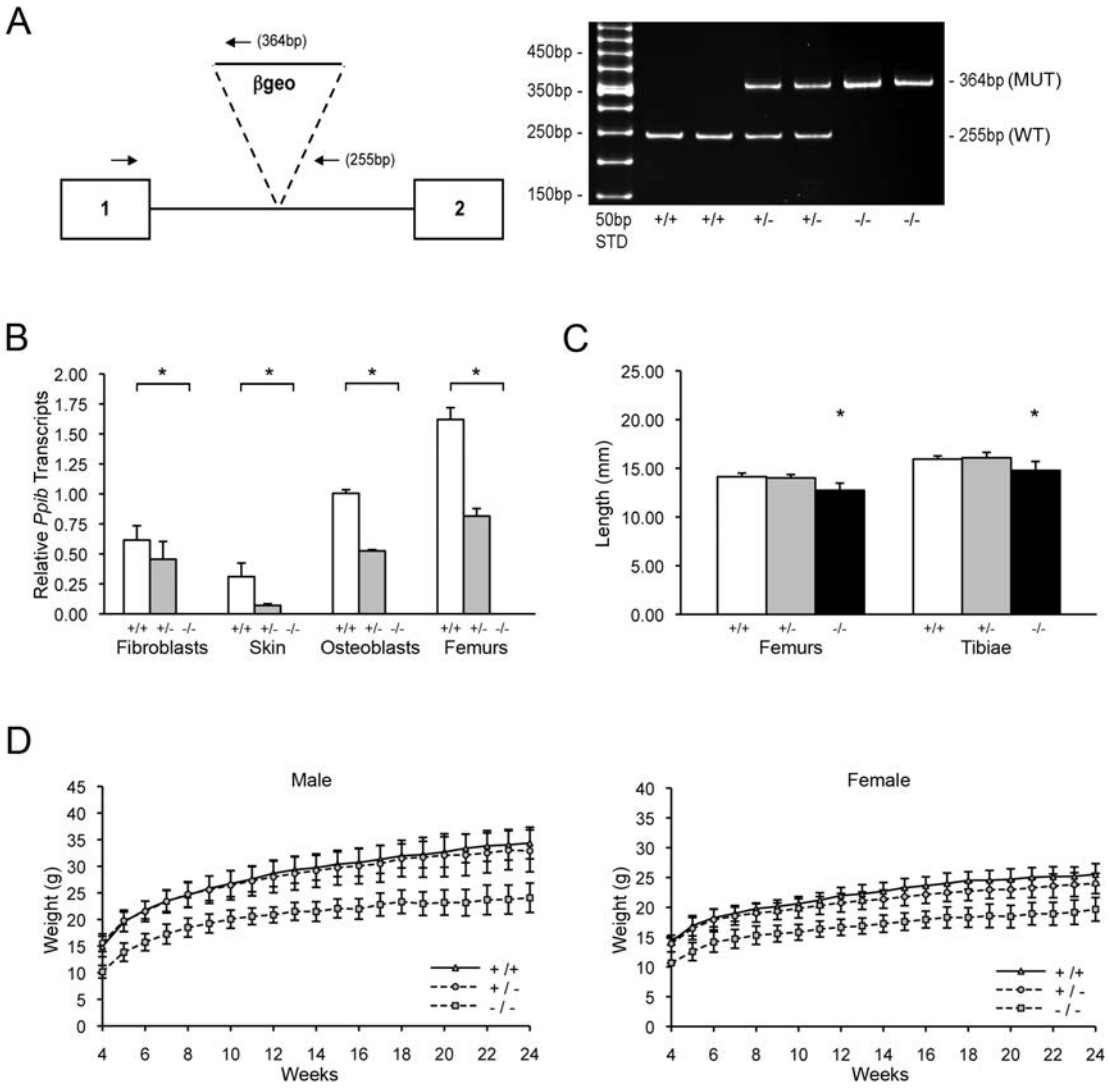

Figure 16. Generation of a murine model of Cyclophilin B deficiency. (A) Diagram of knock-out allele containing β-geo reporter construct inserted into intron 1 of *Ppib* gene, *left*. Hemi-nested primers (arrows) for PCR-based genotyping selectively amplify the wild-type allele (255bp) or the gene-trapped allele (364bp), *right*. Genotypes in offspring of matings between mice heterozygous for the gene-trapped allele are noted as wild-type (+/+), heterozygous (+/-) or homozygous (-/-). (B) Verification of *Ppib* expression by Real-time RT-PCR using total RNA isolated from primary cultures and tissues from newborns and 8-week mice, respectively. Values represent the average of two independent fibroblast and osteoblast cultures, and five independent dermal and femoral samples. (C) Lengths of femora and tibiae from 8-week old wild-type (+/+), heterozygous (+/-) and homozygous (-/-) knock-out mice. Rhizomelia was not detected. (D) Growth curves of male and female mice from weaning (3 weeks) to 24 weeks of age. The gene-trapped ES cell line was generated by Mutant Mouse Regional Resource Center and screened for *Ppib* expression by W. Chang. Genotyping, Real-time RT-PCR, BMD and growth curve analyses were designed and performed by W.A. Cabral.

and tibial lengths were reduced 7% and 10% versus wild-type mice and heterozygotes ($p < 0.00004$ and $p < 0.02$, respectively) (Figure 16C). However, in contrast to *Crtap* and *P3H1* null mice [107,419], *Ppib*$^{-/-}$ mice do not have rhizomelia. The ratio of femoral to tibial length is comparable to wild-type (0.886 ± 0.020 vs 0.861 ± 0.035, $p = 0.07$).

Ppib is Required for Bone Development

Skeletal abnormalities of *Ppib*$^{-/-}$ mice include decreased mineralization and abnormal shape of calvaria, shortened limbs and a deformed and flared rib cage (Figure 17A); these features are accentuated in lethal null pups. Heterozygous mice are essentially normal. At 2 months of age, the rib cage of both heterozygous and homozygous mice has a narrow apex and drooping ribs at the base, providing limited space under the ribs for abdominal contents, which puff out the abdomen (Figure 17B). Homozygous mice also have kyphosis.

Adult knockout mice are osteoporotic; they have decreased aBMD of femora ($p = 0.001$) and vertebrae ($p = 0.02$) (Figure 17C). Femora of 2-month male *Ppib*$^{-/-}$ mice display altered cortical and trabecular structure on μCT analysis (Figure 17D). Their trabecular bone volume is half of wild-type, with reduction of both trabecular number and thickness. *Ppib*$^{-/-}$ trabeculae display the more rod-like architecture found in postmenopausal osteoporotic bone [469,470], as revealed by an increased structural model index (SMI, $p = 0.0003$). Femoral cortical bone is thinner, with decreased bone area and volume, decreasing predicted bone strength (MOI; $p = 0.03$).

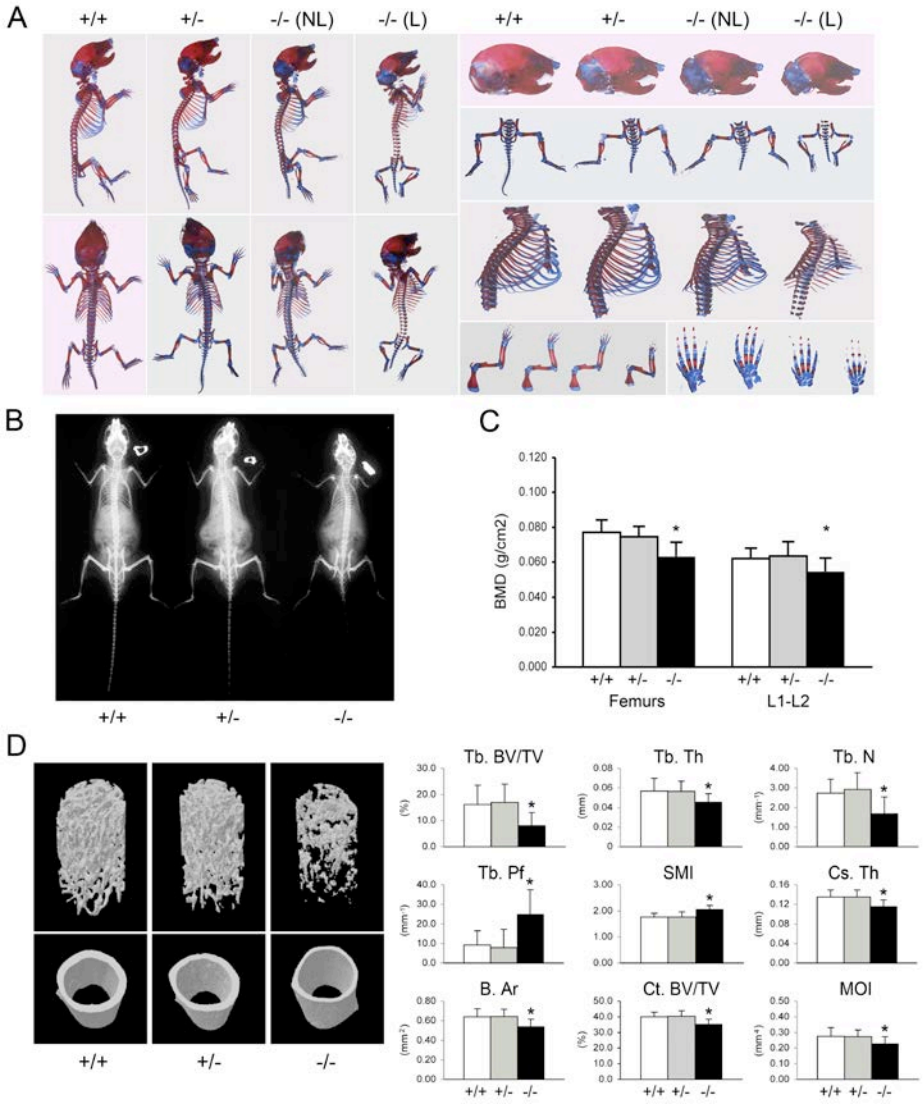

Figure 17. Absence of *Ppib* expression affects bone development. (A) Staining of newborn skeletons with Alizarin red (bone) and Alcian blue (cartilage) reveals undermineralization of calvaria and ribs. Homozygous mice have smaller size of whole skeleton and long bones, and a deformed rib cage. (B) X-rays of 8-week old mice. (C) DXA analysis of 8-week old mice (n=10/genotype). (D) Structural parameters of wild-type (+/+), heterozygous (+/-) and homozygous (-/-) femora at 8 weeks of age characterize reduced bone formation in CyPB-deficient mice (n = 9/genotype). *Left*, 3D reconstructions illustrate reduced trabecular and cortical bone volumes. *Right*, Trabecular parameters are decreased in homozygous (-/-) femora, including reduced bone volume (Tb BV/TV, p = 0.01), thickness (Tb Th, p = 0.04) and number (Tb N, p = 0.01). The increased trabecular bone pattern factor (Tb Pf, p = 0.004) indicates poor connectivity in CyPB-deficient trabecular bone versus wild-type. Structure Model Index (SMI), which indicates the extent of plates (0) and rods (3) is increased in homozygous (-/-) compared to wild-type (+/+) trabecular bone (p = 0.0003). Cortical bone parameters of CyPB-deficient mice are also reduced, with reductions in cross-sectional thickness (Cs Th, p = 0.008), area (B Ar, p = 0.02) and cortical bone volume (Ct BV/TV, p = 0.005). Skeletal staining, Faxitron and DXA were performed and analyzed by W.A. Cabral. Bone structural parameters determined by uCT analysis were provided by A.R. Blissett.

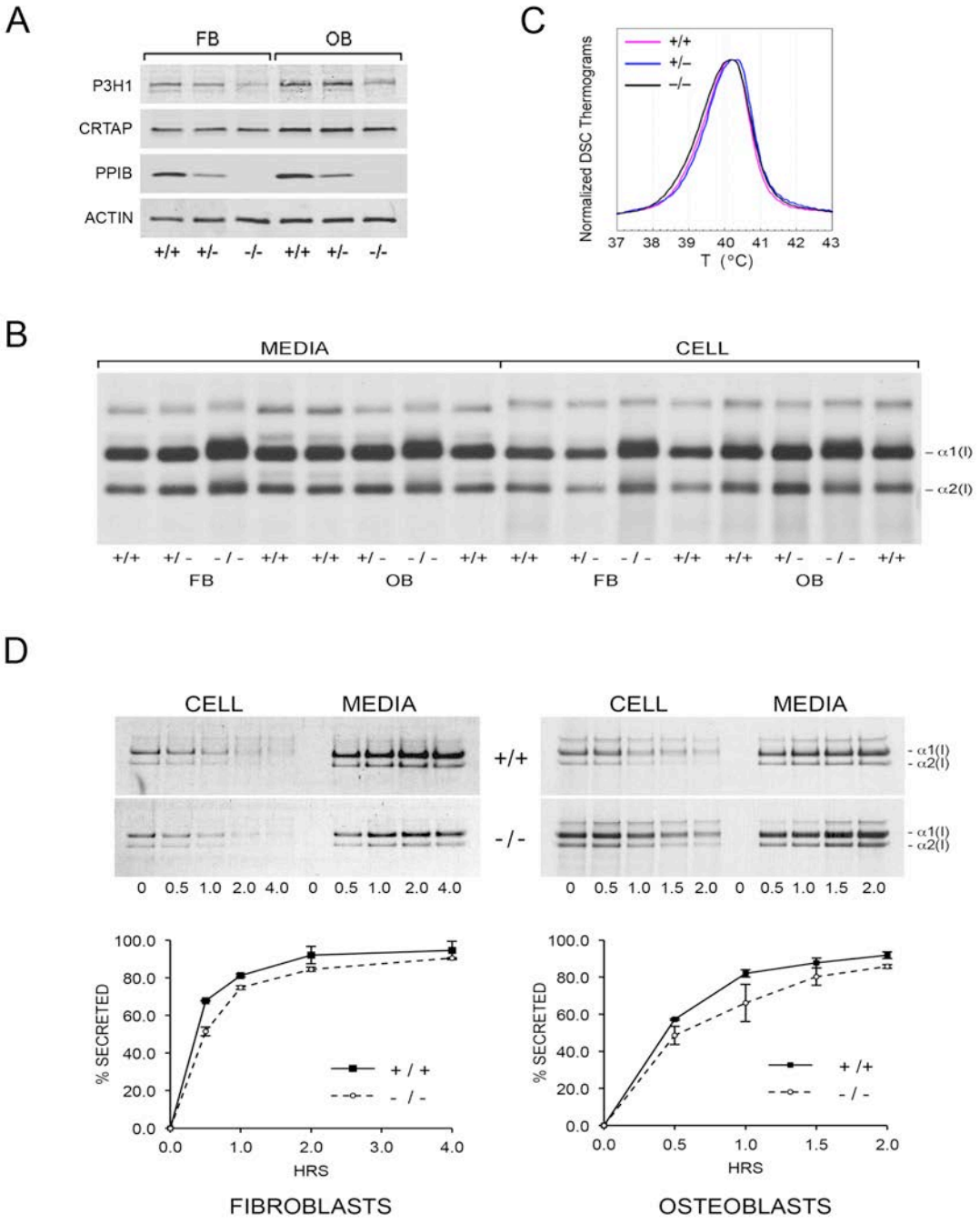

Figure 18. Synthesis of Type I collagen. (A) Western blots of cell lysates with antibodies to collagen 3-hydroxylation complex components. (B) SDS-Urea PAGE analysis of steady-state labeled type I collagen form wild-type (+/+), heterozygous (+/-) and homozygous (-/-) *Ppib*$^{-/-}$ fibroblasts (FB) and osteoblasts (OB). (C) Differential scanning calorimetry (DSC) analysis reveals no differences in thermal stability (T_m) of type I collagen secreted by fibroblast cultures. (D) Pulse-chase analysis of type I collagen synthesized and secreted by fibroblasts and osteoblasts in culture. Primary cell cultures, biochemical and Western analyses were generated and performed by W.A. Cabral. DSC was performed by E. Makareeva.

CyPB stabilizes prolyl 3-hydroxylation complex structure and function

We examined the stability and function of the collagen 3-hydroxylation complex. In *Ppib*$^{-/-}$ FB and OB, complete absence of CyPB was associated with a 50% reduction in P3H1 (Figure 18A), but did not alter CRTAP levels. Although CyPB was decreased in heterozygous cells, P3H1 and CRTAP were unchanged. Despite substantial levels of P3H1 and CRTAP in *Ppib*$^{-/-}$ cells, 3-hydroxylation of α1(I)P986 residues was severely reduced. In *Ppib*$^{-/-}$ mice, only 5-11% of OB and FB and 1-2% of dermal and bone tissue collagen α1(I) P986 residues were modified, in contrast to wild type and heterozygous cultures and tissues in which there is nearly complete collagen 3-hydroxylation (Table 6).

Table 6. Type I Collagen Post-Translational Hydroxylation

	Hyl / (Hyl+Lys)	4-Hyp / (Hyp+Pro)	% α2(I) P707 3-Hyp	% α1(I) P986 3-Hyp	% Telo Hyl (NH2, COOH)
+ / +					
fibroblast	20.0	49.5	0-10	97	ND
skin	12.9	40.2	6-10	97	ND
osteoblast	46.4	50.9	24	97	ND
femur	ND	ND	13-21	98	85, 85
humerus	25.2	44.6	ND	ND	ND
+ / –					
fibroblast	21.7	48.5	0-10	89	ND
skin	12.8	39.8	6-11	98	ND
osteoblast	ND	ND	24	97	ND
femur	ND	ND	10-12	98	ND
humerus	29.0	43.9	ND	ND	ND
– / –					
fibroblast	20.1	51.2	0-10	11	ND
skin	2.3	41.9	1-2	1-2	ND
osteoblast	43.7	52.7	13	5	ND
femur	ND	ND	5-7	2	95, 83
humerus	28.9	43.9	ND	ND	ND

ND, Not determined

Samples were generated by W.A. Cabral, and analyzed by M. Terajima, M. Yamauchi, D.R. Eyre and M.A. Weis.

The secondary prolyl 3-hydroxylation site was also undermodified; *Ppib*-null OB and femora have half the α2(I) P707 3-hydroxylation of wild-type and heterozygous samples.

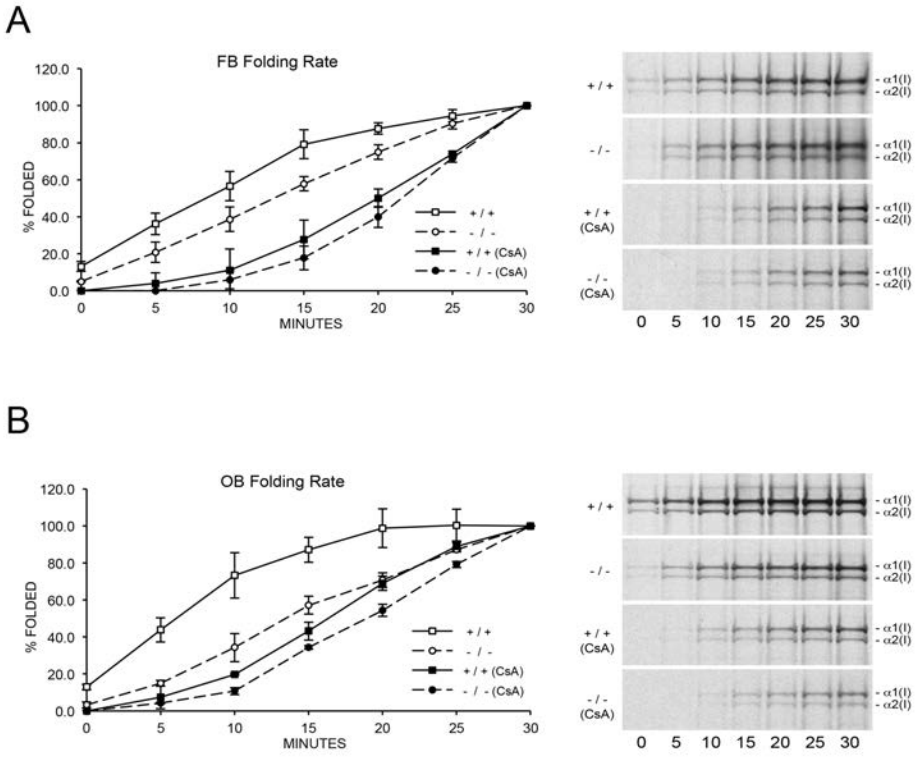

Figure 19. Cyclophilin B catalyzes folding of type I collagen. (A) Assay for intracellular folding of type I collagen in fibroblast cultures. Data represents the average from two independent cell lines for each genotype. (B) Assay for intracellular folding of type I collagen in calvarial osteoblast cultures. Assays performed and analyzed by W.A. Cabral.

CyPB role in collagen folding and helical modification

Type I collagen alpha chains from *Ppib*$^{-/-}$ FB and OB have delayed and broadened gel migration, consistent with delayed folding and overmodification of lysyl residues (Figure 18B). Collagen overmodification in the absence of a structural defect is expected to increase thermal stability [109,110]. However, type I collagen secreted by wild-type, heterozygous and homozygous null cells had equivalent melting temperatures (T_m) by differential scanning calorimetry (Figure 18C). In agreement with T_m results, amino acid

analysis of type I collagen secreted by *Ppib*$^{-/-}$ FB and OB revealed normal proportions of hydroxylated lysine and proline residues (Table 6).

To resolve these apparent discrepancies, we utilized a direct intracellular collagen folding assay. Accumulation of intracellular protease-resistant collagen was significantly slower in *Ppib*-null FB than in wild-type cells (Figure 19A). In *Ppib*$^{-/-}$ OB, the delay in type I collagen folding compared to wild-type was nearly double that in FB (Figure 19B). Notably, in both cell cultures, addition of the inhibitor CsA further delayed the rate of collagen folding in *Ppib*$^{-/-}$ as well as wild-type cells. These data suggest that, although CyPB is the primary PPIase involved in catalyzing folding of the type I collagen helix, other PPIases may also be capable of supporting this function.

