## Title

Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype—phenotype relationships



## Journal

Human Molecular Genetics, 18(3)

## ISSN

0964-6906

## Authors

Bodian, Dale L
Chan, Ting-Fung
Poon, Annie
et al.

## Publication Date

2009-02-01

## DOI

10.1093/hmg/ddn374

*Human Molecular Genetics, 2009, Vol. 18, No. 3* **463–471**
*doi:10.1093/hmg/ddn374*
*Advance Access published on November 7, 2008*

# Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype–phenotype relationships

**Dale L. Bodian[1,‡], Ting-Fung Chan[2,†,‡], Annie Poon[2], Ulrike Schwarze[3], Kathleen Yang[3], Peter H. Byers[3,4], Pui-Yan Kwok[5,6] and Teri E. Klein[1,*]**

[1]Genetics Department, School of Medicine, Stanford University, Stanford, CA 94305-5120, USA, [2]Cardiovascular Research Institute, University of California, San Francisco, CA 94143-0793, USA, [3]Department of Pathology, University of Washington, Seattle, WA 98195-7470, USA, [4]Department of Medicine, University of Washington, Seattle, WA 98195-7470, USA, [5]Department of Dermatology, Cardiovascular Research Institute, and [6]Institute for Human Genetics, University of California, San Francisco, CA 94143-0793, USA

Received October 3, 2008; Revised and Accepted November 4, 2008

**Osteogenesis imperfecta (OI), also known as brittle bone disease, is a clinically and genetically heterogeneous disorder primarily characterized by susceptibility to fracture. Although OI generally results from mutations in the type I collagen genes, *COL1A1* and *COL1A2*, the relationship between genotype and phenotype is not yet well understood. To provide additional data for genotype–phenotype analyses and to determine the proportion of mutations in the type I collagen genes among subjects with lethal forms of OI, we sequenced the coding and exon-flanking regions of *COL1A1* and *COL1A2* in a cohort of 63 subjects with OI type II, the perinatal lethal form of the disease. We identified 61 distinct heterozygous mutations in type I collagen, including five non-synonymous rare variants of unknown significance, of which 43 had not been seen previously. In addition, we found 60 SNPs in *COL1A1*, of which 17 were not reported previously, and 82 in *COL1A2*, of which 18 are novel. In three samples without collagen mutations, we found inactivating mutations in *CRTAP* and *LEPRE1*, suggesting a frequency of these recessive mutations of ~5% in OI type II. A computational model that predicts the outcome of substitutions for glycine within the triple helical domain of collagen α1(I) chains predicted lethality with ~90% accuracy. The results contribute to the understanding of the etiology of OI by providing data to evaluate and refine current models relating genotype to phenotype and by providing an unbiased indication of the relative frequency of mutations in OI-associated genes.**

## INTRODUCTION

Osteogenesis imperfecta (OI), also known as brittle bone disease, is a heterogeneous disorder characterized by susceptibility to fracture. The disease varies in severity from mild (OI type I) to perinatal lethal (OI type II) and exhibits both autosomal dominant and recessive inheritance patterns. The dominantly inherited forms that account for ~90% of infants and adults with OI result from heterozygous mutations in *COL1A1* or *COL1A2*, the genes that encode the proα1(I) and proα2(I) chains of type I collagen, the major structural protein of bone. Recessively inherited OI results in many cases from homozygous or compound heterozygous mutations in *CRTAP* and *LEPRE1* (1–3), which encode cartilage-associated protein and leucine proline-enriched proteoglycan (leprecan; prolyl-3-hydroxylase-1), respectively. These two proteins are part of a complex that hydroxylates a single proline residue at position 986 of the triple helical domain of the proα1(I) chain of collagen during its biosynthesis. Mutations in each of the four genes, *COL1A1*, *COL1A2*, *CRTAP* and *LEPRE1*,

*To whom correspondence should be addressed. Tel: +1 650 736 0156; Fax: +1 650 725 3863; Email: teri.klein@stanford.edu
†Present address: Department of Biochemistry, Faculty of Science, The Chinese University of Hong Kong, Hong Kong.
‡The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

lead to the production of type I collagen molecules that undergo excessive post-translational modification, at least in cultured dermal fibroblasts (1,2,4).

The relationship between the clinical severity of OI and the causative mutations is only partially understood, despite the availability of mutation sequences from >800 individuals with OI (5). The mildest forms of OI generally result from heterozygous mutations that lead to the loss of mRNA from one *COL1A1* allele due to the presence of premature termination codons in the coding sequence of that allele (6–8). Different classes of mutations (single codon nonsense mutations, frameshift mutations and splice site mutations that result in use of out-of-frame cryptic splice sites) can all lead to activation of the nonsense-mediated mRNA decay pathway (6). In the more severe forms of OI in which there is often bone deformity, increased fracture rate, dentinogenesis imperfecta and short stature, the most frequent mutations result in substitution of the glycine residue in almost any of the 338 Gly-X-Y tripeptide motifs found in the triple helical region of each of the proα chains. For these mutations, it appears that the chain in which the mutation occurs, the position of the substituted glycine and the substituting residue all contribute to phenotypic outcome (9). Yet the nature of these relationships is not yet evident. This may reflect the possibility that different mutations result in lethality by distinct pathways (10–12).

Given the diversity of mechanisms underlying the lethal forms of OI, understanding the relationship between genotype and phenotype requires mutation detection and clinical data from a larger number of individuals with OI type II than currently available. To initiate this process, we identified the disease-associated mutations in 62 subjects with OI type II. In 59 cases we found heterozygous mutations in the type I collagen genes, of which 23 of the distinct causative mutations in *COL1A1* and 15 in *COL1A2* are novel. In three samples without collagen mutations, we found homozygous inactivating mutations in *CRTAP* (one) and *LEPRE1* (two individuals homozygous for the same West African allele). Since the subjects were selected based only on phenotype and the production of abnormal type I procollagen molecules, this provides an unbiased indication of the frequency of mutations in these genes in OI type II. The novel collagen mutations provide a valuable addition to the set of known mutations available for studying genotype–phenotype relationships.

