THE MOLECULAR GENETICS OF INHERITED CONNECTIVE TISSUE DISORDERS

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

John Ross Hawkins BSc (Newcastle)

Department of Genetics

University of Leicester

July 1989

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

PUBLICATIONS

Some of this work has been published:

Dalgleish, R., Hawkins, J.R. and Keston, M. (1987) Exclusion of the α2(I) and α1(III) collagen genes as the mutant loci in a Marfan syndrome family.
J. Med. Genet. 24: 148-151

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Raymond Dalgleish for his constant guidance and ideas, Tom Caskey for inspiration and Richard Gibbs much needed technical advice.
I am especially grateful to Dr. Nicola Royle for putting up with me and supplying me constantly with good advice, scientific or otherwise.
Thanks to Alec Jeffreys, Robert Kelly and Justin Thackery for welcome suggestions and comments, and to all the other G19'ers for a friendly and helpful environment.
Thanks also to Helen Dawson for computing assistance.
I must acknowledge the receipt of an SERC studentship.
Finally I would like to express my gratitude to Louise for keeping me happy while writing this.

This thesis is dedicated
to my mother and father

and to the parents of F.B.

ABBREVIATIONS

bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
Ci	Curie
cpm	counts per minute
DEPC	diethylpyrocarbonate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	diaminoethanetetra-acetic acid
EGTA	ethyleneglycol-bis-(β-aminoethylether) N,N,N',N'-tetraacetic acid
IPTG	isopropyl-β-D-thiogalacto-pyranoside
kb	kilo base pairs
MOPS	3-(N-morpholino)propanesulphonic acid
NTP	nucleoside triphosphate. N= adenosine, guanosine, uridine, cytidine
dNTP	2'-deoxynucleoside triphosphate. N= adenosine, guanosine, thymidine, cytidine
ddNTP	2'3'-dideoxynucleoside triphosphate
PBS	phosphate buffered saline
PEG	polyethylene glycol
RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
TEMED	N,N,N',N'-tetramethyl-ethylinediamine
Tris	tris(hydroxymethyl)aminomethane

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 CONNECTIVE TISSUES

The term connective tissue is often used to describe the extracellular matrix and the cells found in it, such as fibroblasts, macrophages and mast cells. The extracellular matrix contains three major fibre-forming proteins: collagen, elastin and fibronectin, which are embedded in a hydrated gel formed by a network of aminoglycan chains.

The amount of connective tissue in the body varies greatly: skin and bone are composed mainly of connective tissue, whereas the brain and spinal cord contain very little. Moreover, the relative amounts of the different types of matrix macromolecules and the way that they are organised within the extracellular matrix vary enormously, with each form adapted to the functional requirements of the particular tissue. The matrix can become calcified to form very hard structures such as bone and teeth, or it can form thin and highly flexible membranes such as the amnion.

1.2 COLLAGEN

Collagen is the most abundant protein in the animal kingdom and is the major protein of the connective tissues. A distinctive feature of collagen is that it forms insoluble fibres that have a very high tensile strength. Its main function is in the provision and maintenance of physical support for extracellular matrices. In addition it is

apparent that the collagens have several physiological and developmental roles.

This chapter contains a brief description of the collagens, collagen genes, diseases apparently caused by abnormal collagen and methods to analyse abnormal collagen genes.

1.2.1 Structural features of the collagens

Detailed reviews of the collagens have been presented by Fessler and Fessler, (1978); Bornstein, (1980); Mayne and Burgeson, (1987) and Miller and Gay (1987).

The fundamental structural unit of collagen is a long, thin (1.5nm diameter) protein that consists of three coiled polypeptide subunits called α chains. Each of the three α chains of the collagen molecule is coiled into a left-handed helix with about three amino acids per turn. The three α chains are then twisted around each other into a right-handed super-helix. This triple-helix is the most important identifying feature of collagen and is often called the 'collagenous domain'. Each α chain contains lengthy sequences of a repeating glycine-X-Y motif in which the X and Y positions are often occupied by proline and hydroxyproline residues respectively. The presence of glycine, the smallest amino acid, in every third position is crucial as the amino acid in this position occupies the restricted space in which the three α chains come together in the centre of the triple helix. Because both proline and hydroxyproline are rigid, cyclic amino acids, they limit the rotation of the polypeptide backbone and thus contribute to the stability of the triple helix. If the collagenous domain consisted solely of the amino acid triplet $(Gly\text{-}Pro\text{-}Hyp)_n$, the molecule would be extremely stable. However, only about one-third of the amino acids in the X and Y positions are

proline and hydroxyproline. Other amino acids may decrease the stability of the triple helix but are essential for the organisation at the next level, which is the assembly of the triple helical molecules into collagen fibrils or non-fibrillar arrays.

The physiochemical properties of the known collagens provide the basis for their division into different collagen types. To date, 13 collagen types have been recognised and characterised to varying degrees. Because of the complexity of the collagen family Miller, (1985) has separated the collagen types into three broad groups (see Table 1.1).

Group 1 comprises collagen types I, II, III, V and XI. These molecules consist largely of a single uninterrupted helical domain approximately 300nm in length. The group 1 molecules are the fibril-forming or 'fibrillar' collagens. The fibrils have a diameter of between 0.2-0.5μm and are formed into bundles termed collagen fibres, the diameters of which may vary between 1-20μm depending on the number of fibrils they contain. The molecular architecture of the fibres is best considered with reference to a microfibrillar unit. X-ray diffraction studies revealed that there are five molecules per microfibrillar unit in cross-section (Fraser et al., 1983). These five molecules are arranged laterally, but staggered by the difference 'D' (each molecule being 4.4D in length). Thus in a lateral arrangement of five molecules there are zones of complete overlap of 0.4D in size and adjacent 'hole' zones of 0.6D, in which there are only four molecules per microfibrillar unit. The overlap and hole zones give rise to the cross-striations (approximately every 68nm) in electron micrographs of collagen I fibrils.

Practically all of the length of a fibrillar collagen molecule is in the triple helical configuration, except for a stretch of about a

TABLE 1.1

COLLAGEN TYPES

COLLAGEN	CHAINS	M_r	MOLECULAR SPECIES
		GROUP 1	
Type I	α1(I)	95,000	$[\alpha1(I)_2\alpha2(I)]$
	α2(I)	95,000	$[\alpha1(I)_3]$
Type II	α1(II)	95,000	$[\alpha1(II)_3]$
Type III	α1(III)	95,000-110,000	$[\alpha1(III)_3]$
Type V	α1(V)	115,000	$[\alpha1(V)_2\alpha2(V)]$
	α2(V)	125,000	$[\alpha1(V)_3]$
	α3(V)	?	[α1(V)α2(V)α3(V)]
	α4(V)?		
Type XI	α1(XI)	?	
	α2(XI)	?	?
	α3(XI)(?)	?	
		GROUP 2	
Type IV	α1(IV)	185,000	$[\alpha1(IV)_2\alpha2(IV)]$
	α2(IV)	170,000	$[\alpha1(IV)_3]$
			$[\alpha2(IV)_3]$
Type VI	α1(VI)	140,000	[α1(VI)α2(VI)α2(VI)]
	α2(VI)	"	$[\alpha1(VI)_3]$
	α3(VI)	"	
Type VII	α1(VII)	>170,000	$[\alpha1(VII)_3]$
Type VIII		(180,000)	?
		GROUP 3	
Type IX	α1(IX)	~70,000	[α1(IX)α2(IX)α3(IX)]
	α2(IX)	"	
	α3(IX)	"	
Type X	α1(X)	49,000	$[\alpha1(X)_3]$
Type XII	α1(XII) ?	?	?
Type XIII	α1(XIII) ?	?	?

dozen residues at each end, the telopeptides. The triple-helical domain contains about a thousand amino acid residues, thus the molecular formula can be approximated as being $(Gly\text{-}X\text{-}Y)_{333}$.

The precursor (procollagen) molecules have more extensive non-helical propeptides at the amino and carboxyl termini. The amino propeptide consists at the terminus of a signal peptide approximately 20 amino acids in length, enabling the passage of the molecule from the endoplasmic reticulum; a globular domain of approximately 70 amino acids which is stabilised by disulphide bridges; and a triple helical domain of approximately 40 amino acids. The carboxyl propeptide consists of approximately 250 amino acids and has a solely globular structure. In the carboxyl-propeptide two cysteine residues form interchain disulphide bridges which help stabilise the three α chains. These bridges are essential in holding the three chains together, enabling the formation of the triple helix. Playing such a crucial role, it is not surprising that regions of the C-propeptide of the fibrillar collagens are strongly homologous to one another (Yamada et al., 1983).

A lack of chemical data has meant that it is not known whether types V and XI collagen do form fibrils. However, sequencing of cDNA and genomic clones of the α2(V) gene (Weil et al., 1987; Woodbury et al., 1989) and the α1(XI) gene (Bernard et al., 1988) justifies the inclusion of these collagen types in group 1 as they appear closely evolutionarily related and have long un-interrupted triple-helical domains.

Group 2 molecules are classified as being of $M_r \geq 95,000$ and having helical domains separated by non-helical segments. This group comprises types IV, VI, VII, and VIII collagen. Because of the presence of the non-helical regions, type IV molecules are highly flexible. They

have a length greater than 400nm and tend to aggregate through end-to-end monomers. Type VI molecules consist of a short (140nm) helical domain plus an extensive globular domain. These molecules are thought to form microfibrillar arrays by end-to-end association of tetrameric units. Type VII molecules are extraordinarily long at 450nm, but only a small portion of this length is apparently triple-helical. It has been speculated that type VII molecules associate as antiparallel dimeric units which may aggregate to form anchoring fibrils (Burgeson et al., 1985). Type VIII molecules seem to consist of three triple-helical domains of roughly equal size interrupted by two short non-helical domains. The nature of the aggregates formed by type VIII molecules is unknown.

Group 3 molecules are defined as short-chain collagens, that is, they have molecular weights of less than 95,000. The best characterised group 3 collagens are types IX and X collagen. Recently two new collagens have been discovered (types XII and XIII) and initial cDNA sequence data would suggest that these are both short-chain collagens (Gordon et al., 1987; Pihlajaniemi et al., 1987).

1.2.2 Collagen types

At least 13 different collagen types have been identified, and are encoded by at least 20 different genes. An outline of the biochemical and functional properties of each collagen type is presented below.

GROUP 1

Type I collagen

Type I collagen is prevalent in several major connective tissues and represents by far, the most abundant collagen in vertebrate

organisms. It generally forms well-structured, cross-striated fibres of substantial size, greater than 400Å in diameter, and possibly over several decimetres in length. The properties of the fibres are influenced by the degree of intra- and inter-molecular cross-linking as well as the orientation and density of the fibrils. Their deposition may be in parallel (tendons and bones) or cross-alignment (skin and organ capsules) depending on the location and functional demands. The distribution of type I collagen indicates that it plays a key role as a supporting element of high tensile strength with very limited elasticity.

Type I collagen is a heterotrimer composed of two identical $\alpha 1(I)$ chains and an $\alpha 2(I)$ chain, though in certain instances homotrimers composed of $\alpha 1(I)$ chains have been observed to exist. The lengthy Gly-X-Y triplet region of type I collagen spans 1014 amino acids, giving an un-interrupted helical domain of about 300nm.

It has been suggested that type I collagen is essential to embryonic development by the insertion of a Mov-13 retrovirus into the $\alpha 1(I)$ gene of mouse embryos, causing disruption of haemopoesis and necrosis of erythropoetic and mesenchymal cells (Lohler et al., 1984).

Type II collagen

Fibres formed with type II collagen molecules exhibit a much more restricted tissue distribution than type I fibres and are found almost exclusively in cartilagenous structures, where they are synthesized by chondrocytes.

Type II collagen makes up 40% of the dry weight of cartilage. Non-expression of type II collagen leads to lethal newborn achondrogenesis, an indication that it is essential to normal development (Eyre et al., 1986).

A cartilage matrix protein, chondrocalcin, has been identified as the carboxyl-propeptide of type II collagen (van der Rest et al., 1986). It has been suggested that it may play a role in endochondral ossification once cleaved from the type II procollagen molecule (van der Rest et al., 1986).

Type III collagen

Fibrils formed from type III collagen molecules are generally found in association with larger type I fibres in the more distensible connective tissues. Skin fibroblasts synthesize types I and III collagen in about a 5:1 ratio (Breul et al., 1980), and the same cells can synthesize both types simultaneously. Type III fibrils are most prevalent in highly differentiated parenchymal organs such as liver, lung and lymphoid tissues. In these structures the loose reticular network composed of type III collagen molecules provides a limited degree of support but does not inhibit rapid diffusion and exchange of metabolites.

In contrast to types I and II procollagens, extracellular processing of type III procollagen often results in retention of the N-terminal propeptide sequences. Failure to remove these sequences from a proportion of type III procollagen molecules results in the presence of constituent chains of M_r 95,000 and 110,000 in molecules resident within the fibrillar aggregates (Table 1.1).

Type V collagen

Aggregates derived from type V molecules occur in a number of tissues where they comprise only a minor fraction of the total collagen pool. Collectively, molecules of the type V system contain at least three unique chains (Table 1.1). The heterotrimer composed of two α1(V)

chains and one α2(V) chain, being the most prevalent. It appears to be the only species of type V collagen present in bone, dermis, corneal stroma and placental membranes.

A fourth chain for type V collagen has been reported to exist, initially named α4(V) it has now been called α1'(V) because of its similarity to the α1(V) chain (Fessler et al., 1985). Whether the proα1'(V) and proα1(V) chains are the products of two similar genes (differing mainly in the non-collagenous peptide regions) or that they arise from one gene by differential splicing of one proα1(V) pre-mRNA is not known.

In contrast to types I, II and III procollagen, the α1 and α2 chains of type V procollagen are synthesized initially as polypeptides of M_r approximately 240,000 and 160,000 respectively. This suggests that the globular termini of these chains are somewhat larger than their types I, II and III procollagen counterparts.

Type XI collagen

Type XI collagen (previously known as type K or 1α2α3α) is likewise a quantitively minor collagen. It is however, apparently limited to cartilagenous structures. A unique feature of type XI collagen is that the α3 chain appears to be identical to α1(II) and distinguished only with respect to elevated levels of lysine hydroxylation and glycosylation (Furuto and Miller 1983). It is not known whether α3(XI) and α1(II) are two closely related genetically distinct variants, or products of the same gene. The α1 and α2 chains are similar, but distinctly different from the type V chains. Thus types V and XI collagen should be thought of as close relatives in either evolutionary or functional terms.

GROUP 2

Type IV collagen

Type IV collagen is a highly specialised form of collagen found only in basement membranes. The molecule is 400nm in length and has a non-collagenous globular (NCI) domain at its carboxyl terminal. Two types of α chain are known to be present in type IV collagen; α1(IV) and α2(IV), though the nature and number of molecular species is unknown. In various tissues [$\alpha1(IV)_2\alpha2(IV)$] heterotrimers and [$\alpha1(IV)_3$] and [$\alpha2(IV)_3$] homotrimers have been observed, the relative chain proportions perhaps varying between different basement membranes.

Type IV forms a network by end-to-end aggregation. Four molecules are bound to one another through their amino-terminal 30nm long triple-helical segments known as the 7S domain. At their opposite ends, the carboxyl-terminal globular NCI domains of two molecules are joined (Duncan et al., 1983). Stabilisation of this alignment is acheived by intermolecular cross-links, including reducible disulphide bridges and non-reducible, presumably lysine derived, bonds located in the 30nm overlap zone (Bailey et al., 1984). In contrast to the fibrillar collagens, the helical regions of the α1(IV) and α2(IV) chains are frequently interrupted by non-helical segments. The non-helical regions appear to be important structural features of the triple helix, necessary for the regulation of not only the flexiblity of the molecules, but also the elastic properties of the macromolecular network of collagen IV.

Type VI collagen

Type VI collagen is widely distributed throughout the connective tissues. It is a disulphide-rich protein with relativley short triple

helical domains and is localised to the boundary zones where interstitial collagen fibres interact with the membranous structures. Type VI collagen is composed of three different chains α1(VI), α2(VI) and α3(VI). It is believed that native type VI molecules exist as heterotrimers of the three chains. However, a homotrimer of α1(VI) chains has been proposed (Furuto and Miller 1981). A model has been proposed for the assembly of type VI collagen (Engvall et al., 1986), in which monomers assemble intracellularly into tetramers and these associate extracellularly into fibrils by end-to-end contact.

Little is known of the biological function of type VI collagen.

Type VII collagen

Type VII collagen is a homotrimer of three identical α-chains. Type VII collagen molecules are thought to associate in an overlapping, antiparallel fashion to form covalently stabilised dimers (Morris et al., 1986). It has been postulated that type VII collagen is a major component of anchoring fibrils (Keene et al., 1987). Anchoring fibrils are necessary structures to the epithelial basement membranes of several tissues which are subjected to considerable frictional and torsional forces. These include skin, oesophagus, cornea, vaginal mucosa and chorioamniotic membranes.

Type VIII collagen

Although there is prelimimary evidence for the presence of type VIII collagen in tissues such as Descemet's membrane in the eye (Kapour et al., 1986), this collagen has been studied primarily as a product of cultured endothelial cells. The constituent chains are disulphide linked and exhibit an apparent M_r of 180,000, 125,000, 100,000 and 60,000. It is not yet clear whether these data signify the presence of

several unique chains, different intermediates in the extracellular processing of a single chain type, or combinations of these possibilities. A model has been formulated for the type VIII molecule (Benya, 1980) which assumes that the 180,000 Da chain represents the typical structure and that the helical region is interrupted at two points at positions one-third and two-thirds along the length of the chain. This would give three helical domains of approximately equal size. According to this model, the type VIII molecule closely resembles types IV and VII molecules in overall size.

GROUP 3

Type IX collagen

Type IX collagen is a cartilage collagen accounting for 5-20% of the collagen in adult hyaline cartilage. The type IX collagen molecule is a heterotrimer of three different chains consisting of three triple helical domains joined by short non-triple helical domains. At the amino and carboxyl ends are non-collagenous domains which do not appear to be homologous to the amino and carboxyl propeptides of the fibrillar collagens.

The identification of trivalent hydroxypyridinium cross-links in some type IX collagen molecules suggests a potential means of cross-linking type IX collagen to type II collagen (van der Rest and Mayne, 1988). However, physical data has not yet shown any such cross-links to occur.

Type X collagen

Type X collagen represents another apparently cartilage-specific protein. Although a quantitatively minor collagen when an entire cartilaginous tissue is examined, it appears to be the major

collagenous product of chondrocytes during their terminal stage of maturation. The accumulation of type X collagen precedes calcification of the tissue. Type X collagen synthesis by fracture callus (Grant et al., 1987) is further evidence in support of its close association with the process of endothelial ossification.

The type X collagen molecule consists of three identical α-chains and is approximately half the length of the fibrillar collagen types I, II, III and V. The nature of the aggregates formed by type X molecules remains unknown.

Type XII collagen

By cDNA cloning, a new collagen gene was identified (Gordon et al., 1987). This gene was designated as encoding type XII collagen. The intron/exon structure of the gene appears to be similar to that of the COL9A1 and COL9A2 genes. The pro-α1(XII) mRNA is found in calvaria, tendon, cornea and sternal cartilage of chick embryos.

Type XIII collagen

Again, discovered by cDNA cloning, type XIII collagen is a short-chain collagen. Little is known of the type XIII collagen protein (Pihlajaniemi et al., 1987).

1.2.3 Collagen biosynthesis

Collagen biosynthesis is characterised by the presence of an unusually large number of co-translational and post-translational modifications. The synthesis of collagen can be considered as occurring in two steps: intracellular steps are required to assemble and secrete the procollagen molecules, and extracellular processing converts the procollagen molecules to collagen which is incorporated into stable,

cross-linked fibrils.

There are five enzymes known to act on the procollagen chains intracellularly. Prolyl 4-hydroxylase, prolyl 3-hydroxylase and lysyl hydroxylase hydroxylate proline or lysine residues. The interaction of these enzymes is influenced by the amino acids flanking each proline or lysine residue. Hydroxylysyl galactosyltransferase and galactosyl-hydroxylysyl glucosyltransferase transfer galactose to some hydroxy-lysine residues and glucose to some galactosylhydroxylysine residues respectively. The importance of the addition of sugars is not understood. Asparagine residues in the propeptides are likewise glycosylated.

The mechanism by which both intrachain and interchain disulphide bonds are formed is poorly inderstood. However some studies suggest that the enzyme protein disulphide isomerase may be involved (Myllyla et al., 1983).

Most of the hydroxylation of proline and lysine residues and glycosylation of hydroxylysine residues occurs as co-translational processing, but continues as post-translational modifications until triple helix formation prevents further processing. Hydroxyproline plays an important role in the stabilization not only of the triple helical collagen molecule, but also in its assembly into microfibrils. The 4-hydroxyproline residue only occurs in the Y position of the Gly-X-Y triplets. This allows better fitting of the triple helices (Nemethy and Scheraga, 1986). Triple helix formation cannot take place until almost all of the proline residues in the Y position have been 4-hydroxylated. Disulphide bonds between the carboxyl-terminal propeptide region of the α-chains direct the association of the three chains and serve as a point of initiation for triple helix formation. Triple helix formation does not require enzymic catalysis and will occur

spontaneously *in vitro*, once the carboxyl propeptide disulphide bonds have been formed.

The procollagen molecule is then translocated across the membrane of the rough endoplasmic reticulum, during or shortly after which, the signal sequences are cleaved from the amino-terminal end. The procollagen molecule is then secreted via the Golgi apparatus.

In the extracellular space, the amino- and carboxy-propeptides are removed by procollagen N-proteinase and procollagen C-proteinase respectively. There are several isozymes of the N-proteinase which cleave different collagen types. It has been observed that the cleaved amino propeptide inhibits collagen synthesis in fibroblast cultures and cell-free systems (Wiestner *et al.*, 1979; Paglia *et al.*, 1979). These results have led to the suggestion that the amino propeptide may be involved in a feedback control of types I and III collagen biosynthesis.

The collagen molecules self-assemble into fibrils in the case of the fibrillar collagens, or other ordered structures for the non-fibrillar collagens. This process does not require any enzymic catalysis or other factors and will occur easily in collagen solutions *in vitro*. For a fibril to achieve its normal strength chemical cross-links must be introduced to link the molecules in the fibre to each other. The cross-linking occurs in at least two stages. The initial event is oxidative deamination of the ε-amino group in certain lysine and hydroxylysine residues to the corresponding aldehyde in a reaction catalysed by lysyl oxidase. The reactive aldehydes then participate in the formation of the various cross-links.

