Review article

Osteogenesis imperfecta: translation of mutation to phenotype

Peter H Byers, Gillian A Wallis, Marcia C Willing

By the end of the recently completed Fourth International Conference on Osteogenesis Imperfecta (Pavia, Italy, 9–12 September 1990) more than 70 mutations in the two genes that encode the chains of type I collagen, the major protein of bone, had been identified as the molecular cause of different forms of osteogenesis imperfecta (OI). Although by no means complete, the set of mutations in hand provides a rough guide to how to predict the phenotypic effects of mutations in type I collagen genes, predicts that certain classes of mutations will give rise to very mild phenotypes that will blend with common disorders, such as osteoporosis, and clarifies the genetic aspects of the widely used clinical classification of OI.1

Clinical classification of OI
For almost a century clinicians have attempted to provide a rationale by which to understand the readily apparent clinical heterogeneity in OI. The early distinction of 'congenita' forms (those with fractures at birth) from 'tarda' forms (those who develop fractures at a later period),2 followed by acknowledgment of further phenotypic variation,3 provided early guides. About a decade ago, these early classifications were largely supplanted, at least for geneticists, by the scheme developed by Sillence et al1 that incorporated genetic, radiographic, and phenotypic criteria. Biochemical and genetic studies have further amended and refined this classification (table). The value of the classification is often questioned because of the difficulty in deciding where the borders of the types of OI lie and how early a subject can be 'classified'. It is clear that there are molecular reasons for the clinical uncertainties and that they reflect the heterogeneous nature of the causative mutations.

Two classes of mutations, those that reduce the amount of type I collagen synthesised by tissues, and those that alter the structure of the molecules made, are responsible for almost all forms of OI (fig 1). In general, the defective synthesis mutations produce the mildest phenotypes. Mutations that alter the structure of the chains of type I procollagen and lead to the formation of abnormal molecules produce a very wide range of phenotypes, from extremely mild to lethal. Because the severity of the phenotype depends on the chain in which the mutation occurs, as well as the nature and location of the mutation, there is considerable overlap of the phenotypic entities. Molecular studies are now designed, in part, to determine how these mutations produce disease and why OI phenotypes occur in about 1 in 5000 to 1 in 10 000 subjects.

Relevant collagen biosynthesis
To understand how mutations in collagen genes result in osteogenesis imperfecta, it helps to describe the genes and their products, and the manner in which the protein is synthesised and processed by the cell (see references 4 and 5 for more complete reviews of collagen biosynthesis and gene structure). Two genes, COL1A1 located on chromosome 17 and COL1A2 located on chromosome 7, encode the proc1(I) and proc2(I) chains of type I procollagen, respectively. These genes each contain more than 50 exons that contain the coding sequences for proteins of about 1400 amino acids. The gene structure reflects the domain structure of the protein. The first five exons encode the signal sequence and an amino-terminal propeptide (the function of which is unclear). The sixth exon encodes a protease cleavage site. The next
**Molecular basis of osteogenesis imperfecta.**

<table>
<thead>
<tr>
<th>OI type</th>
<th>Clinical features</th>
<th>Inheritance*</th>
<th>Biochemical and genetic abnormalities</th>
</tr>
</thead>
</table>
| I       | Normal stature, little or no deformity, blue sclerae, hearing loss in 50%, dentinogenesis imperfecta is rare and may distinguish a subset | AD | Common 'Non-functional' COL1A1 allele  
Rare Substitution for glycine residue in carboxyl-terminal telopeptide of α1(I).  
Substitution of glycine at position 94 of the triple helix in proc1(I).  
Exon deletion in proc2(I) triple helical domain |
| II      | Progressively deforming bones, usually with moderate deformity at birth. Sclerae variable in colour, often lighten with age. Dentinogenesis imperfecta common, hearing loss common. Stature very short | AD (new) | Common Substitutions for glycy1 residues in the triple helical domain of the α1(I) chain and α2(I) chain.  
Rare Rearrangement in the COL1A1 and COL1A2 genes,  
Exon deletions in triple helical domain of COL1A1 and COL1A2.  
Substitutions and small deletions in the non-triple helical carboxyl-terminal propeptide.  
Tripeptide deletion in proc1(I) triple helix |
| III     | Progressively deforming bones, usually with moderate deformity at birth. Sclerae variable in colour, often lighten with age. Dentinogenesis imperfecta common, hearing loss common. Stature very short | AR (rare) | Small deletion in α2(I) on the background of null allele |
| IV      | Normal sclerae, mild to moderate bone deformity and variable short stature, dentinogenesis imperfecta is common and hearing loss occurs in some | AD | Point mutations in the COL1A1 and COL1A2 gene  
AR (uncommon) | Frameshift (4 bp deletion) in COL1A2 that prevents incorporation of proc2(I) chains into molecules  
Point mutations in COL1A1 and COL1A2 gene.  
Exon skipping mutations in COL1A2 |

