LETTER TO JMG

Biochemical screening of type I collagen in osteogenesis imperfecta: detection of glycine substitutions in the amino end of the alpha chains requires supplementation by molecular analysis

W A Cabral, S Milgrom, A D Letocha, E Moriarty, J C Marini

J Med Genet 2006;43:685–690. doi: 10.1136/jmg.2005.040493

Background: The biochemical test for osteogenesis imperfecta (OI) detects structural abnormalities in the helical region of type I collagen as delayed electrophoretic migration of alpha chains on SDS-urea-PAGE. Sensitivity of this test is based on overmodification of alpha chains in helices with a glycine substitution or other structural defect. The limits of detectability have not been reported.

Methods: We compared the collagen electrophoretic migration of 30 probands (types III or IV OI) with known mutations in the amino half of the $\alpha 1(I)$ and $\alpha 2(I)$ chains. Differences in sensitivity were examined by 5% and 6% SDS-urea-PAGE, and with respect to alpha chain, location along the chain, and substituting amino acid.

Results: Sensitivity was enhanced on 5% gels, and by examination of intracellular and secreted collagen. In $\alpha 1(I)$, substitutions in the first 100 residues were not detectable; 7% of cases in the current Mutation Consortium database are in this region. $\alpha 1(I)$ substitutions between residues 100 and 230 were variably detectable, while those after residue 232 were all detected. In $\alpha 2(I)$, variability of electrophoretic detection extended through residue 436. About a third of cases in the Consortium database are located in the combined variable detection region. Biochemical sensitivity did not correlate with substituting residue.

Conclusions: Complete testing of probands with normal type I collagen biochemical results requires supplementation by molecular analysis of cDNA or gDNA in the amino third of $\alpha 1(I)$ and amino half of $\alpha 2(I)$. Mutation detection in OI is important for counselling, reproductive decisions, exclusion of child abuse, and genotype-phenotype correlations.

small glycine side group within the internal aspect of the triple helix. The rate of helical folding determines the extent of modification to the alpha chains, which includes hydroxylation of multiple proline and lysine residues, as well as subsequent glycosylation of hydroxylysines. A delay in the folding process increases modification of all alpha chains in the helix in the portion of the chains that is amino-terminal to the glycine substitution. This can be detected as delayed electrophoretic migration of alpha chains on SDS-urea-PAGE. The biochemical test is most sensitive for detecting structural abnormalities in the carboxyl end of the collagen molecule. Because the helix folds in a carboxyl to amino direction, a substitution in the carboxyl end will lead to overmodification along the full length of the chains and a greater electrophoretic shift. Additionally, cysteine for glycine substitutions in the $\alpha 1(I)$ chain can be readily detected by the presence of $\alpha 1(I)$ dimers. In contrast, substitutions in the amino end of either alpha chain may be undetectable due to the relatively minor extent of overmodification that results from the smaller number of lysine residues on which glycosylation may proceed. Another limitation of biochemical screening involves detection of mutations causing low-abundance alternative splicing. Shorter alpha chains, in which the residues encoded by a single exon are deleted, can be incorporated into a collagen triple helix, but the resulting overmodification of the shorter chain may result in its comigration with normal chains. While electrophoretic mobility shifts have been the basis of the standard biochemical test for structurally abnormal collagen for two decades, the limits of detection are currently undefined.

Another approach for detecting mutations uses enzymatic or chemical cleavage of mismatches in hybrids between synthetic RNA transcripts and proband mRNA or cDNA^{6–8} (RNA/RNA or RNA/DNA hybrids). An advantage of using this approach is that 0.5–1.5 kb of coding sequence can be screened for sequence changes using a single probe. Unlike biochemical screening, detection of single nucleotide mismatches in RNA/RNA or DNA/RNA duplexes can distinguish which alpha chain contains the mutation. Mutation location along the chain can also be predicted, based on the migration of cleavage products on denaturing PAGE. Detection of point mutations in type I collagen by enzymatic or chemical digestion of hybrids has limitations, however. As many as half of all mismatches are undetected because of a preference for cleavage at purines by RNase A (G→A) and at pyrimidines by piperidine (C→T). Low abundance transcripts