CyPB Role in Site-Specific Regulation of Collagen Post-translational Modifications

The apparent discrepancy between slower folding rate and normal total lysyl hydroxylation of collagen from *Ppib*$^{-/-}$ mice led to a more detailed analysis of collagen hydroxylation and subsequent glycosylation in null cells and tissues. Biochemical analysis of *Ppib*$^{-/-}$ skin tissue showed a substantial decrease in lysyl hydroxylation (18% of wild-type) (Figure 20A) and glycosylation (40-50% of wild-type) (Figure 20B) in total collagen extracts. A site-specific analysis of skin and fibroblast lysine modification and crosslink patterns will be presented elsewhere. Decreased modification of skin-derived type I collagen resulted in slightly faster gel migration (Figure 20C). However, *Ppib*$^{-/-}$ bone collagen lysyl hydroxylation was slightly increased in heterozygous ($p = 0.0002$) and homozygous ($p = 0.005$) mice (Figure 20D), with corresponding increases in galactosylhydroxylysine (Figure 20E) and subtly broadened electrophoretic mobility

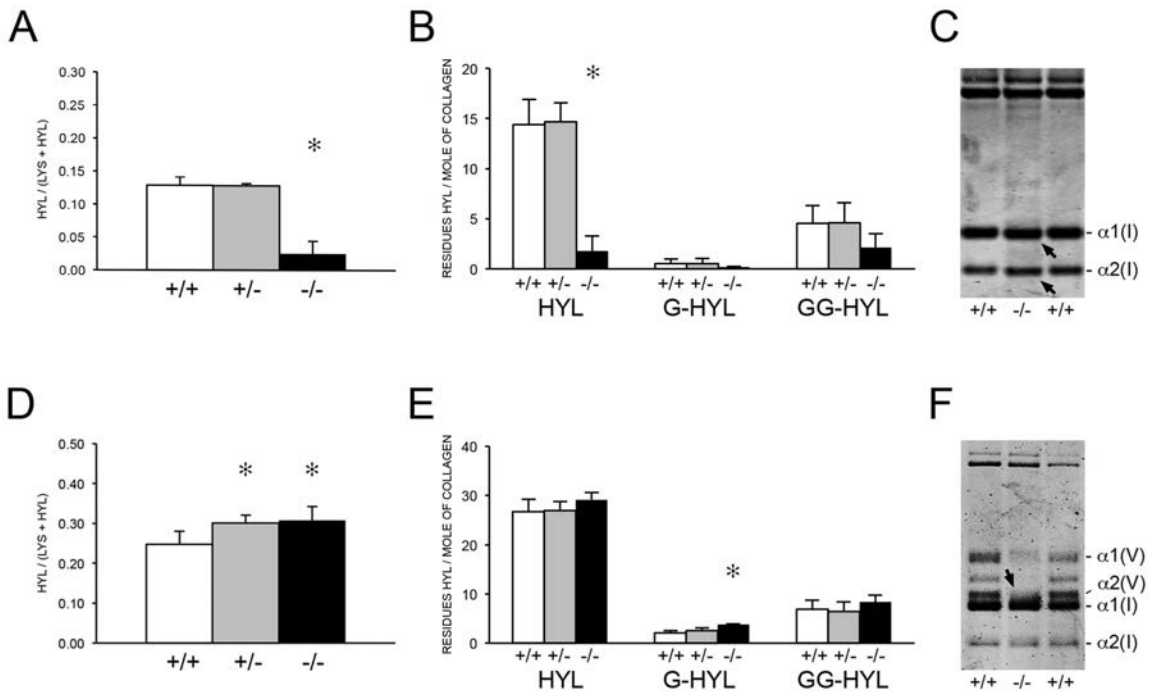

Figure 20. Post-translational modification of Type I collagen from tissues. (A) Total lysyl hydroxylation of type I collagen is dramatically reduced in dermal tissue from *Ppib*$^{-/-}$ mice. (B) Post-translational hydroxylation and glycosylation of type I collagen lysines in dermal tissue from *Ppib*$^{-/-}$ mice. (C) SDS-Urea PAGE analysis of pepsin extracts from dermal tissue demonstrates increased electrophoretic migration of *Ppib*$^{-/-}$ type I collagen alpha chains compared to wild-type. (D) Total lysyl hydroxylation of type I collagen extracted from bone tissue of heterozygous (+/-) and homozygous (-/-) *Ppib*-null mice compared to wild-type (+/+) bone collagen. (E) Analysis of post-translational lysine hydroxylation and glycosylation in *Ppib*$^{-/-}$ bone-derived type I collagen. (F) Type I collagen extracted from bone tissue displays backstreaking of α1(I) chains on SDS-Urea and is consistent with post-translational overmodification. Samples were generated by W.A. Cabral. Modification data generated and analyzed by M. Terajima, M. Yamauchi, D.R. Eyre and M.A. Weis. PAGE analysis performed by W.A. Cabral.

(Figure 20F). In addition, *Ppib*$^{-/-}$ bone extracts show a substantial decrease in type V collagen alpha chains.

Detailed characterization of lysine residues in collagen from OB cultures and bone tissue showed site-specific changes in lysine hydroxylation and glycosylation, and differences between cultured cells and tissues (Table 7; Figure 21A). In *Ppib*$^{-/-}$ OB collagen, only α1(I) K87, α2(I) K87 and K174 are underhydroxylated, with about 20% of K87 residues in *Ppib*$^{-/-}$ collagen unhydroxylated versus <1% of wild-type. However,

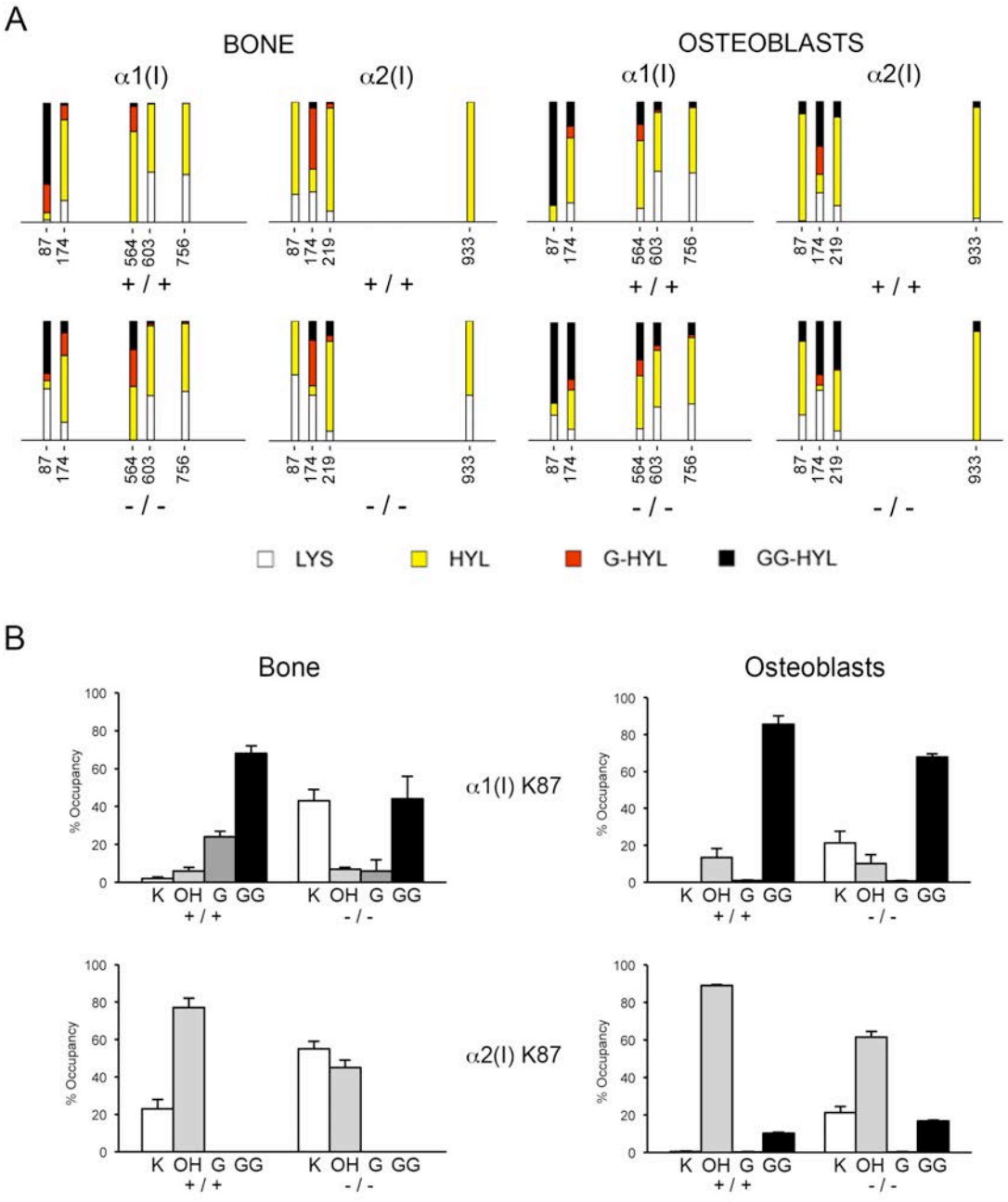

Figure 21. Altered post-translational modification of specific Type I collagen lysine residues in the absence of Cyclophilin B. (A) Quantitation of type I collagen lysine modifications by mass spectrometric analysis. Results were obtained from media of primary osteoblast cultures, or from 3-5 independent tissue samples for each genotype. White, Lys, lysine; Yellow, Hyl, hydroxylysine; Red, G-Hyl, galactosyl-hydroxylysine; Black, GG-Hyl, glucosylgalactosyl-hydroxylysine. (B) Modification of α1(I) and α2(I) K87 residues, important for crosslinking of type I collagen heterotypic fibrils in tissue. K, lysine; OH, hydroxylysine; G, galactosylhydroxylysine; GG, glucosylgalactosylhydroxylysine. Samples were generated by W.A. Cabral. Modification data generated and analyzed by I. Perdivara, under the supervision of K.B. Tomer.

Table 7. Post-translational Modification of Type I Collagen Lysine Residues

		BONE			OB	
		+/+	+/-	-/-	+/+	-/-
α1(I) K87	Lys	2.0 ± 1.0	2.0 ± 0.4	43.0 ± 6.0 *#	0.1 ± 0.1	21.2 ± 6.4 *
	Hyl	6.0 ± 2.0	13.0 ± 5.0	7.0 ± 1.0 #	13.4 ± 4.8	10.1 ± 4.7
	G-Hyl	24.0 ± 3.0	22.0 ± 7.0	6.0 ± 6.0 *#	0.9 ± 0.3	0.8 ± 0.1
	GG-Hyl	68.0 ± 4.0	63.0 ± 11.0	44.0 ± 12.0 *#	85.6 ± 4.6	67.7 ± 1.8 *
α1(I) K174	Lys	18.0 ± 1.0	19.0 ± 3.0	15.0 ± 2.0	15.5 ± 2.7	9.3 ± 1.9 *
	Hyl	67.0 ± 2.0	66.0 ± 2.0	56.0 ± 1.0 *#	54.2 ± 4.2	33.4 ± 1.4 *
	G-Hyl	12.0 ± 2.0	13.0 ± 1.0	19.0 ± 1.0 *#	9.7 ± 3.4	8.9 ± 2.9
	GG-Hyl	2.0 ± 0.0	2.0 ± 0.0	10.0 ± 2.0 *#	20.5 ± 3.5	48.3 ± 2.9 *
α1(I) K564	Lys	0.0	0.3 ± 0.2	0.0	11.0 ± 5.6	9.8 ± 2.4
	Hyl	76.0 ± 4.0	79.0 ± 3.0	45.0 ± 5.0 *#	56.3 ± 1.3	44.9 ± 4.9 *
	G-Hyl	21.0 ± 4.0	17.0 ± 2.0	31.0 ± 3.0 #	13.7 ± 2.5	13.4 ± 2.6
	GG-Hyl	3.0 ± 1.0	4.0 ± 1.0	24.0 ± 4.0 *#	18.9 ± 3.0	31.9 ± 4.7 *
α1(I) K603	Lys	42.0 ± 3.0	40.0 ± 1.0	37.0 ± 2.0	42.0 ± 4.6	28.2 ± 4.3 *
	Hyl	57.0 ± 3.0	59.0 ± 1.0	58.0 ± 2.0	49.1 ± 5.4	48.0 ± 4.8
	G-Hyl	0.6 ± 0.1	0.6 ± 0.0	2.0 ± 0.0 *	1.9 ± 0.1	4.1 ± 0.6 *
	GG-Hyl	0.2 ± 0.1	0.4 ± 0.0	2.0 ± 1.0 *#	7.0 ± 1.4	19.6 ± 2.1 *
α1(I) K756	Lys	40.0 ± 2.0	41.0 ± 1.0	41.0 ± 5.0	40.5 ± 6.8	31.5 ± 3.4
	Hyl	59.5 ± 3.0	58.5 ± 0.8	57.0 ± 6.0	54.5 ± 6.9	57.1 ± 1.9
	G-Hyl	0.4 ± 1.0	0.5 ± 0.2	1.5 ± 1.0	1.2 ± 0.1	2.3 ± 0.3 *
	GG-Hyl	0.1 ± 0.1	0.1 ± 0.1	0.6 ± 0.5	3.7 ± 0.2	10.8 ± 0.9 *
α2(I) K87	Lys	23.0 ± 5.0	20.0 ± 4.0	55.0 ± 4.0 *#	0.6 ± 0.1	21.3 ± 3.1 *
	Hyl	77.0 ± 5.0	80.0 ± 4.0	45.0 ± 4.0 *#	88.9 ± 0.7	61.6 ± 3.0 *
	G-Hyl	0.0	0.0	0.0	0.3 ± 0.1	0.3 ± 0.1
	GG-Hyl	0.0	0.0	0.0	10.2 ± 0.6	16.8 ± 0.4 *
α2(I) K174	Lys	25.0 ± 3.0	20.0 ± 8.0	38.0 ± 6.0 *#	24.1 ± 3.0	42.0 ± 2.4 *
	Hyl	19.0 ± 3.0	20.0 ± 2.0	8.0 ± 3.0 *#	15.3 ± 2.1	4.1 ± 0.6 *
	G-Hyl	51.0 ± 3.0	53.0 ± 6.0	38.0 ± 4.0 *#	23.4 ± 2.3	8.9 ± 0.6 *
	GG-Hyl	5.0 ± 3.0	7.0 ± 2.0	16.0 ± 2.0 *#	37.3 ± 2.3	45.0 ± 2.1 *
α2(I) K219	Lys	9.0 ± 2.0	8.0 ± 2.0	8.0 ± 2.0	13.3 ± 0.6	7.8 ± 0.6 *
	Hyl	86.0 ± 1.0	86.0 ± 2.0	75.0 ± 3.0 *#	73.6 ± 5.8	50.9 ± 1.0 *
	G-Hyl	4.0 ± 1.0	4.0 ± 1.0	5.0 ± 1.0	0.8 ± 0.3	1.2 ± 0.3
	GG-Hyl	1.0 ± 0.6	2.0 ± 0.0	12.0 ± 4.0 *#	12.3 ± 5.9	40.1 ± 1.6 *
α2(I) K933	Lys	0.0	0.0	38.0 ± 3.0 *#	2.9 ± 0.2	0.0 *
	Hyl	100.0	100.0	62.0 ± 3.0 *#	92.0 ± 2.8	91.2 ± 1.9
	G-Hyl	0.0	0.0	0.0	1.0 ± 0.4	0.9 ± 0.2
	GG-Hyl	0.0	0.0	0.0	4.0 ± 2.4	7.9 ± 1.7

* $p < 0.05$ between +/+ and -/-; # $p < 0.05$ between +/- and -/-; no significant difference between +/+ and +/- in bone

Samples were generated by W.A. Cabral. Modification data generated and analyzed by I. Perdivara, under the supervision of K.B. Tomer.

hydroxylation of other helical lysines from *Ppib*$^{-/-}$ OB collagen was normal or subtly increased, i.e. α1(I) K603 and K756 hydroxylation were increased 14% and 9%, respectively (Table 7). Bone tissue collagen from *Ppib*$^{-/-}$ mice has an even more striking increase in the proportion of lysine residues involved in crosslink formation that are unhydroxylated. Hydroxylation of α2(I) and α1(I) K87 residues is decreased 30-40%, respectively, and α2(I) K933 hydroxylysine content is decreased 38%, compared to wild-type bone (Table 7). Decreased hydroxylation is also evident at α2(I) K174, but other residues do not have significant changes.

Glycosylation patterns are also altered in *Ppib*$^{-/-}$ collagen. In wild-type OB collagen, a major helical cross-linking residue, α1(I) K87, has the highest proportion of diglycosylation, while α2(I) K87 is mainly non-glycosylated (Table 7 and Figure 21B), as we previously reported [124,471]. In *Ppib*$^{-/-}$ collagen, mono- and di-glycosylation of α1(I) K87 is decreased in tissue and OB culture collagen, and α2(I) K87 di-glycosylation is only slightly increased in OB culture collagen, despite slow helical folding (Figure 21B). In OB and bone-derived *Ppib*$^{-/-}$ collagen, all other lysine residues assayed have substantial increases in glycosylation, corresponding with delayed intracellular collagen folding. We observed increased glycosylation at α1(I) K174, K564, K603, K756 and α2(I) K219 in *Ppib*$^{-/-}$ OB collagen, and to a lesser extent in *Ppib*$^{-/-}$ bone-derived collagen, (Table 7), in agreement with the increased total lysyl hydroxylation and broad gel migration (Figure 20D, E).

Collagen helical lysine modifications are catalyzed primarily by lysyl hydroxylase 1 (LH1) for hydroxylation, GLT25D1, for galactosylation of hydroxylysine [121], and LH3, which harbors both lysyl hydroxylase and glucosyltransferase activities [83,86,126,471]. By

real-time RT-PCR, we found a modest increase in transcript levels for all three enzymes in *Ppib*$^{-/-}$ FB cultures, but not in skin, OB or femoral tissue (Figure 22A-C). Importantly, the protein levels of these enzymes were normal on western blots of FB and OB lysates (Figure 22D). These data verify that expression levels of the modifying enzymes do not account for the alterations in lysine modification demonstrated in collagen from *Ppib*$^{-/-}$ cells and tissues.

Absence of CyPB affects collagen synthesis and secretion

The combined effects of CyPB absence, slower collagen folding and abnormal modification, impact procollagen synthesis and secretion. Pulse-chase experiments show increased (nearly double) total collagen synthesis by CyPB-deficient OB, while fibroblasts produce about half the amount of collagen as wild-type cells (Figure 18D, upper). The kinetics of collagen secretion is only slightly delayed for *Ppib*$^{-/-}$ FB and OB (Figure 18D, lower). This finding is similar to our previous report on FB with null mutations in *LEPRE1* [110], but distinct from the reported increased rate of secretion in FB with null mutations in *CRTAP* [107], and of uncertain physiological significance.

Absence of CyPB Affects Collagen Matrix Deposition and Crosslinking

To determine the consequences of altered collagen post-translational modification on extracellular matrix, we analyzed collagen deposition into matrix in culture. Sequential extraction of the incorporated labeled collagen revealed that collagen deposited into maturely cross-linked matrix by OB (Figure 23A, left) and FB was decreased 80% and 70%, respectively, compared to wild-type cultures. Raman spectroscopy of OB-derived

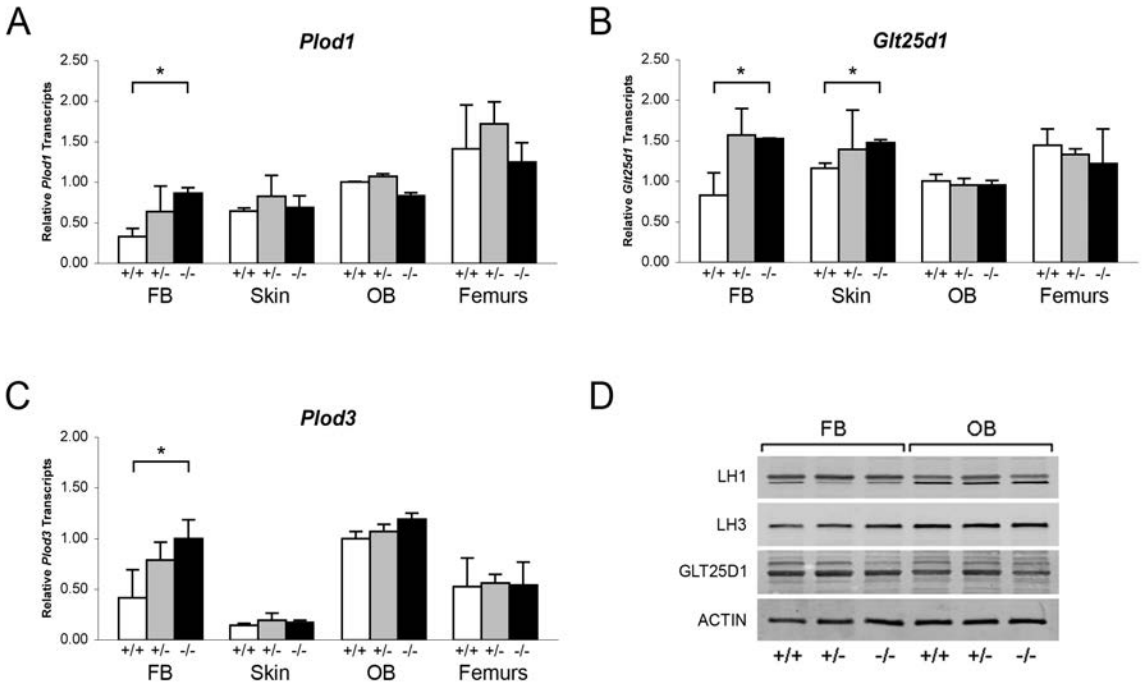

Figure 22. Expression of ER resident collagen helical lysine modification enzymes. (A-C) Real-time RT-PCR of total RNA from two independent cell cultures of newborn fibroblasts (FB) and calvarial osteoblasts (OB), and 5 independent skin and femoral samples for each genotype at 8 weeks of age. (D) Western blots of cell lysates probing for lysyl hydroxylase 1 (LH1/PLOD1), lysyl hydroxylase 3 (LH3/PLOD3), glycosyltransferase 25 domain containing 1 (GLT25D1), and β-actin in wild-type (+/+), heterozygous (+/-) and homozygous (-/-) *Ppib*-null cells and tissues. Samples generated and analyzed by W.A. Cabral.

matrix corroborated the biochemical analysis (Figure 23A, right), indicating a two-thirds reduction in matrix collagen content compared to wild-type.