## RESULTS

### *COL1A1* and *COL1A2* SNPs

A total of 203 distinct genomic sequence variations were identified in the type I collagen genes among the 63 subjects with OI type II, 98 in *COL1A1* and 105 in *COL1A2* (Table 1). This includes 142 SNPs (Supplementary Material, Tables S1 and S2) and 61 mutations and non-synonymous sequence variants of unknown significance (Table 2). Of the 142 SNPs, 35 are novel: 17 in *COL1A1* and 18 in *COL1A2*. Of the 76 common SNPs with minor allele frequency (MAF) >0.05 listed in dbSNP which are located within the genomic regions of the resequencing effort (25 in *COL1A1* and 51 in *COL1A2*), all but three (rs41317349 and rs2586494 in *COL1A1* and rs17073 in *COL1A2*) are seen in our sample set. The SNPs rs41317349 and rs2586494 are particularly common in the African-American and Asian populations, respectively (13), so their absence may reflect the ethnicities in our OI patient cohort that is largely Caucasian. Approximately 90% of the identified SNPs in each gene are intronic (53 of 60 in *COL1A1* and 73 of 82 in *COL1A2*), and only one non-synonymous SNP was recognized in each gene, rs1800215 (p.Ala1075Thr) in *COL1A1* and rs42524 (p.Ala549Pro) in *COL1A2*. The rs42524 SNP is present at frequencies of 17.7, 26.0, 9.38 and 30.2% in African-American, Hispanic, Chinese and Caucasian populations of healthy individuals, respectively (13). In the same study, rs1800215 was found at a frequency of 6.25% in African-American population (negligible in other ethnic groups).

Three intronic SNPs in *COL1A2* with MAF ~0.05 or greater among the OI individuals were not seen previously in our healthy cohort (13) (Table 3). One, with an MAF of 4.1%, is a novel SNP, whereas the other two were reported previously. rs34026686 has only one submitter listed in dbSNP and no allele frequency information, suggesting that it could be a rare allele. However, in our OI sample set it has an MAF of 14.4%. rs10228528 has an MAF of 4.8% among the OI subjects and is listed with an MAF of only 0.8% among the Caucasians (CEU), although it is known to be more common in Han Chinese (CHB) and Yoruba African (YRI) populations (14).

**Table 1.** Distribution of distinct genomic sequence variations in type I collagen genes

| | | COL1A1 Previously reported | Novel | COL1A2 Previously reported | Novel |
|---|---|---|---|---|---|
| Mutations | Coding[a] | 11 | 21 | 7 | 12 |
| | Splicing | 0 | 6 | 0 | 4 |
| SNPs | Synonymous | 4 | 2 | 5 | 3 |
| | Non-synonymous | 1 | 0 | 1 | 0 |
| | Intronic | 38 | 15 | 58 | 15 |

[a]Includes causative mutations and sequence variants of unknown significance.

### *COL1A1* and *COL1A2* mutations

We identified mutations in type I collagen genes in 59 of the 63 samples (Table 2). In one sample (C2) in which protein studies indicated a rearrangement in *COL1A1*, no mutation was detected in genomic DNA but cDNA analysis revealed a nine-exon deletion (exons 40–48). Of the 59 samples in which collagen mutations were identified, 37 had mutations in *COL1A1* and 22 in *COL1A2*. Among the 34 distinct *COL1A1* causative mutations, 26 resulted in substitution for a glycine within the Gly-X-Y triplet domain of the triple helix, four altered splice sites and resulted in exon skipping, one resulted in exon skipping from a deletion spanning coding and intronic sequence, one was a nine-exon deletion from genomic DNA, one was a duplication of nine nucleotides and the last (p.Asp1413Asn) altered a single residue in the carboxyl-terminal propeptide. Our unpublished studies suggest that this last mutation interferes with chain associ-

**Table 2.** *COL1A1* and *COL1A2* mutations

| Subject | COL1A1 cDNA | Protein | Triple helix | Exon | COL1A2 cDNA | Protein | Triple helix | Exon | References[a] |
|---|---|---|---|---|---|---|---|---|---|
| Missense mutations | | | | | | | | | |
| F1 | c.1058G>A | p.Gly353Asp | 175 | 17 | | | | | |
| E6 | c.1103G>T | p.Gly368Val | 190 | 17 | | | | | |
| E1 | c.1273G>A | p.Gly425Ser | 247 | 19 | | | | | (5,28)[b] |
| H2 | c.1364G>A | p.Gly455Asp | 277 | 21 | c.700C>T | p.Arg234Cys | 144 | 15 | |
| F5 | c.1409G>T | p.Gly470Val | 292 | 21 | | | | | |
| D8 | c.1526G>T | p.Gly509Val | 331 | 23 | | | | | |
| G4 | c.1643G>C | p.Gly548Ala | 370 | 24 | | | | | |
| A1 | c.1804G>A | p.Gly602Arg | 424 | 26 | | | | | (29) |
| G1 | c.1804G>A | p.Gly602Arg | 424 | 26 | | | | | (29) |
| D7 | c.1814G>A | p.Gly605Asp | 427 | 26 | | | | | |
| G7 | c.1840G>C | p.Gly614Arg | 436 | 27 | | | | | |
| A7 | c.2218G>C | p.Gly740Arg | 562 | 32 | | | | | |
| | c.1168G>A | p.Ala390Thr | 212 | 18 | | | | | |
| E4 | c.2425G>A | p.Gly809Ser | 631 | 36 | | | | | (5,30) |
| E8 | c.2470G>C | p.Gly824Arg | 646 | 37 | | | | | |
| H4 | c.2533G>C | p.Gly845Arg | 667 | 37 | | | | | (5,31) |
| H5 | c.2542G>C | p.Gly848Arg | 670 | 37 | | | | | |
| B1 | c.2596G>A | p.Gly866Ser | 688 | 38 | | | | | (5,32,33)[b] |
| B7 | c.2623G>A | p.Gly875Ser | 697 | 39 | | | | | (5) |
| | c.863A>T | p.Glu288Ala | 110 | 13 | | | | | |
| E5 | c.2650G>A | p.Gly884Ser | 706 | 39 | | | | | |
| G2 | c.2650G>A | p.Gly884Ser | 706 | 39 | | | | | |
| E3 | c.2687G>A | p.Gly896Asp | 718 | 40 | | | | | |
| C5 | c.2839G>T | p.Gly947Cys | 769 | 41 | | | | | (5) |
| | c.2563A>C | p.Asn855His | 677 | 38 | | | | | |
| G3 | c.2930G>A | p.Gly977Asp | 799 | 41 | | | | | |
| B3 | c.3001G>T | p.Gly1001Cys | 823 | 42 | | | | | |
| B2 | c.3065G>T | p.Gly1022Val | 844 | 43 | | | | | (5,34) |
| F3 | c.3065G>T | p.Gly1022Val | 844 | 43 | | | | | (5,34) |
| B5 | c.3164G>A | p.Gly1055Asp | 877 | 44 | | | | | |
| G8 | c.3280G>A | p.Gly1094Ser | 916 | 46 | | | | | |
| C6 | c.3299G>A | p.Gly1100Asp | 922 | 46 | | | | | (5) |
| | c.436C>A | p.Pro146Thr | na | 5 | | | | | |
| D6 | c.4237G>A | p.Asp1413Asn | na | 51 | | | | | (17) |
| F2 | | | | | c.847G>C | p.Gly283Arg | 193 | 17 | |
| F8 | | | | | c.1190G>A | p.Gly397Glu | 307 | 21 | |
| D4 | | | | | c.1360G>T | p.Gly454Cys | 364 | 24 | (5) |
| A3 | | | | | c.1369_1370GG>CT | p.Gly457Leu | 367 | 24 | |
| F7 | | | | | c.1577G>A | p.Gly526Glu | 436 | 27 | (5,35)[c] |
| E7 | | | | | c.1685G>T | p.Gly562Val | 472 | 29 | |
| A6 | | | | | c.2215G>C | p.Gly739Arg | 649 | 37 | (5)[c] |
| F6 | | | | | c.2243G>T | p.Gly748Val | 658 | 37 | |
| H3 | | | | | c.2369G>A | p.Gly790Asp | 700 | 39 | (36) |
| B8 | | | | | c.2567G>T | p.Gly856Val | 766 | 41 | |
| E2 | | | | | c.2845G>A | p.Gly949Ser | 859 | 44 | (5,37,38)[b] |
| H7 | | | | | c.2864G>A | p.Gly955Asp | 865 | 44 | |
| A4 | | | | | c.3080G>A | p.Gly1027Glu | 937 | 46 | (5) |
| Insertions and deletions | | | | | | | | | |