Type I collagen is the primary structural component of the extracellular matrix of skin, tendon, and bone. It is a heterotrimer consisting of two $\alpha 1(I)$ and one $\alpha 2(I)$ chains, encoded by the COL1A1 and COL1A2 genes. Quantitative defects of COL1A1 expression have been well established to cause the mild form of osteogenesis imperfecta (OI) (type I) or early onset osteoporosis.¹ Qualitative defects of type I collagen occur in both $\alpha 1(I)$ and $\alpha 2(I)$ chains.^{2,3} About 80% of these structural mutations are substitutions for glycine residues in a collagenous Gly-X-Y triplet; most of the remaining 20% of structural changes are exon splicing defects. Structural defects result in the clinically significant forms of OI (types II, III, and IV), characterised by skeletal fragility and deformity, scoliosis, and short stature.^{4,5}

Glycine substitutions disrupt the linear folding of the type I collagen heterotrimer, due to steric requirements for the

Abbreviations: CSGE, confirmation-sensitive gel electrophoresis; dHPLC, denaturing high performance liquid chromatography; DMEM, Dulbecco's modified eagle medium; OI, osteogenesis imperfecta; RT-PCR, reverse transcription polymerase chain reaction

may also remain undetected due to autoradiographic background of radiolabelled antisense probes or from degradation of alternatively spliced transcripts. False positive mismatch detection is infrequent, although cleavage of SNPs in the collagen coding sequence can occur.

Mismatch detection has been updated with the use of CSGE (confirmation-sensitive gel electrophoresis) or dHPLC (denaturing high performance liquid chromatography) of PCR fragments spanning all collagen exons and intron-exon boundaries. CSGE is more sensitive than single strand conformation polymorphism analysis⁹ and detects approximately 80% of sequence changes in type I collagen. dHPLC is useful for high-throughput applications and is reported to have up to 93% detection rate for sequence variations.^{10 11} This sensitivity results in frequent detection of exonic and intronic sequence variants. Both CSGE and dHPLC require further characterisation by direct sequencing of any gDNA PCR products in which heteroduplexes are detected^{12 13} to identify the sequence change as a mutation or polymorphism. Neither CSGE nor dHPLC detects sequence changes that do not introduce conformational changes or are larger than a single exon.

A more comprehensive approach involves direct sequencing of PCR fragments covering each of the 52 exons in COL1A1 and COL1A2, including splice recognition sites, or sequencing of patient cDNA covering the entire coding region of each alpha chain.

In the investigation presented here, we studied the sensitivity of the biochemical test for mutations in the amino half of the $\alpha 1(I)$ and $\alpha 2(I)$ chains. We compared the collagen electrophoretic migration of a series of OI probands with known mutations, examining differences in biochemical detection with respect to alpha chain, location along the chain, and substituting amino acid residue. The results indicate that complete detection of mutations in this region of type I collagen requires supplementation of biochemical testing with additional molecular approaches, such as CSGE or dHPLC and, ultimately, direct sequencing.

METHODS

Cell culture

Skin fibroblast cultures were established from dermal punch biopsies. Fibroblasts were cultured in Dulbecco's modified eagle medium (DMEM) containing 10% fetal bovine serum and 2 mM glutamine in the presence of 5% CO₂.

Steady state collagen analysis

To label procollagens, confluent fibroblast cultures of probands and control cells (ATCC 2127) were incubated for 2 h in serum-free medium containing 50 µg/ml ascorbic acid, followed by incubation with 260 µCi/ml of 3.96 TBq/mmol L-[2,3,4,5-³H]proline in serum-free medium for 16 h. Procollagens were harvested from media and cell layer and precipitated with ammonium sulfate, as previously described.¹⁴ Collagens were prepared by pepsin digestion (50 µg/ml) of procollagen samples and electrophoresed on 5% and 6% SDS-urea-PAGE.

RNA:DNA hybrid analysis

Total RNA was isolated from cultured fibroblasts of probands and control cell lines using Tri Reagent (Molecular Research Center, Cincinnati, OH) according to the manufacturer's protocol.¹⁵ The $\alpha 1(I)$ and $\alpha 2(I)$ collagen mRNA was amplified by reverse transcription polymerase chain reaction (RT-PCR). Total RNA (1 µg) was reverse transcribed with 20 U of murine leukaemia virus reverse transcriptase, oligo(dT), and an RNA PCR core kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Following cDNA synthesis, transcripts were amplified by PCR.¹⁶

Proband cDNA was screened for mutations by RNase A digestion of RNA/DNA hybrids as previously described.⁶

Subcloning and sequencing of proband cDNA and gDNA

Leukocyte and fibroblast genomic DNA was isolated from probands using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). RT-PCR and gDNA PCR products were cloned into the pCR2.1 plasmid (Invitrogen, Carlsbad, CA). Subclones were sequenced using the dideoxy chain termination method¹⁷ and the Sequenase 2.0 DNA Sequencing Kit (Amersham, Cleveland, OH). Sequencing of subclones utilised a sense primer corresponding to the pUC M13 forward primer. Sequence variations were confirmed by digestion of gDNA PCR products with appropriate restriction enzymes.

