Molecular mechanisms and clinical manifestations of rare genetic disorders associated with type I collagen

Yanqin Lu1,2*, Shie Zhang1,2, Yanzhou Wang3, Xiuzhi Ren4, Jinxiang Han1,2

1 Key Laboratory for Biotech-Drugs of National Health Commission, Key Laboratory for Rare & Uncommon Diseases of Shandong Province, Shandong Medicinal Biotechnology Centre, Shandong First Medical University & Shandong Academy of Medical Sciences, Ji‘nan, China; 2 School of Medicine and Life Sciences, University of Jinan-Shandong Academy of Medical Sciences, Ji‘nan, China; 3 Shandong Provincial Hospital Affiliated to Shandong First Medical University, Ji‘nan, China; 4 Department of Orthopaedic Surgery, The People’s Hospital of Wuqing District, Tianjin, China.

Summary
Type I collagen is an important structural protein of bone, skin, tendon, ligament and other connective tissues. It is initially synthesized as a precursor form, procollagen, consisting of two identical pro-α1(I) and one proα2(I) chains, encoded by COL1A1 and COL1A2, respectively. The N- and C- terminal propeptides of procollagen are cleavage by N-proteinase and C-proteinase correspondingly, to form the central triple helix structure with Gly-X-Y repeat units. Mutations of COL1A1 and COL1A2 genes are associated with osteogenesis imperfecta, some types of Ehlers-Danlos syndrome, Caffey diseases, and osteogenesis imperfect/Ehlers-Danlos syndrome overlapping diseases. Clinical symptoms caused by different variations can be variable or similar, mild to lethal, and vice versa. We reviewed the relationship between clinical manifestations and type I collagen – related rare genetic disorders and their possible molecular mechanisms for different mutations and disorders.

Keywords: Type I collagen, biosynthesis, osteogenesis imperfecta, Ehlers-Danlos syndrome, Caffey disease, N- and C- propeptide, mutation

1. Introduction
Type I collagen, as an important structural protein, is classified as one of the major fibrillar collagens (1). It is the most common collagen found in the human body and other vertebrates. Type I collagen is distributed widely, including bones, ligaments, tendons, cartilage, skin, liver, heart valves, cornea, lungs and other connective tissues (2,3). It is synthesized as procollagen precursor, which predominately consists of two identical procollagen precursor chains, encoded by COL1A1 and COL1A2 gene, respectively. Three α-chains fold into a 300 nm-long triple helix with short nonhelical terminal peptides-telopeptides. The triple helical region of each α-chain has the obligatory Gly-X-Y repeats with glycine in every third position. The amino acids in X and Y position of collagen are often proline and hydroxyproline, respectively (4,5).

The heterotrimer is the dominant form of type I collagen. Homotrimers of three α1(I) chains are identified in fetal tissues, tumors and some fibrotic lesions and this isoform is more resistant to cleavage by collagenases than heterotrimers (6-9). In a recessively homozygous oim mice with G deletion at nucleotide 3983 of COL1A2 gene, and recessively cardiac valvular type Ehlers-Danlos syndrome (cvEDS), collagen molecules exist as homotrimers of three α1(I) in the extracellular matrix and proα2(I) chains are not incorporated into procollagen molecules (10-12). Mutations in COL1A1 and COL1A2 genes cause Caffey disease (OMIM 114000) (13), the arthrochalasia type Ehlers-Danlos syndrome (aEDS) (OMIM 130060, 617821) (14-16), cvEDS is an autosomal recessive genetic form (OMIM 225320) (12) and autosomal dominant osteogenesis imperfecta (OI) (OMIM 166200, 166210, 259420, 166220), postmenopausal osteoporosis (OMIM 166710) (17). Symptomatic patients with different or similar clinical phenotypes are described

*Address correspondence to:
Dr. Yanqin Lu, Shandong First Medical University & Shandong Academy of Medical Sciences, 18877 Jingshi Road, Ji‘nan 250062, China.
E-mail: yanqinlu@yahoo.com
in different disorders and mutations. Subclinical or asymptomatic individuals with these inherited diseases may express an overlap syndrome, or different penetrance and expressivity in patients with the same pathogenic mutations.

This review summarizes the biosynthesis process of type I collagen and investigates the correlationship between phenotype and genotype and mechanisms of rare genetic disorders related to \textit{COL1A1} and \textit{COL1A2}.