*Continued*

**Table 2.** *Continued*

| Subject | COL1A1 cDNA | Protein | Triple helix | Exon | COL1A2 cDNA | Protein | Triple helix | Exon | References[a] |
|---|---|---|---|---|---|---|---|---|---|
| H6 | c.3148_3156dupGCTCCTGGT | p.1050_1052dupAlaProGly | 874 | 44 | | | | | |
| C7 | | | | | c.1380_1397delCCCC GCTGGAAAAGAAGG | p.461_466delProAlaG lyLysGluGly | 371_376 | 24 | |
| G5 | | | | | c.2113_2121del GCTGGTCCT | p.705_707delAlaGlyPro | 615_617 | 35 | (16) |
| F4 | | | | | c.2391_2393dupCCC | p.Pro798dup | 708 | 39 | |
| C8 | | | | | c.2415_2432delCCCT CCTGGTCCCCCTGG | p.806_811delProProG lyProProGly | 716_721 | 40 | |
| H1 | | | | | c.3171_3188delTCCT TCTGGCCCTGCTGG | p.1058_1062delProSe rGlyProAlaGly | 968_972 | 48 | |
| Splicing | | | | | | | | | |
| C4 | c.957+5G>A | IVS14+5G>A | | Skip exon 14 | | | | | |
| A5 | c.2509_2559+9del | | | Skip exon 37 | | | | | |
| B6 | c.3261+1G>A | IVS45+1G>A | | Skip exon 45 | | | | | |
| H8 | c.3423+2T>A | IVS47+2T>A | | Skip exon 47 | | | | | |
| A2 | c.3424-1G>C | IVS47-1G>C | | Skip exon 48 | | | | | |
| C2 | | | | Genomic deletion of exons 40–48 | | | | | |
| B4 | | | | | c.1701_1719+6del | | | Skip exon 29 | |
| G6 | | | | | c.1720-2A>G | IVS29-2A>G | | Unknown–?skip exon 30 | |
| C1 | | | | | c.2673+1G>A | IVS41+1G>A | | Skip exon 41 | |
| A8 | | | | | Not resolved | | | Moderate level of skip exon 46 | |

[a]References are listed for previously reported mutations; this field is left blank for novel mutations.
[b]Reported in both lethal and non-lethal OI cases.
[c]Reported as OI type III.

**Table 3.** Common SNPs in *COL1A2*

| Genomic position | refSNP cluster ID[a] | Major allele | Minor allele | MAF in OI subjects | MAF in CEU[a] | MAF in CHB[a] | MAF in YRI[a] |
|---|---|---|---|---|---|---|---|
| 93863673 | Novel | A | G | 0.041 | No data | No data | No data |
| 93878223 | rs34026686 | G | A | 0.144 | No data | No data | No data |
| 93883010 | rs10228528 | G | A | 0.048 | 0.008 | 0.31 | 0.25 |

[a]According to dbSNP build 129.

ation. Among the 22 causative mutations we identified in *COL1A2*, 13 resulted in substitution for triple helix glycine residues, three were deletions of different 18 bp fragments, two were splice site mutations that resulted in exon skipping, one was a 3 bp duplication and one a 9 bp deletion. Two additional mutations resulted in exon skipping: one deletion spanning coding and intronic sequence and the other one not yet resolved. In sample H2, we identified a mutation in *COL1A1* that is most likely responsible for the OI type II phenotype c.1364G>A, p.Gly455Asp, Gly277Asp in the triple helix and, in addition, a second mutation, c.700C>T, p.Arg234Cys, Arg144Cys in the triple helical domain of *COL1A2*. Substitutions of arginine by cysteine in the triple helical domain of proα1(I) chains has been associated with classical forms of Ehlers–Danlos syndrome (15). We have not been able to study parental samples to determine if there is a phenotype associated with the *COL1A2* sequence alteration. The 38 novel, causative mutations represent a 24% addition to the set of 159 distinct mutations in *COL1A1* and *COL1A2* associated with lethal forms of OI published by the OI consortium (99 and 44 unique substitutions for glycine in αl(I) and α2(I), respectively, and seven and nine unique splicing mutations in *COL1A1* and *COL1A2*, respectively) (5).