1.2.4 Collagen genes

In vertebrates, the thirteen or more distinct collagen types are coded for by at least 20 genes. These genes are semi-dispersed throughout the genome. Those genes so far assigned chromosomal localisations are listed in Table 1.2. The loci COL1A1, COL1A2, COL2A1 and COL11A1 are dispersed on different chromosomes (Retief et al., 1985; Henderson et al., 1983; Law et al., 1986; Henry et al., 1988, respectivley). However the loci COL3A1 and COL5A2; COL4A1 and COL4A2; COL6A1 and COL6A2 are clustered and share the same map locations. COL3A1 and COL5A2 have been mapped to 2q31-32.3 (Emanuel et al., 1985a; Huerre-Jeanpierre et al., 1986). Interestingly types III and V collagen are usually co-expressed (Bornstein and Sage, 1980). In addition, the COL6A3 locus has been mapped to 2q37 (Weil et al., 1988a) and the fibronectin (Henry et al., 1985) and elastin (Emanuel et al., 1985b) loci to 2q34-36 and 2q31-qter respectively. So five extracellular matrix genes have been mapped to the distal region of the long arm of chromosome 2. Hence it would be of interest to further investigate whether these proteins are evolutionarily related and whether the genes are coordinately regulated. The close map locations of these genes is quite striking, but not as extraordinary as the arrangement of the COL4A1 and COL4A2 genes. These two genes have been mapped to 13q34 (Boyd et al., 1988). The two genes are encoded on opposite strands of the DNA and are so closely situated that the transcription initiation sites of the two genes could be as little as 42bp apart (Soininen et al., 1988; Poschl et al., 1988). The first exons of the two genes are a mere 126bp apart. This is the first description of two structural genes from a complex organism that code for two polypeptide chains of the same protein molecule that have overlapping 5'-flanking regions.

The COL6A1 and COL6A2 loci have too been mapped to the same

TABLE 1.2

COLLAGEN GENE MAP LOCATIONS

COLLAGEN TYPE	CHAIN	GENE	MAP LOCATION
Type I	α1(I)	COL1A1	17q21.3-22.1
	α2(I)	COL1A2	7q21.3-22.1
Type II	α1(II)	COL2A1	12q14.3
Type III	α1(III)	COL3A1	2q31-32.3
Type IV	α1(IV)	COL4A1	13q34
	α2(IV)	COL4A2	13q34
Type V	α2(V)	COL5A2	2q31-32.3
Type VI	α1(VI)	COL6A1	21q22.3
	α2(VI)	COL6A2	21q22.3
	α3(VI)	COL6A3	2q37
Type XI	α1(XI)	COL11A1	1p21

position: 21q22.3 (Weil et al., 1988a), but in this case their proximity is not yet known. The apparent structural homology of the α1 and α2 chains and lack of homology to the α3 chain suggests an early divergence of the α3(VI) gene from a common ancestor, followed more recently by a duplication of the gene at 21q22.3.

It may be expected that despite the apparent similarity between the α1(VI) and α2(VI) chains, their fine structure may be quite different, serving to reduce the likelihood of unequal crossing over between the two genes. It has been suggested (Solomon et al., 1985; Griffin et al., 1987) that some selective pressure may exist to encourage the dispersal of the collagen genes, a process that would act to reduce unequal exchanges. Therefore those collagen genes residing close to others, may be sufficiently different to minimise the chance of unequal exchange; recently duplicated; or under the effect of the same cis-acting regulatory system.

Collagen genes have one of the most highly interrupted gene structures kmown in man. The genes for the fibrillar collagens are typically divided into 50 or more exons. There is considerable variation in intron length from gene to gene resulting in overall gene lengths that vary from 18kb for COL1A1 (Barsh et al., 1984) to 35kb for COL1A2 (Dickson et al., 1985).

Early sequencing of random exons of type I collagen genomic clones (Yamada et al., 1980; Wozney et al., 1981) indicated that the genes consisted predominantly of 54bp exons, encoding exactly 18 amino acids or rather 6 Gly-X-Y units, with each exon beginning at the start of a glycine codon. Any exceptions were still multiples of 9bp, namely 45, 99, 108 and 162bp. These observations led to the suggestion that the primordial building block from which the collagen genes evolved was a unit of 54bp. This 54bp building block was believed to have been

amplified by duplication, the 45 and 99bp exons being due to recombinational and mutational events resulting in the loss of one Gly-X-Y unit. This hypothesis was attractive on the basis of its simplicity.

Earlier work by McLachlen (1976) on the amino acid sequence of the α1(I) gene indicated that regions of internal homology in the α1(I) chain were repeated at the 'D' distance, the stagger distance between collagen molecules in a fibril. Then analysis of the α1(I), α2(I) and α1(II) chains (Hofman et al., 1980) revealed smaller repeat units of D/3, D/6, D/11 and more significantly D/13 - 18 amino acids. This data may better reflect the original genetic unit, as the 'D' distance repeat may be a result of selection pressure directed towards improved aggregation of the molecules into fibrils. Also derived from such data is the fact that the homologies within one repeat are more conserved across species than from one repeat to another within an α-chain. Thus the occurrence of early intrachain duplication with subsequent duplication of the whole chain is the likely evolutionary history of the collagen gene family.

An alternative view is that introns were somehow introduced into a gene that had previously consisted of a single continous open reading frame. Such a development would help reduce the likelihood of unequal recombinational exchange changing the length of the α-chain.

The isolation and characterisation of types IV and X collagen genes has caused the views of collagen gene evolution to be modified. In these genes, exons with sizes of 64, 71, 78, 123, 147, and 182 were found. Not only did this run contrary to the 'multiples of 9bp rule' but intron exon boundaries were found to interrupt codons. The type X collagen gene contains a single open reading frame of 2000bp without introns (Ninomiya et al., 1986). The molecules produced from these

genes do not form fibrils and the genes may therefore experience less selection against interruptions of the triple helix and any size changes. If the presence of introns in fibrillar collagens prevents recombinational rearrangments, and so maintains the precise length and distribution of charged and hydrophobic residues along the chains, it might be predicted that such events occur more frequently in the genes of the non-fibrillar collagens. If this were true, one would expect greater diversity among each non-fibrillar collagen type in different animal species than among the fibrillar collagens.

1.2.5 Collagen gene expression

A variety of observations suggest that expression of the fibrillar collagen genes is subjected to intricate regulatory events at both the transcriptional and post-transcriptional levels. Modulation in the level of expression of the fibrillar collagen genes has been reported in a variety of developmental situations. The RNA levels for types I, II and III collagen show a 10-fold increase during the development of chick embryos (Merlino et al., 1983). Differentiating myoblasts in culture demonstrate a 15-fold increase in the levels of pro-α1(I) and pro-α2(I) mRNAs and a 10-fold increase in the level of pro-α1(III) mRNA (Gerstenfeld et al., 1984).

During chondrogenesis, type I procollagen synthesis is replaced by type II procollagen (Kosher et al 1986; Castagnola et al., 1988). However, substantial amounts of type I collagen mRNAs persist, indicating that control of type I collagen gene expression is exerted at both transcriptional and post-transcriptional levels (Kosher et al., 1986; Focht and Adams, 1984; Allenbach et al., 1986).

The three genes for type I collagen (situated on different chromosomes) have been shown to be coordinately controlled during both

rapid and slow growth of human fibroblasts (Miskulin et al., 1986). Studies on the expression of fibrillar collagen genes in cells after exposure to various compounds; such as ascorbic acid (Lyons et al., 1984), glucocorticoids (Walsh et al., 1987), growth factors (Ignotz et al., 1987) and cytokines (Goldring et al., 1987) all show these compounds to influence the levels of transcription.

A uniquely conserved sequence of about 50bp has been identified around the translational start of the COL1A1, COL1A2 and COL3A1 genes that has a negative regulatory effect upon translation. This sequence contains an inverted repeat capable of forming a stem-loop structure. Rossi and de Crombrugghe (1987a) have demonstrated that this sequence can form intermolecular RNA dimers. They suggest that equilibrium between monomers and dimers could be regulated to control the rate of transcription.

Transcriptionally active chromatin exhibits conformational modifications that are revealed by the increased susceptibility of its DNA to nuclease digestion (Weintraub and Groudine, 1976). The insertion of a single Moloney murine leukaemia virus into the first intron of the mouse COL1A1 gene results in the loss of a DNase-hypersensitive site and a 20- to 100-fold reduction in transcription of the gene (Hartung et al., 1986), indicating that the DNase-hypersensitive site is a prerequisite for, rather than a consequence of, gene activity.

Functional deletion studies, using an expression assay in Xenopus laevis oocytes, demonstrated that in close proximity to the DNase-hypersensitive site in the first intron of the human COL1A1 gene is a 782bp cis-acting region with enhancer activity (Rossouw et al., 1987). Using transient transfection experiments in chicken fibroblasts, Bornstein et al. (1987), identified a 274bp fragment from the first intron of the human COL1A1 gene, that has a significant orientation-

specific negative regulatory activity. More recently, this element has been found to be flanked by sequences that both neutralise the inhibitory effect and impart an overall positive effect upon transcription (Bornstein and McKay, 1988). It would thus appear that expression of the COL1A1 gene is controlled by several intronic elements that function coordinately with 5'-flanking and promoter elements.

An orientation independent transcriptional enhancer has been identified in the first intron of the mouse COL1A2 gene (Rossi and de Crombrugghe, 1987b). This enhancer displays cell specificity, since it functions in NIH3T3 fibroblasts but is completely inactive in a lymphoid cell line. However, transgenic mice experiments show that the COL1A2 promoter (or rather the region -2000 to +54) alone is sufficient to direct stage- and tissue-specific activity of the gene (Khillan et al., 1986).

Whether transriptional enhancers and other regulatory elements within the first intron are a general feature of collagen genes, will require more functional analyses.

That all regulatory elements of the COL2A1 gene lie close to or within the gene has been shown by the production of transgenic mice carrying the entire human COL2A1 gene along with 4.5kb 5' and 2.2kb 3' to the gene (Lovell-Badge et al., 1987). These mice showed correct stage- and tissue-specific expression of the foreign gene.

1.3 INHERITED CONNECTIVE TISSUE DISORDERS

Collagen being the major constituent of the extracellular matrix was always a likely candidate for the abnormal protein in inherited connective tissue disorders. As a direct consequence of the complex gene structure and biosynthetic pathway of collagen, there are many steps at which errors can occur and so result in a disease state. These obviously affect primarily those tissues in which the proper development and integrity of connective tissue is of paramount importance.

Those inherited connective tissue disorders that are attributed to defective collagen include the many variants of osteogensis imperfecta (OI), the Marfan syndrome (MS), the Ehlers-Danlos syndrome (EDS) and several other related disorders. These disorders are very heterogeneous and each disease is a grouping of associated phenotypes. Moreover, the line of definition between the diseases is a grey one, especially in the case of OI and EDS; where very mild OI and EDS may essentially encompass the same clinical manifestations.

In this section a brief description of only those diseases in which defective collagen metabolism is strongly implicated will be presented; and in the next section, the apparent underlying genetic abnormalities will be reviewed.

1.3.1 The Marfan syndrome

The Marfan syndrome (Pyeritz and McCusick, 1979, 1981; and Pinnel and Murad, 1983) is an autosomal dominant trait. The disease is very heterogeneous, not only between families but within them, the variability within families being due to variable penetrance. Estimates of prevalence range widely from 1.5/100,000 to 4/100,000 with about 5% of cases being due to new mutations (Dalgleish *et al.*, 1987a).

In its typical form, the Marfan syndrome (MS) consists of skeketal disproportion, arachnodactyly, ocular disorders (myopia and ectopic lens), and cardiovasular problems (principally aortic incompetence and dissecting aneurysm of the ascending aorta). Early death frequently results from these cardiovasular abnormalities, both in the patients with the fully penetrant disease and in those who have no other obvious signs of the disease. It has been shown that there is no correlation between the severity of the cardiac and oculoskeletal abnormalities.

The pressing need for genetical or biochemical diagnosis is not for prenatal diagnosis, as it is in so many other inherited disorders, but to provide a precise diagnosis early in life, so those at risk can be identified and properly cared for.

1.3.2 The Ehlers-Danlos syndrome

The Ehlers-Danlos syndrome (EDS) is the most heterogeneous of the connective tissue disorders. The common abnormalities in this group of disorders are joint laxity, skin hyperextensibility, easy bruisibility and tissue fragility. It is probably the most prevalent of the connective tissue disorders and is underdiagnosed, however incidence figures were not available. The syndrome has been classified into at least 11 clinical types, with autosomal dominant, autosomal recessive and X-linked recessive modes of inheritance suggested.

1.3.3 Osteogenesis imperfecta

Osteogenesis imperfecta (OI) has an overall incidence of about 1 in 20,000 births and inheritance is seen as both autosomal dominant and autosomal recessive.

Bone fragiliy is the cardinal clinical manifestation of the OI

syndromes. In addition short stature, joint laxity, easy bruising, blue sclerae, presenile hearing loss and dentinogenesis imperfecta may also be present. This phenotypic diversity has prompted several attempts at classification, the most recent and widely accepted one having been proposed by Sillence et al., (1979). Accordingly OI is divided into four phenotypic groups, although many affected inividuals do not fit such rigid categorisation.

The subtypes of OI are:-

Type I: Dominantly inherited with blue sclerae.

Type II: Perinatally lethal.

Type III: Progressively deforming with normal sclerae.

Type IV: Dominantly inherited with normal sclerae.

These 'Sillence' classifications will be used throughout this thesis.

Type I OI is the largest group and is further defined by postnatal onset of fractures, joint laxity, and often pre-senile hearing loss and dentinogenesis imperfecta. The minimum incidence is approximately 3.5/100,000 live births with a new mutation rate of approximately $0.25\text{-}0.3 \times 10^{-5}$ (Sillence et al., 1979)

In type II OI the bone fragility is extreme, resulting in death in utero, during or soon after birth. These babies characteristically have broad crumpled femora, and beaded ribs; the result of in utero fractures. The frequency of live births is approximately 1.6/100,000. As such a phenotype cannot be inherited in a Mendelian dominant fashion, the inheritence until recently was assumed to be recessive, at least in most cases. However it is now apparent that most or even all cases are new (dominant) mutations, with an approximate 7.7% risk of recurrence (Thompson et al., 1987) being due to germline mutations. Thus the new mutation rate in type II OI is the same as the disease incidence.

Type III OI is the most severe survivable OI. It is characterised by severe bone fragility, short stature, skeletal deformity, dentinogenesis imperfecta and hearing loss. Like type II OIs, the family history is usually negative and thus is again probably caused by new mutations. The minimum incidence of type III OI is quoted as 1.45/100,000 live births (Sillence et al., 1979).

Type IV is the mildest type of OI, and is characterised by moderate bone fragility, short stature, normal sclerae, and often hearing loss and dentinogenesis imperfecta. Incidence figures for type IV OI were unavailable.

1.4 MOLECULAR BASIS OF INHERITED COLLAGEN DISORDERS

1.4.1 The Marfan syndrome

The fibrillar collagens, being the major stress-bearing components of the tissues most affected in MS have long been suspected of being the defective components in this disorder. However evidence that collagen is defective has largely been circumstantial.

Several reports have documented that an increased fraction of tissue collagen is extractable by non-denaturing solvents in MS (Laitinen et al., 1968; Priest et al., 1973). An abnormality of cross-linking has long been the simplest explanation for these observations. Some patients (but not all) have increased levels of urinary hydroxyproline, an indication of accelerated collagen turnover. The phenomenon of abnormal cross-linking in MS was observed to be due specifically to a lack of hydroxylysyl-derived cross-links (Boucek et al., 1981), suggesting incomplete maturation of collagen causes the defective collagen organisation in MS. In addition, animals fed nitriles or copper-deficient diets, both of which interfere with collagen cross-linking, have aortic abnormalities which are quite similar to those of patients with MS.

One question that cannot readily be answered is whether abnormal collagen is the underlying defect or whether abnormal collagen is a secondary effect of the (true) inherited abnormality. The possibility exists that other connective tissue components may be abnormal as well. Elastin, another major component of the aortic wall, has been demonstrated to be present at decreased levels in the aortic wall of some MS patients (Halme et al., 1985). Increased uinary excretion of desmosine, an amino acid unique to elastin and decreased levels of desmosine in elastin from aortic tissue have been described by

Abraham et al., (1982). In addition, mucopolysaccharide and hyaluronic acid levels are raised in the the aortic media and cultured fibroblasts from these patients (Pyeritz and McKusick, 1981). Thus there is a problem, with all this diverse information, of cause and effect, with the underlying defect being camouflaged by the array of secondary effects.

Hope was given to the collagen afficianados that MS was 'their' disease, when Byers et al., (1981), described a patient with an apparent heterozygous insertion of approximately 20 amino acids into the α2(I) collagen chain. It was assumed that the apparent insertion was the cause of the increased level of collagen extractability observed in this patient. However genomic cloning of the pro-α2(I) gene from this patient by Henke et al., (1985), failed to reveal any coding abnormality. A 38 base pair insertion was found in an intron, but was later found to be a neutral polymorphism (Dalgleish et al., 1986). Recently experiments by R.J. Wenstrup (Duke University, USA) have identified an amino acid substitution, which is speculated to have given the effect of the apparent insertion (P. Byers, personal communication to R. Dalgleish). Further work needs to be done to support the hypothesis that this is the disease-cause in this patient.

In contrast, linkage analysis of the pro-α2(I) gene has excluded this gene as the mutant locus in several families (Tsipouras et al., 1986; Dalgleish et al., 1987a; Ogilvie et al., 1987a), suggesting perhaps that the pro-α2(I) gene is not commonly, or ever, the disease causing gene in MS.

1.4.2 The Ehlers-Danlos Syndrome

In all the varieties of EDS in which biochemical abnormalities have been observed, the affected molecules are collagen.

EDS types I, II and III are all accompanied by larger than normal dermal collagen fibrils, frequent composite fibrils and smaller than normal fibrils. Given the autosomal dominant mode of inheritance, these disorders appear to result from heterozygosity for an abnormal structural macromolecule.

EDS IV has been one of the most extensively studied types and is now defined on the basis of abnormal synthesis, secretion, or structure of type III collagen. In two families with type IV EDS, the type III collagen gene has been seen to segregate with the disease (Nicholls et al., 1988; Tsipouras et al., 1986b) giving lod scores ($\theta = 0$) of 3.86 and 2.1 respectively. The first mutation identified in COL3A1 was described by Superti-Furga et al., (1988). In this case a patient with type IV EDS was seen to have a 3.3kb deletion from the gene, shortening the mRNA by approximately 600 base pairs. Recently Tromp et al., (1989) became the first workers to fully characterise a mutation in the type III collagen gene. Using a complicated combination of genomic cloning, cDNA cloning and PCR amplification, the mutation was found to be a G to A transition in the first base position of the codon, converting glycine 907 to arginine.

Types V and IX EDS both show linkage to the X-chromosome. The recessive inheritance suggesting a defective collagen modifying enzyme. In both disorders the level of lysyl oxidase activity is affected.

The recessively inherited EDS VI phenotype results from decreased lysyl hydroxylase activity (Pinnel et al., 1972). It has been demonstrated that the hydroxylation of types I, II and III collagen is markedly decreased, while that of types II, IV and V collagen is normal. This finding has raised questions regarding the molecular basis of hydroxylation, the types of lesions which might result in alterations of substrate specificity, and whether there are multiple

forms of lysyl hydroxylase.

EDS VII has come to be seen as a disorder of processing of the amino-terminal propeptide of type I collagen. Characterisation of EDS VII mutations is just beginning. Eyre et al., (1985) described a patient with an apparent deletion of the N-propeptidase cleavage site of the pro-α2(I) chain. Cole et al., (1986) described a patient in which 24 amino acids were were deleted from the pro-α1(I) chain. These amino acids correspond to exon 6, the exon encoding the N-propeptidase cleavage site, although whether the absence of exon 6 derived peptide was due to a deletion or a splicing error is not known. Weil et al., (1988b) have described a patient in which the 5' splice sequence of the pro-α2(I) intron 6 was altered by a base substitution from GT to GC. This had the effect of the splicing out of exon 6 from the RNA. Recently D'alessio et al., (1988) have described a bizarre compound heterozygote. This patient has the last codon of exon 6 in both COL1A1 and COL1A2 mutated by a single base mutation from ATG to ATA. This has the effect of the COL1A1 exon 6 being skipped, while the COL1A2 exon 6 is correctly spliced in nearly 10% of RNA molecules.

Some patients, for which there was no family history of EDS, had been believed to have recessive versions of the disease and thought to be caused by defective N-propeptidase. With more in-depth studies this opinion is changing to one of new dominant mutations in type I collagen.

At present there is very little biochemical information on EDS types VIII and X.

On the basis of knowledge of the abnormalities in types IV and VII EDS, it would appear that to some extent, the affected gene can be predicted from the phenotype and hence the phenotypic classifications reflect the genotype. Perhaps with the new approaches to mutation

detection and mapping (see section 1.5) the relationship between genotype and phenotype in EDS will soon be established.

1.4.3 Osteogenesis Imperfecta

With type I collagen being a major component of bone, it is not surprising that linkage studies of types I and IV OI have linked OI to type I collagen (Sykes et al., 1986; P. Tsipouras, personal communication). However some families have been identified in which OI is not linked to type I collagen. One such family was a consanguinous pedigree in which the disease was inherited in an autosomal recessive fashion (Aitchison et al., 1988). It is likely in this case that a collagen processing enzyme is responsible for the disease. Generally though it can be considered that OI is a disease of type I collagen. To date 16 mutations in OI have been characterised. These range from gross deletions to point mutations. Despite each mutation being different, a vague picture is beginning to emerge of the relationship between genotype and phenotype in this heterogeneous group of disorders. Most of the mutations described thus far, lie in the (1014 amino acid) helical domain of the type I procollagen molecule.

One of 8 possible amino acid substitutions resulting from a mutated glycine codon is a cysteine residue, an amino acid not normally present in the helical domain. To date, in COL1A1 seven mutations have been shown to cause cysteine for glycine substitutions. The first of these to be described was in a patient with type II OI (Steinmann et al., 1984). The procollagen molecules in this infant had decreased thermal stability, delayed secretion, and excessive lysyl hydroxylation and hydroxylysyl glycosylation along the entire length of the helix. The mutant gene was cloned and sequenced revealing a G to T transversion in the first base position of amino acid codon 988 in

COL1A1 (Cohn et al., 1986). Interestingly this patient's mother had the Marfan syndrome. However, until the Marfan locus or loci have been mapped, it cannot be stated whether the Marfan allele contributed to the disease phenotype in any way.

