Eight mutations including 5 novel ones in the COL1A1 gene in Czech patients with osteogenesis imperfecta
Lucie Hruskova, Igor Fijalkowski, Wim Van Hul, Ivo Marik, Geert Mortier, Pavel Martasek, Ivan Mazura

Background and Aim. Osteogenesis imperfecta (OI), also called brittle bone disease, is a clinically and genetically heterogeneous disorder characterized by decreased bone density. Autosomal dominant forms result from mutations in either the COL1A1 (collagen type I alpha-1 chain) or COL1A2 (collagen type I alpha-2 chain) genes encoding the type I collagen. The aim of this study was to identify mutations and allelic variants of the COL1A1 gene in patients with osteogenesis imperfecta (OI).

Methods and Results. Molecular genetic analysis of the COL1A1 gene was performed in a cohort of 34 patients with OI. The DNA samples were analysed by PCR and Sanger sequencing. DNA changes in coding sequences of the gene were compared with Type 1 Collagen Mutation Database. Genetic variants resulting in either quantitatively or structurally defective protein production were found in 6 unrelated patients. Four identified mutations are connected to decreased protein production (Tyr47X, Arg131X, Arg415X, Gln1341X), 2 result in amino acid substitution (Cys61Phe, Pro1186Ala) and the last affects splicing (c.1057-1G>T). Further, one silent mutation (Gly794Gly) was detected. No protein analysis was performed.

Conclusion. Of the 8 identified mutations, 5 were novel and have not been reported before. Only one causes substitution of glycine located within the Gly-X-Y triplets in the triple helical domain. Two mutations are located in major ligand binding regions (MLBR) which are important for bone strength and flexibility. Although the genotype-phenotype correlation is still unclear, our findings should contribute to elucidating this relationship in patients diagnosed with OI.

Key words: collagen type I, COL1A1, mutations, osteogenesis imperfecta

INTRODUCTION

Osteogenesis imperfecta (OI) is a disorder of human connective tissue characterized by low bone mass and decreased bone material strength resulting in a higher risk of fractures. Affected patients show variable clinical signs such as blue sclera, hearing loss, joint hypermobility, skin hyperlaxity, normal to low stature and dentinogenesis imperfecta (DI). Skeletal radiographs reveal osteopenia and pathological fractures of long bones and/or vertebrae. Based on clinical and genetic features fourteen types of the disease (type I to XIV) have been delineated to date. Depending on the genetic basis, two forms of OI – collagenous (types I to IV) and non-collagenous (types V to XIV), are described1. About 90% of OI cases are due to dominant mutations in either the COL1A1 (collagen type I alpha-1 chain) gene or COL1A2 (collagen type I alpha-2 chain) gene both of which encode the type I collagen protein. These mutations result in either reduced production (haploinsufficiency) of the protein or in the synthesis of structurally abnormal collagen12. Haploinsufficiency results from mutations creating stop codons (premature termination codon) indicating transcription termination of affected gene. Such shortened alpha chains miss C-propeptide domain which is important for alpha chains assembling into triple helix and are destroyed by a process called nonsense-mediated decay. This process eliminates mutant alpha chains and leads to the synthesis of reduced amount of normal protein. Decreased production (haploinsufficiency) of collagen type I is typical for OI type I (ref.1,3,4). Patients affected by OI type I (Dominantly inherited OI with blue sclerae) are of normal stature and have blue sclera (probability of blue sclerae occurrence is near to 100% (ref.1)). Bone fractures occur especially during childhood and moderate bone deformities, especially of the lower limbs and spine, could be observed. Other clinical sign observed in OI type I individuals is hearing impairment which usually results in hearing loss5. Based on presence of dental abnormalities, type I is divided into A and B subtypes (with and without dentinogenesis imperfecta, respectively) (ref.2,6). In general, patients affected with OI type IA are of normal stature and have lower incidence of fractures whereas OI type IB patients have higher frequency of fractures and more skeletal deformities resulting in shorter stature5,7. Synthesis of abnormal collagen type I is described in OI types II, III and IV. Mutations affecting the structure of alpha chains include amino acids substitutions, exon-skip-
ping mutations and more complex gene rearrangements. OI type II (Lethal perinatal OI with radiographically crumpled femora and beaded ribs) is a lethal type of OI with high percentage of perinatal mortality. In rare cases, patients survive few days. Typical clinical signs include multiple intrauterine fractures, severe skeletal deformities, blue sclera, low birthweight and low birth crown-heel length. The third type of OI (Progressively deforming OI with normal sclera) is the most severe form characterised by progressive deformities, high frequency of fractures, short stature due to severe skeletal deformities, triangular face, hearing impairment, basilar impression and dentinogenesis imperfecta. In some cases, infants die due to cardiac or pulmonary insufficiency. The last type of OI with origin in defective collagen type I production, OI type IV (Dominantly inherited OI with normal sclerae), is the most heterogeneous form of the disease. Individuals exhibit mild to moderate phenotype and are of short stature. First fractures occur at birth, moderate deformities (especially of long bones of the lower and upper limbs) are presented. Teeth and hearing impairment could be presented. Like the first OI type, subtypes A and B (with and without DI, respectively) are distinguished. In the remaining 10% of cases, OI is the result of recessive mutations in IFITM5 (Interferon-induced transmembrane protein 5), OI type V; bone formation co-participant (ref.10), SERPINF1 (Serpin peptidase inhibitor, clade F, member 1), OI type VI; bone mineralization participation (ref.11), CRTAP (Cartilage associated protein, OI type VII; 3-hydroxylation complex component) (ref.12), P3H1 (Prolyl 3-Hydroxylase 1, OI type VIII; 3-hydroxylation complex component) (ref.12,13), PP1B (Peptidyl-prolyl isomerase 1 (Cyclophilin B), OI type IX; 3-hydroxylation complex component) (ref.13). SERPINH1 (Serpin peptidase inhibitor, clade H, member 1), OI type X; collagen type I chaperone (ref.12), FKBP10 (FK506-binding protein 10, OI type XI; collagen type I chaperone) (ref.13), SP7 (Osteoblast-specific transcription factor 7, OI type XII; cell differentiation co-participant) (ref.14), BMP1 (Bone morphogenetic protein 1, OI type XIII; embryogenesis and skeletogenesis process co-participant) (ref.13). Recently, mutations in the WNT1 gene (Wingless-type MMTV integration site family, member 1, bone mass regulator) have been described in patients diagnosed with OI type IV (ref.16).