As seen previously, the most common mutations are missense mutations that result in substitution for triple helical glycine residues and represent 80 and 60% of the observed mutations in *COL1A1* and *COL1A2*, respectively, in this set. Twice as many were observed in *COL1A1* as in *COL1A2* (26 versus 13), consistent with the set of mutations reported by the OI consortium, in which 67% of the lethal glycine missense mutations occurred in *COL1A1* (5). We also identified mutations affecting splice sites, insertions and deletions in the triple helical region of multiples of three amino acids and mutations in repetitive regions, all of which have been discussed previously regarding OI (5,16). Sixteen of the 18 mutations previously seen were reported in at least one case of lethal OI. Although c.3423+2T>A (IVS47+2T>A) in *COL1A1* has not been described previously, mutation of the adjacent G of the obligate GT at the donor site (c.3423+1G>A, IVS47+1G>A) was found in one patient with lethal OI (5).

The data set also includes some unique features. The *COL1A2* mutation c.1369_1370GG>CT, p.Gly457Leu is the first observed leucine for glycine substitution in a type I collagen gene and the only known example of a dinucleotide mutation in a glycine codon in this gene. The c.700C>T, p.Arg234Cys mutation is the first mutation in which arginine is replaced by cysteine in the triple helical domain of proα2(I). Four additional individuals with substitutions for glycine in proα1(I) chains also had a second amino substitution that affected a non-glycine residue in proα1(I) chains, three in the triple helical region (p.Glu288Ala, Glu110Ala in the triple helix, p.Ala390Thr, Ala212Thr in the triple helix and p.Asn855His, Asn677His in the triple helix) and one in the N-terminal propeptide (p.Pro146Thr). None of these variations have been reported previously although the p.Ala390Thr was seen in one other individual in the diagnostic setting (unpublished data). The contribution of the non-glycine substitution to the phenotype is not known; however, p.Glu288Lys (c.862G>A) was observed in a patient with OI type I who also had a p.Asp1219Glu in the C-propeptide cleavage site, as well as in the patient's unaffected parent (17), suggesting that substitution of the glutamic acid, at least with lysine, does not cause lethal disease. It is not known whether the pairs of mutations are co-allelic.

### Phenotype prediction for glycine substitutions in proα1(I)

We are developing a computational method for predicting OI phenotype of collagen mutations.

The mutations reported here provide an important opportunity to evaluate the current implementation of the model, which was trained on the OI consortium mutations (5), on new data. Of the 29 patients with mutations resulting in substitution of a triple helical glycine in proα1(I) chains, the clinical outcome in 26 (90%) is predicted correctly (Table 4). Of the 17 unique, novel mutations not included in the training set, 15 (88%) were classified correctly. The model can also be applied to serine and cysteine substitutions for glycine in the triple helical region of proα2(I) but no novel mutations of this type were identified. The three *COL1A1* mutations with incorrect predictions are (i) p.Gly353Asp, Gly175Asp in the triple helix, which lies one glycine N-terminal to the model's cutoff between lethal and non-lethal, (ii) a rare lethal alanine substitution and (iii) p.Gly866Ser, Gly688Ser in the triple helix, which is non-lethal in five of six previously observed cases.

### *LEPRE1* and *CRTAP* mutations

We found no mutations in type I collagen genes in four samples and in three of those we identified mutations in genes in which mutations were recently recognized to give rise to recessive forms of severe/lethal OI. In two of those samples, D1 and D3, there is homozygosity for a recently recognized splice site mutation in the *LEPRE1* gene (c.1080+1G>T, IVS5+1G>T) that is of West African origin and is a cause of lethal OI in the African-American community in the USA (1). The ethnic origin of the two families was not stated. In a third infant, C3, there was homozygosity for a splice site mutation in the *CRTAP* gene (c.471+2C>A, IVS1+2C>A). No consanguinity was reported in the parents.

Based on the small number of patients in our set with mutations in the *LEPRE1* and *CRTAP* genes, an estimate of the proportion of OI type II patients with recessive mutations in these two genes is 4.8% (90% confidence interval

**Table 4.** Prediction of lethality for *COL1A1* glycine missense mutations

| Subject | Mutation | Triple helix position | Lethality in training set[a] | Prediction | Correct? |
|---|---|---|---|---|---|
| F1 | p.Gly353Asp | 175 | na | Non-lethal | No |
| E6 | p.Gly368Val | 190 | na | Lethal | Yes |
| E1 | p.Gly425Ser | 247 | Lethal | Lethal | Yes |
| H2 | p.Gly455Asp | 277 | na | Lethal | Yes |
| F5 | p.Gly470Val | 292 | na | Lethal | Yes |
| D8 | p.Gly509Val | 331 | na | Lethal | Yes |
| G4 | p.Gly548Ala | 370 | na | Non-lethal | No |
| A1 | p.Gly602Arg | 424 | na | Lethal | Yes |
| G1 | p.Gly602Arg | 424 | na | Lethal | Yes |
| D7 | p.Gly605Asp | 427 | na | Lethal | Yes |
| G7 | p.Gly614Arg | 436 | na | Lethal | Yes |
| A7 | p.Gly740Arg | 562 | na | Lethal | Yes |
| E4 | p.Gly809Ser | 631 | Lethal | Lethal | Yes |
| E8 | p.Gly824Arg | 646 | na | Lethal | Yes |
| H4 | p.Gly845Arg | 667 | Lethal | Lethal | Yes |
| H5 | p.Gly848Arg | 670 | na | Lethal | Yes |
| B1 | p.Gly866Ser | 688 | Non-lethal | Non-lethal | No |
| B7 | p.Gly875Ser | 697 | Lethal | Lethal | Yes |
| E5 | p.Gly884Ser | 706 | na | Lethal | Yes |
| G2 | p.Gly884Ser | 706 | na | Lethal | Yes |
| E3 | p.Gly896Asp | 718 | na | Lethal | Yes |
| C5 | p.Gly947Cys | 769 | Lethal | Lethal | Yes |
| G3 | p.Gly977Asp | 799 | na | Lethal | Yes |
| B3 | p.Gly1001Cys | 823 | Lethal | Lethal | Yes |
| B2 | p.Gly1022Val | 844 | Lethal | Lethal | Yes |
| F3 | p.Gly1022Val | 844 | Lethal | Lethal | Yes |
| B5 | p.Gly1055Asp | 877 | na | Lethal | Yes |
| G8 | p.Gly1094Ser | 916 | na | Lethal | Yes |
| C6 | p.Gly1100Asp | 922 | Lethal | Lethal | Yes |

[a]na, not applicable.