The next mutation to be characterised was again in an infant with type II OI with similar procollagen abnormalities to that above (Vogel et al., 1987). The mutation was characterised by sequencing cloned cDNAs and demonstrated a cysteine for glycine substitution at residue 748 of the pro-α1(I) chain, again a G to T transversion in the first base of the codon. Interestingly this patient also showed a defect in the processing of the procollagen molecule to pC-collagen by N-proteinase (shown to not be a secondary defect in N-proteinase). Thus the introduction of a cysteine in this position of the molecule has produced a conformational change far away in the amino-terminus, 776 residues along (Vogel et al., 1988).

Recently Constantinou et al., (1988), described a patient with type II OI. Sequencing of cDNA clones demonstrated, again, a G to T transversion causing a glycine to cysteine substitution at residue 904 of the α1(I) chain. A complication in this case exists; although the glycine to cysteine mutation was sporadic, the patient's mother's fibroblasts synthesised procollagen in which about one-quarter of the pro α-chains were over modified and had a lower melting temperature. It is unknown whether this variant, which causes no observed phenotypic abnormality in the mother, contributes to the phenotype of the child.

These three mutations all lie in the α1(I) CB6 and CB7 cyanogen bromide peptide fragments, the two helical region fragments nearest the carboxyl terminus of the polypeptide. More recently three glycine to cysteine mutations have been characterised nearer the amino terminus of the molecule, none of which are lethal.

A glycine to cysteine substitution at amino acid residue 526 of α1(I) (Starman et al, 1989) is seen in a patient with the moderately severe type I disease phenotype. Glycine to cysteine substitutions at residues 175 and 94 were observed in patients with type IV and very mild OI resectively, (P. Byers, personal communication to Dr. Dalgleish).

Thus there appears to be a gradient effect which depends upon the position of the substitution in the molecule; the closer it is to the carboxyl terminus, the more severe it is.

Hydroxylation of prolyl and lysyl residues begins on nascent chains and ceases when a stable triple helical conformation is achieved. Mutations which result in the substitution of a glycine (or small deletions or insertions that result in the disruption of the Gly-X-Y sequence motif) could produce an unstable or slowly assembling triple helical domain and, as a consequence, cause the intracellular retention and excessive modification of affected molecules.

It is likely that triple helix formation, which begins normally at the carboxyl-terminus, and procedes towards the amino-terminus, proceeds at the normal rate until the region of the mutation is reached. Thereafter, the rate of stable helix formation is delayed, and all three chains in abnormal molecules are available to the post-translational modifying enzymes for longer than normal. So for a heterozygous mutation it would be expected that three-quarters of procollagen molecules with mutations in α1(I) and one half of the procollagen molecules with mutations in α2(I) would show overmodification. In the three cases described in detail above, the positions of the mutations are correlated precisely with the boundary of overmodification. It is likely therefore that the severity of the disease is determined, at least in part, by the extent of

overmodification.

An appparent exception to this hypothesis is a patient with a dominantly inherited form of OI, in which a cysteine residue has been localised to the carboxyl-terminal cyanogen bromide peptide α1(I)CB6, (Steinmann et al., 1986). Molecules containing mutant chains from this patient are stable, secreted normally and have normal levels of post-translational modification and intracellular degradation. The lack of overmodification correlates with the disease phenotype. However, the presence of the cysteine in α1(I)CB6 contrasts with those cases described above.

It was suggested that this could be explained by the mutation occurring in the X or Y position of the Gly-X-Y motif rather than the glycine position.

The mutant region was later cloned from PCR-amplified material and DNA sequencing demonstrated the mutation to indeed be a glycine to cysteine substitution, but in the C-telopeptide, not the triple helix, (Cohn et al., 1988). Again this mutation is a G to T transversion at the first position of the codon (1017). It is speculated that the disease maybe caused by the presence of an interchain disulphide bond, giving rigidity to a normally flexible part of the protein and may affect the structure of the carboxy-terminal telopeptide.

To date, three glycine to arginine substitutions have been characterised. Protein sequence data from a baby with type II OI revealed a glycine to arginine substitution at residue 391 of the CB8 region of the α1(I) chain. For a single base change this would have to be caused by a G to C transversion of the first base of the GGC codon for residue 391 (Bateman et al., 1987).

The lysyl hydroxylation pattern for the procollagen from this infant showed over modification in the CB8 and the adjoining CB3

peptides, but not in the carboxyl-terminal CB7 and CB6 peptides.

In another type II OI patient, a more basic charge distribution was observed in the CB7 peptide of α1(I), (Bateman et al., 1988). Sequencing of cloned cDNAs demonstrated a G to A transition substitution causing a glycine to arginine substitution at position 664. The lysyl hydroxylation pattern in this infant is consistent with the model outlined above.

More recently, a glycine to arginine substitution in the last triplet of the α2(I) chain has been observed (Wenstrup et al., 1988). By sequencing gene clones, a G to C transversion was seen in the first base position of codon 1012. This patient unlike the two above, had the moderate type IV OI. Despite the mutant molecules being overmodified along their entire length, this patient survives, supporting the view that similar mutations are less disastrous in α2(I) than in α1(I). This may simply be because with mutations in α2(I), only one half of the molecules incorporate an abnormal chain, rather than the three-quarters with α1(I) mutations. Secondly for an α2(I) mutation there is never more than one abnormal chain per mutant molecule. It is thought that molecules with one mutant chain are secreted more poorly than those with two, (de Vries and de Wet, 1986; P. Byers, personal communication). Thus for an α2(I) mutation fewer mutant molecules would be secreted than for an α1(I) mutation. Thirdly, deviations from the Gly-X-Y motif in type I procollagen may better be tolerated in α2(I) than in α1(I) because of the structural differences in the two chain types.

From these single amino acid substitution mutations it would seem that mutations in different chains and in different positions on the same type I procollagen chain, have different phenotypic effects; not only do similar mutations in similar positions in α1(I) and α2(I)

have different degrees of severity, but the same amino acid substitution in different positions on an α-chain leads to varying degrees of severity, perhaps due to the extent of overmodification of the triple helix.

In this 'gradient of severity' there must be threshold points at which the type of OI changes; most obviously - the point at which an amino acid substitution is lethal or not. This threshold point of lethality would appear to have been roughly mapped for glycine to cysteine substitutions in the α1(I) chain.

Glycine to cysteine substitutions at residues 988 (Cohn et al., 1986), 904 (Constantinou et al., 1988) and 748 (Vogel et al., 1987) all result in lethal phenotypes. However similar substitutions at residues 526 (Starman et al., 1989), 175 and 94 (P. Byers, personal communication to Dr. Dalgleish) result in surviving infants with varying severities of disease. Thus the threshold of lethality in α1(I) for helical glycine to cysteine substitutions would appear to lie somewhere in CB3 or CB7. Mutations towards the amino terminus of this point would be expected to not only be compatible with life, but also be less severe the closer they lie to the amino terminus. There may even be a further threshold point near the amino terminus at which mild OIs become EDSs.

So already a gradient of severity map can be begun to be drawn (with some inaccuracy) for the two type I procollagen chains.

The substitution of glycine to arginine seems to be even more damaging than cysteine; glycine to arginine substitutions at residues 391 and 664 (Bateman et al., 1987 & 1988) are both lethal. Thus the threshold point of lethality for glycine to arginine substitutions in α1(I) is further towards the amino-terminus than glycine to cysteine substitutions.

There are however, two reports that do not immediately conform to this rule: Nicholls et al., (1984) and de Vries & de Wet, (1986), have described surviving patients with moderately severe OI, with cysteine residues in the CB6 and CB8 peptides of their α1(I) chains. It is probable then that these mutations are in the X or Y positions of the Gly-X-Y motif and are more easily tolerated.

A third kind of amino acid substitution in OI has recently been described by Baldwin et al., (1989). In this case, a lethal variant, by sequencing cDNA clones, was seen to be a G to A transition mutation in the second base position of the codon, converting glycine 907 of the α2(I) chain to aspartate.

Several deletions have been observed in OI patients and the same 'gradient of severity' hypothesis probably applies to these too.

The first deletion to be identified was one of approximately 500bp from the COL1A1 gene of a patient with type II OI (Chu et al., 1983). Half the pro-α1(I) mRNAs were shortened by approximately 250 nucleotides in which the reading frame was retained, thus shortened pro-α1(I) chains were synthesised. Like procollagen molecules with .α1(I) amino acid substitutions, the population of type I procollagen molecules was three-quarters abnormal and unable to form stable triple helical molecules. Such molecules were rapidly degraded intracellularly or after secretion (Williams and Prockop, 1983). It was later demonstrated (Chu et al., 1985b) that the end points of the deletion are contained within introns resulting in the elimination of three exons of the triple helical domain (exons 27-29). Interestingly the termini of the deletion are located within two short inverted repeats, hence an intron-intron recombination event was speculated to be the mechanism by which the deletion was generated.

In this patient there was also an elevated type III collagen

mRNA content. It has been suggested that this is a compensatory effect (Williams and Prockop, 1983), implying that type III collagen can substitute for type I. But as this patient suffered from type II OI, and hence was lethally affected, this cannot be substantiated. However, Bateman et al., (1986), in a study of 17 type II OI cases, localised type III collagen to the uncalcified matrix of OI bone. It is likely therefore, that type III collagen was produced as part of the fracture repair process or as the consequence of the coordinately controlled expression of types I and III collagen (Miskulin et al., 1986).

A large deletion in a type I collagen gene has recently been characterised in the α2(I) gene of a patient with type II OI (Willing et al., 1988). Gene cloning and sequencing revealed that in this patient a 4.5kb section of the a2(I) was deleted resulting in the loss of helical region exons 34-40, encoding 180 amino acids. Again this was an intron to intron deletion, but this time no inverted or direct repeats were found as evidence of homologous recombination.

Sippola et al., (1984), described a patient with an autosomal dominantly inherited mild variant of OI that had shortened α2(I) chains. This was mapped α2(I)CB4 and appeared to be an in-frame deletion of about 30 amino acids. The procollagen from cultured fibroblasts of this patient showed a decreased thermal stability and resistance of the mutant molecules to digestion by procollagen N-proteinase.

Genomic cloning and sequencing of the mutant gene (Kuivaniemi et al., 1988) revealed this not to be an in-frame deletion as such, but an RNA splicing defect caused by a 19bp deletion, taking out the 3' end of intron 10 and 5' end of exon 11; hence losing the 5' splice junction of exon 11. At each end of the deleted region is a 4bp direct repeat, preventing precise definition of the boundaries of the deletion.

The consequence of this deletion is to generate an RNA transcript that is spliced from the last codon of exon 10 to the first codon of exon 12. Thus in effect the whole of exon 11 has been 'deleted', and because the exons of the α-chains of fibrillar collagens begin with a complete codon, the 'deletion' is in-frame.

This mutation is of further interest because the affected members of the patient's family show variable phenotypic expression. One possible explanation of this is that the efficiency of exon skipping may be different in different individuals. More complex mechanisms however, may be involved in explaining the pleiotropism of single-base mutations of type I procollagen seen in other families such as the family with moderate OI with a cysteine residue in the α1(I) chain (de Vries and de Wet, 1986).

This point raises a word of caution; we may be able to guess the primary effect of a mutation in type I collagen, but the actual phenotype may be influenced by the background genetic constitution of any affected individual.

All the mutations described thus far, have been heterozygous. All the lethal OI's characterised are believed to have been due to spontaneous mutations - germline or early somatic, and all non-lethal cases have been inherited as autosomal dominants. Only one case of a proband with lethal OI has yet been described demonstrating different mutations on both alleles of a type I collagen gene. Skin fibroblasts from the infant showed all pro-α2(I) chains to be shortened (de Wet et al., 1983b). This shortening was mapped to $CB3,5^{A}$ and was of about 20 amino acids. Neither parent had shortened pro-α2(I) chains, but the father however, showed a reduced rate of α2(I) synthesis. Thus it appeared that the infant had inherited a null α2(I) allele from its father and had recieved a mutated allele from its mother.

R-loop mapping of DNA:RNA hybrids formed between mutant mRNAs and normal pro-α2(I) genomic clones and also S1 mapping of the mutant pro-α2(I) mRNAs (de Wet et al., 1986) mapped the mutation to exon 28. In addition this approach showed that the paternal α2(I) allele was being transcribed but probably not translated or if so, very inefficiently.

Tromp and Prockop (1988), cloned the mutant gene and sequenced the region of interest. The mutation was discovered to be a single base mutation substituting an adenine for a guanine, changing the universal consensus sequence for the 3' splicing site of of intron 27 from -AG- to -GG-. This efficiently causes the splicing out of exon 28, giving the effect of a whole exon deletion. Tromp and Prockop failed to fully address the question of the contribution of the null allele to the mutant phenotype of this individual. In the light of the characterisation of the mild OI (Kuivaneimi et al., 1988), as a splicing defect removing exon 11, it is likely that the absence of a normal fully expressed allele from the infant's father, has caused what we might expect to be a mild OI to be lethal. This infant can therefore be described as being a 'compound heterozygote'.

The case also illustrates the point of what has been termed 'protein suicide' in collagen biosynthesis (Prockop, 1984): A non-functioning allele for COL1A1 (eg. Barsh et al., 1982) or COL1A2 (eg. de Wet et al., 1983b) is likely to be less disastrous than a structural mutation in a normally expressed type I collagen gene. The mutant chain is usually allowed to participate in triple-helix formation giving the effect of a wholly mutant procollagen protein. Whereas in the case of a non-functioning allele, all protein molecules formed will be normal.

The only characterised example of a homozygous defect in type I

procollagen is a child with a moderately severe OI, of consanguinous parents. Cultured fibroblasts from this individual were observed to lack pro-α2(I) chains and the type I procollagen consisted of trimer pro-α1(I) (Deak et al., 1983). α2(I) mRNA levels in the child were observed to be normal, thus the lack of α2(I) chains was not due to a decrease in α2(I) transcription. S1 mapping studies revealed a homozygous defect in the carboxy-propeptide coding region of the pro-α2(I) mRNA (Dickson et al., 1984).

Cloning and sequencing of the mutant region of COL1A2 revealed a 4bp deletion in exon 51 changing the last 33 amino acids of the carboxyl propeptide (Pihlajaniemi et al., 1984). The normal stop codon was lost and a new one was created 4bp 3' to the original site; making the α-chain with the deletion, exactly the same length as the normal α-chain. The asymptomatic parents were seen to be heterozygotes for this defect. The complete failure of the child to produce type I procollagen containing α2(I) chains shows the crucial role of the carboxyl-propeptide sequence in chain assembly and goes some way to explain the evolutionary conservation of the carboxyl-propeptide (Yamada et al., 1983).

Only one example of an insertion has been recorded in an OI patient (Byers et al., 1988). In this patient 50-60 amino acids were inserted into half the pro-α1(I) chains synthesised, probably by tandem duplication of an approximately 600bp segment of COL1A1. Characterisation of the mutation in this patient is eagerly awaited.

All the 'deletions' described, except the 4bp carboxy-terminus deletion (Pihlajaniemi et al., 1984) have been in-frame. This has not been due to the deletions being multiples of 3bp, but because whole exons have been deleted or spliced out of the RNA molecule. Because the fibrillar collagen gene helical domain exons always begin at a glycine

codon, the exons act as 'cassettes' and can be deleted without interrupting the helix folding, as the glycine residues amino-terminal to the deletion are still in-register. However, it should be pointed out that the patient with the deletion of exons 34-40 (Willing et al., 1988) had type I procollagen with a lower than normal thermal stablilty and increased post-translational modification amino-terminal to the deletion junction. Thus, in this case, propagation of the triple helical structure may not be be normal or the structure formed is not sufficiently stable to prevent further modification.

What is disturbed in all deletion mutants though, is the distribution of charged residues in the X and Y positions of the Gly-X-Y motif, which is normally precisely matched between the α1(I) and α2(I) chains. Amino-terminal to a deletion, the charge pattern will have been interrupted and become staggered. This will probably cause a decrease in the thermal stablitiy of the procollagen molecule.

Another effect of a deletion is that the unequal lengthed chains will produce a ragged amino terminus of the helix, disturbing the recognition site of N-terminal proteinase (see Sippola et al., 1984).

Thus across the spectrum of OI phenotypes, there is rapidly accumulating data on the heterogeneity of defects behind this complex group of disorders. Despite this heterogeneity and the incompleteness of the data, the relationship between cause and effect in OI is beginning to make sense.

1.4.4 Other inherited connective tissue disorders

Recently two rarer inherited connective tissue disorders have been shown to be caused by defective collagen. In two families with the Stickler syndrome, the type II collagen gene been shown to be tightly

linked to the disease (Francomano et al., 1987).

In one family with spondylepiphiseal dysplasia a single exon was found to have been deleted from one allele of the type II collagen gene (COL2A1) (Lee et al., 1989).

1.5 THE DETECTION OF MUTATIONS IN GENETIC DISEASES

To date, most defects identified in mutant genes have been based upon large size differences, as detected by Southern or Northern blotting or by sequencing of cloned DNA fragments. It is probable that a large proportion of disease-causing mutations are point mutations. Because of the rarity of polymorphic markers in linkage disequilibrium with mutant genes, and point mutations that create or destroy restriction endonuclease recognition sites, it is of fundamental importance that a technique be developed which will identify point mutations. Despite considerable effort, no general method of detecting all possible mutations in genomic DNA is yet available. However several techniques have been developed that go some way towards this goal.

One method is based upon the melting behavior of DNA, another on the hybridisation dynamics of DNA, and others on the susceptibility of mismatches in nucleic acid heteroduplexes to enzymic or chemical cleavage.

1.5.1 Gradient denaturing gel electrophoresis

DNA fragments differing by single base substitutions can be separated from each other by electrophoresis in polyacrylamide gels containing an ascending gradient of the DNA denaturants urea and formamide (Fischer and Lerman, 1983).

DNA fragments of identical size but differing by a single base pair, will migrate through the gel initially at an identical rate. As they migrate into a critical concentration of denaturant, specific regions or 'domains' within the fragment melt to produce partially denatured DNA. Melting of a domain is accompanied by an abrupt decrease in mobility. The position in the gel at which the decrease in mobility is observed corresponds to the melting temperature (Tm) of the domain.

Since a single-base substitution within the melting domain results in a Tm difference, partial denaturation of the mutant and wild-type DNA fragments will occur at different positions in the gel. DNA molecules can therefore be separated on the basis of very small differences in the Tm values of their melting domains.

However if a single-base substitution lies in the highest melting temperature domain, the fragment undergoes complete strand dissociation, and the resolving power of the gel is lost. Thus, mutations in the melting domain with the highest Tm cannot be separated from the wild-type DNA fragment. In cloned DNA, this limitation has been overcome by attaching a GC-rich DNA sequence, called a 'GC-clamp', next to the DNA fragment of interest (Myers et al., 1985a; 1985b). This GC-clamp makes it possible to detect nearly all single-base substitutions in all melting domains of an attached DNA fragment by preventing complete strand dissociation as it proceeds through the gel. An important point is that most single-base substitutions that do not alter the base composition (e.g. AT to TA transversions) can be readily separated on the gel. Thus, the changes in Tm due to differences in base stacking alone are sufficient to allow discrimination.

A significant increase in the power to detect single-base changes in genomic DNA has been achieved by examining heteroduplexes between wild-type and mutant DNA fragments on gradient denaturing gels (Myers et al., 1985c). The greater sensitivity is presumably a result of the instability caused by the disruption of the helix at the point of mismatch. This method has been simplified by Noll and Collins, (1987), who circumvented the need to remove unbound probe with S1 nuclease, by adding an excess of complementary M13 DNA, allowing only genomic DNA:probe hybrids to migrate into the gel. The simple production of RNA probes has enabled the gradient denaturing gel system

to be extended to RNA:DNA and RNA:RNA heteroduplexes for the detection of single base substitutions (Smith et al., 1986).

1.5.2 Allele-specific oligonucleotides

Common characterised single-base substitutions can be detected by differential hybridisation (Wallace et al., 1979). Synthetic oligonucleotides (usually 19-mers) are prepared that correspond to either the normal or the mutant DNA sequence. At a defined stringency of hybridisation, in a simple 'dot-blot' analysis, the allele identical to the oligonucleotide probe will bind the probe, and the other allele will not. Such allele-specific oligonucleotides (ASOS) have permitted the accurate detection of point muations in the β-globin gene (Conner et al., 1983), but the procedure has suffered from high backgrounds due to non-specific binding. However, the advent of the polymerase chain reaction (PCR) has been used to markedly improve the ASO detection method (Saiki et al., 1986). PCR primers are designed that hybridise to DNA flanking the mutation, and the fragment is then amplified. In the subsequent dot-blot analysis, the amount of other genomic DNA is tiny, thus preventing any non-specific probe hybridisation.

Although this method has the potential to reveal any mutation, the sequence of the mutation must already be known. In addition, the number of inherited diseases caused by common mutations is likely to be very few.

1.5.3 S1 mapping

S1, a single-strand-specific nuclease can be used to cleave mismatches in DNA:DNA or DNA:RNA heteroduplexes (Shenk et al., 1975; Berk and Sharp, 1977). The technique reliably detects differences of 5 base pairs or more between the probe and the target DNA or RNA molecule

(Cheboulne et al., 1984). It had been suggested by Shenk et al., (1975) that single-base substitutions may be detected by S1 mapping, but this possibility has never been substantiated. It would thus seem that S1 mapping, being a simple and reliable method for detecting length differences in nucleic acid heteroduplexes, is not sensitive enough to detect single base mismatches.

1.5.4 RNase A mapping

That single-base mismatches in RNA:RNA heteroduplexes could be cleaved by ribonucleases was first observed by Freeman and Huang, (1981) in a study mapping temperature sensitive mutations in heteroduplexes of ^{32}P-labelled vesicular stomatis virus mRNA and virion RNA. The ability to generate RNA probes (Green et al., 1983; Zinn et al., 1983: Melton et al., 1984) enabled this strategy to be applied to other organisms.

Around the same time Winter et al., (1985) and Myers et al., (1985d) used RNA probes to detect single-base substitutions in human genes. Winter et al. used RNase A to cleave single base mismatches in heteroduplexes of c-Ki-ras antisense RNA (probe) and tumour-derived c-Ki-ras mRNA. Myers et al. used RNase A to cleave single-base mismatches between wild-type ß-globin sense and antisense RNA (probe) and mutant genomic and cloned DNA.