2. Biosynthesis of type I collagen

The process of type I collagen biosynthesis undergoes multiple intracellular and extracellular steps (Figure 1) (18). First, procollagen is synthesized on the ribosome and transported into the rough endoplasmic reticulum (rER), where they undergo a series of post-translational modifications. Hydroxylation of prolyl and lysyl residues are catalyzed by prolyl 4/3-hydroxylases (P4H, P3H) and lysyl hydroxylases (LH), respectively (19). O-linked glycosylation of hydroxylysines is followed by the attachment of galactose and then glucose to hydroxylysine residues. These modifications are processed simultaneously with the synthesis and folding of procollagen chains and occur only within the unfolded region of the chains (20-22). Unlike most proteins, which begin to fold stepwise from N-terminal end to C-terminal end while the rest of the chain is still being synthesized.

The folding process of procollagen initiates from association of two pro\alpha1(I) and one pro\alpha2(I) peptide chain within C-propeptide region, then stepwise to N-terminal end of the chain. C-propeptide folds in a zipper-like manner only after chain synthesis is completed (21). Chain recognition sequences (CRSs) of the C-propeptide is involved in the process of folding. It interacts with the CRSs of a neighboring chain during intracellular trimerization of procollagen to ensure correct homotrimeric and heterotrimeric chain stoichiometry (23). CRSs is variable and responsible for selective assembly of different types of collagen (23,24). Meanwhile, protein disulfide isomerase (PDI) catalyzes the formation and rearrangement of intra- and inter-molecular disulfide bonds (25). There are many other proteins involved in the procollagen assembly. ER chaperones like glucose regulated proteins GRP78 (BiP) and GRP94 (GP96) bind to unfolded and partially folded C-propeptide. Misfolding procollagen leads to sequestration of BiP and GRP94, activates conventional unfolded protein response (UPR) signaling, and leads to ER associated degradation (ERAD) of misfolded protein by proteasomes (26-29). Heat shock protein 47 (HSP47),
as a collagen-specific molecular chaperone, makes the triplex-helix folding favorable (30). It regulates LH2 activity through association to peptidyl-prolyl cis-trans isomerase (FKBP65) and LH2 to form HSP47-FKBP65-LH2-BiP complex. FKBP65, interacting with HSP47, functions to balance the lysyl hydroxylation. BiP increases affinity of the lysyl hydroxylation complex (31).

After assembly, procollagen is transported into Golgi through endoplasmic reticulum to Golgi intermediate compartment (ERGIC), or vesicular tubular clusters (VRC) in a coat protein complex II (COPII)-dependent process. ER membrane protein TANGO1, acts as the receptor for large cargoes- procollagen and organizer of ER exit sites. It is retained in the ER and exported out of the ER in large COPII –coated carriers, and retrieved back to the ER by retrograde coat, COPI. TANGO1 interacts with the C-terminal RDEL retrieval sequence of HSP47 and is exported from ER to Golgi through interacting proteins cTAGE5 and SEC12 (32-34).

In Golgi, N-propeptide and C-propeptide cleavage have been identified in the secretory pathway inside the cell (35). Nevertheless, a predominant fraction of procollagen is secreted with intact N- and C-propeptides into extracellular matrix (36). Then, N-terminal and C-terminal globular regions of propeptides are removed by a disintegrin and a metalloproteinase with thrombospondin repeats (ADAMTS) and bone morphogenetic protein 1 (BMP1)/Tolloid-like proteinases from procollagen to generate triple-helix fibril monomers named tropocollagen (37). Finally, monomers assemble into macromolecular fibers by covalent crosslinks between certain lysine and hydroxylysine residues (19,38). Fiber assembly is tightly regulated \textit{in vivo} by a variety of molecules. Disruption in collagen fiber formation, structure and function by mutations leads to different bone pathology.