The incidence of OI ranges from 1:25000-40000 (non-lethal forms) to 1:50000 (lethal forms) live births in the world. Currently, no data are available on OI prevalence in the Czech population.

The aim of this study is identification of mutations and allelic variants of COL1A1 gene. Type I collagen is a heterotrimer, composed of two alpha-1(I) and one alpha-2(I) chain encoded by COL1A1 and COL1A2 respectively. Alpha-1(I) chain represents 2/3 of this heterotrimer. Generally, in the random selection of abnormal chains 75% of mutant heterotrimers contain one or more defective alpha-1(I) chains. Further, possible genotype-phenotype relationship of identified mutations will be discussed. Determination of possible effect of identified mutations on clinical picture of affected patients could be helpful for providing well-timed and helpful treatment of affected individuals.

**MATERIALS AND METHODS**

**Material**

Included in the study were 34 patients with a diagnosis of OI. There were 11 males and 23 females aged 7 to 57 years. The diagnosis was based on evaluation of clinical, radiological and biochemical features. In 19 patients, a diagnosis of OI type IA was made, 6 individuals were affected by type III, in 4 cases OI type IVA was identified and 5 patients suffered from type IVB. No case of OI type II was included in the study.

This study was performed in accordance with principles of the Declaration of Helsinki and approved by the Ethics Committee of General University Hospital in Prague (project 83/14). Participants provided their written informed consent with involvement in the study. The purpose and procedures of the research were explained, rights to refrain from the study under any circumstances and for any reason were emphasized. Informed consent was ensured by the Ambulant Centre for Defects of Locomotor Apparatus in Prague and was carried out according to “Informed approbation in medical genetics” by the committee of the Society of Medical Genetics (SMG) Czech medical Association of J. E. Purkyne.