The technique is based upon the ability of RNase A, a single-strand-specific ribonuclease to cleave mismatches in RNA:RNA or RNA:DNA heteroduplexes. The ^{3} ^{2}P-labelled single-stranded RNA probe is synthesised from a wild-type DNA template cloned into a plasmid vector containing a bacteriophage transcription system. The RNA probe is hybridised to denatured genomic DNA or total cellular RNA in solution. The resulting RNA:RNA or RNA:DNA hybrid is then treated with RNase A.

The probe strand of the hybrid is then analysed by electrophoresis in a denaturing gel. If the target RNA or DNA is wild-type, a single band is observed on the autoradiograph of the gel, since the hybrid has not been cleaved by the RNase. However if the target RNA or DNA contains a mutation, two additional bands will be present. If the mutation is a single-base substitution, very small deletion or an insertion of any size; the sum size of the two bands will equal the size of the wild-type band. If the mutation is a deletion the two bands may not add up to the wild-type size or there may only be one additional band. For heterozygous mutations, the wild-type band will remain, representing the normal allele.

In all recent studies using RNase A cleavage mapping, the target substrate used has been RNA. This has the advantage that only the coding regions of the gene are analysed, allowing the often large and frequent introns of human genes to be excluded from analysis. An improvement has been made to the RNA:RNA hybrid option by including a mRNA purification step after hybridisation to remove excess probe, thus reducing background radioactivity (Gibbs and Caskey, 1987).

Unfortunately only some mismatches are cleaved by RNase A digestion. Using a collection of β-globin mutations, Myers *et al.*, (1985d) examined all 12 possible types of mismatch in RNA:DNA hybrids. Recognition of each type of mismatch varied between 0 and 100%, indicating that the flanking sequences have an effect on the ability of RNase A to cut at a given mismatch. Approximately 25% of all possible single-base substitutions were detected in using one species of RNA probe (*i.e.* sense or antisense).

It is therefore likely that only about 25% of single-base substitutions can be detected in RNA:RNA hybrids. However in their study of 14 Lesch-Nyhan patients, Gibbs and Caskey, (1987), identified

mutations in 5 patients using the RNA:RNA option. It is possible that the figure could be raised as high as 80% , if additionally, sense RNA probes were hybridised to cDNA (perhaps amplified by PCR). RNase A cleavage, like gradient denaturing gel electrophoresis will only identify some point mutations. However RNase cleavage has the advantage that it will not only identify point mutations, but map them too. The use of overlapping RNA probes allows very fine mapping of point mutations. In addition to the example of Gibbs and Caskey, the technique has been used in the characterisation of the mouse 'sparse fur' mutation (Veres et al., 1987), the incidence of K-ras gene mutations in human colon tumours (Forrester et al., 1987), the analysis of variability in the infuenza virus (Lopez-Galindez et al., 1988) and the identification of mutations at the RB1 locus in retinoblastoma tumours (Dunn et al., 1988).

1.5.5 Chemical cleavage of DNA heteroduplexes

The ability of chemicals to react with mismatched regions in DNA:DNA heteroduplexes has begun to be exploited in the detection and mapping of point mutations.

Novak et al., (1986) developed a method in which the chemical 1-cyclohexyl-3-(2-[4-(4-methyl)morpholinyl]ethyl)carboniimide was used to 'tag' mismatched base pairs in DNA:DNA heteroduplexes. The presence of this tag has the effect of retarding duplexes with mismatches when electrophoresed on polyacrylamide gels. However results are not always easy to interpret, and not all mismatches are detected.

Cotton et al., (1988) further investigated the ability of chemicals to bind mismatched regions of DNA. Two reagents, osmium tetroxide and hydroxylamine, were found to potentially recognize all variants as they react with mismatched thymidine and cytosine

respectively. This method in addition to its possible universal recognition, also maps mutations, because piperidine is employed to cleave the DNA at the modified mismatch. One apparent disadvantage of the method is that so far only small probes of less than 500 bases have been used.

1.5.6 Summary

Five methods have been described that can identify mutations not detectable Southern and Northern blotting.

Three methods (gradient denaturing gels, RNase A and chemical heteroduplex cleavage) have the potential to identify unknown single base substitutions. Gradient denaturing gel electrophoresis theoretically, should be able to detect upto 50-70% of all single base substitutions in genomic DNA. RNase A cleavage mapping of genomic DNA may be able to detect over 50% of single base changes, but probably only 25-35% of single base changes in mRNA. The recent development of chemical modification followed by chemical cleavage may be able to detect all single base substitutions in genomic DNA.

Because of the size and intron/exon structure of collagen genes; for a study of mutations in OI, any method that used RNA as a substrate would be a favourite. The RNase A and chemical cleavage methods have the advantage over gradient denaturing gels in that they have the power to map any mutations detected.

In this thesis a study of RNase A mapping in OI is described. Had such a study been initiated now, the potential to identify all mutations by the chemical cleavage method of Cotton et al., (1988), may have led it to be chosen, despite the necessity to analyse genomic DNA rather than the RNA.

If the chemical cleavage method is successfully employed as a

screening tool, it may become the first general method for detecting single base substitutions.

1.6 Object of research

It is apparent that defects in collagen genes are responsible for some of the inherited connective tissue disorders.

A study of genetic linkage was undertaken to determine if any of the major fibril-forming collagens were defective in a large Marfan syndrome family.

Another aim was to improve the understanding of the relationship between genetic defect and phenotype in OI. This was approached by using a technique that should redress the probable bias of known defects in OI.

CHAPTER 2

MATERIALS AND METHODS

2.1 DNA, RNA AND CELL LINES

Control genomic DNA and RNA samples were prepared from fresh placentae obtained from the Royal Infirmary, Leicester. Marfan pedigree DNAs were prepared and supplied by Dr. Marion Keston (MRC clinical and population cytogenetics unit, Edinburgh).

The type II O.I. cell lines; F.S. and F.W. were donated by Drs. P.J. Batstone and D. Goudie, respectively (Duncan Guthrie Institute of Medical Genetics, Glasgow). The type II O.I. cell lines; GM6260, GM5747, GM2328 were purchased from the National Institute of General Medical Sciences, USA. The characterised type II O.I. cell line, A.K. and the uncharacterised type II OI cell lines, M.G., P.E., P.C., F.B. and N.A. were gifts from Dr. Beat Steinmann (Universitats-Kinderklinik, Zurich). DNAs from the parents of F.B. were supplied by Dr. Andrea Superti-Furga (Universitats-kinderklinik, Zurich).

2.2 PLASMIDS AND BACTERIAL STRAINS

2.2.1 Recombinant Plasmids

The human α1(I) collagen cDNA clones; Hf677 and Hf404 (Chu et al., 1982) were gifts from Dr. F. Ramirez (Morse Institute of Molecular Genetics, New York, USA). The COL1A1 genomic clones CG102 and CG103 (Barsh et al., 1984) were provided by Dr. R. Gelinas (Fred Hutchinson Research Center, Seattle, USA). The COL1A2 genomic clones NJ3' (Tsipouras et al., 1983), λHpro.α2(I)-1 (Tajima et al., 1984) and HpC1 (Dalgleish et al., 1982) were provided by Dr. F. Ramirez, Dr. S. Pinnell (Duke University, USA), and Dr. R. Dalgleish (Universtiy of

Leicester). The COL2A1 genomic clone cosHcol1 (Weiss et al., 1982) was provided by Dr. E. Solomon (ICRF, London). The COL3A1 genomic clone Idf17 (Chu et al., 1985) was provided by Dr. F. Ramirez.

2.2.2 Nonrecombinant Plasmids

Nonrecombinant plasmids pUC13 (Messing, 1983), Bluescribe and Bluescript SK$^-$ (Stratagene Cloning Systems, San Diego, USA), pBR328 (Soberon et al., 1980), pTZ18R and pTZ19R (Mead et al., 1986) were used in constructing recombinant clones.

2.2.3 Bacterial Strains

The E. coli stain used was a recA derivative of JM83 (Messing and Vieira, 1982) produced by Mark Matfield (ICI Joint Laboratory, University of Leicester) and provided by Prof. W.J. Brammar (of the same laboratory). This strain has the following genotype: ara, X(lac-pro), strA, thil, Φ80dlacIq, ZM15, rpsL, recA.

2.3 ENZYMES, ANTIBIOTICS, CHEMICALS AND REAGENTS

The source of the most important of these are listed below. Other chemicals not listed were all analytical grade.

	Source
Ampicillin	Sigma
Chloramphenicol	"
Tetracycline	"
Dithiothreitol	"
Isoproyl-β-D-thiogalacto-pyranoside (IPTG)	"
Polyethylene glycol (PEG) grade 6000	"
Herring sperm DNA	"

Diethylpyrocarbonate (DEPC)	"
N-2-Hydroxyethlpiperazine -N-2-ethane-sulphonic acid (HEPES)	"
Piperazine-N-N'-bis[2-ethane-sulphonic acid] (PIPES)	"
RNase A	"
Proteinase K	"
Orange G	"
Formamide	"
DNase I	Pharmacia
Ficoll 400	"
Ribonucleotide triphosphates	"
DNA polymerase I (Klenow fragment)	"
AMV Reverse transcriptase	"
Bovine serum albumin	"
T7 RNA polymerase	"
Phenol	Fisons
Polyvinylpyrrolidine	"
3-(N-morpholino) propane sulphonic acid (MOPS)	"
2-amino-2-(hydroxymethyl) propane-1,3-diol (TRIS)	"
Formaldehyde	"
T4 DNA ligase	Bethesda Research Labs
Restriction endonucleases	"
Calf intesstinal phosphatase	Boehringer Manheim
Calf liver tRNA	"
Agarose	FMC
Acrylamide	Serva
Amberlite resin	"
Bisacrylamide	Uniscience

Urea	BDH
Dichlorosilane	"
N,N,N',N'-Tetramethylethylenediamine (Temed)	Bio-Rad
Ammonium persulphate	"
5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-gal)	Anglian
Dulbecco's modified eagles medium (DMEM)	Gibco/BRL
Foetal calf serum	"
Trypsin EDTA	"
Phosphate buffered saline	"
RNase inhibitor	Promega
Taq DNA polymerase	Cambio

2.4 SOLID AND LIQUID MEDIA

2.4.1 For bacteria

Luria broth (10g Difco Bacto Tryptone; 5g Difco Bacto yeast extract; 5g NaCl per litre of distilled water). Luria agar plates were prepared by solidifying liquid medium with 15g Difco Bacto-agar per litre.

2.4.2 For cell culture

Dulbecco's modification of Eagles minimum essential medium (DMEM, see Morton, 1970) was used to culture all fibroblast cells and was obtained as a working strength solution containing 4.5g glucose per litre and stored at 4°C. DMEM was supplemented with foetal calf serum to the appropriate percentage. Serum was stored at -20°C in aliquots.

Penicillin and streptomycin were added to DMEM at concentrations of 100 units per ml and 100μg per ml respectively.

Phosphate buffered saline (PBS) was obtained as a 10x working strength solution and Trypsin EDTA was obtained as a 1x strength solution.

2.5 GENERAL TECHNIQUES FOR HANDLING DNA AND RNA

2.5.1 Phenol extraction

DNA and RNA solutions were mixed with 1 volume of phenol; chloroform; isoamyl alcohol; 8 hydroxyquinoline (100:100:4:0.1, w:v:v:w) saturated with TE, pH7.5 (10mM Tris-HCl, 1mm EDTA) and briefly centrifuged to separate the phases. The upper aqueous phase, containing the DNA or RNA was removed carefully to avoid disturbing the interphase and the phenol layer was re-extracted with an equal volume of TE, pH7.5. The phenol was AR grade and not re-distilled.

2.5.2 Ethanol Precipitation

DNA and RNA were typically precipitated from solution by the addition of 0.1 volume of 4M NaCl and 2.5 volumes of ethanol. The solution was mixed well and chilled at -20°C for approximately 20 minutes. Precipitated DNA or RNA was pelleted by centrifugation at either 10,000 rpm for 20 minutes at 0°C (Sorvall HB4 rotor), or in a MSE Micro-Centaur centrifuge for 10 minutes at maximum speed. The supernate was discarded and the pellet rinsed with 70% ethanol, centrifuged for 2 minutes and the 70% ethanol removed. DNA and RNA pellets were vacuum-dried and resuspended in an appropriate solution for further manipulation.

2.5.3 Restriction endonuclease digestion of DNA

DNAs at a final concentration of 0.2mg/ml were incubated in the manufacturers recommended buffer at 37°C for 1 hour, unless otherwise

stated. Digests of genomic DNA however, were incubated for 5 hours at 37°C. Spermidine trichloride was added to DNA that proved to be difficult to digest, at a final concentration of 4mM, to enhance the efficiency of restriction endonuclease digestion (Bouche, 1981; Pingoud, 1985). For DNA that was to be electrophoresed, a 0.1 volume of Orange G dye (0.1% w/v Orange G, 20% w/v ficoll 400, 100mM EDTA) was added and the mixture loaded onto an agarose gel of the appropriate size and concentration.

2.5.4 Alkaline phosphatase treatment of DNA

500ng DNA was digested with the appropriate restriction endonuclease. The DNA was then heated to 65°C for 5 minutes to inactivate the enzyme. Then 0.01 units of calf intestinal phosphatase were added and incubated at 37°C for 30 minutes. The DNA was then electrophoresed on a low gelling temperature agarose gel and prepared for ligation as described in section 2.5.5.

2.5.5 Ligation of DNA fragments into Plasmids

This method was based upon that of Crouse *et al.* (1983).

DNA to be subcloned was prepared by appropriate restriction endonuclease digestion, followed by electrophoresis on a 0.6% low gelling temperature agarose gel: 500ng of vector DNA was digested and a sufficient amount of DNA to be subcloned was digested so that an equimolar amount of DNA could be ligated. If necessary, vector DNA was dephosphorylated as described before. The fragment and vector DNA were excised on a U.V. transilluminator, placed into separate pre-weighed microfuge tubes and weighed. The volume of each gel slice was estimated by assuming that the gel has a density of 1g/ml.

The gel slices were melted by heating at 70°C for 10 minutes and cooling to 37°C for 5 minutes. Ligation reactions containing 100ng of DNA

were set up in a 100µl volume of ligase buffer (30mM Tris-HCl, 10mM $MgCl_2$, 5mM DTT, 1mM ATP, pH 7.6; stored as a 5x concentrate at -20°C). 0.5U of T4 DNA ligase was added and the mixture incubated at 15°C overnight. Blunt-end ligations were incubated at 4°C.

2.5.6 Note on handling RNA

Any solutions, glassware or plasticware which were to come into contact with RNA were made up with, or pre-treated with, a 0.05% solution of diethylpyrocarbonate (DEPC) in distilled water. Solutions and equipment were incubated overnight at room temperature before DEPC removal by autoclaving. Solutions containing Tris were not DEPC treated, but were made up with DEPC-treated distilled water.

2.5.7 Determination of DNA and RNA concentrations

Following genomic DNA, large scale plasmid DNA and placental RNA preparations, the nucleic acid concentration was determined by U.V. absorbance: A known dilution of DNA or RNA was analysed by a Unicam SP1800 U.V. spectrophotometer at 260nm. The concentration (in µg/ml) was determined by multiplying the absorbance by 50 for double-stranded DNA; 35 for single-stranded DNA and RNA and by the dilution factor.

Concentrations of quick plasmid DNA and fibroblast RNA preparations were determined visually by comparing nucleic acid intensity, on a U.V. transilluminator, against DNA or RNA of known concentration.

2.6 AGAROSE GEL ELECTROPHORESIS

2.6.1 DNA gels

Horizontal agarose gels with loading slots 6mm wide were prepared and run in electrophoresis buffer (40mM tris-acetate, 2mM EDTA, pH 7.7)

containing ethidium bromide at 0.5µg/ml. The gel size varied according to the number of samples to be run and the resolution required. The concentration of agarose was varied between 0.5 and 2% (w/v), depending upon the DNA fragment sizes to be separated. Molecular weight markers were usually λ/HindIII and ΦX174/HaeIII. Marker DNA samples were heated at 65°C for 5 minutes to dissociate the cohesive ends of λ. Prior to loading, DNA samples were mixed with 0.1 volume of Orange G dye. Gels were run at room temperature at 60-100 volts for 1-2 hours or 15-45 volts overnight, as appropriate, until the Orange G dye had reached the end of the gel.

2.6.2 RNA gels

The quality of RNA samples was tested by electrophoresis in horizontal agarose gels as for DNA samples but with the addition of a heat shock step at 65°C for 3 minutes prior to loading the RNA on the gel. RNA was separated sufficiently well on 1% agarose gels.

2.6.3 Recovery of DNA from agarose gels

DNA was electrophoresed on an agarose gel as described before. A sheet of dialysis membrane was cut to give a single layer 1cm wider than the loading slot and 1mm deeper than the gel. The membrane was boiled for 10 minutes in TE. While using a UVSL-58 hand-held long wave U.V. wand (Ultra-Violet Products Inc., California, USA) to visualise the DNA, the gel was cut just in front of the DNA to be collected . Using a pair of tweezers, the membrane was inserted into the cut, until it rested on the gel base-plate. The upper projecting part of the membrane was folded over, so that it rested on the gel surface pointing towards the loading slot. If necessary, another sheet of membrane was inserted behind the DNA to be collected, to prevent the migration of unwanted DNA. The

DNA was electrophoresed onto the membrane at 150-200 volts. When fully loaded, with the current still on, the cut was extended so that the upper part of the gel could be prised away from the lower part to open up a gap of 1-3mm, such that the DNA face of the membrane is not touching the gel. The gel is left in this state for a few minutes to ensure complete loading. Then with the current still on, the membrane was grasped with a pair of tweezers and transferred in a single movement to a microfuge tube. This was always done as rapidly as possible, and must take no longer than 1 second. With the end of the membrane trapped in the lid of the tube , the tube was centrifuged for 1 minute at maximum speed in a MSE MicroCentaur centrifuge. The membrane was then re-wetted with 30µl of distilled water and centrifuged again. The DNA was recovered by phenol extraction and ethanol precipitation.

2.6.4 Gel photography

Gels were visualised and photographed on a Chromato-Vue C-63 U.V. light transilluminator (Ultra-Violet Products Inc., California, USA). The camera was a Kodak MP-4 land camera and the film used was Polaroid type 55, Kodak Tri-X or Kodak professional film. The Kodak films were developed in M&B Qualitol developer and fixed in Kodak rapid fixer.

2.7 ACRYLAMIDE GEL ELECTROPHORESIS

To resolve DNA fragments in the region of 1000 base pairs or less, acrylamide gels were used in preference to high percentage ordinary agarose and sieving agarose gels.

Standard gels were 12% (w/v) acrylamide and prepared as below:

40% Acrylamide stock	12ml
10x TBE	2ml
Distilled water	26ml
	40ml

40% acrylamide stock solution: 38g acrylamide and 2g N,N'-methylenebisacrylamide were dissolved in a final volume of 100ml distilled water. 5g Amberlite resin was added and the mixture stirred for 1 hour. The resin was removed by filtration and the 40% acrylamide stock stored at 4°C. 10x TBE is: 1M Tris base, 0.83M boric acid, 30mM EDTA.

140µl 10% ammonium persulphate and 30µl TEMED were added to the gel solution and mixed in thouroughly. The mix was poured using a syringe into a 1mm thick 16cm x 20cm Bio-Rad Protean II gel mould, with 6mm wide, 2cm deep well-formers. The gel was allowed to stand for at least 30 minutes to polymerise. Once set, the gel was mounted into the gel apparatus and the top and bottom buffer chambers both filled with 0.5x TBE.

DNA samples were of 100-500ng in 12µl volumes. 3µl loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 40% (w/v) sucrose) was added and mixed prior to loading. the gel was electrophoresed at 25mA with cold water circulating through the cooling tank. Gels were run until the xylene cyanol dye was approximately two-thirds of the way down the gel. This usually took between 2 and 3 hours.

Following electrophoresis, the gel was stained in 0.5µg/ml ethidium bromide, with gentle shaking for 20 minutes. The gel was briefly de-stained in distilled water for 5 minutes to reduce background staining. The gel was then placed on a U.V. transilluminator to be photographed.

2.8 PREPARATION OF PLASMID DNA

2.8.1 Plasmid 'mini-preps'

This is the method of Ish-Horowitz and Burke (1981), but starting with a harvest of 1.5ml of overnight culture.

2.8.2 'Quick' plasmid preps

A 6ml culture of the plasmid-containing E.coli JM83recA was grown with shaking at 37°C in Luria broth overnight, with the appropriate antibiotic to provide selection for the plasmid. 4ml of the culture was used to inoculate 100ml of pre-warmed Luria broth and grown with shaking at 37°C to an optical density (O.D.) of 1.5 at 600nm. 20mg of chloramphenicol was then added to increase plasmid copy number, and the culture allowed to grow overnight.

The bacteria were harvested by centrifugation at 7,000 rpm for 10 minutes at room temperature in a Sorvall GS3 rotor and Sorvall RC5B centrifuge. The supernate was removed and the pellet resuspended in 6ml solution I (25mM Tris-HCl, 50mM glucose, 10mM EDTA, pH 8.0) and incubated for 5 minutes at room temperature. 12ml solution II (freshly prepared 0.2N NaOH, 1% SDS) was then added and gently mixed. To precipitate chromosomal DNA and most of the proteins, 9ml cold solution III (5M potassium acetate, pH 4.8) was added and the solution was vortexed and incubated on ice for 5 minutes. The tube was then centrifuged at 10,000 rpm at 4°C for 10 minutes in a Sorvall HB4 rotor in a RC5B centrifuge. The supernate was filtered through polyallomer wool. A 0.6 volume of propan-2-ol was added with mixing to precipitate the nucleic acids. After a 10 minute incubation at room temperature, the mixture was centrifuged at 10,000 rpm as before. The pellet was washed in 70% ethanol, centrifuged at 10,000 rpm for 5 minutes as before and resuspended in 3ml TE. 1.5ml 7.5M ammonium acetate was then added to precipitate any residual protein. After a 10 minute incubation at room temperature the mixture was centrifuged at 10,000 rpm for 10 minutes at room temperature. The supernatant was then transferred to a clean tube and mixed with 2.5 volumes of ethanol and allowed to stand for 5 minutes. The nucleic acids were pelleted by centrifugation at 10,000 rpm for 10 minutes at 4°C,

again using the Sorvall HB4 rotor and RC5B centrifuge. The pellet was washed in 70% ethanol and centrifuged for 5 minutes as above. The nucleic acids were then resuspended in 300µl TE. As for plasmid 'mini-preps', the large amounts of RNA present were removed by digestion with a 0.1 volume of 2µg/ml boiled RNase A during restriction endonuclease digests of the DNA.

2.8.3 Large scale plasmid preps

This method is similar to the 'quick' plasmid prep, but includes a caesium chloride gradient centrifugation step.