*AD, autosomal dominant; AR, autosomal recessive.

43 exons encode the core 1014 amino acid triple helical domain of the chain, that consists of repeating tripeptide units with the sequence of Gly–X–Y in which X and Y exclude cysteine and tryptophan. The final four exons encode an additional protease cleavage site and the globular carboxyl-terminal domain of the protein that is involved in molecular assembly. The genes are coordinately transcribed, the mRNAs are processed and transported to the cytoplasm where they are translated on membrane bound polysomes. The precursor proteins are inserted through the membrane of the rough endoplasmic reticulum and the signal sequences are cleaved. During chain elongation, all the prolyl residues (about 100 in each chain) that precede glycine in the triple helix and some of the lysines (less than 10 in each chain) are hydroxylated by the enzymes prolyl 4-hydroxylase and lysyl hydroxylase, respectively. Some hydroxylsine residues are glycosylated and a single asparagine linked oligosaccharide is added to each chain in the carboxyl-terminal propeptide.

Each procollagen molecule contains two proc1(I) chains and a single proc2(I) chain. The correct assembly of molecules is directed by interaction of the proα1 chains through their carboxyl-terminal globular domains. After aggregation in the 2:1 ratio the chains fold the triple helix from the carboxyl-terminal end towards the amino-terminus. The triple helix requires a glycine in every third position to fold correctly. Although the modification by the hydroxylating and glycosylating enzymes is initiated during chain elongation, it ceases only with the formation of a stable triple helix, at which point the molecule is no longer a substrate. The completed molecule is transported out of the cell through the Golgi apparatus and deposited in the pericellular environment where the two propeptides are cleaved from the molecule and the collagen molecules spontaneously aggregate to form fibril structures. The formation of fibrils depends on the uniformity of subunits and abnormal molecules have highly deleterious effects on fibril structure.

**Point mutations are common**

Despite the repetitive nature of much of the collagen gene (which might be suspected to lead to recombinational events), point mutations are the most common events that lead to OI phenotypes. These mutations may result in substitutions for glycine residues within the triple helix, may produce exon skipping events if they occur in the consensus splice donor or acceptor sites, or may alter the ability of chains to aggregate into molecules if they occur within
the domains that encode the carboxyl-terminal propeptide.

**Point mutations in the triple helix: genotype-phenotype relationship**

As a rule, substitutions for glycine residues in the triple helix of the \(\alpha 1(I)\) chain have the most severe phenotypic consequences when they occur towards the carboxyl-terminal end of the molecule and are mildest towards the amino-terminal end. However, there is a marked influence of the nature of the substituting amino acid and there may be important effects of local domain structure (fig 2).

Cysteine for glycine substitutions have been the most readily characterised because the presence of cysteine in the triple helical domain of \(\alpha 1(I)\), from which it normally is excluded, is detected at the protein level by the formation of a disulphide bonded dimer of \(\alpha 1(I)\) chains. Substitutions of cysteine for glycine at positions 988, 604, 748, 718, 69 and 6910 are lethal, those at 526 and 415 produce an OI type III phenotype, those at 382, 223, 205 (Pruchno et al, unpublished data), 178, 12 and 175 produce an OI type IV phenotype, and those at 94, 946 (Starman et al, unpublished data), 43, 14 and 19 produce even milder phenotypes. Of the two major aberrations from this general picture, a mild phenotype in a woman with a substitution at 904, and a lethal phenotype with substitution at residue 244, only the former is explained by somatic mosaicism for the mutation which was lethal in her child. The lethal outcome in a child with a cysteine for glycine substitution at 244 is unexpected, given the fairly smooth 'phenotypic gradient', but might be explained by an unidentified structural feature of the molecule in that region, or by death from causes unrelated to OI.