Consistent with the marked reduction in hydroxylation of the helical crosslinking lysine in bone in the absence of CyPB, there was a significant difference in the crosslinking pattern. For divalent crosslinks, there was a nearly three-fold increase in the helical lysine-involved crosslink, hydroxylysinonorleucine (HLNL), in homozygous humeral bone ($p = 1.9 \times 10^{-8}$), while the helical hydroxylysine-involved dihydroxylysinonorleucine (DHLNL) is comparable to wild-type bone (Figure24A). The increase in HLNL crosslinks decreases the DHLNL/HLNL ratio in *Ppib*$^{-/-}$ humerii (p= 0.0004). For the trivalent mature crosslinks, hydroxylysylpyridinoline (HP) crosslinks

were unchanged in *Ppib*$^{-/-}$ bone relative to wild-type. However, the helical lysine-involved crosslink, lysylpyridinoline (LP), was markedly increased by four to five-fold in *Ppib*$^{-/-}$ humerii ($p = 2.8 \times 10^{-9}$) and femoral tissue ($p = 0.0001$), respectively (Figure 24B). The resulting HP/LP ratio in CyPB-deficient bone is decreased 4.25-fold in humeral ($p = 0.00004$) and 5.6-fold in femoral bone tissue ($p = 0.003$). It is also quite noteworthy, in the context of both the reduced collagen deposition into *Ppib*$^{-/-}$ osteoblast matrix and the increase in non-hydroxylated forms of crosslinks in tissue, that total bone collagen crosslinks were increased in null mice versus wild-type at 2 months of age (DHLNL+HLNL, $p=0.002$; HP+LP, $p = 0.00004$ and 0.003 for humerus and femora, respectively).

The abnormal collagen modification and crosslinking affects the structure and organization of collagen fibrils. In dermal fibrils of *Ppib*$^{-/-}$ mice, we noted the presence of disorganized aggregate forms, as well as a 25% increase in the average fibril diameter ($p = 1.7 \times 10^{-15}$), and a broader distribution of fibril diameters ($p = 4.0 \times 10^{-8}$) (Figure 23B). In contrast to dermal fibrils, collagen fibrils from femoral tissue were less densely packed and had visibly decreased diameters compared to wild-type. Bone fibril diameters were not quantitated because multiple bundle orientations in each field does not allow a clear analysis of fibril cross-sections.

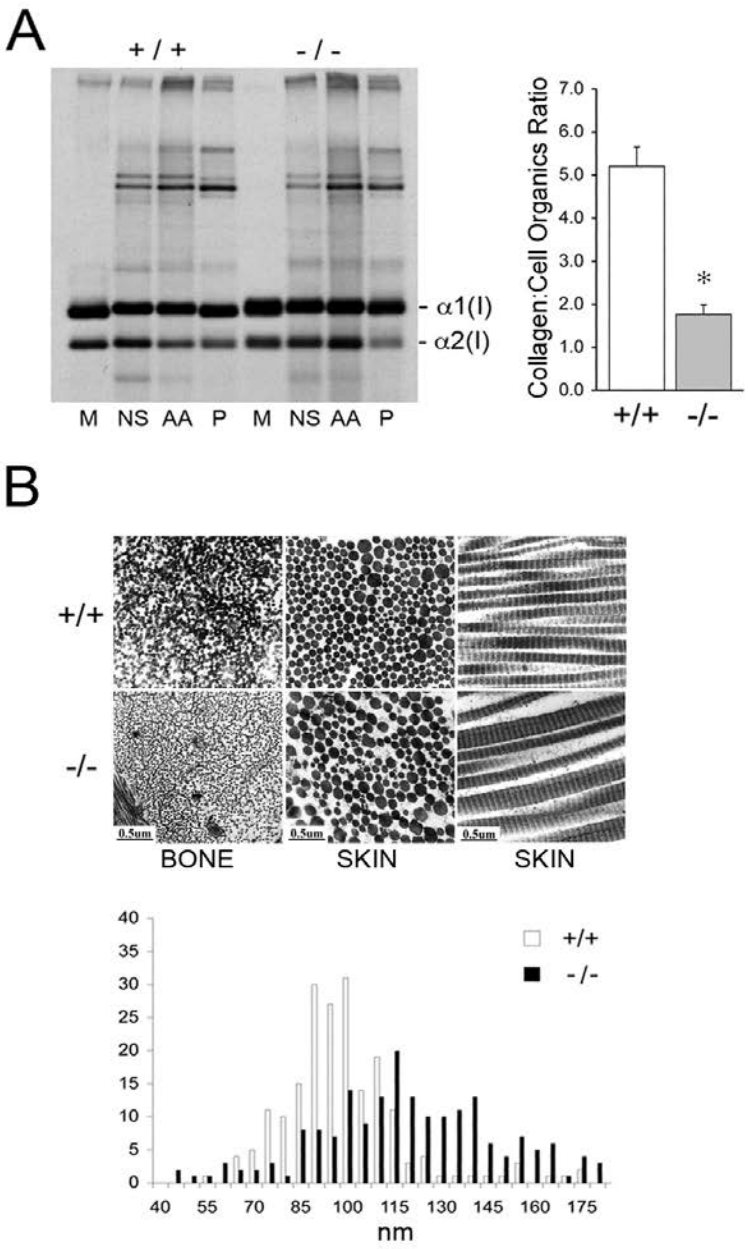

Figure 23. Dysregulation of collagen deposition and fibril assembly. (A) *Left*, Deposition of type I collagen by osteoblasts into extracellular matrix in culture. Post-confluent cultures were pulsed for 24 hr, followed by serial extraction of incorporated collagens from the media (M), neutral salt (NS), acid soluble (AA, immaturely crosslinked) and pepsin soluble (P, maturely crosslinked) fractions of the matrix. *Right*, Matrix collagen to cell organics ratio from Raman micro-spectroscopy shows decreased collagen content in matrix deposited by homozygous *Ppib*-null (-/-) versus wild-type (+/+) osteoblasts in culture ($p = 0.002$). (B) Transmission electron micrographs of femoral and dermal collagen fibrils from 8 week-old wild-type (+/+) and CyPB-deficient mice (-/-). Diameters of 200 dermal fibrils were measured for each sample and plotted, right. Samples generated by W.A. Cabral. PAGE and fibril diameter distribution analysis performed by W.A. Cabral. TEM performed by J.F.E. Enterprises. Raman spectroscopy data generated and analyzed by E. Mertz.

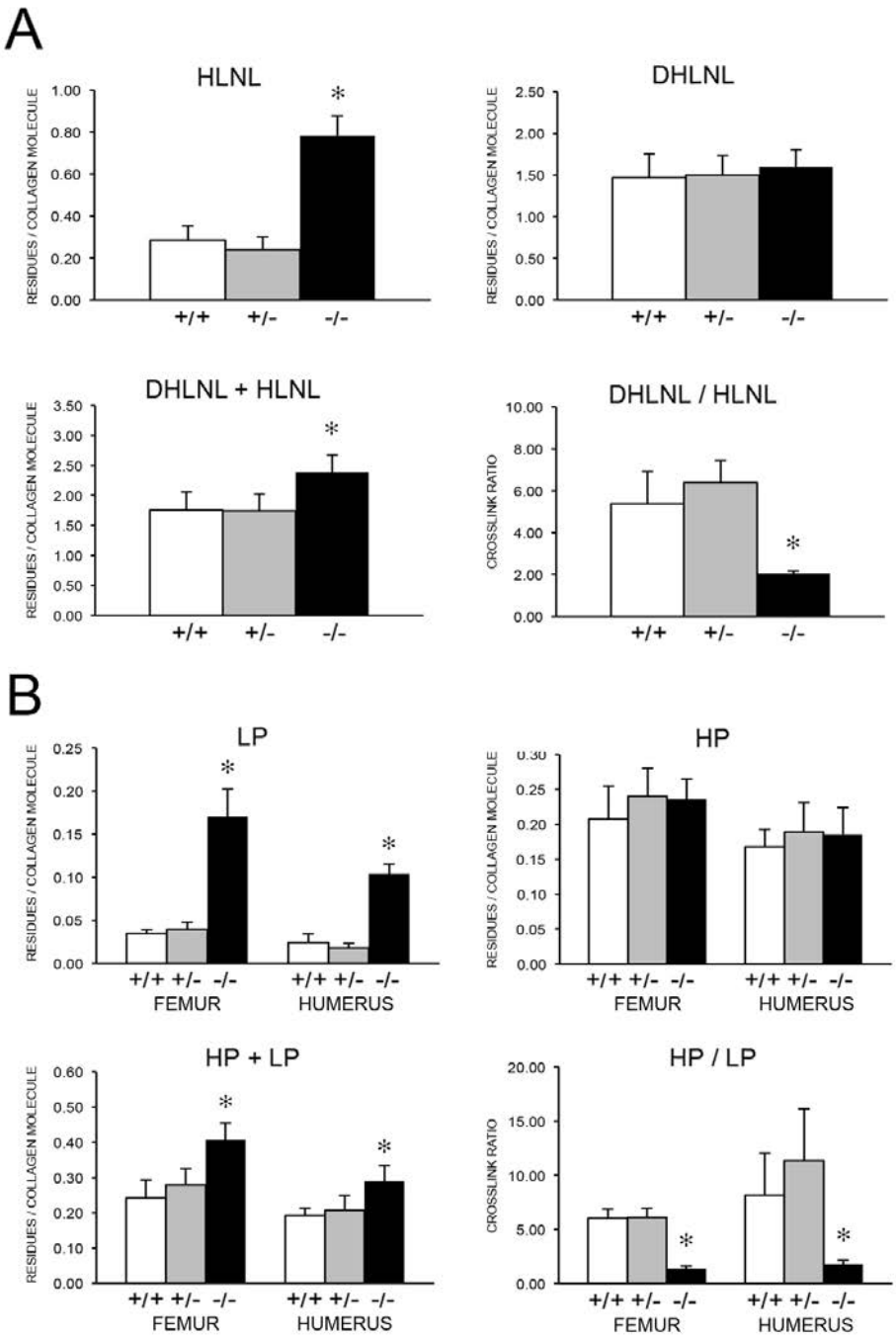

Figure 24. Abnormal crosslink profile in *Ppib*-null mouse bone. (A) Quantitation of divalent crosslinks in murine humerii reveals increased HLNL (hydroxylysinonorleucine) crosslinks, which require helical lysine residues, but no change in DHLNL (dihydroxylysinonorleucine) crosslinks, which involve helical hydroxylysine residues. (B) Quantitation of trivalent crosslinks in murine humerii and femora. Total pyridinoline crosslinks are increased due to an increase in lysyl pyridinoline (LP), but not hydroxylysyl pyridinoline (HP) crosslinks in bone. Samples were generated by W.A. Cabral. Crosslink data generated and analyzed by M. Terajima, M. Yamauchi, D.R. Eyre and M.A. Weis.

DISCUSSION

To investigate the role of CyPB/PPIB in post-translational modification and folding of type I collagen, we generated a CyPB knockout mouse using a *Ppib* gene-trapped ES cell line. We confirmed complete absence of *Ppib* transcripts in cultured cells, as well as dermal and bone tissue of homozygous null mice. Furthermore, CyPB was undetectable in immunoblots of both fibroblast and osteoblast cultures. The resulting murine phenotype reproduces the clinical findings in patients with *PPIB* deficiency, including growth deficiency with bone deformities, reduced bone mineral density and decreased trabecular and cortical bone volume. Collagen folding is delayed in $Ppib^{-/-}$ cells, and further folding delay by CsA inhibition provides support for an additional collagen PPIase. Unexpectedly, CyPB deficiency also results in tissue- and site-specific alterations of post-translational hydroxylation and glycosylation of type I collagen, which secondarily affect fibril assembly, crosslinking in matrix, and bone mineralization.

Our initial goal in generating this mouse model of CyPB deficiency was to address the inconsistent findings of type I collagen lysyl and P986 hydroxylation among patients with type IX OI. In our *Ppib* KO mice, α1(I) P986 3-hydroxylation is nearly absent from dermal and bone collagen, demonstrating that CyPB is required *in vivo* for 3-hydroxylation complex activity in tissues. This is consistent with the 3-hydroxylation status reported in bone and skin collagen in a previous *Ppib* murine model, as well as with <1% modification of P986 residues in α1(I), α1(II) and α2(V) chains from bone, skin and cartilage of *Crtap*-null mice [107,112,405]. P986 3-hydroxylation of type I collagen was absent in tail tendon and bone and was reduced, but not absent, in skin of P3H1-null mice [419,420]. Furthermore, the finding that absence of CRTAP and CyPB, but not P3H1,

completely abrogates α1(I) P986 hydroxylation raises the possibility that there is tissue-specific functional redundancy of P3H isoforms in mice. We also observed a 50% reduction of 3-hydroxylation at the secondary α2(I) P707 site in collagen from skin and bone, suggesting that P3H1 or other P3H isoforms can modify this site without requiring CyPB for activity. For other sites and/or collagen alpha chains, the requirement for CyPB may be quite stringent. Delayed electrophoretic migration of collagen alpha chain, caused by increased post-translational modification, is considered to be a reliable surrogate assay for slower helix folding. Steady-state collagen from *Ppib*$^{-/-}$ cultured fibroblasts and osteoblasts has delayed gel mobility, consistent with the significant delay in helical folding of type I collagen demonstrated in direct intracellular folding assays in both cell types. Collagen folding in *Ppib*$^{-/-}$ fibroblasts and osteoblasts was delayed 5 to 8 minutes, respectively. By comparison, fibroblast cell lines from OI patients harboring typical structural defects in type I collagen demonstrated 5 to 60 minutes delay in formation of the collagen helix versus control cells, depending on the glycine substitution [356].

Treatment of control fibroblast and osteoblast cultures with cyclosporin A (CsA), a non-specific cyclophilin inhibitor, resulted in a 10-12 minute delay in collagen folding, similar to published experiments [99,108]. Unexpectedly, CsA treatment of CyPB-deficient cells further increased the collagen folding delay by 6-8 minutes beyond that seen in untreated cells. These data suggest that another CsA-sensitive PPIase is capable of partial compensation for CyPB deficiency, and may be involved in collagen folding under normal conditions. We previously reported a type IX OI patient with total absence of CyPB due to a start codon substitution [421]. Fibroblast collagen from this patient had normal helical modification suggestive of normal collagen folding, prompting us to

propose redundancy of collagen *cis-trans* isomerases. The identity of this PPIase is open to speculation. Only one cyclophilin, but at least six FK506 binding proteins (FKBPs), occur in the rER [93,458]. Until recently, FKBP65, which binds gelatin and is partially inhibited by CsA *in vitro,* was an attractive candidate [101]. However, addition of FK506 to cell cultures does not delay collagen folding, and absence of FKBP65 in patients with OI, Bruck syndrome and Kuskokwim disease has negligible effects on collagen helical folding [103,472,473], limiting FKBP65 to at most a compensatory role in the absence of CyPB. Similarly, mutations in *FKBP14*, which encodes FKBP22, were identified in an Ehlers-Danlos like-syndrome similar to the kyphoscoliotic type of EDS (EDS VIA) caused by LH1 deficiency, but their types I, III and V collagens showed normal electrophoretic migration [474].

One of the striking findings of this study was the difference in post-translational modification in collagen isolated from cell cultures versus dermis and bone. *Ppib*$^{-/-}$ fibroblast and osteoblast steady-state type I collagen has significant electrophoretic delay, but normal total hydroxylysine content, which suggested that increased glycosylation delayed gel mobility. In contrast, collagen extracted from dermal tissue of CyPB-deficient mice has an 80% decrease in total helical hydroxylysine content, leading to more rapid gel migration of alpha chains than collagen from wild-type skin. The decreased hydroxylation of collagen in dermal tissue of the *Ppib*$^{-/-}$ mouse is similar to the dermal tissue data from American quarter horses with hyperelastosis cutis caused by a homozygous *Ppib* missense mutation that does not impair its isomerase function *in vitro*. The significantly decreased hydroxylysine content of dermal collagen in the horse is proposed to result from loss of a CyPB-LH1 interaction required for LH1 activity [427].

Type I collagen extracted from humerii of *Ppib*$^{-/-}$ mice had different post-translational modification than both osteoblasts and dermal tissue. Bone-derived collagen alpha chains demonstrated subtle broadening and backstreaking on PAGE analysis, consistent with the increased global hydroxylation and glycosylation of lysyl residues determined by amino acid analysis. However, detailed mass spectrometric analysis revealed site-specific alterations in lysine post-translational modification of bone tissue collagen. We found dramatic underhydroxylation of lysine residues involved in the formation of intermolecular crosslinks in tissue, specifically at α1(I) K87 (57% Hyl in null vs 98% in WT), α2(I) K87 (45% Hyl in null vs 77% in WT), and α2(I) K933 (62% Hyl in null vs 100% in WT). Significant decreases in hydroxylation of these residues are also seen in CyPB-deficient osteoblast collagen. It is generally accepted that hydroxylation of helical lysine residues is catalyzed by LH1. However, LH1 stability does not appear to be affected in *Ppib*$^{-/-}$ tissues based on western blot analysis (Figure 22). Whether CyPB contributes to LH1 folding, or stabilizes its activity by binding, are among the possibilities to be investigated. Thus absence of CyPB from bone has critical direct and indirect effects on collagen, affecting both collagen folding and the activity of a major collagen-modifying enzyme, respectively. This gives CyPB a pivotal role in determining the structure of secreted collagen.

The site-specific changes in lysine hydroxylation seen in CyPB-deficient cells and tissues raises the possibility of an additional regulatory mechanism in bone, where there is apparent functional redundancy of LH activity directed at helical lysines not involved in crosslinking. Primary LH1 and LH2 deficiency demonstrated that LH1 hydroxylates lysines in the collagen triple helix while LH2 functions as the telopeptidyl lysyl

hydroxylase [81,475-477]. However, collagen demonstrated variable decreases in Hyl content in bone from Ehlers-Danlos Type VIA patients (10-43% of normal) and a *Plod1* KO mouse (75% of wild-type), although specific helical lysines involved in crosslinks were always underhydroxylated based on decreased HP/LP crosslink ratios [72,478,479]. Thus, although LH1 appears to be required as the primary hydroxylase for helical domain crosslinking lysines in bone collagen, the relative roles of LH1, LH2 and LH3 at other helical sites and in other tissues is less well understood. LH3, which has LH, galactosyltransferase and glucosyltransferase activity *in vitro* and in culture [83,84,471,480], could be the source of helical Hyl and even increased lysine glycosylation, in *Ppib*$^{-/-}$ bone [481]. Both a KO murine model for LH3 and a mouse with inactivated LH3 hydroxylation have faster gel migration of fibroblast type I collagen [482]. In the single human case of LH3 deficiency reported, the disaccharide derivative of pyridinoline crosslinks was absent in the patient's urine [126].

The intracellular hydroxylation of collagen helical K87 and telopeptidyl lysine residues determines the collagen crosslink pathway in bone matrix. Crosslinks between collagen molecules in extracellular matrix are crucial to skeletal function because they contribute to matrix stability, bone strength and elasticity. In both dominant OI caused by primary collagen defects and recessive OI caused by CRTAP and P3H1 deficiency, the overhydroxylation of collagen helical lysines leads to increased divalent DHLNL/HLNL and trivalent HP/LP collagen crosslink ratios [115,483-486]. In contrast, in *Ppib*$^{-/-}$ mice the underhydroxylation of K87 residues in tissue collagen results in substantial decreases in DHLNL/HLNL and HP/LP ratios in bone. Notably, these changes are similar to findings in primary LH1 deficiency, in which bone-derived urinary peptides reflect decreased

hydroxylation of collagen lysine residues involved in crosslink formation [475], and decreased HP/LP ratios and increased total pyridinoline crosslinks were observed [75,89,475]. The cause of increased total crosslinks seen in *Ppib*$^{-/-}$ bone collagen is not clear. Possibly, decreased K87 hydroxylation and glycosylation may favor binding or activity of lysine oxidase at this site [124,487,488].

Eyre and colleagues proposed that the collagen 3-hydroxylation modification supports matrix supramolecular assembly by fine-tuning the intermolecular alignment of collagen molecules to facilitate crosslink formation [115]. Our finding of severe reduction in collagen deposition into matrix is consistent with this hypothesis. However, we now understand that collagen secreted by CyPB-deficient osteoblasts has site-specific alterations in hydroxylation and glycosylation as well as absent P986 3-hydroxylation, which could also alter matrix assembly in addition to changes in collagen crosslinking and fibril structure. Furthermore, although this investigation focused on type I collagen, extracts from *Ppib*$^{-/-}$ bone show a substantial decrease in the quantity of type V collagen alpha chains (Figure 20F).

The effect of altered collagen crosslinking in *Ppib*$^{-/-}$ bone on mineralization remains to be explored. Bone from OI caused by collagen structural defects or CRTAP deficiency, whose crosslink pattern is opposite to this *Ppib*$^{-/-}$ mouse, is paradoxically hypermineralized. Collagen crosslinking and bone mineral crystallinity were strongly correlated in long bones of several congenic mouse strains [489]. On the other hand, increased HP/LP ratios correlate with the ultimate compressive strength of trabecular bone, but are independent of BMD [490,491]. Comparison of mineralization in *Ppib*$^{-/-}$ and classical OI bone will provide insight into the interaction of bone crosslinks and

mineralization, and could point to collagen modification as an intracellular pathway by which osteoblasts can actively influence bone mineralization.

ACKNOWLEDGEMENTS

The authors wish to thank Jyoti Rai at the University of Washington for expert technical assistance.

AUTHOR CONTRIBUTIONS

All experiments were conceived and designed by WAC, IP, MAW, WC, SL, DRE, MY and JCM. WAC, IP, MAW, MT, ARB, WC, ENM and EM conducted the experiments. The manuscript was written by WAC and JCM. All authors participated in data analysis and critical review of the manuscript prior to submission.

Chapter IV: Conclusions

Mechanism of 3-Hydroxylation Defects: Recessive OI

The recent discovery that defects in the collagen prolyl 3-hydroxylase complex cause recessively inherited bone dysplasia has created a paradigm shift in osteogenesis imperfecta. Once considered the phenotypic outcome of collagen structural defects, OI is now thought of as a collagen-related disorder also encompassing the cellular machinery involved in collagen biosynthesis. Characterization of the distinct and shared metabolic pathways that are affected in dominant and recessive OI has dramatically altered our understanding of bone development, improved OI detection and diagnosis, and offered exciting new potential for therapeutic approaches to treating this disease.