The $\alpha 1(I)$ G34R and G193S mutations were identified at the Center for Gene Therapy, Tulane University Medical Center, using CSGE¹² of PCR amplification products from each exon and its surrounding intronic sequences to identify fragments with sequence variations. Products with abnormal CSGE were directly sequenced using an ABI Prism 377 DNA Sequencer.

RESULTS

We compared the electrophoretic migration of normal fibroblast collagen with collagen from 30 patients with clinical OI (types III or IV) and known mutations in the amino half of the $\alpha 1(I)$ or $\alpha 2(I)$ chain of type I collagen (tables 1 and 2). Type I collagen overmodification was analysed by 5% and 6% SDS-urea-PAGE. Overmodification was detected as delayed electrophoretic migration of alpha chains (backstreaking, increased band width, or baseline shift) compared to normal type I collagen. Alpha chains migrate as tighter bands on 6% acrylamide, allowing for visualisation of subtle baseline shifts in migration. Greater discrimination of overmodified alpha chains, however, was obtained on 5% acrylamide as band broadening and backstreaking.

Electrophoretic migration of collagens with a structural defect in $\alpha 1(I)$ is shown in fig 1. Electrophoretic abnormalities were not detected in type I collagens with glycine substitutions in the first 100 residues of $\alpha 1(I)$ on either 5% or 6% acrylamide gels, with the exception of a fast migrating component in the $\alpha 2(I)$ chain of cell layer collagen from the G13D proband, previously shown to be due to a pepsin-sensitive site.^{18 19} Although mutations at the extreme amino end of $\alpha 1(I)$ may interfere with processing of the aminopropeptide,¹⁸ the pN-collagen band is difficult to use for diagnosis of these cases during screening because the pepsin used for collagen screening digests pN-collagen and examination of undigested procollagen may yield artifactual false positive partial processing bands. Detection of overmodified type I collagen with substitutions between residues 121 and 232 was variable. Secreted and cell layer collagen from fibroblasts of probands with G121D, G136R, G154R, G187A, and G193S substitutions showed baseline shifts of alpha chains on 6% acrylamide gels and more extensive band broadening on 5% acrylamide gels, while collagens with G148R, G220A, and G232S substitutions showed no detectable electrophoretic abnormalities on 5% or 6% SDS-urea-PAGE. In the G121–232 region, detectability of overmodification did not correlate with particular substituting residues. Arginine, alanine, and serine are each variably detectable; in addition, each is associated with overmodification at a residue amino-terminal to a position with normal migration. For example, G148R, G220A, and G232S are normal, but overmodification is detected in the G136R, G187A, and G193S samples. Overmodification of alpha

Table 1 Phenotypic characteristics of individuals with COL1A1 glycine substitutions

Mutation	OI type	Current age	Sex	Z score†‡	Fracture (n)	Age at diagnosis	LE rods	Scoliosis	Spine rods	DI	BI	Hearing	Cardiac§
gly13asp¶	IV	17	F	-4.9	50	Infancy	-	+	+	+	+	-	-
gly25val	IV	19	F	-4.0	>20	Birth	+	+	+	-	-	-	-
gly34arg	III	3	M	*	*	Birth	-	+	+	*	*	*	*
gly76glu	III	17	F	-6.9/-4.7	>50	Birth	+	+	-	-	-	+	+
gly88glu	IV	24	F	-3.0	>40	Birth	+	+	+	-	+	+	-
gly121asp	III	9	M	*	*	Birth	+	*	*	*	*	*	*
gly136arg	IV	21	F	-2.5	*	Infancy	-	-	-	-	+	-	-
gly148arg	IV/I	45	F	-1.9	~6	Childhood	-	-	-	-	*	-	-
gly154arg	III	15	M	-4.93/-3.2	>50	Birth	+	+	+	+	-	+	-
gly187ala	III/IV	8	F	-6.9	>25	Birth	+	+	+	+	-	+	-
gly193ser	III	13	F	-6.6	>50	Birth	+	+	-	+	-	-	-
gly220ala	IV/I	17	M	*	2	2	-	*	*	+	*	-	+
gly232ser	III	14	F	-6.8/-4.5	25-30	Birth	+	+	-	+	-	-	-
gly313cys	III	7	M	*	Several	Infancy	-	*	*	+	*	*	+
gly349cys	III	13	F	*	4	Infancy	-	*	*	*	*	*	*
gly352ser	IV	20	F	-2.2	15-20	Birth	+	+	-	+	+	-	-
gly448ser	IV	9	M	-5.2	20-30	Birth	+	-	-	+	-	-	-
gly523cys	IV	24	F	-5.2	~50	Birth	+	+	-	+	+	-	-
gly589ser	III/IV	12	M	-5.6/-3.7	>25	Birth	+	-	-	+	+	-	-