3. Rare genetic disorders caused by \textit{COL1A1} and \textit{COL1A2} genes

3.1. Osteogenesis imperfecta

3.1.1. Phenotypic and genotypic characteristics of OI

Osteogenesis imperfecta, also known as brittle bone disease, is a genetic connective tissue disorder with genetic and phenotypic diversity (39,40). Skeletal features for OI are an increased susceptibility to bone fractures and reduced bone density, short stature, scoliosis and skull deformities. Extraskeletal characteristics include hearing and sight impairment, dentinogenesis imperfecta, blue sclera, involvement of connective tissue, and cardiovascular and neurological abnormalities to some extent (Figure 2) (41,42). OI has a broad clinical spectrum, ranging from mild to lethal forms. According to Silence classification system, OI is divided into four types. Type I is the mildest form, patients with type I OI usually have family history, blue sclera, hearing impairment, near normal stature and
rarely have dentinogenesis imperfecta. Type II OI is perinatal lethal, clinical characteristics include defective cranial ossification and severe skeletal deformity due to intraterine rib, long bone fractures, and respiratory failure. Type III OI is the most severe form, affected individuals have triangular faces, frontal bossing, coxa vara, basilar invagination, early onset scoliosis, short stature, multiple fractures and long bone deformities (43,44). Type IV is the moderately severe OI form, with overlapping phenotypes ranging from type I to type III. Generally, affected individuals experience broken fracture, mild short stature, dentinogenesis imperfecta, hearing loss, sclerae hue. Nearly 85% OI patients belong to type I to type IV, which is induced by COL1A1 and COL1A2 genes encoding proα1(I) and proα2(I) of type I procollagen in an autosomal-dominant inherited form (45,46). Type VI to XIX are predominated by an autosomal-recessive form, and the number of pathogenic genes are multiple, while the total percentage for OI is no more than 20%, the pathogenic genes for autosomal-recessive and X-linked recessive forms includes CRTAP, LEPRE1, PPIB, BMP1, FKBP10, SERPINH1, SERINF1, WNT1, SP7, SERPINF1, TMEM38B, SPARC, MBTPS2 and TENT5A (39,47-49).

3.1.2. COL1A1 and COL1A2 gene mutations of OI

Generally, there’s a relationship between genotype and phenotype for dominant OI caused by type I collagen genes. The molecular defect for type I OI is a null COL1A1 allele due to frame-shifts or premature termination codons (PTCs), resulting in a decrease in the amount of structurally normal type I collagen. Splice site mutation can lead to mis-splicing with subsequent PTCs (50).

Substitutions and deletions in the triple helix region of type I collagen is the main mutation pattern for type II to type IV OI, predominated by substitutions of obligatory Gly-X-Y units. Substitutions of residues in this unit make procollagen folding more difficult, for these mismatches will further destabilize the triple helix. Triple-helix destabilization promotes procollagen misfolding, leads to retention of misfolded procollagen in the ER and HSP47 is unregulated (51). ER stress mediated by these mutations affect these cells producing type I collagen to some extent, resulting in structural and qualitative abnormal type I collagen (52-54). Glycine to serine substitution is most prevalent in all kinds of glycine changes. Meanwhile, Glycine to serine or alanine substitutions cause less damage than those of arginine, glutamic acid, aspartic acid or valine. Triple-helix folding of the later is much delayed. Glycine substitutions located in α1(I) tend to be more severe than the α2(I) chain (55,56). Glycine substitutions at C-terminal of collagen chains are more severe than that of N-terminal (57).

COL1A1 or COL1A2 mutations of OI are predominated by variations in the triple-helix region. OI Mutations in N- and C-propeptide or nearby are rare and commonly patients with these terminal mutations have overlapping clinical characteristics with other syndromes. Mutations in the C-propeptide cleavage

The author confirm this is right with high bone mineral density (58,59) and moderate to severe forms of OI are identified (60). For mutations in N-propeptide region, uncleaved N-propeptides will be enrolled into collagen fiber, and hence alter the fiber size, structure and strength (61,62). For individuals who have mutations at these terminal regions of type I collagen, possible symptoms with other syndrome should be considered.

3.2. Type I collagen gene related EDS

3.2.1. Cardiac-valvular type Ehlers-Danlos syndrome

Ehlers-Danlos syndrome is a group of connective tissue disorders with phenotypic and genotypic variability. The modified classification based on Villefranche recognizes 13 different subtypes, most of which are linked to mutations in one of the genes encoding collagen proteins, enzymes for posttranslational modifications, myomatrix proteins and glycosaminoglycans (63,64). Type I collagen-related EDS are rare, with no more than 50 patients in total (65,66). Cardiac-valvular type EDS (cvEDS) is characterized by cardiac valvular defects, joint hypermobility (generalized or restricted to small joints) and skin hyperextensibility, with atrophic scars, thin skin, and easy bruising. Minor criteria for diagnosis include inguinal hernia, pectus deformities, joint dislocations, and foot deformities (63,67-69). The absence of pro2(I), and α2(I) are observed in all reported seven cvEDS patients (10,12,70,71). Biallelic COL1A2 mutations lead to nonsense-mediated mRNA decay and loss of function was suggested to be the molecular mechanism of cvEDS (10,71). Homozygous pro2(I) deficiency was reported also in type III OI patient, with recurrent fractures including rib fractures, popcorn deformities of the knee, wormian bone, generalized osteopenia, short stature, and hypermobile fingers. Collagen fibrils were mostly normal except for focal areas which were disorganized in size and shape (72). Patients with homozygous mutation of COL1A2 at intron 46 produced non-functional pro2(I), presenting EDS/OI phenotypes including generalized joint hypermobility and foot deformities, pale blue sclerae and a mild increase in bone fragility (73). In all, deficiency of pro2(I) leads to different diseases and phenotypes and its mechanisms need to be further studied.