Null mutations in *CRTAP* and *LEPRE1* (P3H1) cause severe to lethal recessive types VII and VIII OI with the distinctive features of extrememely low BMD ($z \leq -6$), undertubulated long bones, bulbous metaphyses and rhizomelia (relative shortening of the proximal segment of the limbs) [354,492]. Types VII and VIII OI are considered osteochondrodystrophies that affect both bone and cartilaginous tissues. This is likely due to the putative roles of secreted CRTAP and P3H1 proteins in these tissues, as well as to the intracellular roles of the P3H complex as a collagen folding chaperone and in 3-hydroxylation of α1(I) and α1(II) P986 residues in osteoblasts of bone and chondrocytes of cartilage, respectively. Furthermore, the overlapping phenotypes resulting from CRTAP and P3H1 deficiency can also be attributed to their mutual stabilization in the ER; absence of one component results in complete loss of both [457]. Thus all cases of CRTAP and P3H1 deficiency share the biochemical feature of type I collagen overmodification that is commonly seen in moderate to severe forms of dominant OI [354].

In dominant OI overmodification of collagen alpha chains is clearly due to a delay in helical folding and subsequent increased exposure to the post-translational modifying enzymes in the ER. The same increased hydroxylation and glycosylation of type I collagen helical lysyl residues occurs in CRTAP and P3H1 deficiency, and may reflect a role for the P3H complex as a collagen folding chaperone.

However, it is still unclear if absence of the complex or absence of the collagen modification is the underlying defect of collagen prolyl 3-hydroxylase deficiency in types VII and III OI. Indeed, the collagen 3-hydroxylation modification itself has been proposed as a passive marker of intracellular complex chaperone activity [456], a local stabilizing modification to the triple helix [114,493], a non-collagenous protein binding site [494], and as a recognition sequence that mediates intermolecular alignment required for efficient collagen fibril assembly [115]. Generation of mouse models expressing either type I or type II collagens unable to be modified by the P3H complex should help to clarify the pathophysiology of these disorders.

Frequency of Recessive OI

Osteogenesis imperfecta occurs with a frequency of 1 in 15-20,000 births in Caucasian American and Northern European populations. Of these cases, approximately 85% are due to dominant mutations of *COL1A1* and *COL1A2* causing Classical OI. Homozygosity or compound heterozygosity for collagen mutations is extremely rare, with only three reported instances, and in most populations familial recurrence of OI is attributed to autosomal dominant parental mosaicism [358]. However, familial recurrence of

severe forms of OI due to recessive inheritance of collagen 3-hydroxylation defects may account for as much as 5% of OI cases in North America and Europe. In contrast to the low frequency of recessive OI types in Caucasian populations, a few isolated populations have been identified with high carrier frequencies of recessive alleles that cause types VII and VIII OI.

The first of the recurring mutations was identified in an isolated First Nations community in northern Quebec, affected individuals of which were first characterized with type VII OI [107]. Although the carrier frequency of the *CRTAP* c.472-1021C>G mutation within this small community (approximately 2500 individuals) has not been established, haplotype analysis of one large three-generation pedigree, including 8 affected, suggests a founder mutation six to eight generations ago [407]. A recurring *LEPRE1* mutation (c.232delC) also occurs in the genetically isolated Irish Travellers [409], which according to one estimate may cause recessive OI in as many as 1% of births in this population [495]. This rate of OI in Irish Travellers predicts that 1 in 5 individuals may be carriers for the c.232delC allele.

The most recent data reveals that over one quarter (22/82) of the disease-causing alleles reported for *LEPRE1* are represented by the c.1080+1G>T mutation. Commonly referred to as the West African allele, this mutation was initially identified in four of five children first identified with type VIII OI [110]. Each of the probands harbouring this mutation were children of African American or West African émigré families, suggesting the presence of a recurring mutation among individuals of West African descent. Using anonymized samples, we determined an African American carrier frequency of 0.4% by screening genomic DNA from more than 3000 infants of the Mid-Atlantic region of the

United States [413]. The carrier frequency among African Americans predicts an incidence of type VIII OI due to this mutation alone at 1 in 250,0000 births, which is in the range of typical rare recessive conditions. Thus among the nearly 600,000 African American infants born in the U.S. each year, only two or three cases of recessive OI due to homozygosity for the *LEPRE1* c.1080+1G>T mutation would be expected to occur. However, given the estimated rate of lethal OI as 1 in 40-60,000 births in North America, the carrier frequency suggest that 15-23% of cases of lethal OI in African Americans will be caused by homozygosity for the West African allele. Thus, while the *LEPRE1* c.1080+1G>T mutation does not contribute significantly to overall African American perinatal deaths in the U.S., the frequency of carriers makes it important for genetic counseling purposes.

An additional 3000 genomic DNA samples from contemporary African cohorts provided convincing evidence for our hypothesis of a West African origin for this allele; the mutation was detected in nearly 1.5% of samples screened from Ghana and Nigeria, but was not identified among additional cohorts from other African regions [413]. The high carrier frequency (1.5%) for this mutation among cohorts from sub-Saharan West Africa predicts that recessive type VIII OI will occur with the same incidence as dominant OI (1 in 18,260 births) in these regions. These numbers represent the potential for a substantial public health burden in West African countries, since dominant and recessive forms of OI are expected to occur with equal frequency in this population. Furthermore, a recent report suggests that severe to lethal forms of OI are as common as mild to moderate forms among Nigerians, and is consistent with the mild to lethal range of phenotypes in dominant OI and the severe to lethal range seen in P3H1 deficiency [496,497].

The West African allele is the only recurring OI-causing mutation that has been characterized in a geographically widespread population, in this case due to the African Diaspora. The occurrence of this mutation within a conserved haplotype among carriers not only confirms a founder effect, but also is consistent with a mutation that occurred between 650 and 900 years ago, prior to the Atlantic slave trade [413]. Thus it was not surprising to find additional carriers in regions affected by the slave trade. For example, we have also identified this mutation in 5/863 (0.58%) Jamaican newborn genomic DNAs that were screened, estimating an occurrence of Type VIII OI due to homozygosity for *LEPRE1* c.1080+1G>T at 1/109,958-119,163 births (unpublished data).

It should be noted that Irish Travellers and individuals of West African descent are not the only ethnic populations affected by molecular defects in *LEPRE1* [498]. While the West African allele is a substantial factor, the number of recessive *LEPRE1* alleles, as well as the number of reported cases of Type VIII OI, outnumber those that have been reported for types VII and IX OI, due to mutations in *CRTAP* and PPIB, combined [499]. Deficiency of cyclophilin B, caused by mutations in *PPIB*, is the least frequently occurring cause of recessive OI. However, the small number of human cases and animal models have provided insights into its direct and indirect roles in the collagen biosynthetic process.

The *Ppib*-Null Mouse: A Unique Model of Recessive OI

Although recessive OI has now been sufficiently investigated to draw some general conclusions about the mechanism of recessive OI in cases of CRTAP and P3H1 deficiency, the picture has been less clear for PPIB/Cyclophilin B deficiency. The loss of α1(I) P986 modification and presence of collagen alpha chain overmodification is a defining biochemical feature of CRTAP and P3H1 deficiency, and cases of types VII and VIII OI are uniformly severe to lethal in phenotype.

At the present time only 8 cases of type IX OI due to *PPIB* mutations have been reported (Table 8), only 4 of which have been investigated beyond mutation identification [421-423]. These cases differ from those with CRTAP or P3H1 deficiency in that rhizomelia is absent, although they share the features of broad, undertubulated long bones and white sclerae. Two of these cases with moderate OI phenotypes have mutations (c.2T>G, p.Met1Arg; c.343+1G>A, p.Gly115delins10) that result in complete

Table 8. Reported Cases of PPIB Deficiency in Type IX Osteogenesis Imperfecta

PPIB Mutation	OI Severity	mRNA	protein	collagen	α1(I) 3Hyp986	3OH Complex
c.26T>G/c.26T>G p.M1R/p.M1R	moderate	55%	none	normal	98%	50% reduced
c.343+1G>A/c.343+1G>A p.G115GfsX16/p.G115Gdelins10	severe/moderate	minimal	none	overmodified	100%	decreased
c.120delC/c.313G>A p.V42SfsX16/p.G105R	lethal	minimal	residual	overmodified	33%	decreased
c.556_559del/c.556_559del p.K186QfsX8/p.K186QfsX8	severe/lethal	normal	ND	overmodified	33%	decreased
c.414_423del/ c.414_423del p.S139TfsX21/p.S139TfsX21	lethal/severe	10%	none	overmodified	ND	decreased
c.451C>T/c.451C>T p.Q151X/p.Q151X	severe	ND	ND	ND	ND	ND
c.497A>C/c.497A>C p.H166P/p.H166P	severe	ND	ND	ND	ND	ND
c.563_566del/c.563_559del p.D188AfsX5/p.D188AfsX5	lethal	ND	ND	ND	ND	ND

absence of cyclophilin B protein on immunoblots, but do not cause a reduction of collagen α1(I) P986 3-hydroxylation [421,423]. The remaining two cases with lethal OI demonstrate decreased (30%), but not absent, 3-hydroxylation of collagen resulting from their mutations (c.120delC/c.313G>A, p.Val42Serfs*16/p.Gly105Arg; c.556_559del, p.Lys186Glnfs*8) [421,422]. Importantly, in the cases in which partial hydroxylation of P986 residues was observed, the PPIB mutations result in either normal levels of transcripts or residual levels of cyclophilin B protein. Accordingly, these data suggest the presence of a truncated cyclophilin B protein that functions in a dominant negative manner within the P3H complex, resulting in a dysfunctional complex that partially modifies the substrate and causes collagen overmodification. Thus, regardless of which components of the 3-hydroxylation complex are absent, all cases of types VII-IX OI that have complete absence of P986 hydroxylation result in collagen overmodification. In contrast, the proband with the start codon mutation completely lacked cyclophilin B protein, had normal P986 3-hydroxylation and normal collagen modification [421]. As discussed more fully below, this may be due to redundancy for cyclophilin B function by an unknown ER-resident protein.

Two animal models for PPIB deficiency have been previously reported. The first of these is HERDA, a progressive skin fragility syndrome observed in American Quarter horses caused by a naturally occurring missense mutation (c.115G>A, p.G6R) [427]. This mutation results in a substitution at the amino terminus of processed PPIB that does not alter the substrate-binding site. Although the *in vitro* isomerase activity of mutant cyclophilin B was intact, collagen folding was delayed in fibroblasts. This horse model of PPIB deficiency exhibits normal P986 3-hydroxylation and decreased collagen post-

translational modification resulting in a skin phenotype without apparent bone dysplasia. The surprising finding in this study was the decreased hydroxylysine content of equine dermal tissue, presumably due to a loss of physical interaction between HERDA cyclophilin B and LH1 that was demonstrated in immunoprecipitation experiments. In contrast, a previously reported knock-out mouse model of PPIB deficiency demonstrates absent 3-hydroxylation and delayed electrophoretic migration consistent with increased post-translational modification, similar to the collagen biochemistry of CRTAP and P3H1 deficiency [112]. As with the human cases, the information from the animal models demonstrates that the presence of an abnormal cyclophilin B protein results in at least partial modification of the P986 residue. Unlike the human cases, however, complete absence of cyclophilin B abrogates prolyl 3-hydroxylation of the collagen substrate. Both animal models resulted in collagen post-translational modification that differed from each other, as well as the human cases. Therefore we developed a *Ppib*-null mouse with the goal of addressing the inconsistencies between the human cases, as well as the differences between the previously reported animal models and human cases.

Ppib expression is completely blocked in the gene trap mutant allele used in our mouse model, compared with a previously reported *Ppib*-knockout mouse that retains 25% of wild-type transcripts. Total absence of *Ppib* expression in the mouse results in growth deficiency and reduced femoral and tibial lengths, as seen in other mouse models with deficiency of P3H complex components [107,112,419]. In contrast to the *Crtap*$^{-/-}$ and *P3h1*$^{-/-}$ mice, however, *Ppib*$^{-/-}$ mice do not have rhizomelia. Surviving *Ppib*$^{-/-}$ mice have a severely osteoporotic bone phenotype, accentuated by reduced structural parameters, decreased bone mineral density, and significantly reduced bone strength. Furthermore,

the 40-50% perinatal lethality seen in *Ppib*$^{-/-}$ mice probably reflects the full range of phenotypes seen in humans with CyPB deficiency. Thus our mouse model of CyPB deficiency faithfully reproduces the phenotypic features found in human cases of type IX OI [421-423].

The critical role of CyPB for 3-hydroxylation of both type I collagen sites (i.e., α1(I) P986 and α2(I) P707) in mice is evident from the nearly complete loss of modification of α1(I) chains and the 50% reduction in 3-hydroxylation of α2(I) chains. These data suggest that the remaining P3H complex (CRTAP and P3H1) components do not function in the absence of CyPB in mice for the α1(I) P986 site, but that a partially redundant system is involved in modification of the α2(I) P707 site. Thus it seems likely that functional redundancy by a P3H isoform that does not require CyPB can compensate for α2(I) P707 modification, or another unidentified protein compensates for CyPB's role.

We have verified CyPB's role as a peptidyl-prolyl *cis-trans* isomerase involved in type I collagen folding. In the absence of CyPB, the formation of trypsin-resistant collagen helices was significantly delayed, seemingly recapitulating the delayed folding of collagen in the presence of cyclosporine A that was demonstrated in earlier studies [99,108]. However, we have demonstrated that addition of cyclosporine A to *Ppib*$^{-/-}$ cells further increases the delay in collagen folding, suggesting the presence of an additional cyclosporine-sensitive factor involved in catalyzing collagen folding. Although CyPB has long been considered to be the unique isomerase to catalyze the rate-limiting step in helical folding, our data demonstrate an addition factor of unknown identity that is able to function in this manner.

The most striking findings in our study are the differences in modification of type I collagen synthesized in cultured cells versus tissues. Specifically, modification of collagen synthesized in culture does not necessarily reflect the extent of modification seen in tissues. For instance, PAGE and amino acid analysis of type I collagen alpha chains synthesized by both fibroblasts and osteoblasts demonstrated normal lysyl hydroxylation with increased glycosylation, resulting in delayed electrophoretic migration. However, type I collagen extracted from dermis was undermodified and showed increased electrophoretic mobility, while collagen extracted from bone tissue was overmodified. These findings highlight the tissue-specific roles of collagen-specific modifying enzymes in the ER. First, the loss of lysyl hydroxylation in the absence of CyPB in dermis supports the recent hypothesis that LH1 requires CyPB for modification of the collagen substrate [427]. Second, the presence of overmodified type I collagen in bone tissue demonstrates that osteoblasts either do not require the CyPB-LH1 interaction for modification of lysyl residues, or an alternative isomerase interacts with LH1 for its activity, or an alternative lysyl hydroxylase isoform can compensate for loss of LH1 function in bone tissue. Currently, LH3 is a prime candidate for this compensatory function, since it is expressed in bone and has been demonstrated to exhibit hydroxylase, galactosyl-, and glucosyltransferase activities *in vitro* and in culture [83,84,471]. It remains to be determined whether loss of LH activity towards type I collagen in the absence of CyPB is due to loss of LH1 activity catalyzed by CyPB, or if LH1 is unable to recognized the collagen substrate with peptidyl-prolyl bonds in the *cis* conformation.

Surprisingly, the loss of post-translational modification was most notable at amino terminal helical lysine residues involved in intermolecular crosslinks in extracellular

matrix. Thus the loss of hydroxylation at K87 residues, which are involved in crosslink formation, results in an alteration of crosslink profiles. We found an increase in the LP/HP ratio in bone tissue, similar to the crosslink profile seen in bone of patients with Type VI Ehlers-Danlos Syndrome due to LH1 deficiency. However, although the bone phenotype in EDS VI patients can be described as osteopenic, the major phenotypic findings in these individuals are hyperextensible skin and joints [76]. While these features are noted in the CyPB-deficient mouse, the phenotypic features of this mouse model are primarily that of a bone dysplasia.

Controversies pertaining to the role of cyclophilin B in collagen synthesis remain. First, Pyott and colleagues suggest that inhibition or absence of cyclophilin B results in loss of peptidyl-prolyl isomerase activity important for the collagen C-propeptide conformation necessary for chain association [423]. The authors hypothesized that the cyclosporin A-induced delay in the appearance of trypsin-resistant collagen helices seen in direct intracellular folding assays was misinterpreted as delayed folding, but was instead a result of delayed heterotrimer formation. However, we have been unable to detect any delays in procollagen chain incorporation in our mouse model, so this putative function of cyclophilin B remains to be independently verified. Furthermore, given the short delay in chain association which was observed (2-3 minutes), the physiological relevance of these findings is in question.

A second controversy has risen over whether cyclophilin B is the unique *cis-trans* isomerase that catalyzes the rate-limiting step in collagen folding. We have demonstrated that cyclosporine A further increases the delay in the appearance of trypsin-resistant helices in direct intracellular folding assays performed on *Ppib*-null cells, consistent with

the presence of an additional collagen-interacting cyclosporine-sensitive PPIase in the ER. Importantly, our study has revealed tissue-specific differences in collagen biochemistry in the absence of cyclophilin B which are relevant for the bone phenotype exhibited in recessive OI. The absence of cyclophilin B in murine dermis and subsequent decreased hydroxylation and glycosylation of collagen lysine residues is consistent with a loss of interaction required for LH1 activity, thus implying that LH1-collagen interaction requires CyPB. However, in murine bone tissue the absence of cyclophilin B is accompanied by an increase in collagen post-translational modification. Taken together, these studies are consistent with a major, but not unique, role for cyclophilin B as a collagen *cis-trans* isomerase, and agree with our previous proposal of functional redundancy based on findings of normal collagen post-translational modification in the proband with a *PPIB* start codon mutation [421]. Seven PPIases are located in the ER, including six FK506-binding proteins (FKBPs) and cyclophilin B [91,93]; only cyclophilin B and FKBP65 have been demonstrated to physically interact with collagen in the ER [101]. However, the genetic and biochemical evidence that exists suggests that FKBP65 does not have a role in helical modification or folding. FKBP65-deficient cells do not demonstrate a delay in collagen folding in direct folding assays (data not shown), and FKBP65 activity is only minimally inhibited by CsA and FK506 [99]. Formal verification, however, requires *Ppib-Fkbp10* compound knock-out cells, or knock-down studies in *Fkbp10*- and *Ppib*-null cells.

Future studies on our *Ppib*-null mouse model will address the question of collagen isomerase redundancy by creation of double knockouts or knockdown of candidate cyclophilins and FKBPs in order to replicate the increased delay of folding seen upon

cyclosporin A treatment of cyclophilin B-deficient cells. Also, the effect of altered collagen crosslinking in *Ppib*$^{-/-}$ bone on mineralization remains to be explored. Bone from patients or mouse models of OI caused by collagen structural defects or CRTAP deficiency contains a crosslink pattern that is opposite to this *Ppib*$^{-/-}$ mouse due to increased hydroxylysine derived crosslinks, and is paradoxically hypermineralized [406,483,484,486,500]. On the other hand, increased HP/LP ratios correlate with the ultimate compressive strength of trabecular bone, but are independent of BMD [490,491]. Analysis of the process of mineralization, as well as the organization of the mineral component in *Ppib*$^{-/-}$ and classical OI bone will provide insight into the role of bone crosslinks in mineralization, and could point to collagen modification as an intracellular pathway by which osteoblasts can actively influence this process. In addition, since CyPB is known to directly interact with several ER-resident proteins [501], its role in ER stress and differentiation of osteoblasts will be analyzed.

References

1. Nicholls, J.J., Brassill, M.J., Williams, G.R. & Bassett, J.H. The skeletal consequences of thyrotoxicosis. *J Endocrinol* **213**, 209-21 (2012).

2. Marini, J.C., Cabral, W.A., Barnes, A.M. & Chang, W. Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development. *Cell Cycle* **6**, 1675-81 (2007).

3. Gilbert, S.F. *Developmental Biology*, (Sinauer Associates, Inc., Sunderland, MA, 2006).

4. Wiesmann, H.P., Meyer, U., Plate, U. & Hohling, H.J. Aspects of collagen mineralization in hard tissue formation. *Int Rev Cytol* **242**, 121-56 (2005).

5. Beniash, E. Biominerals--hierarchical nanocomposites: the example of bone. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **3**, 47-69 (2011).

6. Kadler, K.E., Holmes, D.F., Trotter, J.A. & Chapman, J.A. Collagen fibril formation. *Biochem J* **316 (Pt 1)**, 1-11 (1996).

7. Einhorn, T.A. Bone strength: the bottom line. *Calcif Tissue Int* **51**, 333-9 (1992).

8. Ishikawa, Y. & Bachinger, H.P. A molecular ensemble in the rER for procollagen maturation. *Biochim Biophys Acta* **1833**, 2479-91 (2013).

9. Myllyharju, J. & Kivirikko, K.I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* **20**, 33-43 (2004).

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

11. Eyre, D.R., Weis, M.A. & Wu, J.J. Advances in collagen cross-link analysis. *Methods* **45**, 65-74 (2008).

12. Chu, M. & Prockop, D.J. Gene Structure. in *Connective Tissue and its Heritable Disorders* (eds. Royce, P. & Steinmann, B.) 223-248 (Wiley-Liss, Inc., New York, 2002).

13. Persikov, A.V., Ramshaw, J.A., Kirkpatrick, A. & Brodsky, B. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry* **44**, 1414-22 (2005).

14. Xu, K., Nowak, I., Kirchner, M. & Xu, Y. Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to

the disruption of a set of interchain salt bridges. *J Biol Chem* **283**, 34337-44 (2008).

15. Vuust, J., Abildsten, D. & Lund, T. Control of type I collagen synthesis: evidence for pretranslational coordination of pro alpha 1 (I) and pro alpha 2 (I) chain synthesis in embryonic chick bone. *Connect Tissue Res* **11**, 185-91 (1983).

16. Brownell, A.G. & Veis, A. Intracellular location of triple helix formation of collagen. Enzyme probe studies. *J Biol Chem* **251**, 7137-43 (1976).

17. Veis, A. & Kirk, T.Z. The coordinate synthesis and cotranslational assembly of type I procollagen. *J Biol Chem* **264**, 3884-9 (1989).

18. Yamada, Y., Mudryj, M. & de Crombrugghe, B. A uniquely conserved regulatory signal is found around the translation initiation site in three different collagen genes. *J Biol Chem* **258**, 14914-9 (1983).

19. Cai, L., Fritz, D., Stefanovic, L. & Stefanovic, B. Binding of LARP6 to the conserved 5' stem-loop regulates translation of mRNAs encoding type I collagen. *J Mol Biol* **395**, 309-26 (2010).