*No data available.

†When two z scores are available for a patient, the first z score is from before pamidronate infusions and the second is the z score after pamidronate infusions.

‡DEXA z scores for gly25val and gly88glu were derived from measurements at the radius.

¶G13D has an OI/EDS phenotype 7. Unlike G13A7, G13D has had >50 long bone fractures in childhood and scoliosis requiring spinal fixation.

§Cardiac findings include MV prolapse (G76E), VSD with R ventricular diverticulum and abnormal tricuspid valve (G313C), and ASD with mild R heart dilatation (G220A).

LE, lower extremity

chains was detectable in type I collagen with an $\alpha 1(I)$ substitution carboxyl to amino acid residue 313 using 5% and 6% acrylamide gels; cell layer collagen tended to be more overmodified than secreted collagen due to intracellular retention of mutant collagen. In general, 5% acrylamide gels were more sensitive than 6% acrylamide gels for detection of collagen overmodification caused by glycine substitutions beyond the first 100 residues of the $\alpha 1(I)$ chain.

Electrophoretic migration of collagens with a structural defect in $\alpha 2(I)$ is shown in fig 2. The pattern of type I collagen overmodification due to glycine substitutions in the amino half of the $\alpha 2(I)$ chain differs from that of $\alpha 1(I)$ substitutions. Detection of overmodification caused by glycine substitutions in the amino half of the $\alpha 2(I)$ chain was more subtle than for the $\alpha 1(I)$ chain, and substitutions without detectable biochemical abnormalities extended further along

the chain. No glycine substitutions in the first 100 residues were available for testing. Slight broadening of $\alpha 1(I)$ bands in cell layer collagen was observed for G103D, G121D, G217D, G238S, and G244S on 5% and 6% acrylamide gels, with no detectable abnormalities of secreted collagen. G247S showed no electrophoretic abnormalities, however G247C migration was slightly retarded for both alpha chains. Variable detection of overmodification was observed for collagen with an $\alpha 2(I)$ glycine substitution in the second quarter of the chain. Broadening of $\alpha 1(I)$ chains was detected in media and cell fractions on 5% and 6% acrylamide gels for G268S, G337S, G370S, and G421D, but was better detected on 5% acrylamide. G436R had no detectable abnormality. The media fraction for G511S was normal, but overmodification was detected in cell layer collagen on 5% and 6% acrylamide gels. With the exception of the G421D sample, which contained an

Table 2 Phenotypic characteristics of individuals with COL1A2 glycine substitutions

Mutation	OI type	Current age	Sex	Z score†‡	Fracture (n)	Age at diagnosis	LE rods	Scoliosis	Spine rods	DI	BI	Hearing	Cardiac§
gly103asp¶	III	24	M	*	>50	Birth	+	+	-	+	-	-	*
gly121asp	IV	6	M	-3.5	~20	Infancy	-	-	-	-	-	-	-
gly217asp	III/IV	16	M	*	5-10	Birth	-	-	-	+	*	*	*
gly238ser	IV	10	F	-6.7/-2.3	20	Prenatally	+	-	-	+	-	-	-
gly244ser	IV	12	F	-5.7/-3.5	10-15	Infancy	-	-	-	-	+	-	-
gly247cys	III	21	M	-7.1	30	Birth	+	+	-	+	+	+	-
gly247ser	III	43	M	*	22	Birth	+	*	-	*	*	*	*
gly268ser	IV	23	F	*	>50	Birth	+	+	+	+	-	-	-
gly337ser	III	16	F	-6	>50	Birth	+	+	-	+	+	+	-
gly370ser	III	20	M	-5.9	5-10	Birth	+	+	+	+	-	+	+
gly421asp	II	Deceased	M	*	10	Birth	-	*	*	*	*	*	*
gly436arg	IV	23	M	*	None	12	-	-	-	-	-	+	*
gly436arg	IV	19	F	*	None	7	-	+	-	-	-	-	*
gly511ser	IV	16	M	*	Multiple	*	+	+	-	*	*	*	*
gly511ser	I	17	F	*	1	*	-	+	-	*	*	*	*