3.2.2. The arthrochalasia type Ehlers-Danlos syndrome

The arthrochalasia type of EDS, formerly known as VIIa and VIIb, with pathogenic genes of COL1A1 and
COL1A2, respectively. It is different from other types of EDS by major diagnostic criteria of typical congenital hip dislocation, generalized joint laxity with multiple dislocations/subluxations, and skin hyperextensibility. Minor criteria for diagnosis include tissue fragility, easy bruising, muscle hypotonia, kyphoscoliosis and radiologically mild osteopenia (63,68,74,75).

Generally, COL1A1 and COL1A2 gene mutations for aEDS patients are clustered around the exon/intron 5 and 6 region. Heterozygous in frame deletion of N-telopeptide of either proα1(I) or proα2(I) are a common genetic basis for aEDS. Splice site mutations of intron 5 and 6 of COL1A1 or COL1A2 gene lead to whole or partial skipping of exon 6 (76,77). N-telopeptide links N-propeptide to the triple-helix domain. It contains cleavage site for procollagen-N-proteinase and cross-linking lysyl residue. Thus, deletion of N-telopeptide results in incomplete cleavage of procollagen and defective cross-linking. N-propeptide is incorporated within fibrils to form accumulated pNcollagen (77,78).

Until now, over 30 aEDS have been reported and most mutations are caused by COL1A2 mutation (14,65,66,74,77-87). It seems that aEDS caused by COL1A1 have more severe phenotypes than COL1A2 gene. Collagen fibrils in aEDS patients are more loosely and irregular organized, vary widely in diameter and present with ragged cross-sections in outline. This is extremely severe in patients with EDS VIIa compared to EDS VIIb patients (75).

Patient with exon 5 and exon 6 deletion of COL1A1 presented with bilateral dislocation of the hips, extreme joint hypermobility, femoral fracture and normal skin at birth (81). The 3’ acceptor splice site COL1A1 IVS5 -1G>A results in alternative splicing of exon 6. Patients with this mutation have overlapping phenotypes resembling OI, including multiple fractures, wormian bones, generalized dentinogenesis imperfecta, severe kyphoscoliosis, and relative short stature. The skin was generally lax, redundant, and hyperextensible, but without scarring or bruising (81). Mutation of c.543G>A (p.Met181Ile) in the last nucleotide of exon 6 induced abnormal pre-mRNA splicing without exon 6 (13). One of the probands with this mutation presented with typical aEDS symptoms and a fracture history (86). Patients with EDS VIIb are more prevalent than VIIa type and fractures are rarely documented (88). Clinically, there is some overlap between OI and EDS type VII, especially for EDS type VIIa as mentioned above. In reverse, hypermobility is also seen in OI patients (88), overlapping with aEDS.

3.2.3. OI/EDS

Patients with OI/EDS combined clinical symptoms of OI as well as EDS. OI/EDS mutations are restricted to the region of N-terminal and C-terminal of type I collagen. The N-terminal of type I collagen triple helix region, corresponds to exons 7-14 (89-94). The first 85 amino acids at N-terminal of each chain of type I collagen helix are called the N-anchor, which is vital for proper folding and stability of N-terminal end of the triple helix. Mutations at this region lead to incomplete or delayed N-propeptide cleavage and association of pNcollagen with fibrils (89). OI/EDS patients with this type of mutation usually presented mainly by generalized joint hyperlaxity and skin hyperextensibility, early progressive scoliosis as well as mild to lethal OI symptoms including relatively short stature, blue sclerae, infrequent or frequent bone fracture and different levels of osteopenia (Figure 3) (89-91,93). We found ptosis

![Figure 3. Clinical symptoms (A-D) and radiographic characteristics (E-G) for OI/EDS.](www.irdrjournal.com)
in one individual with family history existed in OI/EDS with the mutation of c.3521C>T (Ala1174Val) in COL1A1; this symptom is also displayed in type XV OI caused by the Wnt1 gene (95-97).