20. McLaughlin, S.H. & Bulleid, N.J. Molecular recognition in procollagen chain assembly. *Matrix Biol* **16**, 369-77 (1998).

21. Koivu, J. Disulfide bonding as a determinant of the molecular composition of types I, II and III procollagen. *FEBS Lett* **217**, 216-20 (1987).

22. Koivu, J. Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen. *FEBS Lett* **212**, 229-32 (1987).

23. Doyle, S.A. & Smith, B.D. Role of the pro-alpha2(I) COOH-terminal region in assembly of type I collagen: disruption of two intramolecular disulfide bonds in pro-alpha2(I) blocks assembly of type I collagen. *J Cell Biochem* **71**, 233-42 (1998).

24. Uitto, J. & Prockop, D.J. Rate of helix formation by intracellular procollagen and protocollagen. Evidence for a role for disulfide bonds. *Biochem Biophys Res Commun* **55**, 904-11 (1973).

25. Berg, R.A. & Prockop, D.J. Purification of (14C) protocollagen and its hydroxylation by prolyl-hydroxylase. *Biochemistry* **12**, 3395-401 (1973).

26. Kozlov, G., Maattanen, P., Thomas, D.Y. & Gehring, K. A structural overview of the PDI family of proteins. *Febs J* **277**, 3924-36 (2010).

27. Appenzeller-Herzog, C. & Ellgaard, L. The human PDI family: versatility packed into a single fold. *Biochim Biophys Acta* **1783**, 535-48 (2008).

28. Freedman, R.B., Hirst, T.R. & Tuite, M.F. Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem Sci* **19**, 331-6 (1994).

29. Freedman, R.B. & Klappa, P. Protein disulphide isomerase: a catalyst of thiol: disulphide interchange and associated protein folding. in *Molecular Chaperones and Protein Folding* (ed. Bukau, B.) 437-459 (Harwood Academic Press, London, 1999).

30. Schwaller, M., Wilkinson, B. & Gilbert, H.F. Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase. *J Biol Chem* **278**, 7154-9 (2003).

31. Koivu, J. & Myllyla, R. Interchain disulfide bond formation in types I and II procollagen. Evidence for a protein disulfide isomerase catalyzing bond formation. *J Biol Chem* **262**, 6159-64 (1987).

32. Ko, M.K. & Kay, E.P. PDI-mediated ER retention and proteasomal degradation of procollagen I in corneal endothelial cells. *Exp Cell Res* **295**, 25-35 (2004).

33. Wilson, R., Lees, J.F. & Bulleid, N.J. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *J Biol Chem* **273**, 9637-43 (1998).

34. John, D.C., Grant, M.E. & Bulleid, N.J. Cell-free synthesis and assembly of prolyl 4-hydroxylase: the role of the beta-subunit (PDI) in preventing misfolding and aggregation of the alpha-subunit. *EMBO J* **12**, 1587-95 (1993).

35. Ramshaw, J.A., Shah, N.K. & Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol* **122**, 86-91 (1998).

36. Berg, R.A. & Prockop, D.J. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun* **52**, 115-20 (1973).

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

38. Burjanadze, T.V. & Veis, A. A thermodynamic analysis of the contribution of hydroxyproline to the structural stability of the collagen triple helix. *Connect Tissue Res* **36**, 347-65 (1997).

39. Gorres, K.L. & Raines, R.T. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* **45**, 106-24 (2010).

40. Shoulders, M.D., Hodges, J.A. & Raines, R.T. Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J Am Chem Soc* **128**, 8112-3 (2006).

41. Shoulders, M.D., Satyshur, K.A., Forest, K.T. & Raines, R.T. Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc Natl Acad Sci U S A* **107**, 559-64 (2010).

42. Holmgren, S.K., Taylor, K.M., Bretscher, L.E. & Raines, R.T. Code for collagen's stability deciphered. *Nature* **392**, 666-7 (1998).

43. Bretscher, L.E., Jenkins, C.L., Taylor, K.M., DeRider, M.L. & Raines, R.T. Conformational stability of collagen relies on a stereoelectronic effect. *J Am Chem Soc* **123**, 777-8 (2001).

44. Kivirikko, K.I. & Myllyla, R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol* **82 Pt A**, 245-304 (1982).

45. Tandon, M. *et al.* Substrate specificity of human prolyl-4-hydroxylase. *Bioorg Med Chem Lett* **8**, 1139-44 (1998).

46. Berg, R.A. & Prockop, D.J. Affinity column purification of protocollagen proline hydroxylase from chick embryos and further characterization of the enzyme. *J Biol Chem* **248**, 1175-82 (1973).

47. Pihlajaniemi, T. *et al.* Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO J* **6**, 643-9 (1987).

48. Kukkola, L., Hieta, R., Kivirikko, K.I. & Myllyharju, J. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *J Biol Chem* **278**, 47685-93 (2003).

49. Annunen, P. *et al.* Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J Biol Chem* **272**, 17342-8 (1997).

50. Helaakoski, T., Vuori, K., Myllyla, R., Kivirikko, K.I. & Pihlajaniemi, T. Molecular cloning of the alpha-subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc Natl Acad Sci U S A* **86**, 4392-6 (1989).

51. Nissi, R., Autio-Harmainen, H., Marttila, P., Sormunen, R. & Kivirikko, K.I. Prolyl 4-hydroxylase isoenzymes I and II have different expression patterns in several human tissues. *J Histochem Cytochem* **49**, 1143-53 (2001).

52. Lehmann, H.W. *et al.* Hydroxylation of collagen type I: evidence that both lysyl and prolyl residues are overhydroxylated in osteogenesis imperfecta. *Eur J Clin Invest* **25**, 306-10 (1995).

53. Bornstein, P. The incomplete hydroxylation of individual prolyl residues in collagen. *J Biol Chem* **242**, 2572-4 (1967).

54. Bornstein, P. Comparative sequence studies of rat skin and tendon collagen. I. Evidence for incomplete hydroxylation of individual prolyl residues in the normal proteins. *Biochemistry* **6**, 3082-93 (1967).

55. Holster, T. *et al.* Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. *J Biol Chem* **282**, 2512-9 (2007).

56. Peterkofsky, B. & Udenfriend, S. Enzymatic Hydroxylation of Proline in Microsomal Polypeptide Leading to Formation of Collagen. *Proc Natl Acad Sci U S A* **53**, 335-42 (1965).

57. Kivirikko, K.I. & Prockop, D.J. Enzymatic hydroxylation of proline and lysine in protocollagen. *Proc Natl Acad Sci U S A* **57**, 782-9 (1967).

58. Halme, J., Kivirikko, K.I. & Simons, K. Isolation and partial characterization of highly purified protocollagen proline hydroxylase. *Biochim Biophys Acta* **198**, 460-70 (1970).

59. Hoyhtya, M., Myllyla, R., Piuva, J., Kivirikko, K.I. & Tryggvason, K. Monoclonal antibodies to human prolyl 4-hydroxylase. *Eur J Biochem* **141**, 472-82 (1984).

60. Kaska, D.D., Myllyla, R., Gunzler, V., Gibor, A. & Kivirikko, K.I. Prolyl 4-hydroxylase from Volvox carteri. A low-Mr enzyme antigenically related to the alpha subunit of the vertebrate enzyme. *Biochem J* **256**, 257-63 (1988).

61. Myllyharju, J. & Kivirikko, K.I. Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase. *EMBO J* **18**, 306-12 (1999).

62. Kivirikko, K.I., Myllyla, R. & Pihlajaniemi, T. Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *Faseb J* **3**, 1609-17 (1989).

63. Kivirikko, K.I. & Myllyla, R. Post-translational processing of procollagens. *Ann N Y Acad Sci* **460**, 187-201 (1985).

64. Peterkofsky, B. Ascorbate requirement for hydroxylation and secretion of procollagen: relationship to inhibition of collagen synthesis in scurvy. *Am J Clin Nutr* **54**, 1135S-1140S (1991).

65. Cardinale, G.J. & Udenfriend, S. Prolyl hydroxylase. *Adv Enzymol Relat Areas Mol Biol* **41**, 245-300 (1974).

66. Kivirikko, K.I., Myllyla, R. & Pihlajaniemi, T. Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. in *Post-translational modifications of proteins* (eds. Harding, J.J. & Crabbe, M.J.C.) 1-52 (CRC Press, Boca Raton, 1992).

67. Kivirikko, K.I., Suga, K., Kishida, Y., Sakakibara, S. & Prockop, D.J. Asymmetry in the hydroxylation of (Pro-Pro-Gly) 5 by protocollagen proline hydroxylase. *Biochem Biophys Res Commun* **45**, 1591-6 (1971).

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

69. Ruotsalainen, H., Sipila, L., Kerkela, E., Pospiech, H. & Myllyla, R. Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3, their phylogenetic analysis and tissue-specific expression in the mouse. *Matrix Biol* **18**, 325-9 (1999).

70. Kellokumpu, S., Sormunen, R., Heikkinen, J. & Myllyla, R. Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of luminally-oriented peripheral membrane proteins in the endoplasmic reticulum. *J Biol Chem* **269**, 30524-9 (1994).

71. Takaluoma, K., Lantto, J. & Myllyharju, J. Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences. *Matrix Biol* **26**, 396-403 (2007).

72. Pinnell, S.R., Krane, S.M., Kenzora, J.E. & Glimcher, M.J. A heritable disorder of connective tissue. Hydroxylysine-deficient collagen disease. *N Engl J Med* **286**, 1013-20 (1972).

73. Krane, S.M., Pinnell, S.R. & Erbe, R.W. Lysyl-protocollagen hydroxylase deficiency in fibroblasts from siblings with hydroxylysine-deficient collagen. *Proc Natl Acad Sci U S A* **69**, 2899-903 (1972).

74. Hyland, J. *et al.* A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers-Danlos syndrome type VI. *Nat Genet* **2**, 228-31 (1992).

75. Steinmann, B., Eyre, D.R. & Shao, P. Urinary pyridinoline cross-links in Ehlers-Danlos syndrome type VI. *Am J Hum Genet* **57**, 1505-8 (1995).

76. Yeowell, H.N. & Walker, L.C. Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers-Danlos syndrome type VI. *Mol Genet Metab* **71**, 212-24 (2000).

77. Royce, P.M. & Barnes, M.J. Failure of highly purified lysyl hydroxylase to hydroxylate lysyl residues in the non-helical regions of collagen. *Biochem J* **230**, 475-80 (1985).

78. Bank, R.A. *et al.* Defective collagen crosslinking in bone, but not in ligament or cartilage, in Bruck syndrome: indications for a bone-specific telopeptide lysyl hydroxylase on chromosome 17. *Proc Natl Acad Sci U S A* **96**, 1054-8 (1999).

79. Valtavaara, M. *et al.* Cloning and characterization of a novel human lysyl hydroxylase isoform highly expressed in pancreas and muscle. *J Biol Chem* **272**, 6831-4 (1997).

80. Yeowell, H.N. & Walker, L.C. Tissue specificity of a new splice form of the human lysyl hydroxylase 2 gene. *Matrix Biol* **18**, 179-87 (1999).

81. Uzawa, K. *et al.* Differential expression of human lysyl hydroxylase genes, lysine hydroxylation, and cross-linking of type I collagen during osteoblastic differentiation in vitro. *J Bone Miner Res* **14**, 1272-80 (1999).

82. Pornprasertsuk, S., Duarte, W.R., Mochida, Y. & Yamauchi, M. Lysyl hydroxylase-2b directs collagen cross-linking pathways in MC3T3-E1 cells. *J Bone Miner Res* **19**, 1349-55 (2004).

83. Heikkinen, J. *et al.* Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J Biol Chem* **275**, 36158-63 (2000).

84. Wang, C. *et al.* The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. *Matrix Biol* **21**, 559-66 (2002).

85. Ruotsalainen, H. *et al.* Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J Cell Sci* **119**, 625-35 (2006).

86. Myllyla, R. *et al.* Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J Cell Physiol* **212**, 323-9 (2007).

87. Mercer, D.K., Nicol, P.F., Kimbembe, C. & Robins, S.P. Identification, expression, and tissue distribution of the three rat lysyl hydroxylase isoforms. *Biochem Biophys Res Commun* **307**, 803-9 (2003).

88. Steinmann, B. *et al.* Ehlers-Danlos syndrome in two siblings with deficient lysyl hydroxylase activity in cultured skin fibroblasts but only mild hydroxylysine deficit in skin. *Helv Paediatr Acta* **30**, 255-74 (1975).

89. Takaluoma, K. *et al.* Tissue-specific changes in the hydroxylysine content and cross-links of collagens and alterations in fibril morphology in lysyl hydroxylase 1 knock-out mice. *J Biol Chem* **282**, 6588-96 (2007).

90. Salo, A.M. *et al.* The lysyl hydroxylase isoforms are widely expressed during mouse embryogenesis, but obtain tissue- and cell-specific patterns in the adult. *Matrix Biol* **25**, 475-83 (2006).

91. Galat, A. Peptidylproline cis-trans-isomerases: immunophilins. *Eur J Biochem* **216**, 689-707 (1993).

92. Hamilton, G.S. & Steiner, J.P. Immunophilins: beyond immunosuppression. *J Med Chem* **41**, 5119-43 (1998).

93. Galat, A. Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity--targets--functions. *Curr Top Med Chem* **3**, 1315-47 (2003).

94. Simek, S.L., Kligman, D., Patel, J. & Colburn, N.H. Differential expression of an 80-kDa protein kinase C substrate in preneoplastic and neoplastic mouse JB6 cells. *Proc Natl Acad Sci U S A* **86**, 7410-4 (1989).

95. Coss, M.C., Winterstein, D., Sowder, R.C., 2nd & Simek, S.L. Molecular cloning, DNA sequence analysis, and biochemical characterization of a novel 65-kDa FK506-binding protein (FKBP65). *J Biol Chem* **270**, 29336-41 (1995).

96. Davis, E.C., Broekelmann, T.J., Ozawa, Y. & Mecham, R.P. Identification of tropoelastin as a ligand for the 65-kD FK506-binding protein, FKBP65, in the secretory pathway. *J Cell Biol* **140**, 295-303 (1998).

97. Bruckner, P., Eikenberry, E.F. & Prockop, D.J. Formation of the triple helix of type I procollagen in cellulo. A kinetic model based on cis-trans isomerization of peptide bonds. *Eur J Biochem* **118**, 607-13 (1981).

98. Zeng, B. *et al.* Chicken FK506-binding protein, FKBP65, a member of the FKBP family of peptidylprolyl cis-trans isomerases, is only partially inhibited by FK506. *Biochem J* **330 (Pt 1)**, 109-14 (1998).

99. Bachinger, H.P., Morris, N.P. & Davis, J.M. Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen. *Am J Med Genet* **45**, 152-62 (1993).

100. Murphy, L.A. *et al.* Endoplasmic reticulum stress or mutation of an EF-hand Ca(2+)-binding domain directs the FKBP65 rotamase to an ERAD-based proteolysis. *Cell Stress Chaperones* **16**, 607-19 (2011).

101. Ishikawa, Y., Vranka, J., Wirz, J., Nagata, K. & Bachinger, H.P. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 is a molecular chaperone that interacts with collagens. *J Biol Chem* **283**, 31584-90 (2008).

102. Alanay, Y. *et al.* Mutations in the gene encoding the RER protein FKBP65 cause autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* **86**, 551-9 (2010).

103. Barnes, A.M. *et al.* 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**, 1589-98 (2012).

104. Risteli, J., Tryggvason, K. & Kivirikko, K.I. Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. *Eur J Biochem* **73**, 485-92 (1977).

105. Tryggvason, K., Majamaa, K., Risteli, J. & Kivirikko, K.I. Partial purification and characterization of chick-embryo prolyl 3-hydroxylase. *Biochem J* **183**, 303-7 (1979).

106. Vranka, J.A., Sakai, L.Y. & Bachinger, H.P. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem* **279**, 23615-21 (2004).

107. Morello, R. *et al.* CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* **127**, 291-304 (2006).

108. Steinmann, B., Bruckner, P. & Superti-Furga, A. Cyclosporin A slows collagen triple-helix formation in vivo: indirect evidence for a physiologic role of peptidyl-prolyl cis-trans-isomerase. *J Biol Chem* **266**, 1299-303 (1991).

109. Barnes, A.M. *et al.* Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. *N Engl J Med* **355**, 2757-64 (2006).

110. Cabral, W.A. *et al.* Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat Genet* **39**, 359-65 (2007).

111. Ishikawa, Y., Wirz, J., Vranka, J.A., Nagata, K. & Bachinger, H.P. Biochemical characterization of the prolyl 3-hydroxylase 1.cartilage-associated protein.cyclophilin B complex. *J Biol Chem* **284**, 17641-7 (2009).

112. Choi, J.W. *et al.* Severe osteogenesis imperfecta in cyclophilin B-deficient mice. *PLoS Genet* **5**, e1000750 (2009).

113. Hudson, D.M., Weis, M. & Eyre, D.R. Insights on the evolution of prolyl 3-hydroxylation sites from comparative analysis of chicken and Xenopus fibrillar collagens. *PLoS One* **6**, e19336 (2011).

114. Schumacher, M.A., Mizuno, K. & Bachinger, H.P. The crystal structure of a collagen-like polypeptide with 3(S)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. *J Biol Chem* **281**, 27566-74 (2006).

115. Hudson, D.M., Kim, L.S., Weis, M., Cohn, D.H. & Eyre, D.R. Peptidyl 3-hydroxyproline binding properties of type I collagen suggest a function in fibril supramolecular assembly. *Biochemistry* **51**, 2417-24 (2012).

116. Marini, J.C. *et al.* Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat* **28**, 209-21 (2007).

117. Grassmann, W. & Schleich, H. Uber den Kohlenhydratgehalt des Kollagens II. *Biochem Z* **277**, 329-328 (1935).

118. Spiro, R.G. The structure of the disaccharide unit of the renal glomerular basement membrane. *J Biol Chem* **242**, 4813-23 (1967).

119. Kivirikko, K.I. & Myllyla, R. Collagen glycosyltransferases. *Int Rev Connect Tissue Res* **8**, 23-72 (1979).

120. Liefhebber, J.M., Punt, S., Spaan, W.J. & van Leeuwen, H.C. The human collagen beta(1-O)galactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein. *BMC Cell Biol* **11**, 33 (2010).

121. Schegg, B., Hulsmeier, A.J., Rutschmann, C., Maag, C. & Hennet, T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. *Mol Cell Biol* **29**, 943-52 (2009).

122. Myllyla, R., Risteli, L. & Kivirikko, K.I. Assay of collagen-galactosyltransferase and collagen-glucosyltransferase activities and preliminary characterization of enzymic reactions with transferases from chick-embryo cartilage. *Eur J Biochem* **52**, 401-10 (1975).

123. Risteli, L., Myllya, R. & Kivirikko, K.I. Partial purification and characterization of collagen galactosyltransferase from chick embryos. *Biochem J* **155**, 145-53 (1976).

124. Sricholpech, M. *et al.* Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. *J Biol Chem* **287**, 22998-3009 (2012).

125. Rautavuoma, K. *et al.* Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. *Proc Natl Acad Sci U S A* **101**, 14120-5 (2004).

126. Salo, A.M. *et al.* A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. *Am J Hum Genet* **83**, 495-503 (2008).

127. Bann, J.G., Peyton, D.H. & Bachinger, H.P. Sweet is stable: glycosylation stabilizes collagen. *FEBS Lett* **473**, 237-40 (2000).

128. Mann, K. *et al.* Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen. *J Mol Biol* **261**, 255-66 (1996).

129. Bann, J.G., Bachinger, H.P. & Peyton, D.H. Role of carbohydrate in stabilizing the triple-helix in a model for a deep-sea hydrothermal vent worm collagen. *Biochemistry* **42**, 4042-8 (2003).

130. Nagata, K., Saga, S. & Yamada, K.M. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J Cell Biol* **103**, 223-9 (1986).

131. Nagai, N. *et al.* Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J Cell Biol* **150**, 1499-506 (2000).

132. Clarke, E.P. & Sanwal, B.D. Cloning of a human collagen-binding protein, and its homology with rat gp46, chick hsp47 and mouse J6 proteins. *Biochim Biophys Acta* **1129**, 246-8 (1992).

133. Koide, T., Aso, A., Yorihuzi, T. & Nagata, K. Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47. *J Biol Chem* **275**, 27957-63 (2000).

134. Koide, T. *et al.* 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**, 11177-85 (2006).

135. Koide, T. *et al.* Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation. *J Biol Chem* **281**, 3432-8 (2006).

136. Nishikawa, Y. *et al.* A structure-activity relationship study elucidating the mechanism of sequence-specific collagen recognition by the chaperone HSP47. *Bioorg Med Chem* **18**, 3767-75 (2010).

137. Widmer, C. *et al.* 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**, 13243-7 (2012).

138. Nakai, A., Satoh, M., Hirayoshi, K. & Nagata, K. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J Cell Biol* **117**, 903-14 (1992).

139. Hirayoshi, K. *et al.* HSP47: a tissue-specific, transformation-sensitive, collagen-binding heat shock protein of chicken embryo fibroblasts. *Mol Cell Biol* **11**, 4036-44 (1991).

140. Leikina, E., Mertts, M.V., Kuznetsova, N. & Leikin, S. Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* **99**, 1314-8 (2002).

141. Christiansen, H.E. *et al.* Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta. *Am J Hum Genet* **86**, 389-98 (2010).

142. Ishida, Y. *et al.* Type I collagen in Hsp47-null cells is aggregated in endoplasmic reticulum and deficient in N-propeptide processing and fibrillogenesis. *Mol Biol Cell* **17**, 2346-55 (2006).

143. Thomson, C.A. & Ananthanarayanan, V.S. Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro. *Biochem J* **349 Pt 3**, 877-83 (2000).

144. Bonfanti, L. *et al.* Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* **95**, 993-1003 (1998).

145. Budnik, A. & Stephens, D.J. ER exit sites--localization and control of COPII vesicle formation. *FEBS Lett* **583**, 3796-803 (2009).

146. Venditti, R. *et al.* Sedlin controls the ER export of procollagen by regulating the Sar1 cycle. *Science* **337**, 1668-72 (2012).

147. Peterkofsky, B. Regulation of collagen secretion by ascorbic acid in 3T3 and chick embryo fibroblasts. *Biochem Biophys Res Commun* **49**, 1343-50 (1972).