*No data available.

†When two z scores are available for a patient, the first z score is from before pamidronate infusions, and the second is the z score after pamidronate infusions.

‡DEXA z scores for gly25val and gly88glu were derived from measurements at the radius.

¶Patient diagnosed at birth with severe OI and in utero fractures. Attained household ambulation. 1-2 fractures/3 months and sustained multiple fractures with seizures in adolescence.

§Cardiac findings include MV prolapse and enlarged RA and LA (G370S).

LE, lower extremity

Figure 1 Electrophoretic migration of collagen with structural defects in the $\alpha 1(I)$ chain of type I collagen. The collagen synthesised by dermal fibroblasts from OI probands with glycine substitutions in $\alpha 1(I)$ was isolated from both media and cell layer samples in a steady state labelling. The type I collagen of each proband was electrophoresed on both 5% and 6% acrylamide gels in comparison to normal control collagen (C) and visualised by autoradiography. The location of the glycine substitution along the chain and the substituting residue are indicated below each proband's sample, for example, G13D. In general, 5% acrylamide gels were more sensitive than 6% acrylamide gels for detection of collagen overmodification caused by glycine substitutions beyond the first 100 residues of the $\alpha 1(I)$ chain.

$\alpha 2(I)$ doublet with broadened $\alpha 1(I)$ chains, $\alpha 2(I)$ substitutions were best detected electrophoretically in the $\alpha 1(I)$ chain.

DISCUSSION

In order to better understand the sensitivity of biochemical testing of type I collagen in OI, we have directly compared the electrophoretic migration of collagen synthesised by the dermal fibroblasts of 30 patients with known mutations in the amino-terminal half of the $\alpha 1(I)$ or $\alpha 2(I)$ chain of type I collagen. This selection of probands is by no means exhaustive of the known mutations in this half of the chains and does not create a definitive map of biochemical abnormalities. It does, however, provide an initial insight into the region of the chains in which a substantial number of mutations have been recently identified and in which a high proportion of mutations result in non-lethal phenotypes. A number of patterns emerge from the comparison. Generally, sensitivity of detection was greater for $\alpha 1(I)$ substitutions than for $\alpha 2(I)$ substitutions at about the same helical position. Overmodification was often greater for cell layer than for secreted collagen, presumably due to intracellular retention of abnormal collagen. Overmodification generally decreases the closer the substitution occurs to the amino end of the helix, although a significant number of substitutions do not follow this pattern and raise concern about potential missed diagnoses. From a technical standpoint, sensitivity of biochemical detection of glycine substitutions can be maximised by analysis on 5% SDS-urea-PAGE. Although the tighter bands obtained on 6% SDS-urea-PAGE allowed for detection of subtle baseline shifts of alpha chains, 5% acrylamide gels yielded greater discrimination of overmodified forms.

Substitutions in the first 100 residues of the $\alpha 1(I)$ chain are not biochemically detected. Direct comparison of collagens with substitutions in the first 100 residues of the $\alpha 2(I)$ chains was not possible due to lack of sample availability.

Nevertheless, our data suggest that substitutions in the first 100 residues of the helical region of type I collagen would go undetected by biochemical screening due to the small number of sites at which modification of alpha chains may occur. Detection of overmodification was variable in the $\alpha 1(I)$ G148–232 and the $\alpha 2(I)$ G103–436 regions, and was consistently detected thereafter. For $\alpha 2(I)$ substitutions, in addition to the variable detection region extending further into the chain than for $\alpha 1(I)$, the $\alpha 2(I)$ substitutions are best detected as overmodification of $\alpha 1(I)$.