**Laboratory methods**

Genetic analysis was restricted to the coding (and flanking non-coding) sequences (51 exons) of the COL1A1 gene whose product (alpha-1(I) chain) is two copies in the collagen type I (ref.18). Genomic DNA was extracted from peripheral blood leukocytes using a standard extraction method (QIaAmP DNA Blood Midi Kit, QiaGeN), followed by high resolution melting analysis (HRM), polymerase chain reaction (PCR) amplification and Sanger sequencing. PCR amplification was performed using 47 sets of forward and reverse primers located in sequences of flanking introns (primer sequences are available on request). PCR products were confirmed by horizontal gel electrophoresis. Sequences of these PCR products were obtained on the ABI PRISM 3130xl Genetic Analyzer automated sequencer (Applied Biosystems, Foster City, CA, USA) using the BigDye terminator cycle sequencing chemistry. The data were compared to the wild-type sequence as submitted to Ensembl accession no. ENST00000225964. Nucleotides were numbered from the first base of the start codon (ATG) of the cDNA reference sequence according to GenBank NM_000088.3. Amino-acid residues were numbered from the start codon for methionine (ATG) of the alpha-1 chain of type I collagen (reference sequence GenBank NM_000079.2).

Novel mutations were identified by their absence from the Osteogenesis Imperfecta Variant Database, collagen, type I, alpha 1 (COL1A1) (https://oi.cene.le.ac.uk/home.php?select_db=COL1A1) and the Osteogenesis...
RESULTS

We identified COL1A1 mutations of coding and non-coding sequences in 7/34 patients. None of these mutations were found in the DNA of a healthy control - Caucasian population. Four of the mutations result in premature stop codon (c.141C>A p.Tyr47X, c.391C>T p.Arg131X, c.1243C>T p.Arg415X, c.4021C>T p.Gln1341X), two are missense mutations (c.182G>T p.Cys61Phe, c.182G>T p.Pro1186Ala), one is located in intronic sequence (c.1057-1G>T) and the last one is silent mutation of glycine at position 794 of the COL1A1 gene (Table 1, Fig. 1).

The first patient (OI17) in this study was a 23-year-old woman affected by OI type IA. She is of subnormal height and suffers from joint hypermobility. In this case, we identified the heterozygous p.Tyr47X mutation in COL1A1.

The second case (OI23) was a 45-year-old man of unaffected parents diagnosed with OI type III. He had had multiple fractures since childhood, especially of the
lower limbs, and is confined to a wheelchair. He has a barrel-shaped chest, severe bone deformities and muscle weakness in the lower limbs. We identified homozygous substitution of cysteine by phenylalanine at position p.61 in the DNA sample of this individual. This mutation from the program Alamut was predicted as disease causing \((P = 1.0)\).

The third case (OI42) was a 24-year-old woman suffering from OI type IA. She is of subnormal height and has a mild barrel-shaped chest. DNA analysis identified a premature stop codon at residue p.131 coding for arginine. She has four other affected relatives; unfortunately their DNA samples were not available.

The next case was a 22-year-old young man (OI14) suffering from OI type IA. He shows blue sclerae, has a relatively short barrel-shaped chest and mild deformation of the lower limbs and vertebrae. In this patient, we identified a heterozygous p.Arg415X mutation in exon 19. Even though it is represented in a sequence database (rs72648326), it is disease causing due to production of shortened alpha chains with loss of heterotrimmer formation function. Further, a substitution of proline by alanine at position p.1186 was also identified in heterozygous state. This mutation, using the program Alamut, was predicted as disease causing \((P = 0.998)\).

The patient with an identified mutation in the coding sequence of the COL1A1 gene, p.Gly1341X in heterozygous state, was an 18-year-old boy (OI2) affected by OI IA. The boy is of subnormal height. He has light blue sclerae, barrel-shaped chest and muscle hypotrophy of the limbs. He had suffered from multiple fractures of the spine and both upper and lower limbs.

The heterozygous change c.1057-1G>T was identified in the DNA of a 50-year-old man (OI21) with clinical phenotype corresponding with type IA/IVA of the disease. He suffered from multiple fractures of the limbs since childhood. He has blue sclerae, deformed bones and suffers from hearing loss.

Further, genetic analyses of the 34 patients identified 38 polymorphisms in the COL1A1 gene. This also includes a silent heterozygous variant of glycine at position p.794 of the COL1A1 gene identified in the case of a 24-year-old woman (OI30) diagnosed with OI IVA. She suffered multiple fractures of the lower limbs until the age of 18. The patient is of subnormal height, has deformed lower limb bones and weak muscles. She also showed joint hypermobility. The sclera had been light blue after the birth but had whitened during childhood. This patient was treated with bisphosphonates until the age of 21. Nine of 38 identified polymorphisms were situated in coding sequence, 7 were localised in MLBRs and one occurred in 3' untranslated region of the gene. Allele frequencies of polymorphisms identified in OI patients correlate with Population genetics available on Ensembl database (http://www.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;ENSGL00000108821;r=17:50183289-50201632;t=ENST00000225964).