148. Pacifici, M. & Iozzo, R.V. Remodeling of the rough endoplasmic reticulum during stimulation of procollagen secretion by ascorbic acid in cultured chondrocytes. A biochemical and morphological study. *J Biol Chem* **263**, 2483-92 (1988).

149. Mironov, A.A. *et al.* ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains. *Dev Cell* **5**, 583-94 (2003).

150. Aridor, M. *et al.* The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. *J Cell Biol* **152**, 213-29 (2001).

151. Bard, F. *et al.* Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature* **439**, 604-7 (2006).

152. Saito, K. *et al.* TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. *Cell* **136**, 891-902 (2009).

153. Yu, H. *et al.* Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-45 (1994).

154. Wilson, D.G. *et al.* Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J Cell Biol* **193**, 935-51 (2011).

155. Saito, K. *et al.* cTAGE5 mediates collagen secretion through interaction with TANGO1 at endoplasmic reticulum exit sites. *Mol Biol Cell* **22**, 2301-8 (2011).

156. Malhotra, V. & Erlmann, P. Protein export at the ER: loading big collagens into COPII carriers. *EMBO J* **30**, 3475-80 (2011).

157. Gedeon, A.K. *et al.* Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. *Nat Genet* **22**, 400-4 (1999).

158. Bleasel, J.F. *et al.* Type II procollagen gene (COL2A1) mutation in exon 11 associated with spondyloepiphyseal dysplasia, tall stature and precocious osteoarthritis. *J Rheumatol* **22**, 255-61 (1995).

159. Zankl, A. *et al.* Spondyloperipheral dysplasia is caused by truncating mutations in the C-propeptide of COL2A1. *Am J Med Genet A* **129A**, 144-8 (2004).

160. Tiller, G.E. *et al.* A recurrent RNA-splicing mutation in the SEDL gene causes X-linked spondyloepiphyseal dysplasia tarda. *Am J Hum Genet* **68**, 1398-407 (2001).

161. Scrivens, P.J. *et al.* TRAPPC2L is a novel, highly conserved TRAPP-interacting protein. *Traffic* **10**, 724-36 (2009).

162. Jin, L. *et al.* Ubiquitin-dependent regulation of COPII coat size and function. *Nature* **482**, 495-500 (2012).

163. Caswell, P.T., Vadrevu, S. & Norman, J.C. Integrins: masters and slaves of endocytic transport. *Nat Rev Mol Cell Biol* **10**, 843-53 (2009).

164. Trelstad, R.L., Hayashi, K. & Toole, B.P. Epithelial collagens and glycosaminoglycans in the embryonic cornea. Macromolecular order and morphogenesis in the basement membrane. *J Cell Biol* **62**, 815-30 (1974).

165. Roth, T.F. & Porter, K.R. Yolk Protein Uptake in the Oocyte of the Mosquito Aedes Aegypti. L. *J Cell Biol* **20**, 313-32 (1964).

166. Goldenberg, R. & Fine, R.E. Coated vesicles purified from chick tendon fibroblasts contain newly synthesized type I procollagen. *Exp Cell Res* **157**, 41-9 (1985).

167. Canty, E.G. *et al.* Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. *J Cell Biol* **165**, 553-63 (2004).

168. Canty, E.G. *et al.* Actin filaments are required for fibripositor-mediated collagen fibril alignment in tendon. *J Biol Chem* **281**, 38592-8 (2006).

169. Colige, A. *et al.* Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. *Am J Hum Genet* **65**, 308-17 (1999).

170. Watson, R.B., Holmes, D.F., Graham, H.K., Nusgens, B.V. & Kadler, K.E. Surface located procollagen N-propeptides on dermatosparactic collagen fibrils are not cleaved by procollagen N-proteinase and do not inhibit binding of decorin to the fibril surface. *J Mol Biol* **278**, 195-204 (1998).

171. Adachi, E. & Hayashi, T. In vitro formation of hybrid fibrils of type V collagen and type I collagen. Limited growth of type I collagen into thick fibrils by type V collagen. *Connect Tissue Res* **14**, 257-66 (1986).

172. Woodbury, D., Benson-Chanda, V. & Ramirez, F. Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms to the structural criteria of a fibrillar procollagen molecule. *J Biol Chem* **264**, 2735-8 (1989).

173. Lapiere, C.M., Lenaers, A. & Kohn, L.D. Procollagen peptidase: an enzyme excising the coordination peptides of procollagen. *Proc Natl Acad Sci U S A* **68**, 3054-8 (1971).

174. Kuno, K. *et al.* Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene. *J Biol Chem* **272**, 556-62 (1997).

175. Prockop, D.J., Sieron, A.L. & Li, S.W. Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling. *Matrix Biol* **16**, 399-408 (1998).

176. Tuderman, L., Kivirikko, K.I. & Prockop, D.J. Partial purification and characterization of a neutral protease which cleaves the N-terminal propeptides from procollagen. *Biochemistry* **17**, 2948-54 (1978).

177. Cabral, W.A. *et al.* Mutations near amino end of alpha1(I) collagen cause combined osteogenesis imperfecta/Ehlers-Danlos syndrome by interference with N-propeptide processing. *J Biol Chem* **280**, 19259-69 (2005).

178. Byers, P.H. *et al.* Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the COL1A1 and COL1A2 genes of type I collagen. *Am J Med Genet* **72**, 94-105 (1997).

179. Porter, S., Clark, I.M., Kevorkian, L. & Edwards, D.R. The ADAMTS metalloproteinases. *Biochem J* **386**, 15-27 (2005).

180. Wertelecki, W., Smith, L.T. & Byers, P. Initial observations of human dermatosparaxis: Ehlers-Danlos syndrome type VIIC. *J Pediatr* **121**, 558-64 (1992).

181. Wozney, J.M. *et al.* Novel regulators of bone formation: molecular clones and activities. *Science* **242**, 1528-34 (1988).

182. Kessler, E., Takahara, K., Biniaminov, L., Brusel, M. & Greenspan, D.S. Bone morphogenetic protein-1: the type I procollagen C-proteinase. *Science* **271**, 360-2 (1996).

183. Takahara, K. *et al.* Type I procollagen COOH-terminal proteinase enhancer protein: identification, primary structure, and chromosomal localization of the cognate human gene (PCOLCE). *J Biol Chem* **269**, 26280-5 (1994).

184. Scott, I.C. *et al.* Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian Tolloid-like 2, have differential enzymatic

activities and distributions of expression relevant to patterning and skeletogenesis. *Dev Biol* **213**, 283-300 (1999).

185. Takahara, K., Brevard, R., Hoffman, G.G., Suzuki, N. & Greenspan, D.S. Characterization of a novel gene product (mammalian tolloid-like) with high sequence similarity to mammalian tolloid/bone morphogenetic protein-1. *Genomics* **34**, 157-65 (1996).

186. Kessler, E. & Adar, R. Type I procollagen C-proteinase from mouse fibroblasts. Purification and demonstration of a 55-kDa enhancer glycoprotein. *Eur J Biochem* **186**, 115-21 (1989).

187. Moali, C. *et al.* Substrate-specific modulation of a multisubstrate proteinase. C-terminal processing of fibrillar procollagens is the only BMP-1-dependent activity to be enhanced by PCPE-1. *J Biol Chem* **280**, 24188-94 (2005).

188. Steiglitz, B.M., Keene, D.R. & Greenspan, D.S. PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagen-binding protein differing in distribution of expression and post-translational modification from the previously described PCPE1. *J Biol Chem* **277**, 49820-30 (2002).

189. Hojima, Y., van der Rest, M. & Prockop, D.J. Type I procollagen carboxyl-terminal proteinase from chick embryo tendons. Purification and characterization. *J Biol Chem* **260**, 15996-6003 (1985).

190. Njieha, F.K., Morikawa, T., Tuderman, L. & Prockop, D.J. Partial purification of a procollagen C-proteinase. Inhibition by synthetic peptides and sequential cleavage of type I procollagen. *Biochemistry* **21**, 757-64 (1982).

191. Lindahl, K. *et al.* COL1 C-propeptide cleavage site mutations cause high bone mass osteogenesis imperfecta. *Hum Mutat* **32**, 598-609 (2011).

192. Asharani, P.V. *et al.* Attenuated BMP1 function compromises osteogenesis, leading to bone fragility in humans and zebrafish. *Am J Hum Genet* **90**, 661-74 (2012).

193. Suzuki, N. *et al.* Failure of ventral body wall closure in mouse embryos lacking a procollagen C-proteinase encoded by Bmp1, a mammalian gene related to Drosophila tolloid. *Development* **122**, 3587-95 (1996).

194. Wiestner, M. *et al.* Inhibiting effect of procollagen peptides on collagen biosynthesis in fibroblast cultures. *J Biol Chem* **254**, 7016-23 (1979).

195. Fouser, L., Sage, E.H., Clark, J. & Bornstein, P. Feedback regulation of collagen gene expression: a Trojan horse approach. *Proc Natl Acad Sci U S A* **88**, 10158-62 (1991).

196. Wu, C.H., Walton, C.M. & Wu, G.Y. Propeptide-mediated regulation of procollagen synthesis in IMR-90 human lung fibroblast cell cultures. Evidence for transcriptional control. *J Biol Chem* **266**, 2983-7 (1991).

197. Schlumberger, W., Thie, M., Volmer, H., Rauterberg, J. & Robenek, H. Binding and uptake of Col 1(I), a peptide capable of inhibiting collagen synthesis in fibroblasts. *Eur J Cell Biol* **46**, 244-52 (1988).

198. Smedsrod, B. Aminoterminal propeptide of type III procollagen is cleared from the circulation by receptor-mediated endocytosis in liver endothelial cells. *Coll Relat Res* **8**, 375-88 (1988).

199. Weston, S.A., Hulmes, D.J., Mould, A.P., Watson, R.B. & Humphries, M.J. Identification of integrin alpha 2 beta 1 as cell surface receptor for the carboxyl-terminal propeptide of type I procollagen. *J Biol Chem* **269**, 20982-6 (1994).

200. Davies, D. *et al.* Molecular characterisation of integrin-procollagen C-propeptide interactions. *Eur J Biochem* **246**, 274-82 (1997).

201. Fleischmajer, R. *et al.* Ultrastructural identification of extension aminopropeptides of type I and III collagens in human skin. *Proc Natl Acad Sci U S A* **78**, 7360-4 (1981).

202. Kielty, C.M. & Grant, M.E. The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. in *Connective Tissue and its Heritable Disorders* (eds. Royce, P.M. & Steinmann, B.) 159-221 (Wiley-Liss, New York, 2002).

203. Kadler, K.E., Hojima, Y. & Prockop, D.J. Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. *J Biol Chem* **262**, 15696-701 (1987).

204. Bensusan, H.B. & Hoyt, B.L. The effect of various parameters on the rate of formation of fibers from collagen solutions. *J Am Chem Soc* **80**, 719-724 (1958).

205. Comper, W.D. & Veis, A. The mechanism of nucleation for in vitro collagen fibril formation. *Biopolymers* **16**, 2113-2131 (1977).

206. Gross, J. Influence of time on the reversible association between large molecules: the collagen system. *Nature* **181**, 556 (1958).

207. Wood, G.C. The formation of fibrils from collagen solutions. 2. A mechanism of collagen-fibril formation. *Biochem J* **75**, 598-605 (1960).

208. Veis, A. & Yuan, L. Structure of the collagen microfibril. A four-strand overlap model. *Biopolymers* **14**, 895-900 (1975).

209. Veis, A., Miller, A., Leibovich, S.J. & Traub, W. The limiting collagen microfibril. The minimum structure demonstrating native axial periodicity. *Biochim Biophys Acta* **576**, 88-98 (1979).

210. Na, G.C., Butz, L.J. & Carroll, R.J. Mechanism of in vitro collagen fibril assembly. Kinetic and morphological studies. *J Biol Chem* **261**, 12290-9 (1986).

211. Hulmes, D.J., Miller, A., Parry, D.A., Piez, K.A. & Woodhead-Galloway, J. Analysis of the primary structure of collagen for the origins of molecular packing. *J Mol Biol* **79**, 137-48 (1973).

212. Romanic, A.M., Adachi, E., Hojima, Y., Engel, J. & Prockop, D.J. Polymerization of pNcollagen I and copolymerization of pNcollagen I with collagen I. A kinetic, thermodynamic, and morphologic study. *J Biol Chem* **267**, 22265-71 (1992).

213. Comper, W.D. & Veis, A. Characterization of nuclei in in vitro collagen fibril formation. *Biopolymers* **16**, 2133-42 (1977).

214. Prockop, D.J. & Fertala, A. Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides. Demonstration that assembly is driven by specific binding sites on the monomers. *J Biol Chem* **273**, 15598-604 (1998).

215. Cabral, W.A. *et al.* Procollagen with skipping of alpha 1(I) exon 41 has lower binding affinity for alpha 1(I) C-telopeptide, impaired in vitro fibrillogenesis, and altered fibril morphology. *J Biol Chem* **277**, 4215-22 (2002).

216. Kadler, K.E., Hojima, Y. & Prockop, D.J. Collagen fibrils in vitro grow from pointed tips in the C- to N-terminal direction. *Biochem J* **268**, 339-43 (1990).

217. Silver, D., Miller, J., Harrison, R. & Prockop, D.J. Helical model of nucleation and propagation to account for the growth of type I collagen fibrils from symmetrical pointed tips: a special example of self-assembly of rod-like monomers. *Proc Natl Acad Sci U S A* **89**, 9860-4 (1992).

218. Birk, D.E. & Silver, F.H. Collagen fibrillogenesis in vitro: comparison of types I, II, and III. *Arch Biochem Biophys* **235**, 178-85 (1984).

219. Wenstrup, R.J. *et al.* Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem* **279**, 53331-7 (2004).

220. Wenstrup, R.J. *et al.* Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem* **286**, 20455-65 (2011).

221. Douglas, T., Heinemann, S., Bierbaum, S., Scharnweber, D. & Worch, H. Fibrillogenesis of collagen types I, II, and III with small leucine-rich proteoglycans decorin and biglycan. *Biomacromolecules* **7**, 2388-93 (2006).

222. Ameye, L. & Young, M.F. Mice deficient in small leucine-rich proteoglycans: novel in vivo models for osteoporosis, osteoarthritis, Ehlers-Danlos syndrome, muscular dystrophy, and corneal diseases. *Glycobiology* **12**, 107R-16R (2002).

223. Weber, I.T., Harrison, R.W. & Iozzo, R.V. Model structure of decorin and implications for collagen fibrillogenesis. *J Biol Chem* **271**, 31767-70 (1996).

224. Valli, M. *et al.* "In vitro" fibril formation of type I collagen from different sources: biochemical and morphological aspects. *Connect Tissue Res* **15**, 235-44 (1986).

225. Kagan, H.M. & Li, W. Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. *J Cell Biochem* **88**, 660-72 (2003).

226. Siegel, R.C. & Fu, J.C. Collagen cross-linking. Purification and substrate specificity of lysyl oxidase. *J Biol Chem* **251**, 5779-85 (1976).

227. Bannister, D.W. & Burns, A.B. Pepsin treatment of avian skin collagen. Effects on solubility, subunit composition and aggregation properties. *Biochem J* **129**, 677-81 (1972).

228. Zimmermann, B.K., Timpl, R. & Kuhn, K. Intermolecular cross-links of collagen. Participation of the carboxy-terminal nonhelical region of the 1-chain. *Eur J Biochem* **35**, 216-21 (1973).

229. Tanzer, M.L. Cross-linking of collagen. *Science* **180**, 561-6 (1973).

230. Bailey, A.J., Robins, S.P. & Balian, G. Biological significance of the intermolecular crosslinks of collagen. *Nature* **251**, 105-9 (1974).

231. Reiser, K., McCormick, R.J. & Rucker, R.B. Enzymatic and nonenzymatic cross-linking of collagen and elastin. *Faseb J* **6**, 2439-49 (1992).

232. Yamauchi, M., London, R.E., Guenat, C., Hashimoto, F. & Mechanic, G.L. Structure and formation of a stable histidine-based trifunctional cross-link in skin collagen. *J Biol Chem* **262**, 11428-34 (1987).

233. Eyre, D.R., Koob, T.J. & Van Ness, K.P. Quantitation of hydroxypyridinium crosslinks in collagen by high-performance liquid chromatography. *Anal Biochem* **137**, 380-8 (1984).

234. Henkel, W., Glanville, R.W. & Greifendorf, D. Characterisation of a type-I collagen trimeric cross-linked peptide from calf aorta and its cross-linked structure. Detection of pyridinoline by time-of-flight secondary ion-mass spectroscopy and evidence for a new cross-link. *Eur J Biochem* **165**, 427-36 (1987).

235. Friedman, A.W. Important determinants of bone strength: beyond bone mineral density. *J Clin Rheumatol* **12**, 70-7 (2006).

236. Murray, P.D. & Huxley, J.S. Self-Differentiation in the Grafted Limb-Bud of the Chick. *J Anat* **59**, 379-84 (1925).

237. Frost, H.M. Skeletal structural adaptations to mechanical usage (SATMU): 1. Redefining Wolff's law: the bone modeling problem. *Anat Rec* **226**, 403-13 (1990).

238. Fisher, L.W. & Fedarko, N.S. Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. *Connect Tissue Res* **44 Suppl 1**, 33-40 (2003).

239. Boskey, A.L. Osteopontin and related phosphorylated sialoproteins: effects on mineralization. *Ann N Y Acad Sci* **760**, 249-56 (1995).

240. He, G., Dahl, T., Veis, A. & George, A. Dentin matrix protein 1 initiates hydroxyapatite formation in vitro. *Connect Tissue Res* **44 Suppl 1**, 240-5 (2003).

241. Tye, C.E. *et al.* Delineation of the hydroxyapatite-nucleating domains of bone sialoprotein. *J Biol Chem* **278**, 7949-55 (2003).

242. Addison, W.N., Masica, D.L., Gray, J.J. & McKee, M.D. Phosphorylation-dependent inhibition of mineralization by osteopontin ASARM peptides is regulated by PHEX cleavage. *J Bone Miner Res* **25**, 695-705 (2010).

243. Ou-Yang, H., Paschalis, E.P., Mayo, W.E., Boskey, A.L. & Mendelsohn, R. Infrared microscopic imaging of bone: spatial distribution of CO3(2-). *J Bone Miner Res* **16**, 893-900 (2001).

244. Anderson, H.C. Electron microscopic studies of induced cartilage development and calcification. *J Cell Biol* **35**, 81-101 (1967).

245. Bernard, G.W. & Pease, D.C. An electron microscopic study of initial intramembranous osteogenesis. *Am J Anat* **125**, 271-90 (1969).

246. Raggio, C.L., Boyan, B.D. & Boskey, A.L. In vivo hydroxyapatite formation induced by lipids. *J Bone Miner Res* **1**, 409-15 (1986).

247. Wuthier, R.E. Mechanism of de novo mineral formation by matrix vesicles. *Connect Tissue Res* **22**, 27-33; discussion 53-61 (1989).

248. Hunter, G.K., Hauschka, P.V., Poole, A.R., Rosenberg, L.C. & Goldberg, H.A. Nucleation and inhibition of hydroxyapatite formation by mineralized tissue proteins. *Biochem J* **317 (Pt 1)**, 59-64 (1996).

249. Xiao, Z. *et al.* Analysis of the extracellular matrix vesicle proteome in mineralizing osteoblasts. *J Cell Physiol* **210**, 325-35 (2007).

250. Kirsch, T. *et al.* Annexin V-mediated calcium flux across membranes is dependent on the lipid composition: implications for cartilage mineralization. *Biochemistry* **36**, 3359-67 (1997).

251. Golub, E.E. Role of matrix vesicles in biomineralization. *Biochim Biophys Acta* **1790**, 1592-8 (2009).

252. Wu, L.N., Genge, B.R., Lloyd, G.C. & Wuthier, R.E. Collagen-binding proteins in collagenase-released matrix vesicles from cartilage. Interaction between matrix vesicle proteins and different types of collagen. *J Biol Chem* **266**, 1195-203 (1991).

253. Roschger, P., Fratzl, P., Eschberger, J. & Klaushofer, K. Validation of quantitative backscattered electron imaging for the measurement of mineral density distribution in human bone biopsies. *Bone* **23**, 319-26 (1998).

254. Boivin, G. & Meunier, P.J. Changes in bone remodeling rate influence the degree of mineralization of bone. *Connect Tissue Res* **43**, 535-7 (2002).

255. Arnold, S. *et al.* Quantitative analyses of the biomineralization of different hard tissues. *J Microsc* **202**, 488-94 (2001).

256. Blonder, J., Xiao, Z. & Veenstra, T.D. Proteomic profiling of differentiating osteoblasts. *Expert Rev Proteomics* **3**, 483-96 (2006).

257. Peter, S.J., Liang, C.R., Kim, D.J., Widmer, M.S. & Mikos, A.G. Osteoblastic phenotype of rat marrow stromal cells cultured in the presence of dexamethasone, beta-glycerolphosphate, and L-ascorbic acid. *J Cell Biochem* **71**, 55-62 (1998).

258. St-Jacques, B., Hammerschmidt, M. & McMahon, A.P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev* **13**, 2072-86 (1999).

259. Komori, T. *et al.* Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**, 755-64 (1997).

260. Otto, F. *et al.* Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765-71 (1997).

261. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17-29 (2002).

262. Bonewald, L.F. The amazing osteocyte. *J Bone Miner Res* **26**, 229-38 (2011).

263. Feng, J.Q. *et al.* Loss of DMP1 causes rickets and osteomalacia and identifies a role for osteocytes in mineral metabolism. *Nat Genet* **38**, 1310-5 (2006).

264. Liu, S. *et al.* Pathogenic role of Fgf23 in Hyp mice. *Am J Physiol Endocrinol Metab* **291**, E38-49 (2006).

265. Feng, J.Q., Clinkenbeard, E.L., Yuan, B., White, K.E. & Drezner, M.K. Osteocyte regulation of phosphate homeostasis and bone mineralization underlies the pathophysiology of the heritable disorders of rickets and osteomalacia. *Bone* (2013).

266. Holmbeck, K. *et al.* The metalloproteinase MT1-MMP is required for normal development and maintenance of osteocyte processes in bone. *J Cell Sci* **118**, 147-56 (2005).

267. Zhang, K. *et al.* E11/gp38 selective expression in osteocytes: regulation by mechanical strain and role in dendrite elongation. *Mol Cell Biol* **26**, 4539-52 (2006).

268. Wang, L., Ciani, C., Doty, S.B. & Fritton, S.P. Delineating bone's interstitial fluid pathway in vivo. *Bone* **34**, 499-509 (2004).

