The Human Collagen Mutation Database 1998

Raymond Dalgleish*

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Received October 2, 1997; Accepted October 24, 1997

ABSTRACT

The collagens are a large and diverse family of proteins which are found in the extracellular matrix. In common with one another, the 19 known collagen types have triple-helical domains of variable length but they differ with respect to their overall size and the nature and location of their globular domains. Collagen mutations lead to heritable defects of connective tissues and mutation data for collagen types I and III are presented here. The mutation data are accessible on the worldwide web at http://www.le.ac.uk/genetics/collagen/

INTRODUCTION

The collagens are a complex family of structural proteins comprising at least 19 types (1,2). Some types are abundant and ubiquitous though the majority are minor components with limited tissue distribution and are specialised in their functions. Mutations in the genes encoding the individual collagen α-chains often lead to defects of the extracellular matrix. These mutations are generally dominant and data exist for most of the more abundant collagen types with type I collagen mutations, resulting predominantly in osteogenesis imperfecta (OI), being the most common. This paper updates the data previously presented (3) concerning type I collagen mutations and introduces data for type III collagen where mutations mostly result in Ehlers–Danlos syndrome type IV (EDS IV).

TYPE I COLLAGEN MUTATIONS: AN UPDATE

Since the publication last year of the initial listings of type I collagen mutations (3), a number of additions and corrections have been made. These are summarised in Table 1 and the complete data, summarised in Table 2, can accessed from the web pages.

The additions

The most significant of the additional data are the large number of frameshifts in COL1A1 which lead to the mild type I OI. In each instance, the frameshift results in a premature stop codon being introduced into the normal reading frame. This, in turn, results in translation products lacking a C-terminal propeptide and so preventing their incorporation into collagen triple helices. Such truncated collagen chains are degraded intracellularly. The publication of these data lend strong support to the notion that the majority of cases of type I OI are due to the presence of reduced amounts of normal collagen in the extracellular rather than to the interfering effect of mutant collagen chains (4).

Table 1. Summary of revised data for COL1A1 and COL1A2

<table>
<thead>
<tr>
<th></th>
<th>COL1A1</th>
<th>COL1A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid substitutions</td>
<td>22 + 8</td>
<td>6 + 7</td>
</tr>
<tr>
<td>Exon skipping and other splicing mutations</td>
<td>5 + 4</td>
<td>2 + 1</td>
</tr>
<tr>
<td>Deletions, insertions, duplications and frameshifts</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Polymorphisms</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

In each case the first (or only) number relates to new data where that precise mutation has not been previously reported. The second number represents corrections to data and reports of new instances of mutations that had previously been recognised and reported. For polymorphisms, only the first or definitive report is recorded.

Table 2. Summary of mutation data for COL1A1, COL1A2 and COL3A1

<table>
<thead>
<tr>
<th></th>
<th>COL1A1</th>
<th>COL1A2</th>
<th>COL3A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid substitutions</td>
<td>95 (10)</td>
<td>45 (6)</td>
<td>40 (5)</td>
</tr>
<tr>
<td>Exon skipping and other splicing mutations</td>
<td>17 (4)</td>
<td>19 (2)</td>
<td>28 (4)</td>
</tr>
<tr>
<td>Deletions, insertions, duplications and frameshifts</td>
<td>25</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Polymorphisms</td>
<td>14</td>
<td>35</td>
<td>3</td>
</tr>
</tbody>
</table>

The first (or only) number records the number of uniquely different mutations of each type that have been reported in the literature. The second (in brackets) records the number of those for which more than one instance has been reported. For polymorphisms, only the first or definitive report is recorded.

The other notable additions are the Gly13Ala substitution in COL1A1 (5) and the Gly751Ser substitution in COL1A2 (6). The former is of interest because it resulted in cervical artery dissection in a patient who exhibits few other manifestations of connective tissue apart from blue sclerae and also because neurovascular complications are rare in OI (7). The Gly751Ser mutation in COL1A2 was identified in a family in which there are both heterozygotes and homozygotes, with type III OI being the resulting phenotype in the latter. The heterozygotes had much milder symptoms and it is perhaps surprising that the homozygous form is not lethal given that no normal α2(I) collagen chains and, hence, no normal type I collagen is produced. Such

*Tel: +44 116 252 3425; Fax: +44 116 252 3378; Email: ray@le.ac.uk
families are rare and only two previous accounts are found in the literature (8,9).

Finally, a three base deletion in COL1A1 resulting in a Glu437Asp substitution along with the deletion of Ala438 (10) is worthy of comment. The family in which the mutation has occurred exhibits extreme variation of the clinical manifestations among the affected mother and her four affected offspring and again emphasises that the primary mutation is not the sole determinant of the resulting phenotype.

The corrections
The corrections are of various sorts. In some instances, data that had previously been overlooked are now included. Where more than a single unrelated person harbouring a particular mutation is described in a paper, each individual occurrence is now correctly reported. The skipping of exon 30 in COL1A1 (11) corrects the previously mistaken reporting of it being an exon 22 skip due to use of the now outmoded 3′ to 5′ exon-numbering system. The Pro27Ala substitution in COL1A1 (12) was previously reported as a polymorphism but is almost certainly the cause of the osteopaenia exhibited by the two individuals harbouring the substitution.