With three exceptions, substitutions of serine for glycine carboxyl-terminal to residue 550 of the triple helix are lethal17-21; a substitution at 460 (Willing and Byers, unpublished data) produces an OI type III/IV phenotype. Two mutations, one at 83222 and the other at 844,23 produce OI type III/IV phenotypes and are flanked by mutations that result in lethal phenotypes. Mosaicism has not, apparently, been excluded in either instance but the possibility of a tolerant domain in the molecule has also been suggested. A very severe but non-lethal phenotype that results from a substitution at 100919 may reflect unusual effects of mutations very near the end of the triple helix.

Two mutations that result in substitution of arginine for glycine (at residue 15421 in unrelated subjects) are not lethal while the others that produce the same substitution in different domains24-28 (Willing and Byers, unpublished data) and all those identified so far that substitute valine26-30 or aspartic acid31 32 (Starman et al, unpublished data) for glycine are lethal. The paucity of mutations that result in substitutions of alanine28 for glycine suggests that in many regions of the gene such an event would produce a mild phenotypic effect. Thus, because of the bias toward studying mutations in the lethal phenotype, the substitutions that result in mild phenotypes have not been identified.

In contrast to substitutions for glycine within the triple helix of \(\alpha 1(I)\), for the far smaller number of similar mutations within \(\alpha 2(I)\) there is no semblance of an orderly relationship between phenotype and position or nature of the substituting residue. For example, cysteine for glycine substitutions which produce a mild phenotype (at residue 64633) in dominant families may be flanked by those which are more severe or lethal in effect33-35. Perhaps the nature of the surrounding sequence or of a special domain within the chain explains this finding. Most
other point mutations in the COL1A2 gene have proved to be lethal, although a serine for glycine at 661, a valine for glycine at 586, and an arginine for glycine at 1012 are notable exceptions. Because linkage between mild dominant forms of OI and polymorphic loci in COL1A2 has been shown, point mutations in the COL1A2 gene may prove to be common once the mutations in these families have been identified. It should be noted, however, that a significant number of the mutations in the COL1A2 gene from mildly affected subjects and their families result in exon skipping or in insertion of intron material (see below).

**Exon skipping mutations within the triple helix**

A consequence either of point mutations or of small genomic deletions, exon skipping mutations in the type I collagen genes are the second most common cause of OI characterised (fig 3). For the COL1A1 gene and the proα1(I) chain, the majority of mutations that give rise to exon skipping events are either very severe or lethal (Wallis et al, unpublished data). Mild effects can occur, however, if the mutation involves exons near the 5' end of the gene (for example, exon 8, encoding amino acids 19 to 36 of the triple helix) or if the proportion of chains synthesised by the mutant mRNA is small (for example, exon 17

---

**Figure 2** Map of point mutations in COL1A1 (top) and COL1A2 (bottom) that result in substitutions for glycine within the triple helical domain of the α1(Ⅰ) and α2(Ⅰ) chains, respectively. For each locus, the location of each point mutation is represented along the chain. The substituting residue is listed on the left and the residue number of the substituted glycine is listed along the line. The triple helix contains 1014 amino acids and the first glycine is represented as residue number 1. The phenotype of each mutation is denoted below the mutation.

---

<table>
<thead>
<tr>
<th>COL1A1</th>
<th>Ala</th>
<th>Ser</th>
<th>Cys</th>
<th>Arg</th>
<th>Val</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>46</td>
<td>154</td>
<td>256</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>178</td>
<td>19</td>
<td>298</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>244</td>
<td>94</td>
<td>391</td>
<td>559</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>223</td>
<td>175</td>
<td>550</td>
<td>559</td>
<td>673</td>
</tr>
<tr>
<td></td>
<td></td>
<td>244</td>
<td>125</td>
<td>667</td>
<td>883</td>
<td>883</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OI I</td>
<td>460</td>
<td>OI III/IV</td>
<td>OI III</td>
<td>OI III</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lethal</td>
<td>lethal</td>
<td>lethal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OI IV</td>
<td>OI IV/III</td>
<td>OI IV</td>
<td>OI IV</td>
<td>OI IV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lethal</td>
<td></td>
<td>lethal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>lethal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COL1A2</th>
<th>Ala</th>
<th>Ser</th>
<th>Cys</th>
<th>Arg</th>
<th>Val</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>259</td>
<td>496</td>
<td>586</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>646</td>
<td>472</td>
<td>694</td>
<td>580</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>787</td>
<td>646</td>
<td>1012</td>
<td>907</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lethal</td>
<td>lethal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>lethal</td>
<td></td>
<td></td>
<td>lethal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lethal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OI IV</td>
<td></td>
<td>OI II</td>
</tr>
</tbody>
</table>