269. Boyle, W.J., Simonet, W.S. & Lacey, D.L. Osteoclast differentiation and activation. *Nature* **423**, 337-42 (2003).

270. Tondravi, M.M. *et al.* Osteopetrosis in mice lacking haematopoietic transcription factor PU.1. *Nature* **386**, 81-4 (1997).

271. Yao, Z. *et al.* Tumor necrosis factor-alpha increases circulating osteoclast precursor numbers by promoting their proliferation and differentiation in the bone marrow through up-regulation of c-Fms expression. *J Biol Chem* **281**, 11846-55 (2006).

272. Dai, X.M. *et al.* Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency,

increased primitive progenitor cell frequencies, and reproductive defects. *Blood* **99**, 111-20 (2002).

273. Fukuchi, M. *et al.* High-level expression of the Smad ubiquitin ligase Smurf2 correlates with poor prognosis in patients with esophageal squamous cell carcinoma. *Cancer Res* **62**, 7162-5 (2002).

274. Takayanagi, H. Osteoimmunology: shared mechanisms and crosstalk between the immune and bone systems. *Nat Rev Immunol* **7**, 292-304 (2007).

275. Kanamaru, F. *et al.* Expression of membrane-bound and soluble receptor activator of NF-kappaB ligand (RANKL) in human T cells. *Immunol Lett* **94**, 239-46 (2004).

276. Kawai, T. *et al.* B and T lymphocytes are the primary sources of RANKL in the bone resorptive lesion of periodontal disease. *Am J Pathol* **169**, 987-98 (2006).

277. Takayanagi, H. Mechanistic insight into osteoclast differentiation in osteoimmunology. *J Mol Med (Berl)* **83**, 170-9 (2005).

278. Teitelbaum, S.L. Bone resorption by osteoclasts. *Science* **289**, 1504-8 (2000).

279. Boyce, B.F., Yao, Z. & Xing, L. Osteoclasts have multiple roles in bone in addition to bone resorption. *Crit Rev Eukaryot Gene Expr* **19**, 171-80 (2009).

280. Huelsken, J. & Behrens, J. The Wnt signalling pathway. *J Cell Sci* **115**, 3977-8 (2002).

281. Boyden, L.M. *et al.* High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med* **346**, 1513-21 (2002).

282. Little, R.D. *et al.* A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am J Hum Genet* **70**, 11-9 (2002).

283. Gong, Y. *et al.* LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* **107**, 513-23 (2001).

284. Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J. & Skarnes, W.C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535-8 (2000).

285. Holmen, S.L. *et al.* Decreased BMD and limb deformities in mice carrying mutations in both Lrp5 and Lrp6. *J Bone Miner Res* **19**, 2033-40 (2004).

286. Day, T.F., Guo, X., Garrett-Beal, L. & Yang, Y. Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis. *Dev Cell* **8**, 739-50 (2005).

287. Holmen, S.L. *et al.* Essential role of beta-catenin in postnatal bone acquisition. *J Biol Chem* **280**, 21162-8 (2005).

288. Eyckmans, J., Roberts, S.J., Schrooten, J. & Luyten, F.P. A clinically relevant model of osteoinduction: a process requiring calcium phosphate and BMP/Wnt signalling. *J Cell Mol Med* **14**, 1845-56 (2010).

289. Gaur, T. *et al.* Canonical WNT signaling promotes osteogenesis by directly stimulating Runx2 gene expression. *J Biol Chem* **280**, 33132-40 (2005).

290. Kang, S. *et al.* Wnt signaling stimulates osteoblastogenesis of mesenchymal precursors by suppressing CCAAT/enhancer-binding protein alpha and peroxisome proliferator-activated receptor gamma. *J Biol Chem* **282**, 14515-24 (2007).

291. Van Hul, W. *et al.* Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12-q21. *Am J Hum Genet* **62**, 391-9 (1998).

292. Brunkow, M.E. *et al.* Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. *Am J Hum Genet* **68**, 577-89 (2001).

293. Wergedal, J.E. *et al.* Patients with Van Buchem disease, an osteosclerotic genetic disease, have elevated bone formation markers, higher bone density, and greater derived polar moment of inertia than normal. *J Clin Endocrinol Metab* **88**, 5778-83 (2003).

294. Li, X. *et al.* Targeted deletion of the sclerostin gene in mice results in increased bone formation and bone strength. *J Bone Miner Res* **23**, 860-9 (2008).

295. Li, X. *et al.* Sclerostin binds to LRP5/6 and antagonizes canonical Wnt signaling. *J Biol Chem* **280**, 19883-7 (2005).

296. Sutherland, M.K. *et al.* Sclerostin promotes the apoptosis of human osteoblastic cells: a novel regulation of bone formation. *Bone* **35**, 828-35 (2004).

297. Winkler, D.G. *et al.* Osteocyte control of bone formation via sclerostin, a novel BMP antagonist. *EMBO J* **22**, 6267-76 (2003).

298. Miyazono, K., Hellman, U., Wernstedt, C. & Heldin, C.H. Latent high molecular weight complex of transforming growth factor beta 1. Purification from human platelets and structural characterization. *J Biol Chem* **263**, 6407-15 (1988).

299. Gelb, B.D. Marfan's syndrome and related disorders--more tightly connected than we thought. *N Engl J Med* **355**, 841-4 (2006).

300. Wagner, D.O. *et al.* BMPs: from bone to body morphogenetic proteins. *Sci Signal* **3**, mr1 (2010).

301. Yi, J.J., Barnes, A.P., Hand, R., Polleux, F. & Ehlers, M.D. TGF-beta signaling specifies axons during brain development. *Cell* **142**, 144-57 (2010).

302. Zhou, S. TGF-beta regulates beta-catenin signaling and osteoblast differentiation in human mesenchymal stem cells. *J Cell Biochem* **112**, 1651-60 (2011).

303. Geiser, A.G. *et al.* A new selective estrogen receptor modulator with potent uterine antagonist activity, agonist activity in bone, and minimal ovarian stimulation. *Endocrinology* **146**, 4524-35 (2005).

304. Kinoshita, A. *et al.* Domain-specific mutations in TGFB1 result in Camurati-Engelmann disease. *Nat Genet* **26**, 19-20 (2000).

305. Dunker, N. & Krieglstein, K. Tgfbeta2 -/- Tgfbeta3 -/- double knockout mice display severe midline fusion defects and early embryonic lethality. *Anat Embryol (Berl)* **206**, 73-83 (2002).

306. Qiu, T. *et al.* TGF-beta type II receptor phosphorylates PTH receptor to integrate bone remodelling signalling. *Nat Cell Biol* **12**, 224-34 (2010).

307. Atfi, A. & Baron, R. PTH battles TGF-beta in bone. *Nat Cell Biol* **12**, 205-7 (2010).

308. Urist, M.R. Bone: formation by autoinduction. *Science* **150**, 893-9 (1965).

309. Celeste, A.J. *et al.* Identification of transforming growth factor beta family members present in bone-inductive protein purified from bovine bone. *Proc Natl Acad Sci U S A* **87**, 9843-7 (1990).

310. Jonason, J.H., Xiao, G., Zhang, M., Xing, L. & Chen, D. Post-translational Regulation of Runx2 in Bone and Cartilage. *J Dent Res* **88**, 693-703 (2009).

311. Gori, F., Thomas, T., Hicok, K.C., Spelsberg, T.C. & Riggs, B.L. Differentiation of human marrow stromal precursor cells: bone morphogenetic protein-2 increases OSF2/CBFA1, enhances osteoblast commitment, and inhibits late adipocyte maturation. *J Bone Miner Res* **14**, 1522-35 (1999).

312. Hanada, K., Dennis, J.E. & Caplan, A.I. Stimulatory effects of basic fibroblast growth factor and bone morphogenetic protein-2 on osteogenic differentiation of

rat bone marrow-derived mesenchymal stem cells. *J Bone Miner Res* **12**, 1606-14 (1997).

313. Huang, Z., Ren, P.G., Ma, T., Smith, R.L. & Goodman, S.B. Modulating osteogenesis of mesenchymal stem cells by modifying growth factor availability. *Cytokine* **51**, 305-10 (2010).

314. Noel, D. *et al.* Short-term BMP-2 expression is sufficient for in vivo osteochondral differentiation of mesenchymal stem cells. *Stem Cells* **22**, 74-85 (2004).

315. Tsuji, K. *et al.* BMP2 activity, although dispensable for bone formation, is required for the initiation of fracture healing. *Nat Genet* **38**, 1424-9 (2006).

316. Retting, K.N., Song, B., Yoon, B.S. & Lyons, K.M. BMP canonical Smad signaling through Smad1 and Smad5 is required for endochondral bone formation. *Development* **136**, 1093-104 (2009).

317. Wang, M. *et al.* Smad1 plays an essential role in bone development and postnatal bone formation. *Osteoarthritis Cartilage* **19**, 751-62 (2011).

318. Tan, X. *et al.* Smad4 is required for maintaining normal murine postnatal bone homeostasis. *J Cell Sci* **120**, 2162-70 (2007).

319. Daluiski, A. *et al.* Bone morphogenetic protein-3 is a negative regulator of bone density. *Nat Genet* **27**, 84-8 (2001).

320. Okamoto, M. *et al.* Conditional deletion of Bmpr1a in differentiated osteoclasts increases osteoblastic bone formation, increasing volume of remodeling bone in mice. *J Bone Miner Res* **26**, 2511-22 (2011).

321. Kamiya, N. *et al.* Disruption of BMP signaling in osteoblasts through type IA receptor (BMPRIA) increases bone mass. *J Bone Miner Res* **23**, 2007-17 (2008).

322. Kamiya, N. *et al.* BMP signaling negatively regulates bone mass through sclerostin by inhibiting the canonical Wnt pathway. *Development* **135**, 3801-11 (2008).

323. Kamiya, N. *et al.* Wnt inhibitors Dkk1 and Sost are downstream targets of BMP signaling through the type IA receptor (BMPRIA) in osteoblasts. *J Bone Miner Res* **25**, 200-10 (2010).

324. Heino, T.J., Hentunen, T.A. & Vaananen, H.K. Osteocytes inhibit osteoclastic bone resorption through transforming growth factor-beta: enhancement by estrogen. *J Cell Biochem* **85**, 185-97 (2002).

325. Zhao, S., Zhang, Y.K., Harris, S., Ahuja, S.S. & Bonewald, L.F. MLO-Y4 osteocyte-like cells support osteoclast formation and activation. *J Bone Miner Res* **17**, 2068-79 (2002).

326. Kogianni, G., Mann, V. & Noble, B.S. Apoptotic bodies convey activity capable of initiating osteoclastogenesis and localized bone destruction. *J Bone Miner Res* **23**, 915-27 (2008).

327. Hauge, E.M., Qvesel, D., Eriksen, E.F., Mosekilde, L. & Melsen, F. Cancellous bone remodeling occurs in specialized compartments lined by cells expressing osteoblastic markers. *J Bone Miner Res* **16**, 1575-82 (2001).

328. Everts, V. *et al.* The bone lining cell: its role in cleaning Howship's lacunae and initiating bone formation. *J Bone Miner Res* **17**, 77-90 (2002).

329. Andersen, T.L. *et al.* A scrutiny of matrix metalloproteinases in osteoclasts: evidence for heterogeneity and for the presence of MMPs synthesized by other cells. *Bone* **35**, 1107-19 (2004).

330. Andersen, T.L. *et al.* A physical mechanism for coupling bone resorption and formation in adult human bone. *Am J Pathol* **174**, 239-47 (2009).

331. van der Meulen, M.C., Jepsen, K.J. & Mikic, B. Understanding bone strength: size isn't everything. *Bone* **29**, 101-4 (2001).

332. Roschger, P. *et al.* Alendronate increases degree and uniformity of mineralization in cancellous bone and decreases the porosity in cortical bone of osteoporotic women. *Bone* **29**, 185-91 (2001).

333. Currey, J.D., Brear, K. & Zioupos, P. The effects of ageing and changes in mineral content in degrading the toughness of human femora. *J Biomech* **29**, 257-60 (1996).

334. Lindahl, K. *et al.* COL1 C-propeptide cleavage site mutations cause high bone mass osteogenesis imperfecta. *Hum Mutat* (2011).

335. Mendelsohn, R., Paschalis, E.P. & Boskey, A.L. Infrared spectroscopy, microscopy, and microscopic imaging of mineralizing tissues: spectra-structure correlations from human iliac crest biopsies. *J Biomed Opt* **4**, 14-21 (1999).

336. Carden, A. & Morris, M.D. Application of vibrational spectroscopy to the study of mineralized tissues (review). *J Biomed Opt* **5**, 259-68 (2000).

337. Link, T.M. & Majumdar, S. Current diagnostic techniques in the evaluation of bone architecture. *Curr Osteoporos Rep* **2**, 47-52 (2004).

338. Turner, C.H. Biomechanics of bone: determinants of skeletal fragility and bone quality. *Osteoporos Int* **13**, 97-104 (2002).

339. Crabtree, N. *et al.* Intracapsular hip fracture and the region-specific loss of cortical bone: analysis by peripheral quantitative computed tomography. *J Bone Miner Res* **16**, 1318-28 (2001).

340. Parkinson, I.H. & Fazzalari, N.L. Interrelationships between structural parameters of cancellous bone reveal accelerated structural change at low bone volume. *J Bone Miner Res* **18**, 2200-5 (2003).

341. Kleerekoper, M., Villanueva, A.R., Stanciu, J., Rao, D.S. & Parfitt, A.M. The role of three-dimensional trabecular microstructure in the pathogenesis of vertebral compression fractures. *Calcif Tissue Int* **37**, 594-7 (1985).

342. Weinstein, R.S. & Hutson, M.S. Decreased trabecular width and increased trabecular spacing contribute to bone loss with aging. *Bone* **8**, 137-42 (1987).

343. Ciarelli, T.E., Fyhrie, D.P., Schaffler, M.B. & Goldstein, S.A. Variations in three-dimensional cancellous bone architecture of the proximal femur in female hip fractures and in controls. *J Bone Miner Res* **15**, 32-40 (2000).

344. Burstein, A.H., Zika, J.M., Heiple, K.G. & Klein, L. Contribution of collagen and mineral to the elastic-plastic properties of bone. *J Bone Joint Surg Am* **57**, 956-61 (1975).

345. Landis, W.J. The strength of a calcified tissue depends in part on the molecular structure and organization of its constituent mineral crystals in their organic matrix. *Bone* **16**, 533-44 (1995).

346. Sims, T.J., Miles, C.A., Bailey, A.J. & Camacho, N.P. Properties of collagen in OIM mouse tissues. *Connect Tissue Res* **44 Suppl 1**, 202-5 (2003).

347. Kulak, C.A. & Dempster, D.W. Bone histomorphometry: a concise review for endocrinologists and clinicians. *Arq Bras Endocrinol Metabol* **54**, 87-98 (2010).

348. Sillence, D.O., Senn, A. & Danks, D.M. Genetic heterogeneity in osteogenesis imperfecta. *J Med Genet* **16**, 101-16 (1979).

349. Andersen, P.E., Jr. & Hauge, M. Osteogenesis imperfecta: a genetic, radiological, and epidemiological study. *Clin Genet* **36**, 250-5 (1989).

350. Beighton, P. & Versfeld, G.A. On the paradoxically high relative prevalence of osteogenesis imperfecta type III in the black population of South Africa. *Clin Genet* **27**, 398-401 (1985).

351. Stoll, C., Dott, B., Roth, M.P. & Alembik, Y. Birth prevalence rates of skeletal dysplasias. *Clin Genet* **35**, 88-92 (1989).

352. Marini, J.C. Osteogenesis imperfecta. in *Nelson Textbook of Pediatrics* (eds. Behrman, R.E., Kliegman, R.M. & Jensen, H.B.) 2336-2338 (Saunders, Philadelphia, 2004).

353. Willing, M.C. *et al.* Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am J Hum Genet* **55**, 638-47 (1994).

354. Forlino, A., Cabral, W.A., Barnes, A.M. & Marini, J.C. New perspectives on osteogenesis imperfecta. *Nat Rev Endocrinol* **7**, 540-57 (2011).

355. Bonadio, J., Holbrook, K.A., Gelinas, R.E., Jacob, J. & Byers, P.H. Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. *J Biol Chem* **260**, 1734-42 (1985).

356. Raghunath, M., Bruckner, P. & Steinmann, B. Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J Mol Biol* **236**, 940-9 (1994).

357. Cabral, W.A., Milgrom, S., Letocha, A.D., Moriarty, E. & Marini, J.C. Biochemical screening of type I collagen in osteogenesis imperfecta: detection of glycine substitutions in the amino end of the alpha chains requires supplementation by molecular analysis. *J Med Genet* **43**, 685-90 (2006).

358. Cabral, W.A. & Marini, J.C. High proportion of mutant osteoblasts is compatible with normal skeletal function in mosaic carriers of osteogenesis imperfecta. *Am J Hum Genet* **74**, 752-60 (2004).

359. Edwards, M.J., Wenstrup, R.J., Byers, P.H. & Cohn, D.H. Recurrence of lethal osteogenesis imperfecta due to parental mosaicism for a mutation in the COL1A2 gene of type I collagen. The mosaic parent exhibits phenotypic features of a mild form of the disease. *Hum Mutat* **1**, 47-54 (1992).

360. Constantinou, C.D., Pack, M., Young, S.B. & Prockop, D.J. Phenotypic heterogeneity in osteogenesis imperfecta: the mildly affected mother of a proband with a lethal variant has the same mutation substituting cysteine for alpha 1-glycine 904 in a type I procollagen gene (COL1A1). *Am J Hum Genet* **47**, 670-9 (1990).

361. Forlino, A., Porter, F.D., Lee, E.J., Westphal, H. & Marini, J.C. Use of the Cre/lox recombination system to develop a non-lethal knock-in murine model for osteogenesis imperfecta with an alpha1(I) G349C substitution. Variability in phenotype in BrtlIV mice. *J Biol Chem* **274**, 37923-31 (1999).

362. Daley, E. *et al.* Variable bone fragility associated with an Amish COL1A2 variant and a knock-in mouse model. *J Bone Miner Res* **25**, 247-61 (2010).

363. Chipman, S.D. *et al.* Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. *Proc Natl Acad Sci U S A* **90**, 1701-5 (1993).

364. Nicholls, A.C. *et al.* The clinical features of homozygous alpha 2(I) collagen deficient osteogenesis imperfecta. *J Med Genet* **21**, 257-62 (1984).

365. Schwarze, U. *et al.* Rare autosomal recessive cardiac valvular form of Ehlers-Danlos syndrome results from mutations in the COL1A2 gene that activate the nonsense-mediated RNA decay pathway. *Am J Hum Genet* **74**, 917-30 (2004).

366. Malfait, F. *et al.* Total absence of the alpha2(I) chain of collagen type I causes a rare form of Ehlers-Danlos syndrome with hypermobility and propensity to cardiac valvular problems. *J Med Genet* **43**, e36 (2006).

367. Kozloff, K.M. *et al.* Brittle IV mouse model for osteogenesis imperfecta IV demonstrates postpubertal adaptations to improve whole bone strength. *J Bone Miner Res* **19**, 614-22 (2004).

368. McBride, D.J., Jr., Shapiro, J.R. & Dunn, M.G. Bone geometry and strength measurements in aging mice with the oim mutation. *Calcif Tissue Int* **62**, 172-6 (1998).

369. Misof, B.M. *et al.* Differential effects of alendronate treatment on bone from growing osteogenesis imperfecta and wild-type mouse. *Bone* **36**, 150-8 (2005).

370. Carleton, S.M. *et al.* Role of genetic background in determining phenotypic severity throughout postnatal development and at peak bone mass in Col1a2 deficient mice (oim). *Bone* **42**, 681-94 (2008).

371. Camacho, N.P. *et al.* The material basis for reduced mechanical properties in oim mice bones. *J Bone Miner Res* **14**, 264-72 (1999).

372. Miller, E., Delos, D., Baldini, T., Wright, T.M. & Pleshko Camacho, N. Abnormal mineral-matrix interactions are a significant contributor to fragility in oim/oim bone. *Calcif Tissue Int* **81**, 206-14 (2007).

373. Uveges, T.E. *et al.* Alendronate treatment of the brtl osteogenesis imperfecta mouse improves femoral geometry and load response before fracture but decreases predicted material properties and has detrimental effects on osteoblasts and bone formation. *J Bone Miner Res* **24**, 849-59 (2009).

374. Cabral, W.A., Fratzl-Zelman, N., Roschger, P. & Marini, J.C. Cellular dysregulation of gene expression in response to abnormal extracellular matrix may contribute to matrix hypermineralization in osteogenesis imperfecta. *Bone* **48**, S71 (2011).

375. Phillips, C.L. *et al.* Oim mice exhibit altered femur and incisor mineral composition and decreased bone mineral density. *Bone* **27**, 219-26 (2000).

376. Camacho, N.P., Landis, W.J. & Boskey, A.L. Mineral changes in a mouse model of osteogenesis imperfecta detected by Fourier transform infrared microscopy. *Connect Tissue Res* **35**, 259-65 (1996).

377. Uveges, T.E. *et al.* Cellular mechanism of decreased bone in Brtl mouse model of OI: imbalance of decreased osteoblast function and increased osteoclasts and their precursors. *J Bone Miner Res* **23**, 1983-94 (2008).

378. Paterson, C.R., McAllion, S. & Stellman, J.L. Osteogenesis imperfecta after the menopause. *N Engl J Med* **310**, 1694-6 (1984).

379. Vanleene, M. & Shefelbine, S.J. Therapeutic impact of low amplitude high frequency whole body vibrations on the osteogenesis imperfecta mouse bone. *Bone* **53**, 507-14 (2013).

380. Kalajzic, I. *et al.* Osteoblastic response to the defective matrix in the osteogenesis imperfecta murine (oim) mouse. *Endocrinology* **143**, 1594-601 (2002).

381. Zhang, H., Doty, S.B., Hughes, C., Dempster, D. & Camacho, N.P. Increased resorptive activity and accompanying morphological alterations in osteoclasts derived from the oim/oim mouse model of osteogenesis imperfecta. *J Cell Biochem* **102**, 1011-20 (2007).

382. Li, H. *et al.* Immature osteoblast lineage cells increase osteoclastogenesis in osteogenesis imperfecta murine. *Am J Pathol* **176**, 2405-13 (2010).