A 10ml culture of the relevant E.coli strain was grown overnight in Luria Broth containing the appropriate antibiotic at 37°C with shaking. 8ml of this culture was used to inoculate 400ml of the same broth in a 2L flask. The culture was grown to an optical density (O.D.) of 1.5 at 600nm. Then 80mg chloramphenicol was added and the culture allowed to continue incubating overnight.

The bacteria were harvested by centrifugation at 7,000 rpm in a Sorvall GS3 rotor and Sorvall RC5B centrifuge. The supernate was removed and the pellet resuspended in 12ml solution I (as above) and left at room temperature for 10 minutes. 24ml solution II was added with gentle shaking and left on ice for 5 minutes, followed by the addition of 12ml solution III and a further 15 minute incubation on ice. The mixture was centrifuged at 10,000 rpm at 4°C in a Sorvall HB4 rotor in a Sorvall RC5B centrifuge. The supernate was filtered through polyallomer wool into an acid-washed measuring cylinder. A 0.6 volume of propan-2-ol was added, mixed, allowed to stand for 10 minutes, and centrifuged as in the previous step. The pellet was rinsed in 70% ethanol, centrifuged at 10,000 rpm for 5 minutes and then drained. The pellet was then resuspended in 6ml TE pH 8.0. For each ml of DNA solution 1g CsCl and

0.1ml ethidium bromide (10mg/ml) were added. The mixture was thouroughly mixed and then centrifuged at 10,000 rpm for 10 minutes at room temperature in the RC5B centrifuge. The density of the supernatant was monitored with a refractometer, and adjusted as necessary with CsCl or TE, pH 8.0 to achieve a density of 1.57 g/ml.

The solution was next loaded into a 16mm x 76mm Beckman Quick-Seal polypropylene tube and topped up with a 1.57 g/ml CsCl solution. The tubes were sealed and centrifuged in a Beckman 75Ti rotor 65,000 rpm overnight in a Sorvall OTD65B ultracentrifuge.

The plasmid band was removed with a syringe and an 18 gauge needle and transferred to a Corex glass test-tube. The ethidium bromide was extracted by repeated mixing with equal volumes of propan-2-ol saturated with CsCl, discarding the propanol layer each time until the solution became colourless. The DNA was dialysed extensively against large volumes of TE to remove the CsCl.

The DNA was then ethanol precipitated and dissolved in an appropriate volume of TE.

2.9 TRANSFORMATION OF E.COLI

2.9.1 Preparation of competent cells

This method was adapted form that of Hutchinson and Halverson (1980), and was used to transform E.coli strain JM83recA.

A flask containing 5ml Luria broth was inoculated with a fresh colony of bacteria and grown overnight at 37°C. 1ml of culture was used to inoculate 50ml of fresh medium and grown to an O.D. of 0.36-0.44 at 560nm. The cells were pelleted by centrifugation at 3,000 rpm in a MSE Centaur bench-top centrifuge. The cells were suspended in 20ml of ice-cold sterile 0.05M $CaCl_2$ and placed on ice for 15 minutes. The cells were pelleted by centrifugation, as above, and resuspended in 5ml

ice-cold sterile 0.05M $CaCl_2$, 5% glycerol. The cells were snap frozen in 200µl aliquots in sterile 1.5ml microfuge tubes by placing the tubes in a dry-ice/ethanol bath. The frozen competent cells were either stored at -80°C or in liquid nitrogen. Cells stored at -80°C were best used in few weeks, as competence drops rapidly. Cells stored in liquid nitrogen remained well competent for a few months.

2.9.2 Transformation

10ng of plasmid DNA in 100µl of freshly prepared 10mM Tris-HCl, 10mM $MgCl_2$, 10mM $CaCl_2$, pH 7.4, was added to 200µl of competent cells, having just thawed on ice. The mixture was kept on ice for 25 minutes and then heat-shocked at 37°C for 90 seconds, followed by a 10 minute incubation at room temperature. 1ml of prewarmed Luria broth was added and the cells incubated at 37°C for 1 hour, to allow the cells to recover and begin expressing their antibiotic resistance gene. 20µl and 200µl aliquots of the transformed cells were plated onto Luria agar plates containing the necessary antibiotic. If appropriate IPTG and X-gal were each added at a concentration of 50µg/ml to the agar.

2.10 PREPARATION OF RADIOLABELLED DNA FRAGMENTS

2.10.1 Oligo-labelling of isolated DNA fragments

This method is adapted from that of Feinberg and Vogelstein (1984).

The fragment of DNA to be labelled was separated from its vector by appropriate restriction endonuclease digestion and fractionation by electrophoresis on a 0.6% low gelling temperature agarose gel. The fragment was excised from the gel and the DNA concentration estimated as described in section 2.5.5. The tube was placed in a boiling water bath for 7 minutes to melt the gel and denature the DNA. If the fragment was

to be labelled immediately it was placed in a 37°C water bath for 5 minutes. DNA for use as an oligo-labelled probe was stored at -20°C, and before each reaction was reboiled for 3 minutes and kept at 37°C for 5 minutes.

The reactions were set up using 3µl OLB, 0.6µl BSA, 5-15ng probe DNA, 1µl $\alpha[^{32}P]$dCTP (3000Ci/mmol), 0.6µl Klenow polymerase (1 unit/µl), and distilled water to make the volume upto 15µl.

The labelling reaction was allowed to proceed at room temperature for at least 5 hours, or at 37°C for 1 hour. The reaction was terminated by the addition of 85µl stop solution (20mM NaCl, 20mM Tris-HCl pH 7.5, 2mM EDTA, 0.25% SDS).

OLB is made from the following components and stored at -20°C:

2 volumes solution A; 625µl 2M Tris-HCl pH 8.0, 25µl 5M $MgCl_2$, 18µl β-mercaptoethanol, 350µl H_2O and 5µl each of 0.1M dATP, dGTP and dTTP (each dissolved in 3mM Tris-HCl, 0.2MEDTA, pH 7.0).

5 volumes solution B; 2M HEPES, titrated to pH 6.6.

3 volumes solution C; hexadeoxynucleotides suspended to 90 O.D. units/ml in 3mM Tris-HCl, 0.2M EDTA, pH 7.0.

2.10.2 Determination of probe radioactivity

Prior to hybridisation the extent of radioactive nucleotide incorporation into the probe was estimated.

1µl of the terminated labelling reaction was added to 10µl H_2O and mixed. Two 5µl samples were spotted onto $1cm^2$ pieces of Whatman DE-81 paper. One sample was washed for 6 x 5 minutes in 0.5M disodium hydrogen phosphate, then rinsed for 2 x 1 minute in distilled water and then rinsed for 2 x 1 minute in ethanol. This process washes away any unincorporated nucleotides. After drying the papers, the two were compared by measuring Cerenkov radiation in a Packard Tricarb

scintillation counter. If above 40% of the radioactive nucleotides in the labelling reaction had become incorporated into the DNA, the probe was considered to be of sufficiently high specific activity for use.

2.10.3 End-labelling of DNA fragments

500ng DNA to be end labelled was digested in a 20µl volume with a restriction endonuclease to yield overhanging 5' ends, of which the 3'-most un-base-paired nucleotide was a guanine. Once the digest was thought to be complete 2µCi of [^{32}P]dCTP (3000Ci/mmol) and 2 units of AMV reverse transcriptase were added to the DNA, mixed and incubated at 37°C for 1 hour.

End-labelled DNA fragments for use as electrophoresis size markers were stored at -20°C for upto 2 months.

2.11 TRANSFER OF DNA AND RNA TO HYBOND NYLON FILTERS

2.11.1 Southern blotting

The transfer of DNA from agarose gels to nylon filters (Amersham Hybond-N) is based on the method originally described by Southern (1975).

After being photographed, the gel was soaked in 0.25M HCl with gentle agitation for 7 minutes, to reduce the size of the DNA fragments in situ by partial depurination. After a briefly rinsing the gel in distilled water, the DNA was denatured by washing the gel in 0.5M NaOH, 1.5M NaCl for 30 minutes. The gel was rinsed again in distilled water and washed in 0.5M Tris-HCl, 3M NaCl, pH 7.4 to neutralise.

Gels containing digested genomic DNA were transferred to a blotting apparatus as described in detail by Dalgleish (1987b). Gels containing cloned DNA were simply placed on a glass plate. A sheet of nylon membrane presoaked in 3x SSC (0.45M NaCl, 0.045M Na citrate, pH 7.0) was placed on the gel and overlayed with a sheet on Whatman 3MM

paper (also presoaked in 3x SSC). A 5cm stack of paper towels was placed on the Whatman 3MM paper, and on top of this was placed a glass plate and a 500g weight. The blot was allowed to proceed for at least 2 hours.

After the blot was dismantled, the filter was rinsed in 3x SSC and dried in a 65°C oven. The filter was then wrapped in Saran-Wrap and DNA was covalently bound to the filter by exposure of the wrapped filter to U.V. light on a transilluminator for 30 seconds. The filter was finally cut into lengthwise strips, to fit the hybridisation box.

2.11.2 Northern blots

For Northern blotting, agarose gels were used at a concentration of 1% agarose. For 100ml of gel-mix; 1g agarose was dissolved in 10ml 10x MOPS buffer (0.2M MOPS, 0.01M EDTA, 0.05M sodium acetate, pH 7.0) and 74.1ml distilled water, by heating and stirring. As soon as the dissolved agarose had cooled to 65°C, 17.9ml formaldehyde (38% w/w) was added, gently mixed and poured immediately. To 3-4µg total RNA (dissolved in a volume of 4.8µl DEPC-treated water) was added 10µl formamide (deionised), 2µl 10 x MOPS buffer and 3.2µl formaldehyde. This solution was heated at 65°C for 5 minutes and then cooled on ice. Then 5µl loading dye (50% w/w glycerol, 0.1mg/ml bromophenol blue) and 1µl ethidium bromide (1mg/ml) were added. The sample was then mixed and loaded onto the gel.

Gels were electrophoresed in 1 x MOPS buffer initially at 65 volts, the buffer being circulated by a peristaltic pump. Excess ethidium bromide not bound to the RNA, which migrates to the top of the gel, was excised from the gel to prevent it entering the buffer. The voltage was then reduced to 25v and electrophoresis allowed to continue overnight.

After electrophoresis, gels were photographed and blotted directly onto Hybond-N as for Southern blots. After transfer, 2-4 hours later, the filter was not rinsed, but simply allowed to dry at room temperature. The

RNA was U.V. cross-linked to the filter as for Southern blots.

2.12 FILTER HYBRIDISATION

2.12.1 DNA filters

Two methods were employed, the Marvel method being that of choice in latter times, by way of its simplicity.

i) Denhardt's mix

Filter strips containing genomic DNA's were washed prior to hybridisation, in 3x SSC, 2x Denhardt's solution (0.04% Ficoll, 0.04% polyvinyl pyrrolidine , 0.04% BSA), 200µg/ml sonicated herring sperm DNA (denatured by boiling for 10 minutes prior to addition to the mixture), 6% PEG 6000, 0.1% SDS for 1 hour at 65°C.

Hybridisations were carried out in the same conditions as the pre-hybridisation, except that 5x Denhardt's solution was used. The radio-labelled probe DNA (0.5ng/ml) was boiled to denature for 10 minutes before being added to the hybridisation solution. Hybridisation was then allowed to proceed overnight.

After hydridisation, the filters were rinsed four times and washed twice, for 10 minutes each in 3 x SSC, 0.1% SDS at 65°C. This was followed by four 15 minute washes in 0.5 x SSC, 0.1% SDS at 65°C. Filters containing cloned DNA were washed to a higher stringency, usually 0.1 x SSC, 0.1% SDS at 65°C.

ii) Marvel mix

This method is based upon that of Reed and Mann (1985).

Filters were prewashed in 1.5 x SSPE (1.5mM EDTA, 0.27M NaCl, 15mM NaH_2PO_4, pH 7.7), 1% SDS, 6% PEG 6000, 50ng/ml dried milk powder (Cadbury's Marvel) at 65°C for 1 hour. The dried milk should be thouroughly dissolved in the water prior to the addition of the other components of the prewash solution. This and the pH of the SSPE are

critical to ensure low backgrounds.

Hybridisations were carried out in the same conditions overnight, with a probe concentration of 0.5mg/ml. Post hybridisation washing was as above.

2.12.2 RNA filters

Because of the high level of ribonucleases in dried milk, RNA filters were hybridised using the Denhardt's mix in the same way as DNA filters.

2.12.3 Autoradiography

Flters were blotted to remove to remove excess liquid but were not allowed to dry completely. Filters were wrapped in Saran-Wrap and autoradiographed for an appropriate time, either at room temperature without intensifying screens or at -80°C with intensifying screens.

2.12.4 Probe stripping

Filters were kept moist for autoradiography so that the probe could readily be stripped off and the filters hybridised with another probe. To strip the probe from nylon filters, they were soaked in 0.4M NaOH at 45°C for 10 minutes, followed by washing in 0.1 x SSC, 0.1% SDS, 0.2M Tris-HCl, pH 7.5 at 45°C for 30 minutes.

2.13 PLASMID LIBRARY SCREENING

The library was plated onto nylon filters (Amersham Hybond-N) on 9cm Luria agar plates containing tetracycline at 12.5µg/ml. The plates were incubated at 37°C until the colonies were no more than 1mm in diameter. The colonies were screened by a modification of the method of Hanahan and Meselson (1983).

Two sets of replica filters were made for each plate using filter-to-filter contact. The replica and master filters were marked around the edge with asymmetrically positioned needle holes to allow later alignment of the autoradiographs to the master filters. After replicas had been made the filters were incubated at 37°C until the colonies were approximately 1mm in diameter. Master filters were sealed with parafilm and stored at 4°C. The replica filters were transferred to Luria agar plates supplemented with chloramphenicol at 250 µg/ml, and incubated overnight at 37°C to allow amplification of the plasmids.

The filters were transferred onto Whatman 3MM paper soaked in 0.5M NaOH for 7 minutes to lyse the bacteria and denature the DNA. The filters were then transferred to Whatmann 3MM soaked in neutralising solution; firstly 1 minute in 1M Tris-HCl pH 7.6 and then 5 minutes in 0.5M Tris-HCl pH 7.6, 1.5M NaCl.

The filters were then blotted dry and DNA cross-linked by baking at 80°C for 2 hours. Filters were washed in large volumes of 3x SSC, 0.1% SDS at 65°C for at least 5 hours, with frequent changes of solution to remove cell debris. Pre-hybridisation and hybridisation were carried out using the Marvel mix method (see section 2.12.1).

Areas of the master plates corresponding to a 'positive' colony were scraped off with a sterile loop and resuspended in 1ml Luria broth. The bacteria were diluted by adding 10µl of this stock to 1ml Luria broth. Aliquots of the diluted bacteria were plated out onto nylon filters on Luria agar plates supplemented with tetracycline. Filters with well isolted colonies were replica plated and screened as before. Isolated positive colonies were then picked and grown-up for 'mini-prep' DNA extractions. The plasmid DNA was mapped by the appropriate restriction endonuclease digestions and correct colony-picking confirmed by blotting of the gel followed by hybridisation.

2.14 CELL CULTURE TECHNIQUES

2.14.1 Growth conditions

All fibroblast cell-lines were grown as monolayers of cells at 37°C and in 5% CO_2 in Dulbecco's modification of Eagle's medium (DMEM) supplemented with 10-15% foetal calf serum and 100U/ml penicillin and 0.1mg/ml streptomycin.

All culture manipulations were carried out in a Class II Microbiological safety cabinet (Medical Air Technology Ltd). Glassware to be used for use with cells was rinsed extensively in distilled water before drying and sterilisation at 200°C for 5 hours. All solutions to be used on cells were made up in double distilled water and filter sterilised using 0.2µm Acrodisc filter units (Gelman Sciences Inc.)

2.14.2 Maintenance of proliferating fibroblasts

Subconfluent cells in flasks were rinsed with phosphate buffered saline (PBS) (4ml per 80cm^2 flask) and the rinse discarded. Next 2ml per 80cm^2 flask of trypsin/EDTA were added carefully to the cell monolayer and the flask tilted from side to side to completely cover the cells. The trypsin solution was then removed, taking with it any residual serum, and replaced with a fresh aliquot of trypsin/EDTA. The flask was returned to the incubator for 2-3 minutes until the cells had 'rounded-up' as judged by viewing on a phase contrast microscope (Olympus CK2). The rounded cells were dispersed by repeated pipetting over the cell monolayer. It was always ensured that a single cell suspension was achieved for uniform growth after reseeding.

Pre-warmed medium was added to the 'trypsinised' cells (10ml to a 80cm^2 flask) and gently mixed. the medium was then evenly distributed to between 2 and 4 fresh pre-warmed medium-containing flasks, depending on

the growth characteristics of the cell-line.

For slow growing or low density cells, the medium was removed from the cell monolayer every four days and replaced with a the same amount of fresh pre-warmed medium.

2.14.3 Freezing and thawing of cells

All fibroblasts were frozen in liquid nitrogen in a solution containing 10% DMSO and 20% foetal calf serum. Cells were grown in 25cm^2 flasks until the monolayer covered approximately 70% of the bottom surface of the flask. The cells were trypsinised from the culture flask and added to 9ml of medium in a sterile centrifuge tube. The cells were pelleted by centrifugation at 800 rpm for 10 minutes in an MSE Centaur centrifuge. The medium was then removed from the tube leaving just approximately 100µl medium behind. The cells were resuspended in this small volume of medium by gently tapping the tube. 1ml freezing medium (10ml freezing medium is 7ml DMEM, 2ml foetal calf serum, 1ml DMSO) was added and mixed by gentle pipetting. The cells were transferred to a 1.5ml cryotube and placed in an insulated box. The box was placed at -20°C for 1-2 hours and then transferred to -80°C overnight, before being placed in liquid nitrogen.

Cells were thawed rapidly from liquid nitrogen storage by placing cryotubes directly into a 37°C water bath. When thawed, cells were transferred to an empty 25cm^2 flask. 9ml pre-warmed medium was added; the first ml dropwise, then 0.5ml at a time, gently rocking the flask to mix. This minimises the osmotic shock suffered by the cells. After approximately 6 hours of incubation following thawing, the cells were checked under a phase contrast microscope to have attached to the bottom surface of the flask. If the cells had attached, the medium was replaced with fresh and the cells returned to the incubator.

2.15 PREPARATION OF GENOMIC DNA

2.15.1 Preparation of genomic DNA from blood

This method was used to extract DNA from 10ml fresh or frozen blood that was collected into EDTA tubes.

The tubes were set in a rack at 45° for 2 hours. During this time the red cells settled and the plasma was transferred to a 20ml 'Universal' tube. The cells were pelleted by centrifuging at 1500 rpm for 10 minutes in an MSE Centaur centrifuge.

The supernate was discarded and residual red cells were lysed hypotonically by quickly resuspending the pellet in 5ml distilled water and exactly 30 seconds later adding 5ml 2 x SSC and mixing.

The white blood cells were then centrifuged as before and resuspended in 4ml lysis buffer (0.3M sodium acetate, 20mM Tris-HCl, 1mM EDTA, pH 7.5) and transferred to a 15ml polypropylene tube. The cells were lysed by adding 200µl 10% SDS and thouroughly mixing. The lysate was vigourously phenol extracted, with shaking for 5 minutes and then centrifugation for 10 minutes at 3500 rpm. The aqueous layer was transferred with a wide bore pipette to a fresh tube and the phenol extraction repeated.

The aqueous layer was once more recovered and extracted with an equal volume of chloroform/isoamyl alcohol (24:1 v/v) and centrifuged as before.

The aqueous layer was transferred to a fresh tube and the DNA precipitated by the addition of 2 volumes of ethanol and centrifugation for 10 minutes at 3500 rpm. The pellet was rinsed in 70% ethanol, drained and dissolved in an appropriate volume of TE and stored at 4°C.

DNA recovery was measured by U.V. light absorbance at 260 and 280nm or by comparison with known quantities of phage DNA on agarose

gels.

2.15.2 Preparation of genomic DNA from Placentae

25g of placental material was added to 25ml of 0.15M NaCl, 0.1M EDTA, pH 7.6 and liquidised in a Sorvall Omnimixer. To this mixture 25ml lysis mix (6% tri-isopropylnapthalene sulphonic acid, 8% butan-2-ol, 3% SDS) was added and mixed gently but thouroughly for 5 minutes. Then 5ml 5M sodium perchlorate was added and mixed.

Next 80ml phenol/chloroform/isoamyl alcohol (25:24:1, v/v) was added and mixed thouroughly. The mixture was then centrifuged at 3500 rpm for 10 minutes in an MSE Centaur 2 centrifuge.

The supernatant was removed and added to an equal volume of chloroform/isoamyl alcohol (24:1, v/v). This was mixed thouroughly and centrifuged as before.

The aqueous layer was removed and added to 2 volumes ethanol to precipitate the DNA. The DNA was spooled onto a glass rod and excess ethanol was squeezed out. The DNA was dissolved overnight in 100ml TE by gentle shaking at 4°C. When the DNA had dissolved, a $^{1}/_{10}{}^{th}$ volume of 20x SSC and 0.5ml boiled RNase A (10mg/ml) were added; the mixture was incubated for 1 hour at 37°C with gentle shaking.

Next 1ml 10% SDS and 50µl Proteinase K (20mg/ml) were added and the mixture incubated at 50°C for 3 hours.

The phenol and chloroform extractions were repeated using 60ml of each solvent. The DNA was then precipitated as before, rinsed in 70% ethanol and dissolved in 50ml TE at 4°C with gentle shaking.

DNA concentration was determined by UV light absorbance at 260 and 280nm.

2.15.3 Preparation of genomic DNA from cells in culture

The following method was for the extraction of genomic DNA from 2 confuent 80cm^2 flasks of cultured fibroblasts.

The cells were trypsinised and pelleted as for freezing (see section 2.14.3). The cell pellet was rinsed by resuspension in 1.4ml 1x SSC. The cell suspension was then transferred to a microfuge tube and centrifuged at low speed in a MSE Micro Centaur centrifuge for 2 minutes. The pellet was then resuspended in 200µl cell suspension buffer (0.3M Na acetate, 20mM Tris-HCl pH 7.5, 1mM EDTA). Cells were lysed by the addition of 15µl 10% SDS, which was mixed in by gentle inversions of the tube for 1 minute. 100µl TE was added and the lysate phenol extracted and traces of phenol removed by extracting with an equal volume of chloroform. The nucleic acids were then ethanol precipitated and centrifuged for 1 minute. The pellet was washed with 500µl 70% ethanol and recentrifuged. The supernatant was discarded and the pellet resuspended in 400µl TE.

