Current topic

Molecular abnormalities of collagen in human disease

F M POPE AND A C NICHOLLS
MRC Clinical Research Centre, Harrow, Middlesex

Collagen proteins are a diverse family of structural materials serving a mechanical supportive role in widely varied tissues such as skin, bone, muscle, blood vessels, cartilage, pleuroperitoneal linings, hollow tubular organs, tendons, heart valves, and basement membranes. They may also possess a crucial directive role in embryogenesis and fetal development. The molecules are large, ranging between 50 and 200 Kd. There is ample scope for major mutations, which produce a diverse group of multisystem diseases relevant to all medical specialities. Type I collagen mutations have been implicated in hereditary brittle bone disease, which ranges in severity from a mild premature osteoporosis to lethal childhood dwarfism. Type III collagen mutations occur with Ehlers-Danlos syndrome type IV and related disorders and present with arterial rupture (of large and medium sized arteries), intestinal diverticula, and mitral valve prolapse.

Recent spectacular advances in molecular biology, particularly with recombinant deoxyribonucleic acid (DNA) techniques such as gene cloning, restriction enzyme cleavage of DNA molecules, DNA sequencing, and DNA transfer (by microinjection) into other animal cells, in principle allow the precise dissection and analysis of any mammalian gene.1-3 Even if the gene is unidentified, random DNA sequence variations occur sufficiently close to the disease locus to be used as predictive linkage markers in genetic pedigree analysis of both autosomal dominant and recessive diseases. Such an approach has been brilliantly successful in the molecular dissection and prenatal diagnosis and prevention of the various α and β thalassaemias.4 has allowed accurate predictive diagnosis of Huntington's chorea,4 and most recently has identified several markers close to the cystic fibrosis gene.5 6

McKusick lists over 3000 single human gene disorders,7 and certainly we can expect a golden era for the accurate scientific diagnosis and analysis of a substantial number of single gene human diseases as well as the molecular unravelling of the various disorders of growth and development and the closely related mechanisms for carcinogenesis and disordered growth. Collagen structural genes are now all amenable to molecular analysis of this type and in the next few years we can expect that all the relevant structural genes and proteins will become cloned, sequenced, and located to specific chromosomes. Already the human genes for α1(I), α2(I), α1(II), α1(III), α1(IV), α1(V), α1(IX), and α1(X)14-15 have been at least partially cloned and sequenced and others may be expected to follow rapidly. Various diseases of collagen proteins, including osteogenesis imperfecta, Ehlers-Danlos syndrome types I, II, III, IV, and VII. Marfan’s syndrome, dystrophic epidermolysis bullosa, congenital aneurysms of the circle of Willis, some types of osteopetrosis, osteoarthritis, and probably some forms of congenital heart disease can increasingly commonly be expected to show specific collagen gene and protein mutations. Using a combination of fibroblast cultures from 2-4 mm punch biopsy specimens, leucocyte enrichment from blood samples supplemented occasionally with cartilage, endothelial, and keratinocyte culture, most or all of them may be expected to have specifically identified gene defects in the near future. As such they will then be preventable in principle and in practice.

Collagen genes and proteins

Ten main collagen types have been identified so far and more are likely. The most common and widespread are the so called interstitial collagens, such as types I, II, III, and probably type V proteins,16 which share a tightly coiled triple helical central portion with N and C globular extensions and assemble into cross banded fibres. These substances are widely distributed in bones, skin, tendon, blood vessels, and cartilage. Other atypical proteins with globular elements interspersed along the collagen helix have also been identified. Type IV17 is a good
example, consisting of at least two protein chains, which form the main collagenous components of the lamina densa of basement membrane. Types VI, VII, VIII, IX, and X collagens are fairly minor components of arterial basement membrane, anchoring fibrils and epiphyseal cartilages, respectively.

Type I collagen is a heteropolymer with two genetically distinct component α chains forming substantial proportions of skin, bones, arteries, tendons, and the gastrointestinal tract. Type II collagen occurs only in articular cartilage, and type III collagen is the main collagenous component of arteries, veins, and pleuroperitoneal linings (although it can also be detected in skin, gastrointestinal tract, and heart valves). Type IV collagen is widely distributed in the basement membranes of skin, lungs, gastrointestinal tract, kidneys, and blood vessels.