1.5–12.4%). This extends an earlier estimate that CRTAP mutations cause ~2–3% of cases of lethal OI (2). Because of the measured frequency of *LEPRE1* heterozygosity in the African-Americans in the USA (18,19), these mutations may account for a significant proportion of the severe forms of OI in the USA among that group.

## DISCUSSION

We have identified mutations in samples from 62 subjects with OI type II in whose DNA we sequenced the complete coding regions and flanking intronic regions of the two type I collagen genes, *COL1A1* and *COL1A2*. In these samples, 37 had causative mutations in *COL1A1* and 22 in *COL1A2*, of which 38 distinct mutations had not been reported previously, a 24% addition to the published set of unique missense and splicing mutations in these two genes associated with a lethal phenotype (5). Among the four individuals in whom we did not identify mutations in type I collagen, we found one who was homozygous for a *CRTAP* inactivating mutation and two who were homozygous for a known *LEPRE1* mutation common in the African-American population derived from West Africa. One sample remains incompletely characterized but these cells make abnormal type I collagen molecules. Since the samples were selected by two criteria, a clinical picture consistent with a lethal OI phenotype (although some were identified by ultrasound in the second trimester) and cells that made overmodified type I collagens, these results provide an estimate of the relative frequencies of the spectrum of mutations associated with lethal OI.

The data suggest that there are multiple pathways that lead to lethality. We identified mutations in OI type II subjects likely to cause impaired post-translational modification of procollagen either because of (i) altered sequences in the triple helical domain of the proα chains of type I collagen, (ii) impaired chain association as a result of mutations in the carboxyl-terminal propeptide or (iii) alterations in helix propagation that result from failure to bring the prolyl *cis–trans* isomerase to the molecule (e.g. *CRTAP* and *LEPRE1* mutations) (1–3,18,20). Mutations that alter the sequences of the triple helical domains of the chains of type I procollagen can affect the thermal stability of the molecules (12,21), interfere with secretion efficiency (22), disrupt the chain register (23) or interfere with ligand binding or other interaction sites (5,24).

Models relating genotype to phenotype must capture the diversity of these mechanisms. We allow for this complexity by using a composite predictive model (12). The high accuracy in predicting the lethality of the newly sequenced mutations is consistent with the hypothesis that substitutions of glycine by serine and mutations that substitute bulky amino acids for glycine carboxy-terminal to residue 178 in the triple helical region of proα1(I) differ in the manner or extent to which they disrupt collagen structure or function. Triple helix position 178 was modeled as the end of the predominantly non-lethal N-terminal region based on available data. The newly identified lethal mutation Gly175Asp suggests that position 175 may be a better cutoff. Because the current data set is limited to OI type II mutations, a more complete evaluation of the model requires non-lethal cases. Additional mutations are needed to refine other imprecise aspects of the model including whether the cutoff position depends on the residue replacing glycine and the determinant of lethality of substitutions of glycine by cysteine in the proα1(I) chains. The use of the composite model framework will allow incorporation of other features that may contribute to lethality, such as individual variation and sites of functional importance, when sufficient data are available.

Exceptions to this model may have implications for our understanding of collagen biology. For example, the model misclassifies Gly370Ala in the triple helix of proα1(I) as non-lethal. This suggests that a region around residue 370 may have important function and that its disruption can be lethal. Indeed, substitution of the glycine at 376 by alanine is lethal (5). This region overlaps a binding site for a small proteoglycan (decorin) on collagen molecules (25,26), but it seems unlikely that disturbance of binding of decorin, by itself, is lethal as a decorin knockout mouse does not have OI, despite dramatic alterations in fibrillogenesis (27).

It was proposed recently that some regions of the triple helical domain of type I collagen molecules may lack mutations or have exclusively lethal mutations because disruption interferes with a critical function (although these functions have not been clearly identified) (5). Four of the novel mutations in *COL1A1*, which result in the substitutions Gly706Ser, Gly718Asp, Gly799Asp and Gly916Ser in the triple helical domain, lie within 'lethal-only' regions that overlap the proposed major ligand binding regions MLBR2

and MLBR3 (5,26). Interestingly, Gly688Ser, present in both lethal and non-lethal cases, lies near the edge of MLBR2. In *COL1A2*, three novel mutations were identified, which fall into lethal-only regions 2 (Gly472Val), 6 (Gly766Val) and 7 (Gly865Asp). In contrast, mutations in *COL1A1* that alter residues in triple helix regions 328–346 and 418–436, proposed to be of critical importance due to a lack of observed mutations (5), contain four mutations in our data set (Gly331Val, Gly424Arg, Gly427Arg and Gly436Arg), suggesting that the previous absence of mutations was due to sampling. It is clear that to this point sampling of the almost 2000 possible substitutions of triple helical glycines by other residues in each chain is sufficiently incomplete that it is difficult to draw broad conclusions about the function of particular regions of the triple helical domain based on the absence of mutations.