Next 40µl 20x SSC and 2µl boiled RNase A (10mg/ml) were added, mixed and the tube incubated at 37°C for 1 hour. Then 4µl 10% SDS and 4µl proteinase K (1mg/ml) were added and mixed and the tube incubated at 50°C for 1 hour. The DNA was then phenol extracted, chloroform extracted and ethanol precipitated. The pellet was washed in 500µl 70% ethanol, re-centrifuged and the pellet vacuum-dried. Finally the pellet was resuspended in 80µl TE.

DNA concentration was determined by UV light absorbance at 260nm and 280nm.

2.16 PREPARATION OF TOTAL CELLULAR RNA

2.16.1 Preparation of RNA from Placentae

This method is based upon that of Chomczynski and Sacchi, (1987).

Placentae for RNA preparations were obtained as freshly as possible, and always from elective Caesarian section delivery.

Placentae were put on ice and transferred to the laboratory as quickly as possible. The RNA preparation was either begun immediately or the placenta was cut into pieces and snap-frozen in liquid nitrogen for later attention.

Approximately 12g of fresh or frozen coarsely chopped-up placenta was put in a beaker contianing 80ml denaturing solution (4M guanidinium thiocyanate, 25mM sodium citrate pH 7.0, 0.5% N-lauroylsarcosine, 0.1M β-mercaptoethanol).

The tissue was liquidised with a 'Polytron' blender (Kinematica GmbH, Switzerland) and allowed to settle for 5 minutes. The homogenate was then filtered through polyallomer wool. To each 10ml of filtrate, the following was added sequentially, with mixing between additions:

1ml 2M sodium acetate, pH 4.0

10ml water saturated phenol

2ml chloroform/isoamyl alcohol (49:1 v/v)

This mixture was shaken vigourously for 10 seconds and put on ice for 15 minutes. The mixture was centrifuged in 30ml Corex tubes using a Sorvall HB4 rotor and RC5B centrifuge at 7000 rpm for 30 minutes at 4°C.

The aqueous layer was transferred to fresh Corex tubes, mixed with an equal quantity of propan-2-ol and chilled at -20°C for 2 hours. The mixture was then centrifuged at 10,000 rpm for 20 minutes at 4°C. The tubes were drained and each pellet resuspended in 0.5ml denaturing solution. Once dissolved the RNA was transferred to microfuge tubes and re-precipitated by adding 0.5ml propan-2-ol, mixing and leaving at -20°C for 1 hour.

The tubes were then centrifuged in an MSE Micro Centaur centrifuge for 20 minutes. The pellets were washed in 800µl 70% ethanol, the tubes

drained and each resuspended in 500µl double-distilled water.

RNA concentration was determined by UV light absorbance at 260 and 280nm.

2.16.2 Preparation of RNA from cells in culture

The method for extracting RNA from cultured cells was essentially the same as that above.

Medium was removed from the culture flask, the cells rinsed in PBS and the rinse discarded. The cells were lysed directly by the addition of denaturing solution (4ml per 175cm^2 flask) and transferrred to a Corex tube. The sample was briefly homogenised (as above) to ensure cell lysis was complete.

The method continues as for the placental preparation, with the same reagent proportions. The RNA derived from each 175cm^2 flask was dissolved finally in 50µl double-distilled water.

RNA concentration was estimated by comparing aliquots of the sample with known amounts of placental RNA on agarose gels.

2.17 RNase A MAPPING

This method is adapted from that of Gibbs and Caskey, (1987).

2.17.1 Preparation of probe templates

cDNA fragments were subcloned into either of the transcription vectors: pTZ18R or pTZ19R, in such an orientation as to enable the *in vitro* synthesis of 'antisense' RNA.

DNA was obtained from the recombinant plasmids by the 'large scale plasmid prep' method (see section 2.8.3). Typically 50µg of DNA was linearised with a restriction endonuclease to give 5' overhanging ends. The DNA was phenol extracted three times and traces of phenol removed by extracting with an equal volume of chloroform/isoamyl alcohol (24:1 v/v).

All handling from this point was as for RNA. The DNA was precipitated with a 0.25 volume of 2M sodium acetate pH 5.6, and 2.5 volumes of ethanol and centrifuged in an MSE MicroCentaur centrifuge.

The DNA was washed in 70% ethanol and vaccuum dried. The DNA was then resuspended in TE to an approximate concentration of 0.4µg/µl and stored at 4°C.

2.17.2 RNA probe labelling

The following reaction mix was prepared:

	µl	
10 x NTP's	1	
100mM DTT	1	
5 x buffer	2	
2mg/ml BSA	0.5	
RNasin	0.8	(35U)
T7 RNA pol.	0.7	(10U)
^{32}P-CTP (800Ci/mmol)	...	(40µCi)
Template	0.6	(0.4µg)
Distilled water	...	
	10µl	

This methodology was arrived at after many changes were made to the protocol of Gibbs and Caskey (1987).

10 x NTP's is 5mM GTP, 5mM ATP, 5mM UTP, 75µM CTP.

5 x buffer is 200mM tris-HCl pH 8.25, 30mM $MgCl_2$, 10mM spermidine.

All solutions were made using DEPC-treated double distilled water.

The reaction mix was incubated at 37°C for 2 hours. Then 0.8µl DNase I (6u) was added, and the reaction mix returned to 37°C for a further 15 minutes. The tube was then returned to room temperature,

diluted with 40µl distilled water and phenol extracted. The aqueous layer was transferred to a fresh tube and 50µl 7.5M ammonium acetate and 500µl ethanol was added and mixed. The tube was left at -80°C for 15 minutes before centrifugation in an Eppendorf centrifuge for 15 minutes. The ethanol was then thoroughly drained from the tube and the pellet resupended in 80µl distilled water.

2.17.3 RNA:RNA hybridisation

50µg of total cellular RNA was ethanol precipitated in the presence of the probe, adjusting to 0.1M NaCl with 2M NaCl, adding 6µl probe (approximately 2 x 10^6 cpm) and 2.5 volumes of ethanol. The tube was left on ice for 5 minutes and then centrifuged in an Eppendorf centrifuge for 15 minutes. The tube was drained and resuspended in 160µl deionised formamide. The tube was left at room temperature for 30 minutes to allow the pellet to dissolve completely. Next 40µl 5 x hybridisation buffer was added and mixed (5 x hybridisation buffer is 2M NaCl, 0.2M PIPES pH 6.7). The tube was heated at 85°C for 5 minutes to denature the RNA, before transferring to 65°C for overnight hybridisation.

2.17.4 Polyadenylated RNA purification

With a new scalpel blade and a sterile needle, a piece of Amersham mAP (messenger affinity paper) approximately 4mm x 4mm was cut and soaked in 0.5M NaCl for 5 minutes. The hybridisation tube was returned to room temperature and 100µl 0.5M NaCl added and mixed. The piece of mAP was placed on a 5cm x 5cm square of autoclaved 3MM paper and the RNA spotted onto the mAP a drop at a time. Each drop was allowed to soak through, before the next was applied. When all the RNA had been appplied, the mAP was washed in 0.5M NaCl 3 times for 15 minutes each. The mAP was then removed with a sterile needle and blotted dry by touching the edge of a

tissue and placed in a fresh microfuge tube. 180µl distilled water was added and the tube heated at 65°C for 10-12 minutes. The mAP was then removed from the tube and discarded.

2.17.5 RNase A cleavage

To the mAP eluate was added 20 µl 2.5M NaCl and 100µl 100µg/ml RNase A (100µg RNase A per 1ml of 200mM NaCl, 100mM LiCl, 30mM tris-HCl, 3mM EDTA, pH 7.5; pre-boiled for 10 minutes). The solution was mixed and incubated at room temperature for 30 minutes. Then 20µl 10% SDS and 10µl Proteinase K (10mg/ml) was added, mixed and the tube incubated at 37°C for 20 minutes. Next 1µl tRNA (10mg/ml) was added and the solution phenol extracted. The RNA in the aqueous phase was precipitated by the addition of 750µl ethanol. The tube was chilled on ice for 5 minutes, centrifuged for 15 minutes and drained. The invisible pellet was washed with 200µl 100% ethanol and centrifuged for 5 minutes. The tube was drained again and the pellet vacuum dried.

2.17.6 RNase a mapping gels

Gels for RNase A mapping analysis were 5% acrylamide and made as follows:

Urea	16.8g
10x TBE	2ml
Formamide	8ml
40% Acrylamide stock	5ml
H_2O to a final volume of	40ml

The gel mix was heated to 65°C and stirred to dissolve the urea. The mix was then allowed to cool to room temperature. 280µl 10% ammonium persulphate and 70µl TEMED were added to the gel-mix. The mix was briefly

swirled to mix and then drawn up into a 50ml syringe. The mix was poured with the syringe into a 1mm thick 16cm x 20cm Bio-Rad Protean II gel mould, with 6mm wide, 8mm deep well-formers. The gel was allowed to stand for at least 30 minutes to polymerise. The top buffer chamber of the tank was filled with 1x TBE and the bottom chamber with 0.625x TBE.

Vaccuum dried 'protected' RNA samples were resuspended in 12-25µl tracking dye (stock solution = 15ml deionised formamide, 5mg xylene cyanol, 5mg bromophenol blue, 0.1ml 0.5M EDTA, pH8.0, 0.1% SDS). The samples were heated at 85-90°C for 7 minutes and than rapidly loaded onto the gel (pre-run to warm) along with DNA size markers.

The gel was run at 25mA, the gel being maintained at a warm temperature by a thermostatic water circulator (LKB Multitemp II). Gels were run until the xylene cyanol dye had reached the end of the gel. The gel was 'fixed' to remove urea in a 10% (v/v) methanol, 10% (v/v) acetic acid solution for 20 minutes. The gel was transferred onto Whatman 3MM paper, covered with a sheet of Saran-Wrap and dried on a Bio-Rad gel drier. The gels were autoradiographed for between 12 hours and several days, with or without intensifying screens, depending on the radioactivity of the gel and the quality of the image required.

2.18 POLYMERASE CHAIN REACTION (PCR)

2.18.1 Note on handling PCR solutions

All solutions for use in PCR were made up as cleanly as possible, in an effort to avoid contamination by exogenous DNA. All manipulations were carried out using plasticware directly from the manufacturer's packaging.

2.18.2 Amplifications of genomic DNA

1µg aliquots of human DNA were amplified in 50µl of 45mM tris-HCl

(pH8.8), 11mM $(NH_4)_2SO_4$, 4.5mM $MgCl_2$, 6.6mM β-mercaptoethanol, 4.5μM EDTA, 1mM dATP, 1mM dTTP, 1mM dCTP, 1mMdGTP, 113μg/ml BSA plus 0.6μM of each oligonucleotide primer and 1.4 units of Taq DNA polymerase III.

Reactions performed by an 'intelligent heating block' (Cambio, Cambridge UK) were conducted in 1.5ml microcentrifuge tubes overlaid with 100μl paraffin oil and cycled for 1.3 minutes at 95°C, 1 minute at 65°C and 5 minutes at 70°C for 30 rounds. Reactions performed by a 'programmable Dri-block' (Techne, Cambridge UK) were conducted in 0.5ml microcentrifuge tubes overlaid with 60μl paraffin oil and cycled for 1.5 minutes at 95°C, 1.5 minutes at 60°C and 5.5 minutes at 70°C for 30 rounds. Final amplification cycles were generally chased by a final step of 1 minute at 60°C, to anneal any remaining single stranded DNA with primer followed by an extension phase at 70°C for 5 minutes.

2.18.3 Re-amplifications

Fragments of DNA to be reamplified were cut out of agarose gels and the gel-slice transferred to a microcentrifuge tube. 100μl distilled water was added and the tube freeze-thawed twice, causing some of the DNA in the gel-slice to elute into the water. 13μl of the DNA-containing water was re-amplified for 20 rounds as above.

2.18.4 Analysis of PCR products

The reaction mixes were removed from the paraffin by pipetting through the overlay. The reactions were checked by electrophoresing an aliquot on an agarose gel. Samples for further analysis were phenol extracted, ethanol precipitated and dissolved in 10-20μl distilled water. Restriction endonuclease digestions were carried out on aliquots of the DNA under manufacturer's recommended conditions in the presence of 4mM spermidine-HCl.

2.19 PLASMID DNA SEQUENCING

2.19.1 Plasmid DNA preparation

The method used was adapted from that of Kraft et al., (1988) and is an extension of the 'mini-prep' technique (section 2.8.1) including a PEG precipitation, giving very clean DNA.

A 5ml culture of the recombinant plasmid-containing E.coli JM83recA was grown in Luria broth with shaking overnight, with the appropriate antibiotic. 3ml of the culture was harvested by transferring 1.5ml of culture to a microfuge tube and centrifuging in an MSE Micro Centaur centrifuge for 1 minute. The broth was then decanted and the procedure repeated with a further 1.5ml of culture. The broth was removed from the pellet as completely as possible.

Resuspension of the cells, DNA denaturation and potassium acetate precipitation was conducted exactly as for ordinary 'mini-preps'. Following the potassium acetate preciptation and centrifugation, 400µl of supernate was transferred to a fresh microfuge tube. RNase A was added to a final concentration of 50µg/ml and incubated at 37°C for 30 minutes.

Next the solution was phenol extracted and centrifuged for 2 minutes. The aqueous phase was transferred to a fresh microfuge tube, a 2.5 volume (1ml) of ice-cold ethanol was added, mixed and the tube incubated at -70°C for 20 minutes. The sample was then centrifuged at 4°C for 5 minutes. The supernate was removed, the pellet rinsed with 1ml ice-cold 70% ethanol and recentrifuged for 1 minute. The supernatant was removed and the pellet dried under vacuum.

The pellet was then dissolved in 16.8µl double distilled water. Then 3.2µl 5M NaCl was added and mixed. Next 20µl 13% PEG was added, mixed and the tube incubated on ice in a 4°C room for 45 minutes.

The sample was centrifuged at 4°C for 10 minutes and the supernates discarded. The pellet was rinsed with 1ml ice-cold 70%

ethanol, centrifuged for 1 minute, the supernate removed and the pellet dried under vacuum. The pellet (often invisible) was dissolved in 8µl double distilled water and stored at -20°C.

2.19.2 Sequencing of plasmid recombinants

This method is essentially as described in the Pharmacia 'T7 sequencing' kit instructions.

To the thawed DNA samples; 2µl 2M NaOH was added, vortexed and incubated at room temperature for 10 minutes. Then 3µl 3m Na acetate (pH4.5), 7µl double distilled water and 60µl ethanol were added, mixed and chilled in dry ice for 15 minutes. The tube was centrifuged for 10 minutes at 4°C in an MSE Micro Centaur centrifuge. The pellet was washed in 50µl 70% ethanol and recentrifuged for 5 minutes. The supernatant was removed and the pellet dried under vacuum. The pellet was resuspended in 10µl double distilled water. 2µl annealing buffer (280mM tris-HCl (pH 7.5), 100mM $MgCl_2$, 350mM NaCl) and 2µl primer (5µg/ml) were added and mixed.

The annealing mix was incubated at 37°C for 20 minutes and then at room temperature for 15 minutes.

For each clone four termination reaction tubes were prepared containing 2.5µl of a solution containing all four dNTP's at 150µM, 10mM $MgCl_2$, 40mM Tris-HCl (pH7.5), 50mM NaCl, plus one of each ddNTP at 15µM.

To the annealed DNA was added 2µl labelling mix (dGTP, dCTP, and dTTP, each at 2µM), 1µl [α-^{35}S]dATP (14.8 TBq/mmole, 0.37MBq/µl), 1µl 0.3M DTT and 2µl T7 DNA polymerase (diluted with the manufacturer's dilution buffer to 1.5u/µl). The mixture was mixed by gentle pipetting, quickly centrifuged and incubated at room temperature for 5 minutes.

During this time the termination mixes were warmed to 37°C.

4.5µl of the labelling reaction was transferred to each of the

four termination tubes, mixed by gentle pipetting and incubated at 37°C for 5 minutes. The reaction was stopped by the addition of 5 µl tracking dye (10ml deionised formamide, 10mg xylene cyanol, 10mg bromophenol blue, 0.2ml 0.5M EDTA pH 8.0) This mixture was gently mixed by tapping the tube and briefly centrifuging.

A 3µl aliquot of each solution was transferred to a Titretek multi-well plate (Flow Laboratories, Maryland, USA) and the remainder of the reactions stored at -20°C. The multi-well plate containing the samples to be loaded onto the sequencing gel, was floated without its lid on an 80°C water bath for 2 minutes. The samples were then immediately loaded onto the gel.

2.19.3 Buffer gradient sequencing gels

Preparation and running of 40cm 6% polyacrlamide buffer gradient gels were essentially as described by Biggin et al.,(1983).

The standard mixes used to prepare one gel are given below:

	0.5x	2.5x
40% Acrylamide stock	15ml	3ml
Urea	50g	10g
Sucrose	-	1g
10x TBE	5ml	5ml
H_2O to a final volume	100ml	20ml

A smidgen of bromophenol blue was added to the 2.5x mix, so that the quality of the gradient could be visualised. The solutions were suction filtered through Whatman No.1 filters using a Buchner funnel. Immediately prior to pouring, the following solutions were added to the mixes:

	0.5x	2.5x
10% ammonium persulphate	470µl	175µl
TEMED	32µl	12µl

8ml 0.5x mix and 12ml 2.5x mix were drawn up into a 25ml pipette and poured into the gel mould, which was positioned at an angle of approximately 30° to the horizontal. Then 60ml of 0.5x mix was poured into the mould with a syringe. The gels polymerised within 1 hour and were mounted into the gel apparatus. The top buffer chamber was filled with 0.5x TBE and the bottom chamber with 2.5x TBE.

Gels were pre-run at 2000 volts for 10-20 minutes to warm the gels. Samples were loaded immediately after denaturation. The gel was run at 1400 to 2000 volts for 3-5 hours, the voltage being adjusted to keep the gel hand-hot; approximately 55°C. The length of time the gels were run depended on how far the sequence of interest was from the primer. This was judged by the position of the xylene cyanol dye which travels with a mobility of approximately 120 bases on a 6% gel.

After electrophoresis, the gels were fixed in 10% (w/v) methanol, 10% (v/v) acetic acid for 20 minutes, transferred to Whatman 3MM paper and dried on a Bio-Rad gel drier for 80 minutes. Gels were autoradiographed at room temperature for 12-36 hours, as necessary.

2.20 COMPUTING

Various computing facilities were used in the course of this work. Analyses of DNA sequences and RNA secondary structures were performed using the University of Wisconsin Genetics Computer Group package (Devereux *et al.*, 1984) on the DEC VAX computer at Leicester University.

Data base searches of homologies to oligo-nucleotides was carried out using the Lipman-Pearson package on the DEC VAX computer.

Word-processing was carried out on a Fortune Systems Word Processor computer.

CHAPTER 3

LINKAGE ANALYSIS IN A LARGE MARFAN SYNDROME FAMILY

3.1 Introduction

Many normal genes exist in different allelic forms, for example the blood group loci and the HLA genes. For disease loci, the normal and abnormal gene represent different alleles for the given locus. For a disease locus, the genotype can be deduced in dominant disorders by phenotype and in recessive disorders by family studies. The study of genetic linkage is when the inheritence of two loci are considered simultaneously; in human genetics, usually a disease locus and some other scorable locus or 'marker'. The discovery of restriction fragment length polymorphisms (RFLPs) has dramatically increased the number of scorable loci that can be used as linkage markers. Prior to the advent of RFLPs, linkage studies were restricted by the few markers provided by antigenic variants, protein isoforms and HLA types.

An individual heterozygous at both the disease and the marker locus, produces gametes of four haplotypes in an expected 1:1:1:1 ratio. Of the four haplotypes, the two that each contain one paternal and one maternal allele are called 'recombinants' since a recombination of paternal and maternal alleles must have occurred. The remaining two haplotypes are 'non-recombinants' that is to say that they are the same as the individual's. Whilst, for most pairs of loci, segregation will be independent giving a 1:1 ratio of recombinant and non-recombinant gametes, for some pairs of loci a preponderance of non-recombinants will occur. This phenomenon is known as 'genetic linkage'. The extent of linkage is known as the recombination fraction or 'θ', which is the proportion of recombinant gametes. For genes segregating independently

$\theta = {}^1/_2$. Conversely, for genes completely linked (with no recombinants at all) $\theta = 0$.

The frequency of recombinants is θ and since there are two types of recombinant, the frequency of each type is ${}^{\theta}/_2$, similarly the frequency of each type of non-recombinant is ${}^{1-\theta}/_2$. The relative probability or likelihood [$L(\theta)$] of obtaining a particular pattern of children (or gametes) is the product of the expected frequencies of each genotype:

$$L(\theta) = ({}^{1-\theta}/_2)^n({}^{\theta}/_2)^r$$

where n = no. of non-recombinants

and r = no. of recombinants.

The likelihood of linkage method consists of estimating the most probable value for θ. This is achieved by calculating $L(\theta)$ for the observed segregation pattern, at values of between 0 to 0.5 and comparing the results with $L(\theta = 0.5)$. This yields a relative probability or 'odds ratio'. The data is usually expressed as the $\log_{10}$ of the odds ratio, or the 'lod score' (Z).

Thus $Z(\theta) = \log_{10} [{}^{L(\theta = 0 \text{ to } 0.5)}/_{L(\theta = 0.5)}]$

or more specifically,

$$Z(\theta) = \log_{10} [{}^{(1-\theta/2)^n(\theta/2)^r}/_{(1-0.5/2)^n(0.5/2)^r}]$$

$$= \log_{10} [{}^{(1-\theta)^n.\theta^r}/_{(0.5)^{n+r}}]$$

The calculation is repeated for various values of θ from 0 to 0.5. The odds ratio is then plotted against the various values of θ, and the highest odds ratio gives the maximum likelihood estimate of θ. Conventionally, two loci are taken to be linked only when the odds ratio is greater than 1000 or rather when the lod is ≥ 3.0.

The scenario however for a marker that is not randomly chosen is

somewhat different. In some cases a particular gene, by way of an educated guess, is suspected to be the disease gene. In such cases markers are chosen for linkage analysis that are close to or reside within such 'candidate' genes. In these cases no recombination is expected between the candidate gene and the marker and therefore $\theta = 0$. Hence, as $(1-0)^n=1$ and $0^0=1$, if no recombinants are observed $Z(0) = n.\log_{10}2$. If one or more recombinants are observed, as $0^r=0$, $Z(0) = \log_{10}0 = -\infty$. If recombinants are observed, segregation is said to be 'discordant' and θ cannot equal zero. If no recombinants are observed, segregation is said to be 'concordant' and a lod score is calculated.