The variability of biochemical detection sometimes correlates with the substituting amino acid residue. For example, $\alpha 2(I)$ G247C is overmodified while G247S is not. However, there are several instances of the same substituting residue causing overmodification in one location but not in a relatively more carboxyl position, suggesting that overmodification is also a reflection of folding domains.²⁰ For example, $\alpha 1(I)$ G148R occurs just carboxyl-terminal to a lower stability region of the helix and is not overmodified, perhaps because it does not delay the folding of a region that already has a lower local melting temperature. In contrast, G136R is just amino-terminal to the folding domain; it would delay folding of the downstream high T_m region and is biochemically detectable. The same overmodification pattern was seen for the $\alpha 1(I)$ G220A/G187A and G232S/G193S pairs. All tested substitutions with charged side chains were detectable carboxyl to G154 and G436 in $\alpha 1(I)$ and $\alpha 2(I)$, respectively. Type I collagen with a G121D substitution was available for analysis for each alpha chain; only the $\alpha 1(I)$ substitution resulted in detectable overmodification. Polar substitutions were detectable carboxyl to G193 in $\alpha 1(I)$ and G268 in $\alpha 2(I)$. It should be noted, however, that too small a number of cases was tested to be confident that sidechain patterns will be generally true.

Patients with types I and IV OI have relatively mild clinical and radiographic findings, which can make differentiating between OI and non-accidental injury difficult in the very

Figure 2 Electrophoretic migration of collagen with structural defects in the $\alpha 2(I)$ chain of type I collagen. The collagen synthesised by dermal fibroblasts from OI probands with glycine substitutions in $\alpha 2(I)$ was isolated from both media and cell layer samples in a steady state labelling. The type I collagen of each proband was electrophoresed on both 5% and 6% acrylamide gels in comparison to normal control collagen (C) and visualised by autoradiography. The location of the glycine substitution along the chain and the substituting residue are indicated below each proband's sample, for example, G103D. In general, 5% acrylamide gels were more sensitive than 6% gels for detection of overmodification caused by glycine substitutions in the $\alpha 2(I)$ chain. The pattern of type I collagen overmodification due to glycine substitutions in the amino half of the $\alpha 2(I)$ chain differs from that of $\alpha 1(I)$ substitutions. First, glycine substitutions in $\alpha 2(I)$ are associated with normal collagen electrophoretic migrations further along the amino half of the chain than for $\alpha 1(I)$ substitutions. In addition, with the exception of the G421D sample, which contained an $\alpha 2(I)$ doublet with broadened $\alpha 1(I)$ chains, $\alpha 2(I)$ substitutions were best detected as delayed electrophoresis of the $\alpha 1(I)$ chain.

cases in which it is especially important. Biochemical detection of classic type I OI is generally clear. Classic OI type I has a quantitative defect of COL1A1 expression, due to mutations which introduce a premature termination codon.¹ Biochemically, these cases have an increased type III/type I collagen ratio in fibroblast cultures, as compared to control cells; examples of these cases were not shown in this study. Children with mild type IV OI and some cases of type I OI are at risk for a missed biochemical diagnosis if they have a collagen structural mutation in the regions of the alpha chains in which overmodification is undetectable or variably detectable. Because of the crucial importance of distinguishing non-accidental injury and a genetic disorder, complete molecular analysis of COL1A1 and COL1A2 is necessary to exclude a collagen defect.

For our analysis, we began with clinically unequivocal types III and IV OI, whose collagen mutations had been detected with molecular techniques. All patients had structural abnormalities in collagen, namely glycine substitutions in the amino half of the alpha chains. Systematic evaluation of their alpha chain electrophoresis enabled us to establish the regions of $\alpha 1(I)$ and $\alpha 2(I)$ in which substitutions were undetectable or variably detectable on biochemical assay. In the database of type I collagen mutations compiled by the OI Mutation Consortium,²¹ 114/391 (29%) and 149/291 (51%) of independent glycine substitutions are located in the variable detection regions in the first 230 residues of $\alpha 1(I)$ and the first 436 residues of $\alpha 2(I)$, respectively. Thus, approximately 38% (263/682) of all currently known glycine substitutions occur within the variable detection region. The actual percentage of false negative biochemical tests will be

lower, since many abnormalities in the variable detection region can be biochemically identified. Since approximately 52% (136/263) of OI cases in the variable detection region have mild OI (types I, I/IV, IV, or OI/EDS (Ehlers-Danlos syndrome)), about 20% (136/682) of all cases in which the clinical diagnosis might be questionable are at risk for a missed biochemical detection.