**DISCUSSION**

All of the identified mutations resulting in decreased production of collagen type I (c.141C>A p.Tyr47X, c.391C>T p.Arg131X, c.1243G>T p.Arg415X, c.4021C>T p.Gln1341X) were found in patients affected by OI type I. Except one case (p.Arg131X, OI42), it was the first occurrence of the disease in the family history. Furthermore, in all cases we were able to observe one common clinical sign - subnormal body height. Only the Arg131X had been described in the case of Italian patients with a positive OI family history (the mutant allele was also detected in the mother of the individual). Unfortunately, the clinical data on this individual are not available.

The nucleotide change c.182G>T results in substitution p.Cys61Phe in the COL1A1 gene. The identified mutation is in a homozygous state and is localized on the second exon of the COL1A1 gene which is a part of the N-terminal propeptide domain of the pro-alpha-1(I) collagen chain encoded by the first five exons and by a part of the sixth exon of the gene. The integrity of N-propeptide is important for proper chain association and for subsequent folding of individual alpha chains into a collagen trimer with triple helical configuration (the importance of structure and folding of triple helix is also true for C-propeptide). In this process, cysteine residues play a major role through formation of intra- and interchain disulphide bonds. During proteosynthesis, the N-terminal propeptide is removed from the pro-alpha-1(I) chain by specific proteinase to form the alpha-1(I) chain terminated by N-telopeptide. Because the substitution is not localized close to the N-proteinase cleavage site, its occurrence is not likely to prevent N-propeptide removal. The substitution is also localized in the von Willebrand factor type C (vWF-C) binding region defined by amino acid residues 38-96 (ref.21). Although the mutation could delay vWF-C binding, it is unknown whether or how it affects collagen type I processing. This mutation was first described by Zhang et al. (2012) in the case of a 17-year-old Chinese boy affected by the OI type I (first occurrence of the disease in the family, allelic variant not described). Clinical features described in this case were blue sclera and multiple fractures (more then 10) (ref.19). We found this change in a homozygous state in the patient diagnosed with the progressively deforming form of OI (type III). Individuals have soft, fragile, deformed bones and may suffer from neurologic and respiratory complications. The other clinical picture includes severe scoliosis, blue colour of sclera, hearing loss and dentinogenesis imperfecta. The only clinical sign described in both cases with identified p.Cys61Phe substitution was high frequency of fractures.

Substitution of proline by alanine at position p.1186 of the COL1A1 protein is described in a case identified as OI type IA. The mutation is localized in exon 48 which is part of the third important ligand binding region, called the Major Ligand Binding Region (MLBR). There are three such regions (MLBR1-3) defined by codons 81-200 (MLBR1), 682-830 (MLBR2) and 920-1040 (MLBR3) (codon numbers start at the first amino acid residue of the
helical domain of alpha chains (ref.18,23). These regions produce intermolecular binding of collagen type I with other connective tissue proteins like integrins, COMP (cartilage oligomeric matrix protein), serpinH1, phosphophoryn and others. These interactions result in increasing bone strength and elasticity14. The MLBR3 region is encoded by a part of exon 49, the entire exons 47 and 48 and part of the exon 49 sequence23. Concretely, this region including Pro1186 of the collagen type I molecule, binds amyloid beta A4 precursor protein (APP) that mediates cell-cell interactions15. It is also suggested that amyloid precursor protein is implicated in the regulation of synapse formation, neural plasticity and iron export24-26. Other molecules bound to this part of the alpha-1(I) chain are interleukin 2 (IL2; regulates activity of T-lymphocytes) (ref.18,27) and the collagen binding factor ENDO180 (Mannose receptor, C-type, 2; MRC2) which acts as collagen-binding receptor in the process of endocytosis. It may also play a role in the catabolism of extracellular matrix molecules. Finally, it is suggested this protein may participate in cell-matrix adhesion and cell migration18,28. The identified collagen type I mutations disrupt the binding sequence. A further aspect is that p.Pro1186Ala substitution could negatively affect the stability of collagen type I. Proline and hydroxyproline residues are important amino acids representing 2/3 of all the X and Y positions of the Gly-X-Y motive in alpha chains. High frequency of these two amino acids increases the stability of molecules due to creation of interactions within their side chains and other triple helices29. There is evidence that substitution of proline localized within the triple helical domain of the collagen protein affects thermal stability of the collagen fibril since denaturation temperatures increase with increasing content of proline and hydroxyproline residues30. This variant was also predicted to be damaging using algorithms MutationTester and SIFT.