383. Rauch, F., Travers, R., Parfitt, A.M. & Glorieux, F.H. Static and dynamic bone histomorphometry in children with osteogenesis imperfecta. *Bone* **26**, 581-9 (2000).

384. Marini, J.C. *et al.* Positive linear growth and bone responses to growth hormone treatment in children with types III and IV osteogenesis imperfecta: high predictive value of the carboxyterminal propeptide of type I procollagen. *J Bone Miner Res* **18**, 237-43 (2003).

385. Glorieux, F.H. *et al.* Type V osteogenesis imperfecta: a new form of brittle bone disease. *J Bone Miner Res* **15**, 1650-8 (2000).

386. Cheung, M.S., Glorieux, F.H. & Rauch, F. Natural history of hyperplastic callus formation in osteogenesis imperfecta type V. *J Bone Miner Res* **22**, 1181-6 (2007).

387. Cho, T.J. *et al.* A Single Recurrent Mutation in the 5'-UTR of IFITM5 Causes Osteogenesis Imperfecta Type V. *Am J Hum Genet* **91**, 343-8 (2012).

388. Semler, O. *et al.* A Mutation in the 5'-UTR of IFITM5 Creates an In-Frame Start Codon and Causes Autosomal-Dominant Osteogenesis Imperfecta Type V with Hyperplastic Callus. *Am J Hum Genet* **91**, 349-57 (2012).

389. Tombran-Tink, J. & Barnstable, C.J. Osteoblasts and osteoclasts express PEDF, VEGF-A isoforms, and VEGF receptors: possible mediators of angiogenesis and matrix remodeling in the bone. *Biochem Biophys Res Commun* **316**, 573-9 (2004).

390. Becker, J. *et al.* Exome Sequencing Identifies Truncating Mutations in Human SERPINF1 in Autosomal-Recessive Osteogenesis Imperfecta. *Am J Hum Genet* **88**, 362-71 (2011).

391. Glorieux, F.H. *et al.* Osteogenesis imperfecta type VI: a form of brittle bone disease with a mineralization defect. *J Bone Miner Res* **17**, 30-8 (2002).

392. Drogemuller, C. *et al.* A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta. *PLoS Genet* **5**, e1000579 (2009).

393. Kelley, B.P. *et al.* Mutations in FKBP10 cause recessive osteogenesis imperfecta and type 1 bruck syndrome. *J Bone Miner Res* **2010**, 13 (2010).

394. Barnes, A.M. *et al.* Kuskokwim Syndrome, a Recessive Congenital Contracture Disorder, Extends the Phenotype of FKBP10 Mutations. *Hum Mutat* **34**, 1279-88 (2013).

395. Castagnola, P. *et al.* Cartilage associated protein (CASP) is a novel developmentally regulated chick embryo protein. *J Cell Sci* **110 (Pt 12)**, 1351-9 (1997).

396. Morello, R. *et al.* cDNA cloning, characterization and chromosome mapping of Crtap encoding the mouse cartilage associated protein. *Matrix Biol* **18**, 319-24 (1999).

397. Tonachini, L. *et al.* cDNA cloning, characterization and chromosome mapping of the gene encoding human cartilage associated protein (CRTAP). *Cytogenet Cell Genet* **87**, 191-4 (1999).

398. Wassenhove-McCarthy, D.J. & McCarthy, K.J. Molecular characterization of a novel basement membrane-associated proteoglycan, leprecan. *J Biol Chem* **274**, 25004-17 (1999).

399. Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. & Wadhwa, R. Gros1, a potential growth suppressor on chromosome 1: its identity to basement membrane-associated proteoglycan, leprecan. *Oncogene* **19**, 3576-83 (2000).

400. Tiainen, P., Pasanen, A., Sormunen, R. & Myllyharju, J. Characterization of Recombinant Human Prolyl 3-Hydroxylase Isoenzyme 2, an Enzyme Modifying the Basement Membrane Collagen IV. *J Biol Chem* **283**, 19432-9 (2008).

401. Fischer, G., Bang, H. & Mech, C. [Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides]. *Biomed Biochim Acta* **43**, 1101-11 (1984).

402. Bachinger, H.P. The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. *J Biol Chem* **262**, 17144-8 (1987).

403. Bachinger, H.P., Bruckner, P., Timpl, R., Prockop, D.J. & Engel, J. Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. *Eur J Biochem* **106**, 619-32 (1980).

404. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F.X. Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* **337**, 476-8 (1989).

405. Baldridge, D. *et al.* Generalized connective tissue disease in Crtap-/- mouse. *PLoS One* **5**, e10560 (2010).

406. Fratzl-Zelman, N. *et al.* CRTAP deficiency leads to abnormally high bone matrix mineralization in a murine model and in children with osteogenesis imperfecta type VII. *Bone* **46**, 820-6 (2009).

407. Labuda, M. *et al.* Osteogenesis imperfecta type VII maps to the short arm of chromosome 3. *Bone* **31**, 19-25 (2002).

408. Valli, M. *et al.* Deficiency of CRTAP in non-lethal recessive osteogenesis imperfecta reduces collagen deposition into matrix. *Clin Genet* **82**, 453-9 (2012).

409. Baldridge, D. *et al.* CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. *Hum Mutat* **29**, 1435-42 (2008).

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

411. Moul, A., Alladin, A., Navarrete, C., Abdenour, G. & Rodriguez, M.M. Osteogenesis imperfecta due to compound heterozygosity for the LEPRE1 gene. *Fetal Pediatr Pathol* **32**, 319-25 (2013).

412. Pyott, S.M. *et al.* Recurrence of perinatal lethal osteogenesis imperfecta in sibships: parsing the risk between parental mosaicism for dominant mutations and autosomal recessive inheritance. *Genet Med* **13**, 125-30 (2011).

413. Cabral, W.A. *et al.* A founder mutation in LEPRE1 carried by 1.5% of West Africans and 0.4% of African Americans causes lethal recessive osteogenesis imperfecta. *Genet Med* **14**, 543-51 (2012).

414. Torre-Blanco, A. *et al.* Temperature-induced post-translational over-modification of type I procollagen. Effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self-assembly of collagen into fibrils. *J Biol Chem* **267**, 2650-5 (1992).

415. Byers, P.H. & Cole, W.G. Osteogenesis imperfecta. in *Connective Tissue and Its Heritable Disorders* (eds. Royce, P.M. & Steinmann, B.) 385-430 (Wiley-Liss, Inc., New York, 2002).

416. Kojima, T. *et al.* The retention of abnormal type I procollagen and correlated expression of HSP 47 in fibroblasts from a patient with lethal osteogenesis imperfecta. *J Pathol* **184**, 212-8 (1998).

417. Wilson, R., Freddi, S., Chan, D., Cheah, K.S. & Bateman, J.F. Misfolding of collagen X chains harboring Schmid metaphyseal chondrodysplasia mutations results in aberrant disulfide bond formation, intracellular retention, and activation of the unfolded protein response. *J Biol Chem* **280**, 15544-52 (2005).

418. Lisse, T.S. *et al.* ER stress-mediated apoptosis in a new mouse model of osteogenesis imperfecta. *PLoS Genet* **4**, e7 (2008).

419. Vranka, J.A. *et al.* Prolyl 3-hydroxylase 1 null mice display abnormalities in fibrillar collagen-rich tissues such as tendons, skin, and bones. *J Biol Chem* **285**, 17253-62 (2010).

420. Pokidysheva, E. *et al.* Posttranslational Modifications in Type I Collagen from Different Tissues extracted from wild type and Prolyl 3-hydroxylase 1 Null Mice. *J Biol Chem* **288**, 24742-24752 (2013).

421. Barnes, A.M. *et al.* Lack of cyclophilin B in osteogenesis imperfecta with normal collagen folding. *N Engl J Med* **362**, 521-8 (2010).

422. van Dijk, F.S. *et al.* PPIB mutations cause severe osteogenesis imperfecta. *Am J Hum Genet* **85**, 521-7 (2009).

423. Pyott, S.M. *et al.* Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum Mol Genet* **20**, 1595-609 (2011).

424. Caparros-Martin, J.A. *et al.* Clinical and molecular analysis in families with autosomal recessive osteogenesis imperfecta identifies mutations in five genes and suggests genotype-phenotype correlations. *Am J Med Genet A* **161**, 1354-69 (2013).

425. Van Dijk, F.S., Cobben, J.M. & Pals, G. Osteogenesis Imperfecta, normal collagen folding, and lack of cyclophilin B. *N Engl J Med* **362**, 1941-2 (2010).

426. Price, E.R. *et al.* Human cyclophilin B: a second cyclophilin gene encodes a peptidyl-prolyl isomerase with a signal sequence. *Proc Natl Acad Sci U S A* **88**, 1903-7 (1991).

427. Ishikawa, Y. *et al.* Mutation in cyclophilin B that causes hyperelastosis cutis in American Quarter Horse does not affect peptidylprolyl cis-trans isomerase activity but shows altered cyclophilin B-protein interactions and affects collagen folding. *J Biol Chem* **287**, 22253-65 (2012).

428. Fietzek, P.P., Rexrodt, F.W., Wendt, P., Stark, M. & Kuhn, K. The covalent structure of collagen. Amino-acid sequence of peptide 1-CB6-C2. *Eur J Biochem* **30**, 163-8 (1972).

429. Weis, M.A. *et al.* Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J Biol Chem* **285**, 2580-90 (2010).

430. Marini, J.C., Cabral, W.A. & Barnes, A.M. Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. *Cell Tissue Res* **339**, 59-70 (2010).

431. Dalgleish, R. Osteogenesis Imperfecta Database: LEPRE1. (2009).

432. Wright, B.S. *et al.* Carrier frequency of the RSH/Smith-Lemli-Opitz IVS8-1G>C mutation in African Americans. *Am J Med Genet A* **120A**, 139-41 (2003).

433. Rotimi, C.N. *et al.* A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Diabetes* **53**, 838-41 (2004).

434. Adeyemo, A. *et al.* A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* **5**, e1000564 (2009).

435. Tishkoff, S.A. *et al.* The genetic structure and history of Africans and African Americans. *Science* **324**, 1035-44 (2009).

436. Zeigler-Johnson, C.M. *et al.* Ethnic differences in the frequency of prostate cancer susceptibility alleles at SRD5A2 and CYP3A4. *Hum Hered* **54**, 13-21 (2002).

437. Rannala, B. & Slatkin, M. Methods for multipoint disease mapping using linkage disequilibrium. *Genet Epidemiol* **19 Suppl 1**, S71-7 (2000).

438. Rannala, B. & Bertorelle, G. Using linked markers to infer the age of a mutation. *Hum Mutat* **18**, 87-100 (2001).

439. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841-5 (2010).

440. Ward, L.M. *et al.* Osteogenesis imperfecta type VII: an autosomal recessive form of brittle bone disease. *Bone* **31**, 12-8 (2002).

441. Williams, E.M. *et al.* Phenotypical features of an unique Irish family with severe autosomal recessive osteogenesis imperfecta. *Clin Genet* **35**, 181-90 (1989).

442. Oyinloye, O.O. Osteogenesis imperfecta in a Nigerian neonate: a case report. *West Afr J Med* **27**, 114-6 (2008).

443. Minnis, H., Ramsay, R., Ewije, P. & Kumar, C. Osteogenesis imperfecta and non-accidental injury. *Br J Psychiatry* **166**, 824-5 (1995).

444. Curtin, P.D. *The Atlantic Slave Trade: A Census*, (The University of Wisconsin Press, Madison, 1969).

445. Eltis, D. & Halbert, M. Voyages: The Trans-Atlantic Slave Trade Database. (2009).

446. Eltis, D., Behrendt, S., Richardson, D. & Klein, H. Origins of Enslaved Africans Shipped to North America. in *The Atlantic Slave Trade: A Database on CD-Rom* (Cambridge University Press, 1999).

447. Marshall, D.I. The history of Caribbean migrations: The case of the West Indies. *Caribbean Review* **11**, 6-9, 52-53 (1982).

448. Gordon, A. The new diaspora: African immigration to The United States. *J Third World Studies* **15**, 79-103 (1998).

449. Sandhoff, K., Conzelmann, E., Neufeld, E.F., Kaback, M.M. & Suzuki, K. The GM2 Gangliosidoses. in *The Metabolic Basis of Inherited Disease* (eds. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D.) 1807-1839 (McGraw-Hill, New York, 1989).

450. Zimran, A., Gelbart, T., Westwood, B., Grabowski, G.A. & Beutler, E. High frequency of the Gaucher disease mutation at nucleotide 1226 among Ashkenazi Jews. *Am J Hum Genet* **49**, 855-9 (1991).

451. Sipila, K. & Aula, P. Database for the mutations of the Finnish disease heritage. *Hum Mutat* **19**, 16-22 (2002).

452. Olney, R.S. Newborn Screening for Sickle Cell Disease: Public Health Impact and Evaluation. in *Genetics and Public Health in the 21st Century: Using Genetic Information to Improve Health and Disease* (eds. Khoury, M.J., Burke, W. & Thomson, E.J.) 431-446 (Oxford University Press, New York, 2000).

453. WHO. Sickle-cell anaemia: Report by the Secretariat of the Fifty-ninth World Health Assembly A59/9. (World Health Organization, 2006).

454. Knott, L. & Bailey, A.J. Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance. *Bone* **22**, 181-7 (1998).

455. Eyre, D.R. & Wu, J. Collagen Cross-Links. in *Collagen: Primer in Structure, Processing and Assembly*, Vol. 247 (eds. Brinckmann, J., Notbohm, H. & Muller, P.K.) 207-229 (Springer Berlin Heidelberg, Berlin, 2005).

456. Ishikawa, Y., Wirz, J., Vranka, J.A., Nagata, K. & Bachinger, H.P. Biochemical characterization of the prolyl 3-hydroxylase 1/CRTAP/cyclophilin B complex. *J Biol Chem* **284**, 17641-47 (2009).

457. Chang, W., Barnes, A.M., Cabral, W.A., Bodurtha, J.N. & Marini, J.C. Prolyl 3-hydroxylase 1 and CRTAP are mutually stabilizing in the endoplasmic reticulum collagen prolyl 3-hydroxylation complex. *Hum Mol Genet* **19**, 223-34 (2009).

458. Gothel, S.F. & Marahiel, M.A. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* **55**, 423-36 (1999).

459. Stryke, D. *et al.* BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. *Nucleic Acids Res* **31**, 278-81 (2003).

460. McLeod, M.J. Differential staining of cartilage and bone in whole mouse fetuses by alcian blue and alizarin red S. *Teratology* **22**, 299-301 (1980).

461. Bakker, A.D. & Klein-Nulend, J. Osteoblast isolation from murine calvaria and long bones. *Methods Mol Biol* **816**, 19-29 (2012).

462. Makareeva, E. *et al.* Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta. *J Biol Chem* **283**, 4787-98 (2008).

463. Yamauchi, M. & Katz, E.P. The post-translational chemistry and molecular packing of mineralizing tendon collagens. *Connect Tissue Res* **29**, 81-98 (1993).

464. Yamauchi, M. & Shiiba, M. Lysine hydroxylation and cross-linking of collagen. *Methods Mol Biol* **446**, 95-108 (2008).

465. Eyre, D. Collagen cross-linking amino acids. *Methods Enzymol* **144**, 115-39 (1987).

466. Forlino, A. *et al.* Phenotypic comparison of an osteogenesis imperfecta type IV proband with a de novo alpha2(I) Gly922 --> Ser substitution in type I collagen and an unrelated patient with an identical mutation. *Biochem Mol Med* **62**, 26-35 (1997).

467. Cabral, W.A. *et al.* Type I collagen triplet duplication mutation in lethal osteogenesis imperfecta shifts register of alpha chains throughout the helix and disrupts incorporation of mutant helices into fibrils and extracellular matrix. *J Biol Chem* **278**, 10006-12 (2003).

468. Bateman, J.F. & Golub, S.B. Deposition and selective degradation of structurally-abnormal type I collagen in a collagen matrix produced by osteogenesis imperfecta fibroblasts in vitro. *Matrix Biol* **14**, 251-62 (1994).

469. Hildebrand, T. & Ruegsegger, P. Quantification of Bone Microarchitecture with the Structure Model Index. *Comput Methods Biomech Biomed Engin* **1**, 15-23 (1997).

470. Akhter, M.P., Lappe, J.M., Davies, K.M. & Recker, R.R. Transmenopausal changes in the trabecular bone structure. *Bone* **41**, 111-6 (2007).

471. Sricholpech, M. *et al.* Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. *J Biol Chem* **286**, 8846-56 (2011).

472. Barnes, A.M. *et al.* Kuskokwim Syndrome, a Recessive Congenital Contracture Disorder, Extends the Phenotype of FKBP10 Mutations. *Hum Mutat* (2013).

473. Schwarze, U. *et al.* Mutations in FKBP10, which result in Bruck syndrome and recessive forms of osteogenesis imperfecta, inhibit the hydroxylation of telopeptide lysines in bone collagen. *Hum Mol Genet* **22**, 1-17 (2013).

474. Baumann, M. *et al.* Mutations in FKBP14 cause a variant of Ehlers-Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss. *Am J Hum Genet* **90**, 201-16 (2012).

475. Eyre, D., Shao, P., Weis, M.A. & Steinmann, B. The kyphoscoliotic type of Ehlers-Danlos syndrome (type VI): differential effects on the hydroxylation of lysine in collagens I and II revealed by analysis of cross-linked telopeptides from urine. *Mol Genet Metab* **76**, 211-6 (2002).

476. Uzawa, K. *et al.* Lysine hydroxylation of collagen in a fibroblast cell culture system. *Biochem Biophys Res Commun* **305**, 484-7 (2003).

477. van der Slot, A.J. *et al.* Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *J Biol Chem* **278**, 40967-72 (2003).

478. Eyre, D.R. & Glimcher, M.J. Reducible crosslinks in hydroxylysine-deficient collagens of a heritable disorder of connective tissue. *Proc Natl Acad Sci U S A* **69**, 2594-8 (1972).

479. Pasquali, M. *et al.* Abnormal formation of collagen cross-links in skin fibroblasts cultured from patients with Ehlers-Danlos syndrome type VI. *Proc Assoc Am Physicians* **109**, 33-41 (1997).

480. Passoja, K., Rautavuoma, K., Ala-Kokko, L., Kosonen, T. & Kivirikko, K.I. Cloning and characterization of a third human lysyl hydroxylase isoform. *Proc Natl Acad Sci U S A* **95**, 10482-6 (1998).

481. Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A.H. & Myllyla, R. Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. *J Biol Chem* **279**, 37535-43 (2004).

482. Sipila, L. *et al.* Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. *J Biol Chem* **282**, 33381-8 (2007).

483. Vetter, U., Weis, M.A., Morike, M., Eanes, E.D. & Eyre, D.R. Collagen crosslinks and mineral crystallinity in bone of patients with osteogenesis imperfecta. *J Bone Miner Res* **8**, 133-7 (1993).

484. Bank, R.A. *et al.* Pyridinium cross-links in bone of patients with osteogenesis imperfecta: evidence of a normal intrafibrillar collagen packing. *J Bone Miner Res* **15**, 1330-6 (2000).

485. Fujii, K. & Tanzer, M.L. Osteogenesis imperfecta: biochemical studies of bone collagen. *Clin Orthop Relat Res*, 271-7 (1977).

486. Eyre, D.R. & Weis, M.A. Bone Collagen: New Clues to Its Mineralization Mechanism from Recessive Osteogenesis Imperfecta. *Calcif Tissue Int* **93**, 338-347 (2013).

487. Yamauchi, M., Noyes, C., Kuboki, Y. & Mechanic, G.L. Collagen structural microheterogeneity and a possible role for glycosylated hydroxylysine in type I collagen. *Proc Natl Acad Sci U S A* **79**, 7684-8 (1982).

488. Eyre, D.R. & Glimcher, M.J. Analysis of a crosslinked peptide from calf bone collagen: evidence that hydroxylysyl glycoside participates in the crosslink. *Biochem Biophys Res Commun* **52**, 663-71 (1973).

489. Amblard, D. *et al.* Lower bone cellular activities in male and female mature C3H/HeJ mice are associated with higher bone mass and different pyridinium crosslink profiles compared to C57BL/6J mice. *J Bone Miner Metab* **21**, 377-87 (2003).

490. Banse, X. *et al.* Cross-link profile of bone collagen correlates with structural organization of trabeculae. *Bone* **31**, 70-6 (2002).

491. Banse, X., Sims, T.J. & Bailey, A.J. Mechanical properties of adult vertebral cancellous bone: correlation with collagen intermolecular cross-links. *J Bone Miner Res* **17**, 1621-8 (2002).

492. Marini, J.C., Cabral, W.A. & Barnes, A.M. Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. *Cell Tissue Res* **339**, 59-70 (2009).

493. Mizuno, K., Peyton, D.H., Hayashi, T., Engel, J. & Bachinger, H.P. Effect of the -Gly-3(S)-hydroxyprolyl-4(R)-hydroxyprolyl- tripeptide unit on the stability of collagen model peptides. *Febs J* **275**, 5830-40 (2008).

494. Weis, M.A. *et al.* Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J Biol Chem* **285**, 2580-90 (2009).

495. Callanan, K., Evans, D. & Syron, M. Health Needs of Travellers: Attitudes and Perceptions of the Travelling Community in the Western Health Board. (Traveller Health Unit in Partnership with the Department of Public Health, 2002).

496. Fajolu, I. *et al.* Osteogenesis imperfecta in a set of Nigerian twins - A case report. *Int J Pediat Neonat* **14**(2012).

497. Adeyokunnu, A.A. Spectrum of bone dysplasias in African children: Ibadan Nigerian experience. in *Skeletal Dysplasias* (eds. Papadatos, C.J. & Bartsocas, C.S.) pp 427-439 (Alan R. Liss Inc., New York, 1982).

498. Pepin, M.G. *et al.* Allelic background of LEPRE1 mutations that cause recessive forms of osteogenesis imperfecta in different populations. *Mol Genet Genomic Med* **1**, 194-205 (2013).

499. Dalgleish, R. The Human Collagen Mutation Database 1998. *Nucleic Acids Res* **26**, 253-5 (1998).

500. Roschger, P. *et al.* Evidence that abnormal high bone mineralization in growing children with osteogenesis imperfecta is not associated with specific collagen mutations. *Calcif Tissue Int* **82**, 263-70 (2008).

501. Jansen, G. *et al.* An interaction map of endoplasmic reticulum chaperones and foldases. *Mol Cell Proteomics* **11**, 710-23 (2012).