TYPE III COLLAGEN
The protein
Type III collagen is a homotrimer of α1(III) collagen chains and is expressed in many tissues but is primarily a component of extensible connective tissues such as skin, lungs, gut, vascular system and uterus. It is often co-expressed with type I collagen but is not found in bone (13).

The gene
The α1(III) chains of type III collagen are encoded at the locus COL3A1 which is located at 2q24.3–q31 (14,15). The gene is 44 kb in length and it comprises 52 exons whose organisation are comparable to those of COL1A1 and COL1A2 which encode the α-chains of type I collagen.

Amino acid numbers
As with the α1 and α2 chains of type I collagen, the α1 chain of type III collagen comprises three domains; the N-terminal domain, the collagen domain and the C-terminal domain. With respect to our knowledge of type III collagen mutations, only the collagen domain is of current importance. This domain comprises the 1029 amino acids of the 343 Gly-X-Y repeats with extension N- and C-telopeptides of 14 and 25 amino acids, respectively. The five extra Gly-X-Y repeats compared with the α1 and α2 chains of type I collagen are accommodated as follows: two each in exons 6 and 49 and one in exon 7. By convention, amino acid 1 is the first glycine of the first Gly-X-Y repeat of the collagen domain.

Exon and DNA sequence numbers
The collagen-domain exons of COL3A1 are exons 6–49 and hence are broadly analogous to those of COL1A2, the only differences being those noted above.

The reference cDNA sequence (16) for type III collagen mutations has the accession number X14420 with the first base being numbered 1. The start codon then begins at 103 and the first base of the repeating Gly-X-Y region is 604. This numbering system is at odds with the way in which some type III collagen mutations have been reported in the literature but is used for consistency with the system adopted for type I collagen mutations. For example, the Gly934Glu substitution results from a G→A transition at position 3404 but is reported as being at position 3302 (17). Hence, the positions quoted for mutations in the two numbering systems will differ by 102. Mutations in intron donor or acceptor sequences, leading to exon skipping, are reported relative to the start or end of the intron (e.g., G+A→A or A_2→T).

TYPE III COLLAGEN MUTATIONS
Mutations of type III collagen are dominant negative due to the disrupting effects of the incorporation of one or more defective α1(III) chains into type III collagen triple helices and hence are broadly analogous to those of type I collagen (1,18). However, whereas heterozygous mutations in COL1A1 and COL1A2 result respectively in 75% and 50% abnormal type I collagen molecules, COL3A1 mutations result in an even higher proportion of defective molecules. As type III collagen is a homotrimer, entirely normal molecules will only be assembled in one eighth of cases. Type III collagen mutations also differ from those of type I collagen in that those reported so far are confined to the collagen domain and hence none are analogous to the type I mutations that result in Ehlers–Danlos syndrome types VIIA and VIIB.

Most type III collagen mutations result in the life-threatening connective tissue disorder Ehlers–Danlos syndrome type IV (EDS IV) (19,20). This disorder is characterised by easy bruising and bleeding and in many instances results in sudden and catastrophic rupture of blood vessels or other hollow organs. Outward signs of the disorder include characteristic facies and thin skin that scars easily and allows visibility of underlying blood vessels. Individual cases are often referred to as being mild or severe depending on the symptoms. A few mutations result in disorders whose primary manifestations are arterial aneurysms but the proportion of cases of familial aneurysms, in the absence of other signs of connective disorders, due to type III collagen mutations is small (21,22).

The mutations include amino acid substitutions, alterations to splice donor and acceptor sequences and a range of small and large deletions. Phenotype genotype correlations are no simpler for type III collagen mutations than for type I though, as a generalisation, amino acid substitutions toward the C-terminal end of the collagen domain result in more severe forms of EDS IV. One type III collagen mutation is of particular note. It is a 9 kb internal deletion removing the 15 exons encoding amino acids 586–999 (23). Remarkably, it results in a mild form of EDS IV, suggesting that molecules containing shortened α-chains might be partly functional.

ACCESSING THE DATA
The collagen mutation data may be accessed on the University of Leicester web server at http://www.le.ac.uk/genetics/collagen/ (Fig. 1). Users are encouraged to use this new URL rather than the previously published one. The home page now includes a web
Figure 1. Home page of the Human Collagen Mutation Database.

form to allow users to report errors to or request further information from the database curator. The data are being placed in a relational database which will eventually be linked to the web pages to allow users to submit queries. If you make use of the data from the web server, please cite both this and the previous article (3) in any materials which you prepare for publication.

ACKNOWLEDGEMENTS

I am grateful to Laura Kempton who helped in the initial compilation of the type III collagen data presented here. I would also like to thank those members of the OI Mutation Consortium who have pointed out errors and omissions in earlier reports of the type I collagen data.

REFERENCES