---

*436*

Byers, Wallis, Willing
skipping with a low proportion of the abnormal product.49

For the COL1A2 gene, exon skipping mutations display an orderly gradient of phenotypic expression from lethal to quite mild as they progress from the 3' to the 5' end of the helix coding domain.14 50-54 The transition from lethal to non-lethal appears to occur between exons 28 and 21, but the extent to which instability of the abnormal mRNA contributes to the phenotype is unclear. This trend predicts that mutations even closer to the 5' end than exon 9 may give surprisingly mild phenotypes. In both COL1A1 and COL1A2, skipping of exon 6, which contains the amino-terminal procollagen protease site, produces the Ehlers-Danlos syndrome type VII phenotype,55 56 a dramatic shift from OI.

**Point mutations in the non-triple helical carboxyl-terminal domains**

A small number of mutations has been identified outside the triple helical domain in the COL1A1 gene. Substitution of cysteine for the glycine three residues removed from the carboxyl-terminal end of the triple helix has a mild effect.57 Similarly, substitution for a single cysteine in the carboxyl-terminal propeptide by tryptophan has a mild effect, perhaps because the chain is excluded from molecules17 (see below). In contrast, substitution of arginine for leucine47 in a conserved region of the carboxyl-terminal domain results in a lethal phenotype, apparently because the abnormal chain interferes with secretion of the procollagen molecules into which it is incorporated.

**Multixon rearrangements are lethal when expressed in proteins but shorter deletions may be milder**

Despite the repetitive nature of the collagen genes, rearrangements within COL1A1 and COL1A2 are rare causes of osteogenesis imperfecta and, when they have occurred, have been lethal. A three exon deletion from within the triple helix coding domain of the COL1A1 gene58 59 and a seven exon deletion from within COL1A260 have been identified, although the mechanisms by which they occurred are uncertain. A more complex partial duplication within COL1A1 was also lethal and presumably arose by recombination.51

---

**Figure 3** Map of the location of exon skipping mutations in the COL1A1 (top) and COL1A2 (bottom) genes. The maps of the two genes are drawn roughly to scale but the sizes of the introns are more accurately denoted for COL1A1. Note that in COL1A1 exons 33 and 34 are fused to form a single 108 bp exon and that the COL1A2 gene is about twice the size of the COL1A1 gene. In each case, except for exon 14 in COL1A1, the affected subject is heterozygous for the mutation. The splicing scheme indicates that one allele is processed normally but the other allele results in an mRNA from which the exon indicated has been skipped. There is inefficient splicing of exon 14 in both the COL1A1 alleles in the subject with OI type II that results in about half the mRNA from each allele having the deleted structure.
In contrast, shorter deletions have different effects depending on the mutation. For example, two independent mutations which each remove one Gly-Ala-Pro triplet from the three-fold repeat between 868 and 876 of the α1(I) chain have been lethal. A three amino acid (Gly-Pro-Pro) deletion near the carboxy-terminal end of the triple helix of the α2(I) chain (residues 1009 to 1011) is, however, mild (Wallis et al, unpublished data). Deletion of the three nucleotides that encode the valine at position 255 of the α2(I) chain produces the OI type III phenotype.

Frameshift mutations have differing effects that depend on their location
Frameshift mutations within the triple helical coding domain have proved difficult to identify and none has yet been characterised. In contrast, a frameshift mutation in COL1A2 that alters the coding sequence by insertion of intron sequence in the triple helix, and a frameshift mutation in the carboxy-terminal domain of COL1A2, have, surprisingly, resulted in mild phenotypes. The mild phenotype of the six amino acid insertion between residues 585 and 586 of the triple helix may result from the intracellular retention of molecules that contain the abnormal chain, or from instability of the abnormal mRNA, or both.