The interstitial collagens (I, II, III, and V) are enclosed by large genes, which range from 28–40 Kbp, with 50 intervening sequence (introns), 5 Kbp of coding sequence (exons), and substantial homology at the 3’ (C terminal) and 5’ (N terminal) ends. The precursor procollagens of 120 Kd are subsequently processed into the mature protein by excision of the two terminal extensions by specific C and N terminal propeptidases. Mature triple helical peptides then pack in a quarter D stagger to form the characteristic cross striated interstitial collagen fibrils.

The general structure of the helical protein is a repeating polymer of 330 glycine (Gly) XY sequences. Glycine regularly occurs in every third position to form a highly conserved central spine from which the X and Y residues project (X is often proline (10%) and Y often lysine (4%)). The interstitial genes are highly conserved and the exons are usually coded by multiples of nine base pairs (coding for the ancestral Gly XY tripeptide unit), most often of 54 base pairs, although varying from 45 to 162 base pairs. Presumably the Gly XY repeat is crucial, for the specific physical properties of the rigid collagen triple helix and possibly glycine mutations are either lethal or severely disabling. Particular post-translational modifications include lysine and proline hydroxylation, glycosylation of asparagine residues, and various crosslinkages between hydroxylated lysines. Genetic mutations of collagen α helices resemble those causing the haemoglobinopathies. They include amino acid substitutions, gene insertions and deletions and various regulatory mutations with diminished or absent normal or mutant protein production. Deficiencies of enzymes catalysing collagen cross-linking or excision of either of the terminal pro-peptides have also been documented. So far the only convincing molecular defects have affected the genes and proteins for the α1(I), α2(I), and αⅢ(I) products, although a type VII gene defect is strongly suspected in dystrophic epidermolysis bullosa and a family of large polymorphic deletions identified for the 3′ end of the α1(II) gene. There may be technical problems in identifying specific protein mutations for cartilage collagen and basement membrane proteins (and genes) because of the difficulty of epithelial and cartilage cell culture from conventional biopsy specimens. Perhaps basement membrane and other mutations would act as genetic lethals in heterozygotes and homozygotes and would not be identifiable for this reason.

**Type I collagen mutations**

These characteristically produce inherited bone fragility (osteogenesis imperfecta) but have also been recently associated with Marfan’s syndrome and Ehlers-Danlos syndrome type VII. These diseases share specific clinical features, such as hyperextensible skin and joints, blue sclera, pectus excavatum, scoliosis, and high palatal arches. Important distinguishing features include brittle, easily fractured bones, deafness and short stature (osteogenesis imperfecta), arachnodactyly, lens dislocations, and aortic rupture (Marfan’s syndrome), and short stature and extreme joint extensibility associated with easily bruised fragile skin (Ehlers-Danlos syndrome type VII). Osteogenesis imperfecta (OI) is surprisingly diverse genetically, clinically, and especially biochemically. Silence classified it into four main clinical groups, two of which are mild, autosomal dominant and of late onset, the others being severe, often lethal, and otherwise crippling. They are of early onset and are sometimes autosomal recessive (types II and III OI). Patients with type I OI have blue sclera in contrast to their type IV counterparts, who are white eyed. Both types are fairly mild diseases that are generally manifest after birth (OI tarda) and improving at puberty. Both types have specific variants, which breed true for the presence or absence of dentinogenesis imperfecta: unusually brittle discoloured teeth with disorganised, sparse dental tubules. Bone collagen is almost entirely type I, with traces of types III and V collagens, and all the convincing molecular defects so far identified have affected α1(I) or α2(I) collagen proteins and genes or have been in tight linkage disequilibrium with one of them. There is no consistent correlation between the clinical type of OI and its molecular diseases. Thus cysteine point mutations can cause both lethal type II OI and mild OI tarda.
Similarly, diminished type I collagen synthesis with increased III:1 protein ratios can affect patients with lethal OI and mild OI tarda.\textsuperscript{39, 40} Other defects include a 500 base pair deletion in OI type IIa\textsuperscript{26} and an arginine substitution in the C terminal propeptide associated with severely crippling type III OI (by interfering with mannosylation and secretion of the C propeptide),\textsuperscript{31} a four base nonsense deletion in the C propeptide,\textsuperscript{29, 42, 43} causing severely crippling OI type III (and premature osteoporosis in his heterozygous but clinically normal parents). A structural α2(I) mutation in type II OI\textsuperscript{44} and an N terminal α2(I) deletion in type IV OI have also been documented.\textsuperscript{45}