Mutation in the intronic sequences was identified in one patient. Transversion G to T at position c.1057-1 in the COL1A1 gene was found in a man affected by OI type IVA. Nucleotide change is situated in the acceptor splice site of exon 17 and has been predicted to create a new acceptor splice site localized one base downstream to the first nucleotide of the first codon for glycine (GGT) of the exon 17. Consequently there is a shift of the reading frame of the exon 17 what is completely included in the MLBR1 (ref.19). Several extracellular matrix molecules are known to bind to this part of collagen type I. These include interleukin 2 (regulation of lymphocytes activity), α2β1 integrin which is able to bind other molecules of extracellular matrix (ECM) and GE-decorin (guanidine-extracted decorin) that appears to influence fibrillogenesis and interacts with other molecules of ECM (ref.18,23,29). Further, there are binding sites for osteonectin - a sodium binding glycoprotein that increases bone mineralization, and serpinH1 protein which ensures thermal stabilization of the triple helix domain during procollagen synthesis, helps the folding and assembly of procollagen molecules and participates in the transport of structurally unaffected molecules from the endoplasmic reticulum18,33,34. The c.1057-1G>T mutation results in production of structurally defective collagen type I with delayed binding ability of other ECM molecules increasing bone stability.

The silent c.Gly794Gly variant was identified in the case of a patient affected by OI type IVA. Even though, it is situated in the COMP binding domain defined by codons 582 to 638, the variant does not alter the reading frame and so does not affect the protein production18.

This study is the first comprehensive analysis of patients affected by the collagenous forms of the osteogenesis imperfecta disease. COL1A1 mutations were only found in 20.6% of analysed individuals. The low percentage could be explained by the disease causing mutations occurring in other genes, like COL1A2, PPIB, CRTAP and others, involved in collagen type I processing and synthesis, by the lower probability of detecting C/G and A/T single nucleotide polymorphisms and small homozygous insertions/deletions shift using the HRM method as these may be very close in Tm to the wild type. For some exons, it can also be difficult to interpret HRM results because of the presence of several mutations or polymorphisms35. Additional functional experiments need to be performed to evaluate the effect of identified mutations. Further, analyses of COL1A1 coding sequences excluded by HRM will be prepared.

Currently, COL1A2 genetic screening is performed to provide additional variants in patients diagnosed with one of dominant OI types, 46 of 52 coding (and flanking non-coding) sequences of the gene were analysed. Disease causing mutation of exon 40 (p.Gly814Trp) was identified in a patient affected by the third form of OI (ref.19). Further, genetic analyses identified 14 single nucleotide polymorphisms. 4 were in coding sequence, 1 occurred in MLBR3. Allele frequencies of polymorphisms identified in COL1A2 gene correlate with Population genetics available on Ensembl (http://www.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=eNsG00000164692;r=7:94394561-94431232;t=ENST00000297268).

CONCLUSION

Novel candidate COL1A1 mutations were identified in patients diagnosed with dominant OI forms. Molecular-genetic analyses of another 6/52 COL2A1 coding sequences are currently being prepared. Analyses of other genes involved in the disease origin are also planned as mutations in these genes may account for autosomal recessive forms in the OI patients involved in the study. It could be also helpful to analyse mRNA in patients with identified non-coding sequences mutations to define the effect of DNA splicing and collagen type I production.

Acknowledgement: We are indebted to all patients for their interest and cooperation.

This study was supported by the grants SVV-2013-266517, SVV-2014-260034, SVV-2015-260158, PRVOUK P24/1LF/3 and UNCE 204011 from the Charles University in Prague.

Author contributions: LH: data analysis and interpretation, literature search, manuscript writing; IF, VWH, GM: data analysis and interpretation; IM: clinical data col-
lecion and interpretation; PM: data interpretation; IM: study design, data analysis and interpretation.

Conflict of interest statement: The authors state that there are no conflicts of interest in this study regarding the publication of this article.

REFERENCES