This estimate is reasonably close to the rate reported by Wenstrup *et al* over a decade ago using a biochemical approach without molecular confirmation of collagen mutations. They reported that no type I collagen biochemical abnormality could be detected in 13.5% of all non-lethal OI cases (types I, III, III/IV, and IV) that were screened.²² Since the cases in the Wenstrup study included classic type I OI, which was detected by an altered type III/type I collagen ratio, Wenstrup *et al* underestimate the percentage of type I collagen structural mutations that are not electrophoretically detectable. Our study focuses on the sensitivity of the biochemical test for structural mutations of collagen and does not include probands with quantitative defects of type I collagen (type I OI). Hence, the current best estimate of risk for missed biochemical diagnosis involves between 13 and 20% of structural mutations in the type I collagen helical region.

In addition to differentiating non-accidental injury and mild OI,²³ identification of mutations in OI patients is important for genetic counselling, including the patient's knowledge about the cause of their OI, and reproductive issues such as parental testing for mosaic carrier status or providing reassurance that a mutation occurred *de novo*.²⁴ As more mutations are identified, including those at positions

that would have been missed by biochemical screening, more detailed genotype-phenotype correlations can be made with OI type and with clinically important secondary features of OI such as abnormal pulmonary function, basilar invagination, and hearing loss. Mutation identification also enables correlation of genotype with the highly variable response to drug therapies, such as bisphosphonates. In addition, identifying OI patients without collagen mutations will help define non-collagenous causes of OI.

Biochemical screening detects most structural mutations and provides important functional information. Our results, however, highlight the importance of an accurate clinical diagnosis, as conventional biochemical screening of patients might not provide confirmation of a type I collagen structural abnormality. Sequencing of the COL1A1 and COL1A2 amino end cDNA or genomic DNA of probands in whom no biochemical abnormality is seen can identify probands with (a) glycine substitutions that do not cause overmodification and (b) splice site mutations that yield shortened alpha chains that comigrate with normal chains due to overmodification.²⁵ We suggest that the most prudent diagnostic approach for complete identification of mutations is to supplement a normal biochemical assay with molecular analyses in the amino-third of $\alpha 1(I)$ and the amino-terminal half of $\alpha 2(I)$.

ACKNOWLEDGEMENTS

We are grateful to the children and families of the NIH OI research program for their continued dedication to research on osteogenesis imperfecta, and for their participation in ongoing research.

Authors' affiliations

W A Cabral, S Milgrom, A D Letocha, E Moriarty, J C Marini, Bone and Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD, USA

Competing interests: none declared

Ethics approval: all specimens were obtained from individuals enrolled in IRB approved protocols which included provisions for skin biopsies and molecular testing

Correspondence to: Dr Joan C Marini, Section on Connective Tissue Disorders, Bone and Extracellular Matrix Branch, NICHD, NIH, 9000 Rockville Pike 10/9S241, Bethesda, MD 20892, USA; oidoc@helix.nih.gov