Splice site mutations lead to partial or whole exon skipping similar to aEDS, with atypical or mild form of OI. Mosaic deletion of 9 nucleotides from 3150 to 3158 in the coding region of exon 44 results in in frame deletion of three amino acids was identified in a patient’s cultured fibroblasts. The patient with this mutation has mild clinical EDS and OI symptoms (98). Exon duplication with the addition of 477 amino acids (p.Gly181_Lys657dup) to the triple helical domain of proa2(I) chain leads to relatively mild OI/EDS symptoms (91). Patients with missense mutations of c.563G>A (p.Gly188Asp), c.607G>T (p.Gly203Cys) in COL1A1 and c.326G>A (p.Gly109Asp) in COL1A2 all showed cardiovascular problems, which was not observed in patients with mutations of c.587G>T (p.Gly196Val) in COL1A1 (90). For c.671G>A (p.Gly224Asp) mutation of COL1A1, patients have severe type III OI and moderate EDS, collagen fibrils are organized irregularly, with decreased fibril density and decreased fibril diameter (93).

OI/EDS patients with mutation in C-terminal of type I collagen genes (exon 37 to 51) have broad clinical severity. Substitution of Ala1174Val, Arg1036Cys and Met 1264 Val presents milder OI signs (95,99). Met1264Val substitution located at the proa1(I) of C-propeptide is supposed to impede C-propeptide folding and chain association (99). Also, OI patients in C-terminal of type I collagen genes have various phenotypes from mild to lethal forms.

One special patient with OI/EDS was identified owing to c.4006-1G > A mutation in COL1A1 and biallelic missense variants in TNXB (p.Val1213Ile, p.Gly2592Ser), he presented with severe muscular hypotonia, multiple fractures, and joint hypermobility (100). Homozygous mutation of splice site IVS 46 +2T>C caused non-functional COL1A2 alleles, patient with this mutation showed clinical features of generalized joint hypermobility, foot deformities, muscle hypotonia, blue sclerae and fracture history (73).

3.2.4. Caffey disease

Caffey disease, also known as infantile cortical hyperostosis (ICH), is caused by COL1A1 gene mutation of c.3040C>T and leads to an arginine to cysteine substitution at position 836. It rarely appears after 5 months of age and generally is restored spontaneously by 2 years old. Cortical thickening (hyperostosis) and subperiosteal new bone formation are the main findings of radiography. Tender swelling of long bone and inflammatory reactions often accompany the illness (101). Short stature and persistent bony deformities in 5 individuals from a three-generation Thai family are reported, which expands phenotypes of Caffey disease (102). Lethal or severe prenatal Caffey disease, complied with polyhydramnios, long bone deformity, and rib fractures, which could be misdiagnosed as lethal osteogenesis imperfecta (103-105). Two Caffey disease patients were reported to have overlapping phenotypes of joint hyperlaxity, hyperextensible skin and inguinal hernias resembling type III EDS (13).

Molecular mechanisms for Caffey disease are still unclear. Mutation of c.3040C>T locates in the sequence of interleukin-2 binding site and the amyloid protein precursor (APP), the gap region of major ligand binding regions (106-108). Arg836Cys substitution leads to increased disulfide crosslinking within or between collagen fibrils. Hence, widened and less densely packed collagen fibrils was observed from cultured skin fibroblast from Caffey disease (13). Meanwhile, thermal stability of collagen is reduced when an arginine residue was replaced in the X position of classic Gly-X-Y repeating pattern of triple helix collagen (109). It is also proposed that this typical mutation may regulate COX2/PGE axis mediated by unfolded protein response (110).

4. Conclusion

In conclusion, genetic and phenotypic overlap of OI, aEDS, cvEDS and Caffey diseases caused by type I collagen genes are outlined. Position preference mutations in COL1A1 and COL1A2 genes are highly related with typical phenotypes and diseases with different molecular mechanisms. Mutations of aEDS are limited to intron 5 and 6 of type I collagen genes. OI/EDS mutations locate predominately in the C-terminal and N-terminal region of exon 7 to exon 14. Considering the overlap mutation distribution of OI and EDS, it is a challenge to define disease with atypical phenotypes and atypical mutation distribution in type I collagen genes. Hence, clinical symptoms in detail for these diseases are extremely valuable. Meanwhile, to further analyze the heterozygosity in diseases like Caffey disease, which have a p.Gly188Asp mutation, as well as other diseases presenting clinical manifestations with the same pathogenic gene mutations, penetrance, expressivity and epigenetic regulation inheritance should be considered and combined with phenotype/genotypes of rare genetic syndromes. In all, genotypic and phenotypic relationships and molecular mechanisms for type I collagen- related inherited diseases will be further enriched and revealed as the number of rare diseases cases increases. Novel methods for classification will be useful for identification of rare diseases with atypical symptoms.

Acknowledgements

This project was supported by Grants-in-Aid from...

References

79. Nicholls AC, Oliver J, Renouf DV, Heath DA, Pope


(Received April 8, 2019; Revised May 24, 2019; Accepted May 30, 2019)