Of the substitutions for glycine in the triple helical domain encoded by *COL1A1* identified here, 16 of the 26 distinct mutations are novel. According to published models (5,9,12), almost all substitutions of glycine by valine, aspartic acid, glutamic acid, arginine and tryptophan between residues 178 and 1012 of the triple helix in proα1(I) chains should be lethal. Using codon substitution tables, this suggests that a minimum of 900 of the possible non-synonymous mutations in this region will have a lethal effect. Only 100 distinct lethal glycine substitutions in this region were reported previously, including 49 replacements by alanine, serine and cysteine (5). With the relatively low coverage of these mutations to date, most newly studied infants with lethal OI will have previously unrecognized mutations. Mutations in *COL1A2* are less likely to be lethal, and it has been more difficult to locate the domains in which they occur. Of the mutations identified here, 13 of the *COL1A2* mutations resulted in substitutions for glycine. Of these, six have been previously encountered in infants with lethal OI but the remaining seven were seen for the first time and six of them fell in domains of the triple helix not previously thought to carry lethal mutations. This suggests that there are still too few lethal mutations, or mutations overall, in the *COL1A2* gene to accurately predict the outcome on the basis of any proposed model.

Nearly 10% of the subjects in this study (five of 59) with a type I collagen mutation have a second missense alteration, including one individual with a mutation in both *COL1A1* and *COL1A2*. Although the glycine substitution itself is probably sufficient for lethality in these cases, the additional variations may influence the phenotype resulting from less disruptive mutations. Complete sequencing of both genes and reporting of all sequence variations found will be important for assessing the role of multiple mutations in the development of disease.

These studies indicate that recurrence in families in which a child with lethal OI is born is increased as a result of parental carrier status for recessive mutations in addition to the well-known risk of parental mosaicism for dominant mutations. For mutations in the *COL1A1* gene, it is clear that substitution of glycine residues by large amino acids carboxyl-terminal to position 178 in the triple helical domain of the proα1(I) chains carries a high risk of lethal outcome, something that can be transferred to clinical utility. It is clear that many effects of mutations on behavior of collagen molecules need to be taken into account when determining how they result in lethal mutations. Even in the absence of well-defined pathways and mechanisms, the accumulation of mutations with associated clinical data is an important part of providing the basis for rational counseling of families and represents the first step in forming a truly informative model for how mutations result in phenotypes.

## MATERIALS AND METHODS

### Subject population

We selected 64 DNA samples from unrelated individuals from the Connective Tissue Biopsy Program Repository, an IRB approved activity at the University of Washington that allows further analysis of stored samples. The referral diagnosis for all selected samples was a lethal form of OI. For all samples we had screened cultured fibroblasts for the production of and structure of type I procollagen (22) and cells from all made abnormal proteins. One sample had DNA that failed to amplify (D5). Of the remaining 63, 27 were from fetal samples of $<$24 weeks gestation. In all instances, the ultrasound picture or post-termination or post-spontaneous delivery radiographs were consistent with the diagnosis of OI type II–minimal calvarial mineralization, very short and bowed limbs, multiple rib fractures. Among the remaining 36 samples, there was no clinical information for one, one that left the hospital at 15 days and for whom there was no follow-up information, and for the remainder there were radiographs or clinical photographs consistent with the diagnosis of a lethal form of OI, with the usual diagnosis of OI type II. Patient descriptions are summarized in Supplementary Material, Table S3. Data on ethnic background were absent for virtually all samples.

For two of the samples (C3 and D1) there was a record of familial recurrence of lethal OI. For two others (A3, C2), it was noted that the father had a milder form of OI, OI type III or III/IV. No consanguinity was noted in either family. No prior history of OI was noted in any of the remaining instances.

### Collagen gene sequencing and variant detection

Genomic DNA was isolated from peripheral blood using standard extraction protocols. For each patient, a total of 14 and 23 kb of genomic DNA for *COL1A1* and *COL1A2*, respectively, were sequenced including all coding regions, intron sequences for at least 100 bp on either side of each exon and evolutionarily conserved domains in non-coding regions ($>$75% sequence identity between mouse and human). Primer sequences, PCR conditions and sequence coordinates of each fragment are published elsewhere (13). All primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA).

Sequence traces were aligned with the GenBank reference sequences of the *COL1A1* genomic DNA (AF017178) and cDNA (NM_000088.3), and the *COL1A2* genomic DNA (AF004877.1) and cDNA (NM_000089.3). Sequence variations were identified and confirmed by detection with two different analysis programs: Sequencher v4.8 (GeneCodes

Corp., Ann Arbor, MI, USA) and Mutation Surveyor v3.1 (SoftGenetics LLC., State College, PA, USA). Automatic calls were verified by inspection. All newly identified variants were deposited in dbSNP (14) under the submitter handle UCSF_HG. All mutations identified were confirmed by resequencing of genomic DNA taken from reserved samples.

For those samples in which mutations were not detected in the initial phases of analysis, collagen cDNA was prepared, amplified in overlapping fragments of about 1–1.5 kb and examined by PAGE. For sample C2, the cDNA was amplified in two fragments to identify the extent of the genomic deletion. If no mutation was identified in cDNA, genomic DNA of *CRTAP* and *LEPRE1* was amplified as described previously (18) and sequenced.

Collagen mutations are numbered following the HGVS-approved convention (http://www.hgvs.org/mutnomen/recs.html), which starts with the translation initiator methionine as amino acid +1, and the A of the ATG codon as nucleotide +1. Mutations affecting intronic sequence are referenced to the cDNA sequence. Triple helix positions are provided for amino acids within the triple helical region of each proα chain (residues encoded by codons 179–1192 of the *COL1A1* transcript and 91–1104 of *COL1A2*). The exons of the *COL1A1* gene are numbered consecutively from 1–32. Exon 33 of the gene is referred to as exon 33–34 by tradition, to allow the codons of the helical domain to be similar between the two genes and exons thereafter continue from 35 to 52. Exons of *COL1A2* are numbered consecutively.

Each identified variation was uniquely classified as either a SNP, causative mutation, or non-synonymous sequence variant of unknown significance. All substitutions for glycine within the triple helical domain, mutations that alter consensus splice sites, insertion or deletion of residues within the triple helical domain, and mutations that result in sequence alterations in the carboxyl-terminal propeptides and that had been previously identified in individuals with OI or shown to alter molecular behavior were considered causative mutations. p.Ala390Thr, p.Glu288Ala, p.Asn855His and p.Pro146Thr in the proα1(I) chain are categorized as rare non-synonymous sequence variants of unknown significance. Variations classified as SNPs are detailed in Supplementary Material, Tables S1 and S2. SNPs include synonymous variations and the non-synonymous rs1800215 (p.Ala1075Thr) in *COL1A1* and rs42524 (p.Ala549Pro) in *COL1A2*. SNPs are considered novel if they are not listed in build 129 of dbSNP (14); novel mutations are those not reported in collagen mutation databases (28, Bodian and Klein, manuscript in preparation).