Revised version received 17 February 2006

Accepted for publication 21 February 2006

REFERENCES

- 1 Willing MC, Deschenes SP, Scott DA, Byers PH, Slayton RL, Pitts SH, Arikat H, Roberts EJ. Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am J Hum Genet* 1994;**55**(4):638–47.
- 2 Prockop DJ, Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 1995;**64**:403–34.
- 3 Byers PH, Cole WG. Osteogenesis imperfecta. In: Royce PM, Steinmann B, eds. *Connective tissue and its heritable disorders*. 2nd ed. New York: Wiley-Liss, 2002:385–430.
- 4 Silience DO, Senn A, Danks DM. Genetic heterogeneity in osteogenesis imperfecta. *J Med Genet* 1979;**16**(2):101–16.
- 5 Marini JC. Osteogenesis imperfecta. In: Behrman RE, Kliegman RM, Jenson HB, eds. *Nelson textbook of pediatrics*. 17th ed. Philadelphia: Saunders, 2004:2336–8.
- 6 Grange DK, Gottesman GS, Lewis MB, Marini JC. Detection of point mutations in type I collagen by RNase digestion of RNA/RNA hybrids. *Nucleic Acids Res* 1990;**18**(14):4227–36.
- 7 Saleeba JA, Ramus SJ, Cotton RG. Complete mutation detection using unlabeled chemical cleavage. *Hum Mutat* 1992;**1**(1):63–9.
- 8 Bateman JF, Lamande SR, Hannagan M, Moeller I, Dahl HH, Cole WG. Chemical cleavage method for the detection of RNA base changes: experience in the application to collagen mutations in osteogenesis imperfecta. *Am J Med Genet* 1993;**45**(2):233–40.
- 9 Mackay K, Byers PH, Dalgleish R. 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* 1993;**2**(8):1155–60.
- 10 Giunta C, Steinmann B. Characterization of 11 new mutations in COL3A1 of individuals with Ehlers-Danlos syndrome type IV: preliminary comparison of RNase cleavage, EMC and DHPLC assays. *Hum Mutat* 2000;**16**(2):176–7.
- 11 Biggin A, Holman K, Brett M, Bennetts B, Ades L. Detection of thirty novel FBN1 mutations in patients with Marfan syndrome or a related fibrillinopathy. *Hum Mutat* 2004;**23**(1):99.
- 12 Korkko J, Ala-Kokko L, De Paepe A, Nuytinck L, Earley J, Prockop DJ. Analysis of the COL1A1 and COL1A2 genes by PCR amplification and scanning by conformation-sensitive gel electrophoresis identifies only COL1A1 mutations in 15 patients with osteogenesis imperfecta type I: identification of common sequences of null-allele mutations. *Am J Hum Genet* 1998;**62**(1):98–110.
- 13 Ganguly A. An update on conformation sensitive gel electrophoresis. *Hum Mutat* 2002;**19**(4):334–42.
- 14 Bonadio J, Holbrook KA, Gelinas RE, Jacob J, Byers PH. Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. *J Biol Chem* 1985;**260**(3):1734–42.
- 15 Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;**162**(1):156–9.
- 16 Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;**239**(4839):487–91.
- 17 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;**74**(12):5463–7.
- 18 Cabral WA, Makareeva E, Colige A, Letocha AD, Ty JM, Yeowell HN, Pals G, Leikin S, Marini JC. Mutations near amino end of alpha1(I) collagen cause combined osteogenesis imperfecta/Ehlers-Danlos syndrome by interference with N-propeptide processing. *J Biol Chem* 2005;**280**(19):19259–69.
- 19 Mayer SA, Rubin BS, Starman BJ, Byers PH. Spontaneous multivessel cervical artery dissection in a patient with a substitution of alanine for glycine (G13A) in the alpha 1(I) chain of type I collagen. *Neurology* 1996;**47**(2):552–6.
- 20 Persikov AV, Ramshaw JA, Brodsky B. Prediction of collagen stability from amino acid sequence. *J Biol Chem* 2005;**280**(19):19343–9.
- 21 Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, Milgrom S, Hyland JC, Korkko J, Prockop DJ, De Paepe A, Coucke P, Symoens S, Glorieux FH, Roughley PJ, Lund AM, Kuurila-Svaen K, Hartikka H, Cohn DH, Krakow D, Mottes M, Schwarze U, Chen D, Yang K, Kuslich C, Troendle J, Dalgleish R, Byers P. Consortium for Osteogenesis Imperfecta mutations database of glycine substitutions and exon skipping defects in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Human Mutat* 2006 (accepted for publication).
- 22 Wenstrup RJ, Willing MC, Starman BJ, Byers PH. Distinct biochemical phenotypes predict clinical severity in nonlethal variants of osteogenesis imperfecta. *Am J Hum Genet* 1990;**46**(5):975–82.
- 23 Marlowe A, Pepin MG, Byers PH. Testing for osteogenesis imperfecta in cases of suspected non-accidental injury. *J Med Genet* 2002;**39**(6):382–6.
- 24 Pepin M, Atkinson M, Starman BJ, Byers PH. Strategies and outcomes of prenatal diagnosis for osteogenesis imperfecta: a review of biochemical and molecular studies completed in 129 pregnancies. *Prenat Diagn* 1997;**17**(6):559–70.
- 25 Cabral WA, Fertala A, Green LK, Korkko J, Forlino A, Marini JC. Procollagen with skipping of alpha1(I) exon 41 has lower binding affinity for alpha1(I) C-telopeptide, impaired in vitro fibrillogenesis, and altered fibril morphology. *J Biol Chem* 2002;**277**(6):4215–22.