Except for what Prockop calls suicide mutations, in which one of the three triple helical components disrupts its two normal companions (disturbing the proper winding, assembly, and melting profile of the fibre), there are no satisfactory biophysical explanations of how rather similar mutations can cause both OI tarda and OI lethal variants. An excellent example is the glycine to cysteine substitution, which causes lethal OI with an altered collagen type I melting profile (Tm), compared with a cysteine substitution, causing mild OI and no melting (helical disorganisation) of the collagen triple helix.\textsuperscript{37, 38} Possibly arginine or serine have mutated to cysteine in this instance, producing milder effects on the mature collagen molecule. A third mutation has been described, however, in which a cysteine substitution causes an altered melting profile but a mild form of disease. (de Wet WJ. Personal communication, 1985.) This raises the possibility that sometimes a second (so far undetected) mutation either in the collagen α1(I) or some other major—for example, α1(II)—or minor cartilage collagen component, such as type X or type IV, could be causing a more severe compound disease. Alternatively, the cysteine substitutions are not the direct cause of the disease in this instance. Similar explanations may apply to those forms of lethal OI and mild tarda OI in which reduced III/I collagen protein and messenger ribonucleic acid concentrations cause lethal and mild disease with very similar mutations.

**Type III collagen mutations**

Defects of these genes and proteins are characterised by arterial rupture but have a disease range from a lethal premature aging syndrome (Ehlers-Danlos syndrome type IV—acrogeric pattern) to symptomless relatively normal looking patients with a mild deficiency accompanied by the benign hypermobile syndrome (Ehlers-Danlos syndrome type III) with superadded sacral striae, aortic rupture, and mitral valve prolapse. We have seen very similar biochemical deficiency in one British and two Belgian patients with Ehlers-Danlos syndrome type I. Patients with Ehlers-Danlos syndrome type IV have a strong predisposition to the lethal and unpredictable rupture of large or medium sized arteries. They share the common clinical feature of thin skin, premature aging, a typical facies with thin lips and unusually prominent eyes, abnormally easy bruising, cutaneous venous visibility, and capillary telangiectasia. Type III collagen protein is diminished to 5–50% of expected concentrations as measured by scanning polyacrylamide gel electrophoresis of glycine and proline \(^{1}H\) or \(^{13}C\) labelled skin fibroblast collagens. Yet other variants produce mutant type III proteins\textsuperscript{23} and perhaps a wide range of mutant proteins with amino acid substitutions, gene deletions, and inserts similar to those observed in OI await improved methods of protein fingerprinting and gene analysis. Certainly, phenocopies of type III deficiency with many of the clinical features of Ehlers-Danlos syndrome type IV but normal type III concentrations are common. We have also observed convincing clinical disease associated with normally migrating type III proteins tightly linked to polymorphic markers for the relevant gene (see below).

**Other collagen protein deficiencies**

So far no other disease linked mutations have been detected for a variety of other collagens, including types II, IV, V, VII, VIII, IX, and X proteins. We would expect type II collagen changes to be associated with short limbed dwarfsisms, such as achondroplasia, pseudoachondroplasia, or spondyloepiphyseal dysplasia. The evidence linking the type II gene with achondroplasia is controversial, some authors having found positive linkage, while others have observed negative linkage in this disease.\textsuperscript{46, 47} Type IV and VII collagen defects are good candidates for abnormalities in some lethal forms of epidermolysis bullosa, type X collagen may be associated with some forms of lethal osteogenesis imperfecta, and types VI and VIII collagen deficiency are causal candidates for certain arterial diseases, such as aortic and circle of Willis aneurysms as well as the Marfan’s syndrome.

**Restriction fragment length polymorphisms (RFLP)**

Random variability in DNA sequences close by or within structural genes have been recognised as possible markers for human genetic disease. These are neutral changes that are usually causally unrelated to the disease but sufficiently close to the
mutant gene to avoid meiotic crossing over. An excellent example is the haptoglobin I restriction fragment variable site, which is tightly linked to the sickle cell gene and which will detect this gene reliably in 70% of affected black Americans. This approach can be used whether or not the mutant gene is identified. Two recent exciting examples have linked important disabling genetic diseases such as Huntington’s chorea and cystic fibrosis to RFLP, which segregate close enough to the candidate gene to be closely linked with it. The way is now clear for accurate prenatal diagnosis and prevention as well as the possibility of identification and analysis of the unknown gene in the foreseeable future. This approach has been useful in the following inherited abnormalities of collagen.