Frameshift mutations in the carboxy-terminal domains of the two genes differ in their phenotypic effect. One mutation, a single base insertion in a COL1A1 allele, has a lethal effect, apparently because the molecules that contain the mutant allele are not secreted efficiently, leading to a paucity of normal molecules in the matrix. In contrast, one mutation in the COL1A1 gene and another in the COL1A2 gene have mild phenotypic effects in the heterozygous state because they prohibit incorporation of the abnormal chains into molecules. The COL1A2 frameshift does, however, produce a moderately severe phenotype in the homozygote, the consequence of synthesising type I procollagen molecules that contain only proα1(I) chains.

Translation of mutation to phenotype
The phenotypic consequence of mutations in type I collagen genes reflects the gene in which the mutation occurred, the nature of the mutation, the location of the mutation, and the effect of the mutation on the behaviour of the abnormal chain and on the behaviour of molecules formed by the chain.

Excluded allelic products generally produce mild phenotypes
In general, mutations that prohibit incorporation of an abnormal chain into a molecule, or that simply reduce the amount of the chain available, would be expected to have mild clinical consequences and are found most often among subjects with OI type I (fig 1). Very few such mutations have yet been identified, probably because of the difficulty in detecting mutations that lower the expression of the abnormal allele. These mutations may have to be identified at the genomic level, a formidable task with genes that encompass 18 kb and 38 kb and have more than 50 exons a piece.

Included but abnormal products vary in severity
If abnormal chains are included in molecules and those molecules are at least moderately stable, the phenotypic effects are more deleterious than if an abnormal chain does not aggregate into a molecule. If, however, an abnormal chain leads to very rapid intracellular degradation of molecules that incorporate the chain, the clinical consequences should differ depending on the gene in which the mutation occurs. Mutations in the COL1A1 gene may be highly deleterious and even lethal because they destroy three-quarters of all the type I procollagen molecules synthesised (half of all molecules contain one abnormal proα1(I) chain and a quarter contain two abnormal proα1(I) chains) (fig 1). A similar mutation in the COL1A2 gene could result in loss of only half the molecules made and so might be similar in effect to a null COL1A1 allele (that is, only half the normal amount of type I procollagen molecules are completed in either case) but only if none of the abnormal molecules leaked from the cell.

When an abnormal chain is included in a molecule, the translation of mutation to phenotype depends on several factors. First, almost all molecules that contain chains with mutations in the triple helical domains are less stable than their normal counterparts and are secreted less efficiently. Second, the mutation alters the structure of the mature molecule so that during processing in the rough endoplasmic reticulum the chains remain accessible to the modifying enzymes and undergo additional hydroxylation of lysyl residues in the triple helix and additional hydroxylysyl glycosylation. These molecules may have an extremely long period of residence in the rough endoplasmic reticulum (thus altering cellular architecture and perhaps other functions of the cell). Third, the abnormal molecules that leave the cell are not proteolytically processed as efficiently as the normal molecules and interfere with the normal process of fibril nucletation and growth. Fourth, abnormal fibrils are probably inefficient substrates for mineralisation. Finally, the requirements of bone for molecular structure may be greater than those of skin and other soft tissues to explain, in part, the relative tissue specificity of these mutations.

Point mutations and deletions (large or small) affect the intracellular processing of molecules in much the same way, probably because both introduce phase
shifts in chain registration that reduce stability.16,32 That is, chains in these molecules become over-modified and the molecules are secreted inefficiently. As a result, the phenotypic consequences of diverse mutations are similar. There is not yet a good explanation for the effects of specific substitutions for glycyl residues in the triple helix, but the size of the substituting residue and interactions with nearby residues probably play a role in determining the phenotypic effect.