### *CRTAP* and *LEPRE1* mutation detection

Each exon of the *CRTAP* and *LEPRE1* genes was amplified and then sequenced as previously described (3,18). All mutations were confirmed by reamplification and sequencing of a new genomic DNA sample.

### Lethality prediction

Lethality was predicted using a published method applicable to single missense mutations altering the Gly-X-Y glycines in the triple helical region of the proα1(I) chain (12). One component of the model is a decision tree that classifies mutations within the N-terminal 178 residues of the triple helix as non-lethal. Mutations C-terminal to position 178 are classified by the amino acid substituting for glycine. Arginine, valine, aspartic acid and glutamic acid substitutions are predicted to be lethal, whereas alanine replacements are predicted to be non-lethal. Cysteine substitutions are tentatively classified as lethal C-terminal to triple helix position 688. Serine substitutions for glycine are classified by a second component based on an estimate of the thermostability of the Gly-X-Y triplet C-terminal to the triplet with the substitution. The models were constructed using the OI triple helical glycine substitutions mutations published by the OI consortium (5) as a training set. Predictions for p.Gly884Ser (Gly706Ser) and p.Gly1094Ser (Gly916Ser) based on preliminary sequencing were published previously (12). For subjects with multiple variations, the triple helical glycine substitutions are assumed to be the primary determinant of OI lethality.

### Statistical analysis

Confidence intervals were computed using exact distributions as implemented in R (29).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* Online.

*Conflict of Interest statement*. None declared.

## FUNDING

This work was supported by the National Institutes of Health (AR051582 to T.E.K and T32HL007731 to T.F.C); and the Osteogenesis Imperfecta Foundation and the Children's Brittle Bone Foundation (to D.L.B.). Funding to Pay the Open Access Charge was provided by the National Institutes of Health AR051582.

## REFERENCES

1. Cabral, W.A., Chang, W., Barnes, A.M., Weis, M., Scott, M.A., Leikin, S., Makareeva, E., Kuznetsova, N.V., Rosenbaum, K.N., Tifft, C.J. *et al.* (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat. Genet.*, **39**, 359–365.
2. Barnes, A.M., Chang, W., Morello, R., Cabral, W.A., Weis, M., Eyre, D.R., Leikin, S., Makareeva, E., Kuznetsova, N., Uveges, T.E. *et al.* (2006) Deficiency of cartilage-associated protein in recessive lethal osteogenesis imperfecta. *N. Engl. J. Med.*, **355**, 2757–2764.
3. Morello, R., Bertin, T.K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., Castagnola, P., Rauch, F., Glorieux, F.H., Vranka, J. *et al.* (2006) CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell*, **127**, 291–304.
4. Bonadio, J. and Byers, P.H. (1985) Subtle structural alterations in the chains of type I procollagen produce osteogenesis imperfecta type II. *Nature*, **316**, 363–366.
5. Marini, J.C., Forlino, A., Cabral, W.A., Barnes, A.M., San Antonio, J.D., Milgrom, S., Hyland, J.C., Korkko, J., Prockop, D.J., De Paepe, A. *et al.* (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with