In the study described here, several collagen gene loci were selected as candidates for the Marfan gene. The centre-piece of this study was a large three-generation Marfan kindred in which the grandparents were dead and unavailable for study (Figure 3.1). A clinical description of the family has been presented by Dalgleish et al., (1987a). Four of the nine second generation children had the syndrome, however one died suddenly and was unavailable for study. Two of the second generation individuals have families of their own with more than one child with the syndrome. This allows the possibility of 'water-tight' discordant segregation of any RFLP markers at a candidate locus and the Marfan phenotype (see section 3.6).

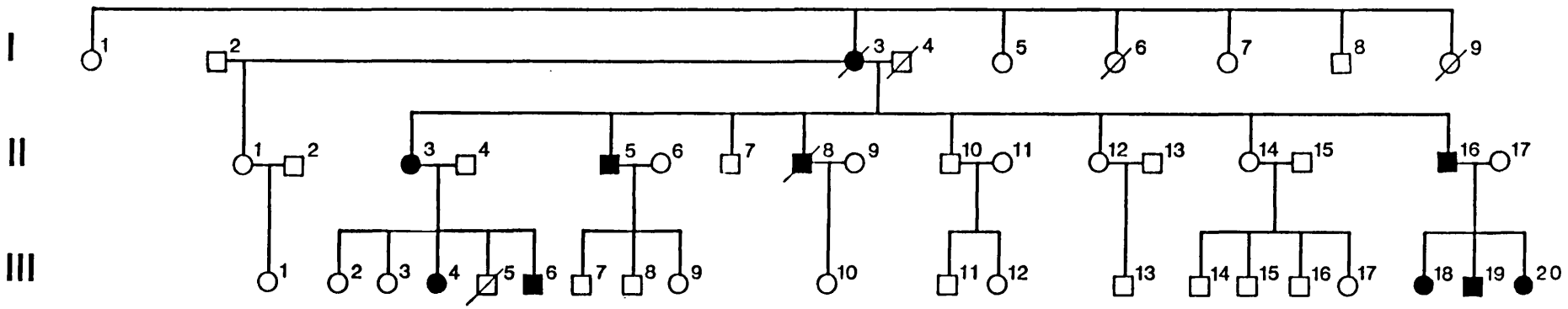
This family in the circumstance of concordant segregation has the potential to yield a lod score of greater than 4. The achievement of such a high lod score would be unlikely using dimorphic RFLPs even as haplotypes, but much more likely using the highly heterozygous multi-allele systems of mini-satellite or VNTR (variable number of tandem repeats) markers.

One potential hazard of linkage studies on the Marfan syndrome

Figure 3.1

Marfan pedigree.

The figure shows the Marfan syndrome family used for this study. Those individuals who certainly have the syndrome are marked as solid circles or squares, for females or males respectively.


I
II
III
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

is the phenomenon of variable penetrance. Any un-diagnosed Marfan individuals will have the potential to make an RFLP marker in the disease-gene appear to be segregating discordantly. In such a case, the individual in question would have to be reassessed and if the phenotype could confidently said to be normal, the recombination fraction (θ) may have to be raised from zero.

With the implication of defective collagen in the Marfan syndrome, (see section 1.4.1), and with the fibrillar collagens being the major stress-bearing components of the tissues most affected, the genes encoding the three major fibrillar collagens (types I, II and III) were selected as candidate genes and used in a linkage study in the family described above. Table 3.1 lists the various RFLPs used as linkage markers.

3.2 Linkage anaysis of COL1A2 markers

The genomic clone, NJ3' (Tsipouras *et al.*, 1983), which lies in the helical-encoding region of the gene, detects DNA fragments of 14kb and 10.5kb when digested with *Eco*RI. All the blood relatives of generation II were homozygous for the absence of the polymorphic site. (Table 3.2). Thus there was no information potential.

M1.6, a 1.6kb *Msp*I subclone of the genomic clone HpC1 (Dalgleish *et al.*, 1982) detects polymorphic DNA fragments of 1.6kb and 2.1kb with *Msp*I. But again all the potentially informative members of the family were homozygous, this time for the presence of the polymorphic site (Table 3.2).

The clone α2(I)-1.2, a 1.2Kb subclone of the genomic clone .λHpro.α2(I)-1 (Tajima *et al.*, 1984) which lies in the helical-encoding region of the gene. This probe reveals *Rsa*I DNA fragments of 2.1kb and 2.9kb in size. These two alleles were termed A and B respectively.

TABLE 3.1

RFLPs used

GENE	ENZYME	REFERENCE
COL1A1	*MspI*	Sykes *et al.* 1986
	RsaI	"
COL1A2	*EcoRI*	Tsipouras *et al.* 1983
	MspI	Grobler-Rabie *et al.* 1985a
	RsaI	Grobler-Rabie *et al.* 1985b
COL2A1	*HindIII*	Driesel *et al.* 1982
COL3A1	*AvaII*	Dalgleish *et al.* 1985
	EcoRI	Tsipouras *et al.* 1986

TABLE 3.2

RFLP Results in COL1A2

Individual	*EcoRI*	*MspI*	*RsaI*
I-5	1-1	1-1	1-2
II-1	1-1	1-1	1-1
III-1	1-2		1-1
II-3	1-1	1-1	1-1
II-4	1-2	1-2	1-2
III-2	1-1		1-2
III-3	1-2		1-1
III-4	1-1		1-2
III-6	1-2		1-1
II-5	1-1	1-1	1-2
II-6	1-2		1-1
III-7	1-1		1-1
III-8	1-1		1-1
III-9	1-2		1-1
II-7	1-1	1-1	1-1
II-9	2-2		1-1
III-10	1-2		1-1
II-10	1-1	1-1	1-1
II-11	1-1		1-1
III-11	1-1		1-1
II-12	1-1	1-1	1-1
II-13	1-1	1-1	1-2
II-14	1-1		1-2
II-16	1-1	1-1	1-2
II-17	1-2	1-1	1-1
III-18	1-1	1-1	1-2
III-19	1-1	1-1	1-1
III-20	1-2	1-1	1-1

Results are listed in Table 3.2. Affected individual II-16 is heterozygous and his normal wife is homozygous for allele A. Individual II-16 passes allele A to 2 of his 3 children, and allele B to the other (Figure 3.2). All three children have inherited Marfan syndrome. Thus it can be seen through the discordant segregation of the α2(I) gene and Marfan syndrome that COL1A2 cannot be the Marfan locus.

3.3 Linkage analysis of COL3A1 markers

AvaII and EcoRI RFLPs were detected with Idf17, a 2kb EcoRI subclone of a 12kb genomic clone also called Idf17 (Chu et al., 1985a) which spans the 3' end of the gene and the region 3' to the end of the gene. The two polymorphic restriction sites detected by this probe both lie outside of the gene. The probe reveals DNA fragments of 6.2kb and 4.5kb with AvaII, and 1.7kb and 2.0kb with EcoRI. Because of their close proximity, these two RFLPs tend to be inherited as a common haplotype.

Restriction fragment segregation of both these RFLPs was seen to be discordant with the Marfan gene (see Table 3.3 and Figure 3.3). For simplicity the AvaII fragments of 6.2kb and 4.5kb were termed alleles C and D respectively.

Exclusion of the α1(III) gene as being the mutant locus in this family is based upon the following criteria: Affected individual II-3 passes allele C to III-2 and III-6, and passes allele D to III-3 and III-4. Individuals III-4 and III-6 both have the Marfan syndrome, but have inherited different COL3A1 alleles from their affected mother. The information derived from the EcoRI RFLP excluded COL3A1 on exactly the same criteria.

Figure 3.2

COL1A2 exclusion.

Autoradiograph demonstrating the exclusion of COL1A2 as the mutant locus in this family.

The exclusion is based upon the II-16/II-17 family unit.

The bands at 2.1 and 2.9kb are termed A and B respectively.

Autoradiographic exposure was for 1 day at -80°C.

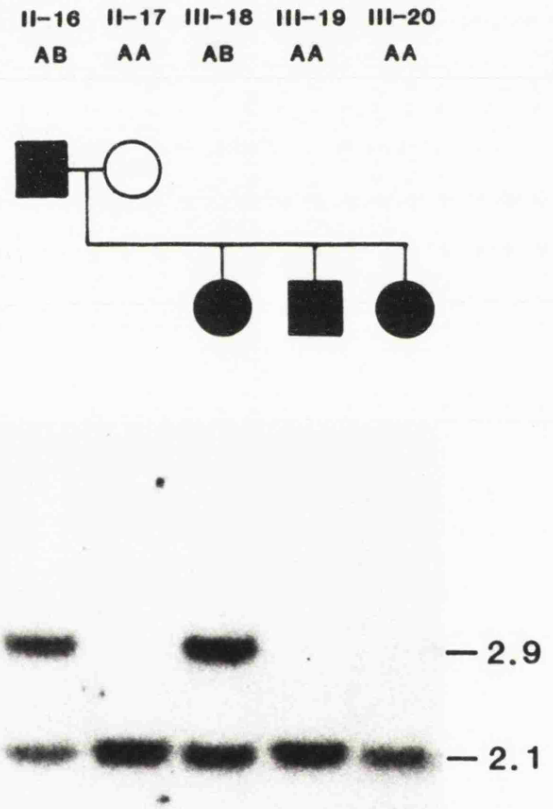

II–16
AB
II–17
AA
III–18
AB
III–19
AA
III–20
AA
— 2.9
— 2.1

TABLE 3.3

RFLP Results in COL3A1

Individual	*AvaII*	*EcoRI*
I-5	1-1	1-1
II-1	1-1	1-1
III-1	1-2	1-2
II-3	1-2	1-2
II-4	1-1	1-1
III-2	1-1	1-1
III-3	1-2	1-2
III-4	1-2	1-2
III-6	1-1	1-1
II-5	1-2	1-2
II-6	1-1	1-1
III-7	1-2	1-2
III-8	1-2	1-2
III-9	1-1	1-1
II-7	1-2	1-2
II-9	1-1	1-1
III-10	1-2	1-2
II-10	1-2	1-2
II-11	1-2	2-2
III-11	1-1	1-2
II-12	1-2	1-2
II-13	1-1	1-1
II-14	1-2	1-2
II-16	1-2	1-2
II-17	1-2	1-2
III-18	1-1	1-1
III-19	1-2	1-2
III-20	1-1	1-1

Figure 3.3

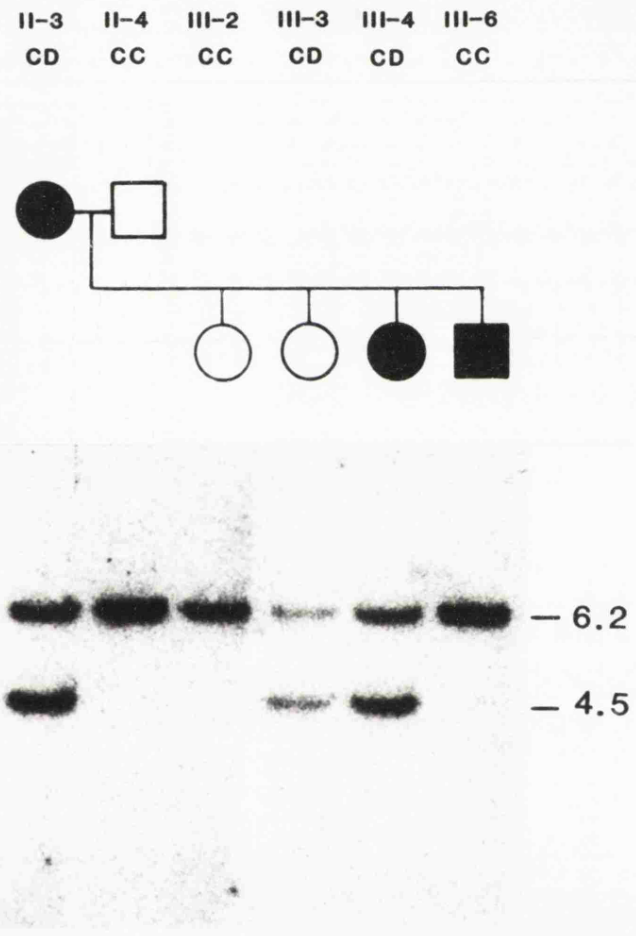
COL3A1 exclusion.

Autoradiograph demonstrating exclusion of COL3A1 as the mutant locus in this family.

The exclusion is based upon the II-3/II-4 family unit.

The bands at 6.2 and 4.5kb are termed C and D respectively.

Autoradiographic exposure was for 3 days at -80°C.


II-3
CD
II-4
CC
III-2
CC
III-3
CD
III-4
CD
III-6
CC
— 6.2
— 4.5

3.4 Linkage analysis of COL2A1 markers

E9.3, a subclone of genomic clone cosHcol1 (Weiss et al., 1982), lies in the helical region of the gene and detects polymorphic DNA fragments of 13.9kb and 7.0 and 6.9kb with HindIII. The 13.9kb fragment has been termed allele E, and the 7.0 and 6.9kb fragments have been called allele F. Results are listed in Table 3.4. Affected individual II-3 is homozygous for allele F, and her affected brother II-5 is homozygous for allele E. From the genotypes of these two individuals, it can be deduced that the parents I-3 and I-4, must both be heterozygotes. The affected mother (I-3) has passed her allele F to affected daughter II-3 and allele E to her affected son (Figure 3.4). Thus it can be seen that these two sibs, who have inherited the Marfan syndrome have inherited different α1(II) alleles from their affected mother. Therefore COL2A1 cannot be the mutant locus.

3.5 Linkage analysis of COL1A1 markers

Until July 1986, no RFLPs had been described in COL1A1 despite extensive searching. Then Sykes et al., (1986) described two RFLPs: one, an RsaI polymorphism lying towards the 5' end of the gene, yielding DNA fragments of 2.6kb and 3.6kb and the other, an MspI polymorphism lying approximately 26kb upstream from the 5' end of the gene, yielding DNA fragments of 2.7kb and 3.0kb. Few details of the probes required to detect these RFLPs were published in this paper, necessitating a search for the relevent DNA fragments. Because of the number of repetitive elements in and around the α1(I) gene, short stretches of the cosmid clone CG102 were subcloned and used as probes to find the RFLPs. The RsaI RFLP was mapped by Dr. Dalgleish, and could be detected with a 'unique' probe that did not detect any repetitive sequences. A 1.6kb HindIII fragment (H1.6) from the 5' end of CG102 was

TABLE 3.4

RFLP Results in COL2A1

Individual	*HindIII*
I-5	
II-1	
III-1	
II-3	2-2
II-4	1-2
III-2	1-2
III-3	2-2
III-4	1-2
III-6	2-2
II-5	1-1
II-6	
III-7	
III-8	
III-9	
II-7	
II-9	
III-10	
II-10	
II-11	
III-11	
II-12	
II-13	
II-14	
II-16	1-2
II-17	1-1
III-18	1-1
III-19	1-1
III-20	1-1

Figure 3.4

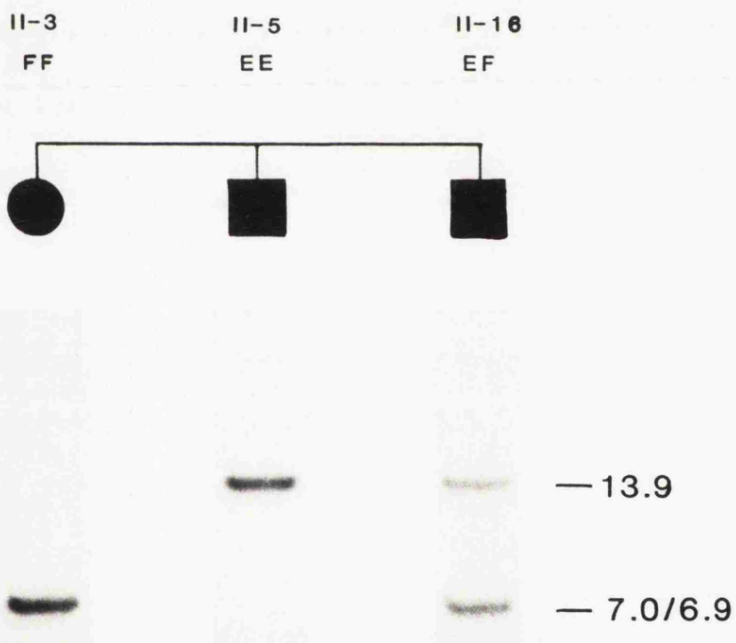
COL2A1 exclusion.

Autoradiograph showing exclusion of COL2A1 as the mutant locus in this family.

The exclusion is based upon the surviving generation-II affected individuals.

The bands at 13.9 and 7.0/6.9kb are termed E and F respectively.

Autoradiographic exposure was for two days at -80°C.


II-3
FF
II-5
EE
II-16
EF
— 13.9
— 7.0/6.9

found to detect the MspI RFLP, though with some difficulty. The probe was not 'unique' and high background hybridisation to dispersed repetitive sequences was obtained even with the use of sheared and denatured human DNA as competitor at 100μg/ml.

Because of the difficulty in obtaining interpretable results with the H1.6 probe, the probe fragment was isolated and its repetitive sequences mapped by restriction mapping and Southern blotting, using total human DNA as a hybridisation probe. The resulting map is shown in Figure 3.5. Two repetitive regions were mapped, and an apparently unique region was mapped to the middle of the fragment. A 700bp RsaI-BglII fragment (RB700) derived from the apparently unique region, was subcloned and used as a hybridisation probe. Although this probe gave better quality results than H1.6, it still contained some repetitive sequences. It was thought that this could be due to a repetitive region that failed to be detected in the mapping experiments or due to a repetitive region extending into an end of the RB700 fragment. The latter possibility was considered more likely as the RsaI site was mapped to being very close to the end of a repetitive region (Figure 3.5).

In an attempt to remove any repetitive DNA from the RsaI end of the fragment, the pRB700 recombinant plasmid was digested with the exonuclease Bal31. RB700 had been subcloned into the HincII and BamHI sites of pUC13. The recombinant plasmid was linearised with HindIII before digestion with Bal31 over a range of incubation times. Aliquots of the reaction were analysed by agarose gel electrophoresis (Figure 3.6). The remainder of the reaction mixes were ethanol precipitated, resuspended in TE and re-ligated.

The DNA from the 16 minute digestion was roughly of the desired size. The ligated DNA from this reaction was transformed into JM83recA.

Figure 3.5

Map of repetitive sequences in probe H1.6.

Figure showing internal restriction sites in the 1600bp HindIII fragment (H1.6) approximately 27kb 5' to the COL1A1 gene, which detects the MspI RFLP.

The boxed regions represent the repetitive sequences as mapped by restriction mapping and Southern blotting.

The 700bp BglII / RsaI fragment was subcloned to produce the probe RB700.

The right-hand repetitive sequence maps to within 140bp of the RsaI site.

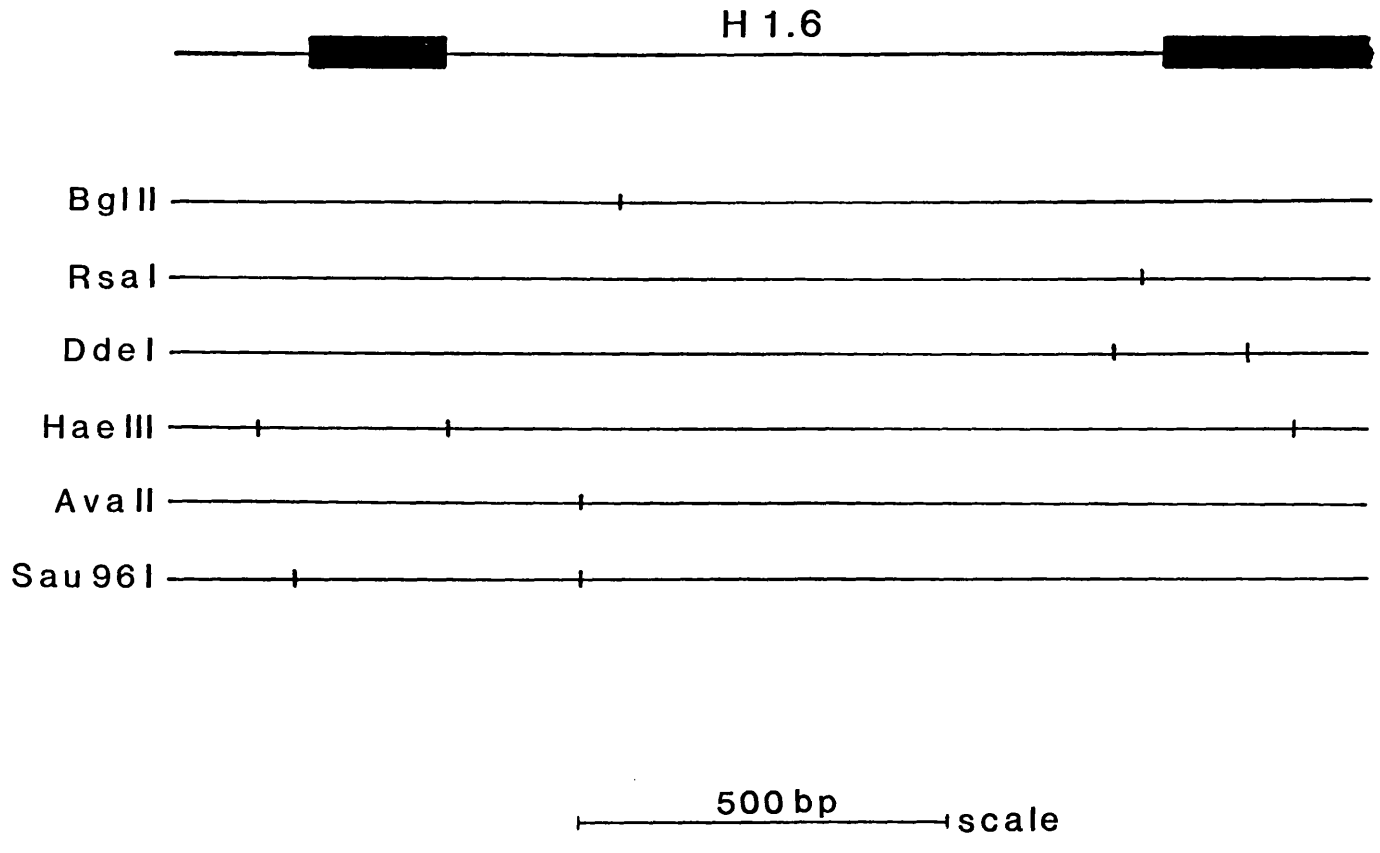

H 1.6
BglII
RsaI
DdeI
HaeIII
AvaII
Sau96I
500 bp
scale

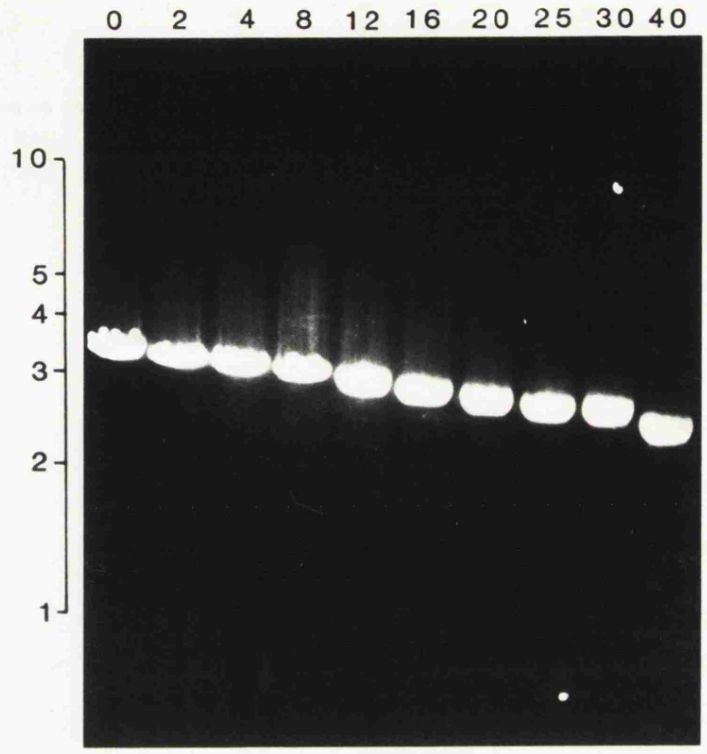
Figure 3.6

Attempt to remove repetitive sequences from the plasmid pRB700 by Bal31 deletion.