Genetics of osteogenesis imperfecta

As a result of biochemical, molecular genetic, and linkage studies there is now little question that the vast majority of subjects with OI are heterozygous for mutations in one of the genes that encode the chains of type I procollagen.25,33,34 There are rare instances of homozygosity35,65 or compound heterozygosity36 for mutations in these genes. There are also uncommon forms of OI that appear to result from mutations in non-collagen genes and are probably inherited in an autosomal recessive fashion.76

There has been controversy about the mode of inheritance of the lethal forms of OI and of some of the deforming varieties. Some clinical studies appeared to provide evidence for autosomal recessive inheritance of lethal OI1 and of recessive inheritance of the progressive deforming varieties of OI.77 More recent clinical studies78 and biochemical and molecular evidence79 have now made it clear that the lethal form of OI almost always results from heterozygosity for mutations in one of the genes of type I collagen (uniparental disomy for an exon skipping mutation provides a remarkable exception, especially as the mutation results in inefficient exon removal49). Biochemical studies, and some molecular studies, provide the same conclusion for deforming varieties of OI.21-23 Despite the molecular evidence of heterozygosity for dominant mutations, recurrence of these phenotypes among sibs born to phenotypically normal parents is not rare. Indeed, for OI type II, the sib recurrence rate is about 6%.73

The modest recurrence risk for OI type II is explained by parental gonadal, and usually somatic, mosaicism in the vast majority of such families.32 The low recurrence risk for OI type III is probably explained by two different mechanisms: parental mosaicism and rare instances of autosomal recessive inheritance, possibly involving non-collagen genes.

The rate of parental mosaicism for collagen gene mutations seems high when compared to other disorders, but comparable studies in other genes have not been completed.

Mutation mechanisms

Collagen genes are probably good reporters of mutations. First, there is a high density of invariant and required glycine residues in the triple helical domain (one-third of all amino acids). Substitution for either of the first two nucleotides of the glycine codon (GGN) changes the amino acid. Thus 22% (two out of nine) of all nucleotides encoding the triple helical domain will probably give rise to a phenotypic change in the heterozygote. Second, the large exon number and sensitivity to exon loss in the protein provides more than 100 additional mutation sensitive sites in each gene. Third, the need to maintain structure in the globular carboxyl-terminal domain to allow for interactions that generate a triple chain molecule provides an unknown number of additional targets. Finally, because type I collagen forms fibrils from identical subunits, appearance of any abnormal molecules interferes with the production of normal structure.

The majority of mutations in collagen genes that produce recognisable forms of OI are point mutations, most of which change a glycine codon to that for another amino acid. Only two of the more than 70 mutations catalogued here have occurred independently in unrelated subjects.21 Both recurrent mutations appeared at CpG dinucleotides and are consistent with deamination of a methyl cytosine to produce thymidine. Because there is considerable bias in the manner in which mutations have been recognised and analysed, the mechanistic basis of mutations is uncertain. It appears, however, that mutations at CpG dinucleotides are not overrepresented, although those sites may be sequences where recurrent mutations are more likely to appear.21

It is surprising to us, given the repetitive structure of the type I collagen genes and the ready identification of large deletions within the type III collagen gene, that large deletions within either gene of type I collagen are uncommon. The multiexon deletions that have been identified occurred as results of intron-intron events and only a single instance of recombination through exon exchange has been recognised. Although it may be that the structure of the gene itself is sufficient to limit recombination, it is possible that other explanations may be necessary to explain the paucity of such events. For example, the presence of other essential genes imbedded within introns or transcripts read from the 'collagen' non-coding strand could limit viability of even heterozygotes for large deletions.

Conclusion

Osteogenesis imperfecta presents a remarkable clinical spectrum, from virtually undetectable to lethal, as a consequence of mutations in the two genes that encode the chains of type I collagen. The phenotypic expression of these mutations reflects the gene in which the mutations occur, the nature of the mutation, the effect of the mutation on the behaviour...
of the gene and abnormal protein, and the consequences of the changes in protein behaviour for assembly of a normal extracellular matrix. The analysis of more than 70 mutations provides insight into the relation of mutation to disease but it is clear that we are still a long way from understanding the grammar by which phenotype is translated from genotype. We estimate that it will take the results of more than 200 mutations to begin to saturate the mutation map and, even then, the finding of unexpected domains of either greater or lesser severity may require additional study.