1. α2(I) Gene polymorphisms: Eco RI, MspI useful in type IV OI and implying the involvement of this gene in that disease; linkage with type I OI has also been recently described. Bgl II and Eco RI markers are also available for this gene.

2. α1(III) Polymorphisms: an Eco RI polymorphism has been described in two American families with lethal autosomal dominant type III collagen deficient Ehlers-Danlos syndrome type IV. A Msp/Ava I polymorphism maps to the same region of the gene and has been successfully used to show linkage in a large Belgian kindred with a disease variant of Ehlers-Danlos syndrome type IV. (DePaepe A, Nicholls AC, Narcisi P, et al. Linkage of a polymorphic marker for the type III collagen gene (Co13A1) to autosomal dominant Ehlers-Danlos syndrome type IV in a large Belgian pedigree. 1986.)

3. α1(II) Polymorphisms: so far polymorphisms have been identified in this large 38 Kb gene, including Hind III and PvuII, polymorphisms of the 9.2 Kb Eco 1 fragment, which has not segregated with either dominant OI or Ehlers-Danlos syndrome type II. Secondly, a common Eco/RBam HI polymorphism localised very close to the 3′ end of the gene shows substantial variability, probably due to deletions of segments of a hypervariable site. The deletions range from 30–300 base pairs with occasional insertions in this site. Deletions of these fragments have segregated both with lethal OI in some white families as well as occurring in 10% of the British Asian populations, presumably for reasons of genetic drift (selective advantage). We have also studied a Bam HI intragenic point polymorphism, which we have observed in over 20% of white patients with osteoarthritis compared with 10% and 0%, respectively, of two separate control samples.

4. Polymorphisms associated with the pro α1(I) gene have been remarkably elusive. Sykes and his colleagues have just published two polymorphisms associated with the 5′ end of the gene. One, an Msp variable site, lies 30 Kb 5′ to the gene and has a population frequency of 0.23, while the other is less common and is an Rsa polymorphism within the 5′ end of the gene.

By using these pro α1(I) together with the pro α2(I) markers mentioned above, Sykes et al have confirmed the association of Silence type I OI with both pro α1(I) and pro α2(I) collagen genes. The former had of course been implicated with this disease on the basis of protein data, while the latter has been suspected both on protein and linkage data for several years.

Other markers

The various cloned collagen genes are dispersed in several autosomal chromosomes, which include chromosome 17–α1(I) 7, α2(I), 2α1(III), α1(V), 12 α2(I), and 13 α1(IV). Markers close to the collagen genes on those chromosomes can also be used to study disease segregation and we can confidently expect that the genes for α1(VI), α1(VII), α1(VIII), α1(IX), and α1(X) will be similarly traced in the foreseeable future.

Prenatal diagnosis and prevention of all the specific molecular defects described above is already a practical possibility and the next few years should witness substantial advances in the management and prevention of common and uncommon diseases caused either directly by collagen gene mutations or lying close to them. These will include osteogenesis imperfecta, various Ehlers-Danlos mutants, some forms of Marfan’s syndrome, certain short limbed dwarfsms, berry aneurysms of the circle of Willis, osteoarthritis, osteoporosis, keloid scarring, elastin gene defects, cystic fibrosis, and many others.

References

Molecular abnormalities of collagen in human disease  527

38 Barsh GS, David K, Byers PH. Type I osteogenesis imperfecta, a nonfunctional allele for pro α(I) chains of type I procollagen. Proc Natl Acad Sci USA 1982;79:3838–42.
528 Pope and Nicholls


Correspondence to Dr F M Pope, Dermatology Research Group, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, England.
Molecular abnormalities of collagen in human disease.

F M Pope and A C Nicholls

Arch Dis Child 1987 62: 523-528
doi: 10.1136/adc.62.5.523

Updated information and services can be found at:
http://adc.bmj.com/content/62/5/523.citation

These include:

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/