collagen binding sites for integrins and proteoglycans. *Hum. Mutat.*, **28**, 209–221.
6. Willing, M.C., Deschenes, S.P., Scott, D.A., Byers, P.H., Slayton, R.L., Pitts, S.H., Arikat, H. and Roberts, E.J. (1994) Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am. J. Hum. Genet.*, **55**, 638–647.
7. Slayton, R.L., Deschenes, S.P. and Willing, M.C. (2000) Nonsense mutations in the COL1A1 gene preferentially reduce nuclear levels of mRNA but not hnRNA in osteogenesis imperfecta type I cell strains. *Matrix Biol.*, **19**, 1–9.
8. Willing, M.C., Deschenes, S.P., Slayton, R.L. and Roberts, E.J. (1996) Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. *Am. J. Hum. Genet.*, **59**, 799–809.
9. Byers, P.H. (1990) Brittle bones—fragile molecules: disorders of collagen gene structure and expression. *Trends Genet.*, **6**, 293–300.
10. Chessler, S.D. and Byers, P.H. (1993) BiP binds type I procollagen pro alpha chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. *J. Biol. Chem.*, **268**, 18226–18233.
11. Chessler, S.D., Wallis, G.A. and Byers, P.H. (1993) Mutations in the carboxyl-terminal propeptide of the proα1(I) chain of type I collagen result in defective chain association and produce lethal osteogenesis imperfecta. *J. Biol. Chem.*, **268**, 18218–18225.
12. Bodian, D.L., Madhan, B., Brodsky, B. and Klein, T.E. (2008) Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry*, **47**, 5424–5432.
13. Chan, T.F., Poon, A., Basu, A., Addleman, N.R., Chen, J., Phong, A., Byers, P.H., Klein, T.E. and Kwok, P.Y. (2008) Natural variation in four human collagen genes across an ethnically diverse population. *Genomics*, **91**, 307–314.
14. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
15. Malfait, F., Symoens, S., De Backer, J., Hermanns-Le, T., Sakalihasan, N., Lapiere, C.M., Coucke, P. and De Paepe, A. (2007) Three arginine to cysteine substitutions in the pro-alpha (I)-collagen chain cause Ehlers–Danlos syndrome with a propensity to arterial rupture in early adulthood. *Hum. Mutat.*, **28**, 387–395.
16. Pace, J.M., Atkinson, M., Willing, M.C., Wallis, G. and Byers, P.H. (2001) Deletions and duplications of Gly-Xaa-Yaa triplet repeats in the triple helical domains of type I collagen chains disrupt helix formation and result in several types of osteogenesis imperfecta. *Hum. Mutat.*, **18**, 319–326.
17. Pollitt, R., McMahon, R., Nunn, J., Bamford, R., Afifi, A., Bishop, N. and Dalton, A. (2006) Mutation analysis of COL1A1 and COL1A2 in patients diagnosed with osteogenesis imperfecta type I–IV. *Hum. Mutat.*, **27**, 716.
18. Baldridge, D., Schwarze, U., Morello, R., Lennington, J., Bertin, T.K., Pace, J.M., Pepin, M.G., Weis, M., Eyre, D.R., Walsh, J. *et al.* (2008) CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta. *Hum. Mutat.*, doi:10.1002/humu.20799.
19. Cabral, W.A., Barnes, A.M., Porter, F.D. and Marini, J.C. (2007) Carrier Frequency of Recurring Mutation Causing Severe/Lethal Recessive Type VIII Osteogenesis Imperfecta in African-Americans. Annual Meeting of the American Society of Human Genetics, San Diego, California. Available from http://www.ashg.org/genetics/ashg07s/index.shtml.
20. Marini, J.C., Cabral, W.A., Barnes, A.M. and Chang, W. (2007) Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development. *Cell Cycle*, **6**, 1675–1681.
21. Beck, K., Chan, V.C., Shenoy, N., Kirkpatrick, A., Ramshaw, J.A. and Brodsky, B. (2000) Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. USA*, **97**, 4273–4278.
22. Bonadio, J., Holbrook, K.A., Gelinas, R.E., Jacob, J. and Byers, P.H. (1985) Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. *J. Biol. Chem.*, **260**, 1734–1742.
23. Willing, M.C., Cohn, D.H., Starman, B., Holbrook, K.A., Greenberg, C.R. and Byers, P.H. (1988) Heterozygosity for a large deletion in the alpha 2(I) collagen gene has a dramatic effect on type I collagen secretion and produces perinatal lethal osteogenesis imperfecta. *J. Biol. Chem.*, **263**, 8398–8404.
24. Sweeney, S.M., Orgel, J.P., Fertala, A., McAuliffe, J.D., Turner, K.R., Di Lullo, G.A., Chen, S., Antipova, O., Perumal, S., Ala-Kokko, L. *et al.* (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.*, **283**, 21187–21197.
25. Yu, L., Cummings, C., Sheehan, J.K., Kadler, K.E., Holmes, D.F. and Chapman, J.A. (1993) In Scott, J.E. (ed.), *Dermatan Sulfate Proteoglycans Chemistry Biology and Clinical Pathology*, Portland Press, London, pp. 183–192.
26. Di Lullo, G.A., Sweeney, S.M., Korkko, J., Ala-Kokko, L. and San Antonio, J.D. (2002) Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J. Biol. Chem.*, **277**, 4223–4231.
27. Danielson, K.G., Baribault, H., Holmes, D.F., Graham, H., Kadler, K.E. and Iozzo, R.V. (1997) Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. *J. Cell Biol.*, **136**, 729–743.
28. Dalgleish, R. (1998) The human collagen mutation database 1998. *Nucleic Acids Res.*, **26**, 253–255.
29. R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.
30. Mackay, K., Byers, P.H. and Dalgleish, R. (1993) An RT-PCR-SSCP screening strategy for detection of mutations in the gene encoding the alpha 1 chain of type I collagen: application to four patients with osteogenesis imperfecta. *Hum. Mol. Genet.*, **2**, 1155–1160.
31. Westerhausen, A., Kishi, J. and Prockop, D.J. (1990) Mutations that substitute serine for glycine alpha 1-598 and glycine alpha 1-631 in type I procollagen. The effects on thermal unfolding of the triple helix are position-specific and demonstrate that the protein unfolds through a series of cooperative blocks. *J. Biol. Chem.*, **265**, 13995–14000.
32. Bateman, J.F., Lamande, S.R., Dahl, H.H., Chan, D. and Cole, W.G. (1988) Substitution of arginine for glycine 664 in the collagen α1(I) chain in lethal perinatal osteogenesis imperfecta. Demonstration of the peptide defect by *in vitro* expression of the mutant cDNA. *J. Biol. Chem.*, **263**, 11627–11630.
33. Horwitz, E.M., Prockop, D.J., Fitzpatrick, L.A., Koo, W.W., Gordon, P.L., Neel, M., Sussman, M., Orchard, P., Marx, J.C., Pyeritz, R.E. *et al.* (1999) Transplantability and therapeutic effects of bone marrow-derived mesenchymal cells in children with osteogenesis imperfecta. *Nat. Med.*, **5**, 309–313.
34. Lund, A.M., Astrom, E., Soderhall, S., Schwartz, M. and Skovby, F. (1999) Osteogenesis imperfecta: mosaicism and refinement of the genotype–phenotype map in OI type III. Mutations in brief no. 242 (Online). *Hum. Mutat.*, **13**, 503.
35. Korkko, J., Kuivaniemi, H., Paassilta, P., Zhuang, J., Tromp, G., DePaepe, A., Prockop, D.J. and Ala-Kokko, L. (1997) Two new recurrent nucleotide mutations in the COL1A1 gene in four patients with osteogenesis imperfecta: about one-fifth are recurrent. *Hum. Mutat.*, **9**, 148–156.
36. Ward, L.M., Lalic, L., Roughley, P.J. and Glorieux, F.H. (2001) Thirty-three novel COL1A1 and COL1A2 mutations in patients with osteogenesis imperfecta types I–IV. *Hum. Mutat.*, **17**, 434.
37. Cohen-Solal, L., Zylberberg, L., Sangalli, A., Gomez Lira, M. and Mottes, M. (1994) Substitution of an aspartic acid for glycine 700 in the alpha 2(I) chain of type I collagen in a recurrent lethal type II osteogenesis imperfecta dramatically affects the mineralization of bone. *J. Biol. Chem.*, **269**, 14751–14758.
38. Rose, N.J., Mackay, K., Byers, P.H. and Dalgleish, R. (1994) A Gly859Ser substitution in the triple helical domain of the alpha 2 chain of type I collagen resulting in osteogenesis imperfecta type III in two unrelated individuals. *Hum. Mutat.*, **3**, 391–394.
39. Nuytinck, L., Wettinck, K., Freund, M., Van Maldergem, L., Fabry, G. and De Paepe, A. (1997) Osteogenesis imperfecta phenotypes resulting from serine for glycine substitutions in the alpha2(I) collagen chain. *Eur. J. Hum. Genet.*, **5**, 161–167.