Collagen genes have proved to be a remarkably robust biological system in which to study mutations because a great deal is known about protein structure, because the protein lives a dual existence in the intracellular and extracellular environments, and because the large collagen gene family means that findings in one system have broad implications. Few investigators had anticipated that substitution of single glycine residues within the triple helix could have such profound effects on folding, modification, secretion, and phenotype. Studies of OI defects have revitalised interest in the intracellular life style of collagen molecules, in the molecules that are involved in their maturation, and on how mutations alter structure and function. At the same time, it has become clear that a more complete understanding is necessary of the structure of bone, the manner in which the protein matrix of bone is built, and how alterations in the amount or structure of the constituent collagens produce aberrant fibrils and, thus, tissues.

The consequences of somatic and germline mosaicism for mutations have been most completely explored in collagen genes but have implications for other genes. Somatic mosaicism at high levels is a result of early embryonic mutation. It is not clear whether collagen genes are at greater risk than other genes for mutation during early embryogenesis, whether entry into the germ cell lineage offers protection from mutation, or whether there is mutation limited to the gene derived from only one parent. It is clear, however, that these studies raise further questions about the timing of new dominant mutations, oblige us to reconsider how to provide genetic counselling in dominantly inherited disorders, and force us to stop taking dogma for granted.

While it is likely that the models developed for understanding the behaviour of type I procollagen molecules will serve for all the fibrillar collagens, we may need to modify our expectations when dealing with collagens that differ in structure. Nevertheless, the methods for identifying and characterising mutations in collagens are now well established and progress in understanding the consequences of mutations in other collagen genes should be rapid.

Finally, we should not underestimate the contribution that heritable mutations in important matrix genes, like collagens, can play in a wide range of late onset, degenerative disorders, and how the study of disorders like OI can point the way to identification of candidate mutations (see references 5, 80, 82, and the last paragraphs of many grant applications for more detailed discussion of these points).

The original work reported here was supported in part by grants from the National Institutes of Health (AR 21557), the March of Dimes Birth Defects Foundation (6–298), the Arthritis Foundation, and by the Michael Geisman Memorial Fellowship from the Osteogenesis Imperfecta Foundation.

Osteogenesis imperfecta: translation of mutation to phenotype


16 Fertala A, Westerhausen A, Morris GM, Rooney JE, Prockop DJ. Two base substitutions in the proalpha1(I) procollagen gene (COL1A1) and COL1A2 that cause lethal osteogenesis imperfecta. The location of glycine substitutions does not in any simple way predict their effects on protein function or phenotype. *Am J Hum Genet* 1990;47:216 (abstract 848).


18 Westerhausen A, Kishi J, Prockop DJ. One mutation that substitutes serine for glycine α1-598 and another mutation that substitutes serine for glycine α1-631 of type I procollagen (COL1A1) both cause lethal osteogenesis imperfecta but have different effects on the thermal stability of the protein. *Am J Hum Genet* 1989;45:228 (abstract 902).


41 Wenzstrup RJ, Cohn DH, Cohen T, Byers PH. Arginine for glycine substitution in the triple-helical domain of the products of one α2(I) allele (COL1A2) produces the osteogenesis imperfecta type IV phenotype. *J Biol Chem* 1988;263:7734–40.


43 Wallis G, Brighton CF, Boyd C, Mathew CG. Mutations associated with the pro-α2(I) collagen gene are responsible for several cases of osteogenesis imperfecta type I. *J Med Genet* 1986;23:411–6.


50 Ganguly A, Baldwin CT, Conway D, Prockop DJ. Use of a carbodiimide reaction to locate single-base mutations. Detection of a single-base mutation in an intron of the type I procollagen
gene (COL1A2) that causes abnormal RNA processing in a proband with lethal osteogenesis imperfecta (type II). Am J Hum Genet 1990;47:217 (abstract 852).


52 Superﬁti-Furga A, Raghunath M, Pistone FM, Comano C, Steinmann B. Deletion of nucleotides +2 to +40 in intron 21 of COL1A2 leading to skipping of exon 21 observed in a boy with short stature, osteoporosis, dentinogenesis imperfecta, but no fractures (OI type IV). Fourth International Conference on Osteogenesis Imperfecta, Pavia, Italy, 9–12 September 1990. Abstracts, 1990:37.


67 Wilting MC, Cohn DH, Byers PH. Frameshift mutation near the 3 end of the COL1A1 gene predicts an elongated pro(1(1)) chain and results in osteogenesis imperfecta type I. J Clin Invest 1990;85:282–90.