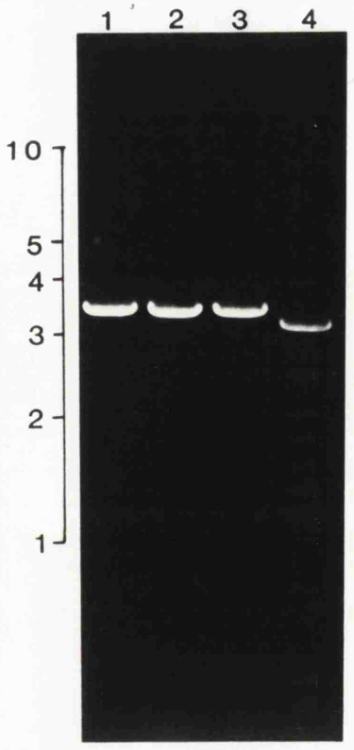
A. Bal31 deletion time course. 0.7% agarose gel containing 0.8µg aliquots of DNA incubated with Bal31 for 0, 2, 4, 8, 12, 16, 20, 25, 30 and 40 minutes. 4µg HindIII digested pRB700 was incubated with 0.15U Bal31 for each of the reaction times. Each reaction was stopped by adding EGTA to a final concentration of 20mM and placed on ice.
Sizes are marked in kilobases.

B. 0.6% agarose gel containing Eco0109 digested 'mini-preps' of colonies transformed with ligated DNA from the 16 minute incubation.
DNA from the 16 minute reaction was ethanol precipitated and 100ng of DNA re-ligated. The ligated DNA was transformed into E.coli JM83 recA.
Four resulting colonies were picked for 'mini-preps'.
Three of the colonies contained re-ligated unaffected plasmid molecules (tracks 1-3).
Track 4 contains clone pJR16'1, which has had approximately 300bp deleted.
Sizes are marked in kilobases.

A



B



The clone chosen as a hydridisation probe, pJR16'1 (Figure 3.6) had approximately 200-300bp deleted and therefore approximately 100-150bp deleted from the RsaI end of the RB700 insert (Figure 3.6). However when used as a probe, JR16'1 gave results no better than RB700. To make matters worse, the Marfan pedigree DNAs proved very difficult to digest with MspI (despite careful cleaning by phenol extraction and ethanol precipitation) and only a few family members could be typed for this polymorphism (Table 3.5).

Attention was then turned to the RsaI RFLP (Sykes et al., 1986). The control DNAs used to map this RFLP were derived from placentae. However, when the Marfan pedigree was investigated for this RFLP, the digests were partial. The experiment was repeated on these and other blood derived DNAs, with and without the addition of spermidine-HCl (Pingoud et al., 1985). The spermidine in this case did not help, and the digests were again partial. However when RsaI from a different manufacturer was tried, the blood derived DNAs were able to be digested to completion. The Marfan pedigree was then investigated for this RFLP. Unfortunately all the generation II affected sibs were homozygous (see Table 3.5), therefore providing no potential for the exclusion of COL1A1 and minimal potential lod score information.

3.6 Discussion

Dimorphic RFLP markers have been used to exclude COL1A2, COL3A1 and COL2A1 from being the possible mutant locus in this Marfan syndrome family. Because of the genetic heterogeneity of the Marfan syndrome, it has long been suspected that there may be more than one 'Marfan gene'. Additionally, because of the high new mutation rate, different families will be unlikely to be distantly related and carrying the same mutation. Hence, linkage information cannot be combined between

TABLE 3.5

RFLP Results in COL1A1

Individual	*MspI*	*RsaI*
I-5		1-1
II-1	1-1	1-1
III-1		
II-3	1-2?	1-1
II-4		2-2
III-2	1-1	1-2
III-3	1-1	1-2
III-4		1-2
III-6		1-2
II-5		1-1
II-6		
III-7		
III-8		
III-9		
II-7		1-1
II-9		
III-10		
II-10	1-2	
II-11		
III-11		
II-12		
II-13		
II-14		
II-16		1-1
II-17		1-1
III-18		1-1
III-19		1-1
III-20	1-1	1-1

families. This means not only that lod scores cannot be used additively, but any gene exclusion applies only to the family used for the study. Only when a candidate gene has been excluded in every or a significantly high number of families studied, can that gene be excluded conclusively from being the 'Marfan gene'.

Because of the variable penetrance of the phenotype in any family, individuals with the 'Marfan gene' may pass undiagnosed or be classed as 'possible Marfan's'. However, those individuals who are diagnosed as having the syndrome are very unlikely to be mis-diagnosed normals. The exclusion of each of the three loci as being the Marfan locus was based only upon the affected members of the family, so that the question of diagnosis is not relevant. One assumption made in these gene exclusions was that the unaffected (unrelated) parent of each family unit used was normal and not contributing to the Marfan phenotype in any way. Given the rarity of the disease this was a reasonable assumption and for the same reason all family relationships were assumed to be genuine.

For reasons described in section 1.4.1, the genes for types I, II and III collagen were reasonable candidates for being the site of the Marfan mutation. The type II collagen gene, encoding a cartilage-specific collagen was unlikely to be the mutant locus, but considered to be worth including for completeness. Unfortunately no information was obtained from the COL1A1 RFLPs for several reasons.

Soon after the work on the COL1A1 RFLPs was initiated it was heard that Dr. B. Sykes (Oxford University) had excluded the 4 loci for types I, II and III collagen in two families. Later, discordance between these loci and the disease was described in those two families and two others by Ogilvie et al., (1987a). Despite the description of a third RFLP in COL1A1 (Ogilvie et al., 1987b), it ceased to be a

priority to exclude COL1A1 as being the Marfan locus in this family.

The results described in this chapter, combined with those of Ogilvie et al., (1987a) argue against the genes encoding the 3 major fibrillar collagens as being the disease-causing genes. However, because of the heterogeneity of the phenotype, further work on other families will have to be done to reinforce this. In addition to the two other known fibrillar collagens (types V and XI), any of the non-fibrillar collagens might possibly be defective in the Marfan syndrome. However, the fact that there is evidence suggesting that collagen is defective in Marfan syndrome does not necessarily mean the underlying defect is in a collagen gene, but that the defective protein could affect collagen structure or metabolism. These genes as they are identified, should be considered as candidates for the 'Marfan gene'. The 11 or more enzymes involved in collagen biosynthesis are unlikely to be the disease cause because the inheritance of enzyme defects is generally recessive.

Once all candidate genes have been excluded, an alternative approach may be adopted using dispersed and anonymous linkage markers, many of which will be more variable and more likely to be informative. However, such a study would require an enormous amount of tedious work and few, if any, groups would want take it on unless it was done as a collaboration between many laboratories, all working on the same family.

CHAPTER FOUR

ESTABLISHMENT OF RNASE A MAPPING IN THE STUDY OF COLLAGEN GENE MUTATIONS

4.1 Introduction

To date, most mutations characterised in OI have been either sizeable deletions or point mutations introducing either a cysteine or an arginine residue into the helix. However, it is unlikely that these mutations represent a true perspective of mutations in OI, but that they are the simplest to detect and therefore represent a biased sample.

It was hoped that the application of RNase A mapping of mutations in OI would broaden the spectrum of mutations seen to cause OI. It is the only technique to be deployed in the detection of collagen gene mutations that does not specifically identify certain types of mutation. RNase A mapping has the potential to identify all types of defect from gross deletions to point mutations. However, being unable to detect all point mutations, the technique will tend to favour the detection of deletions and insertions, however small. Thus the method is not without bias, but does represent the least biased approach deployed so far in the detection and mapping of collagen gene mutations.

At the time of the conception of this project only two papers had been published on RNase A mapping, that of Myers et al., (1985d) concerning single base mismatches between DNA:RNA hybrids and that of Winter et al., (1985) concerning single base mismatches between RNA:RNA hybrids. By the time the project had been seriously initiated two further papers had been published; that of Gibbs and Caskey (1987),

concerning the mapping of mutations at the HPRT locus and that of Forrester et al., (1987), concerning mutations in the K-ras oncogene, both of which employed the RNA:RNA hybrid option.

Because of the lack of published material on the technique and the general fear and scepticism of the technique by fellow scientists, it was decided to set up the technique in steps rather than attempt a full-blown RNase A mapping experiment and have little idea of what was wrong if (and when!) the experiment failed.

4.2 In Vitro Transcription from Cloned cDNA Templates

In vitro synthesis of RNA was directed from cDNA fragments subcloned into the phagemid transcription vectors pTZ18R and pTZ19R (Mead et al., 1986). These vectors contain the highly specific bacteriophage T7 promoter immediately adjacent to the multiple cloning site. Hence transcription of foreign DNA inserts is a purely promoter-driven event requiring no primer.

cDNA fragments were subcloned into these vectors in an orientation such that the T7 promoter is 'downstream' of the insert with respect to its in vivo arrangement. Thus in vitro transcription is directed off the non-coding strand of the DNA, generating 'antisense' RNA.

Recombinant vectors were linearised at the 5' end of the insert (with respect to their in vivo arrangement) to restrict transcription to the insert only. Linearisation was always conducted with restriction endonucleases that leave a 5' overhanging end. Restriction endonucleases that leave a 3' overhanging end were avoided because of the low level of transcription by T7 polymerase that can be promoted by this type of overhang (R. Gibbs, personal communication).

The first experiment in setting up the system of RNase A mapping

was to test that RNA could be in vitro transcribed with equal ease from different sized inserts. Transcription reactions were carried out as described in Table 4.1 part (a) on three subclones; RS355, XE677 and BX887, containing inserts generating 355, 677 and 887 base proα1(I) anti-sense RNAs respectively (see Figure 4.1 and Table 5.1).

The entire reaction mix was electrophoresed on a 0.7% agarose gel with no heating of the samples prior to loading (Figure 4.2). The transcription reaction can be seen to have worked well, yielding µg quantities of RNA equally efficiently from each template.

The DNase I digestion of the template can be seen to have efficiently removed all the template following the transcription reaction. In each track of Figure 4.2, running just above the the major band of RNA is a less intense additional band. Two transcription reactions using the subclone XE677 were carried out as in Table 4.1 part (b) and heated to 65°C for 8 minutes and 80°C for 8 minutes prior to electrophoresis on a 0.7% agarose gel (See Figure 4.2b). The heating reduces the intensity of the of the upper band suggesting that its presence is due to secondary structures in the RNA.

The next stage in the development of the technique was to be able to visualise discrete transcripts and appropriate size markers on a denaturing polyacrylamide gel. The size markers were pUC13 x MspI, pUC13 x TaqI, λ x StyI and M13mp18rf x MspI (See Table 4.2) and end-labelled as described in section 2.11.3. RNA transcripts from RS355, XE677, and BX887 were synthesised in the presence of α ^{32}P-CTP as in Table 4.1 part c. Aliquots of each of the labelled DNA markers and the labelled RNA transcripts were mixed with a one-half volume of sequencing loading dye and electrophoresed on a 5% denaturing polyacrylamide gel, with cold water circulating through the cooling tank of the apparatus. The resulting autoradiograph is shown in Figure

TABLE 4.1

In Vitro Transcription Reactions

	µl			
	a	b	c	d
10mM GTP	1	1	1	1
10mM ATP	1	1	1	1
10mM UTP	1	1	1	1
10mM CTP	1	1	0.5	-
75µM CTP	-	-	-	2
100mM DTT	2	2	2	2
5x Buffer	4	4	4	4
2mg/ml BSA	1	1	1	1
40U/µl RNasin	1	0.8	0.8	1
30U/µl T7 RNA polymerase	0.5	0.5	0.5	0.5
0.8µg/µl Template DNA	1	0.6	0.6	0.6
^{32}P-CTP (800Ci/mmol)	-	-	20µCi	80µCi
H_2O to 20µl				
	20µl	20µl	20µl	20µl

	e
10x NTP(c)	1
100mM DTT	1
5x Buffer	2
2mg/ml BSA	0.5
40U/µl RNasin	0.8
15U/µl T7 RNA polymerase	0.7
0.4µg/µl Template DNA	0.6
^{32}P-CTP (800Ci/mmol)	40µCi
H_2O to 10µl	
	10µl

All incubations were for 2 hours at 37°C, followed by a further 15 minutes after the addition of 1µl DNase I (7.5U/µl) for 20µl reactions or 0.8µl for 10µl reactions.

Figure 4.1

RNA probe map.

Map showing the relative positions of the *in vitro* transcribed RNAs described in sections 4.1 and 4.2.

Marked on the mRNA are the regions corresponding to the protein domains: H = helical domain; T = carboxyl telopeptide; C = carboxyl propeptide; UT = 3' untranslated region.

Hf404 and Hf677 are the cDNA clones from which the subcloned RNA templates were derived.

The horizontal arrows indicate the direction of *in vitro* transcription.

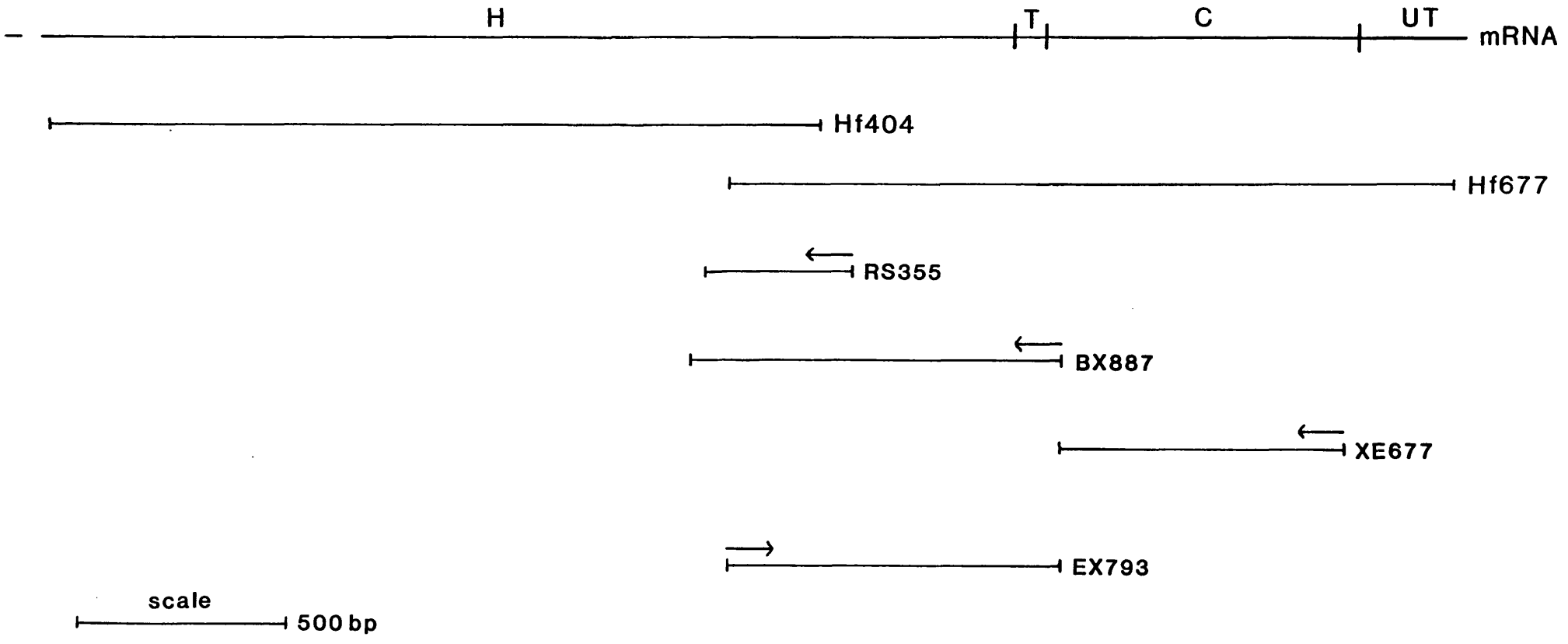

H
T
C
UT
mRNA
Hf404
Hf677
RS355
BX887
XE677
EX793
scale
500 bp

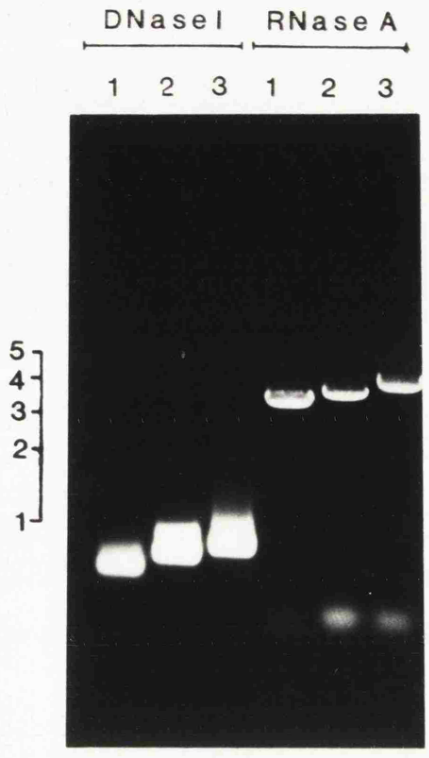
Figure 4.2

In vitro RNA synthesis.

A. 0.7% agarose gel containing in vitro transcription products.
Tracks 1, 2 and 3 contain reaction products of the clones RS355, XE677 and BX887 respectively.
The left-hand three samples were treated with DNase I (at 0.38U/µl) to remove the probe template after the transcription reaction. The DNase digestion is seen to be complete.
Each RNA product is seen as a major band, with a minor band migrating just behind it.
The right-hand three samples were treated after the transcription reactions with RNase A (at 2ng/µl). This proved that it was indeed RNA that was being seen and showed that the DNA templates were not degrading during the reaction.

B. 0.7% agarose gel containing in vitro transcribed RNAs heated before loading. The RNAs were heated to 65°C and 85°C for 8 minutes prior to loading. This has the effect of reducing the intensity of the minor band migrating just behind the major band.

A



B

TABLE 4.2

DNA SIZE MARKERS

pUC13/*MspI*	pUC13/*TaqI*	λ/*StyI*	M13/*MspI*
501	1444	1489	1596
489	736	925	829
404	476	421	818
326			651
242			542
190			472
147			454
110			357
109			215
67			176
34			156
26			130
			123
			60

Un-resolvable sizes not included

4.3. The RNA transcripts are seen as discrete homogeneous bands, showing that only full-length transcripts are generated from each of the different sized templates. It was thus felt that the same conditions could probably be employed for labelling transcripts of 350 bases to 1000 bases in size.

The DNA size markers are seen to relate closely enough to the RNA sizes for the intended purpose. The inserts of RS355, XE677 and BX887, along with co-transcribed polylinker DNA will produce transcripts of approximately 405, 705 and 920 bases respectively. These are roughly sized by the size markers to be 410, 710 and 1000 bases respectively. Thus sizing of RNA transcripts with DNA size markers is feasable, but for larger fragments should be used only as a guideline rather than as an accurate sizing.

4.3 RNase A protection of an artificial RNA

In view of the anticipated technical difficulty in producing 'protected' fragments of human RNA, it was decided firstly to attempt to protect an abundant *in vitro* transcribed RNA molecule.

A 793bp EcoRI-XhoI fragment derived from the cDNA clone Hf677 was subcloned into the EcoRI and SalI sites of pTZ18R. This recombinant plasmid was called 'EX793' (see Figure 4.1). Subcloning into pTZ18R meant that the *in vivo* direction of transcription was maintained, enabling the synthesis of 'sense' RNA from the template.

The RNA was synthesised as in Table 4.1 part b. Overlapping labelled antisense RNAs were generated from RS355 and BX887 as in Table 4.1 part (c). 10µl EX793 sense RNA (half the reaction) was added to 10µl of each probe or 'protector' RNA (again, half the reaction). Each tube was ethanol precipitated and the invisible pellet washed in 70% ethanol. The hybridisation, RNase A digestion and recovery of

Figure 4.3

Labelled DNA size markers and RNA transcripts.

Tracks 1-5 contain ^{32}P labelled DNA size markers. Each track contains 1µl (approximately $^{1}/_{20}^{th}$) of the labelling reaction (see section 2.10.3).

Track 1 is pUC13/MspI; track 2 is pUC13/TaqI; track 3 is λ/StyI; track 4 is M13mp18rf/MspI; track 5 is M13mp18rf/MspI plus pUC13/MspI. Sizes of these markers are listed in Table 4.2.

Tracks 6-8 contain *in vitro* synthesized ^{32}P labelled RNA transcripts. Each track contains $^{3}/_{2000}^{ths}$ of the labelling reaction (see Table 4.1 part c).

Track 6 contains the transcript generated from clone RS355.

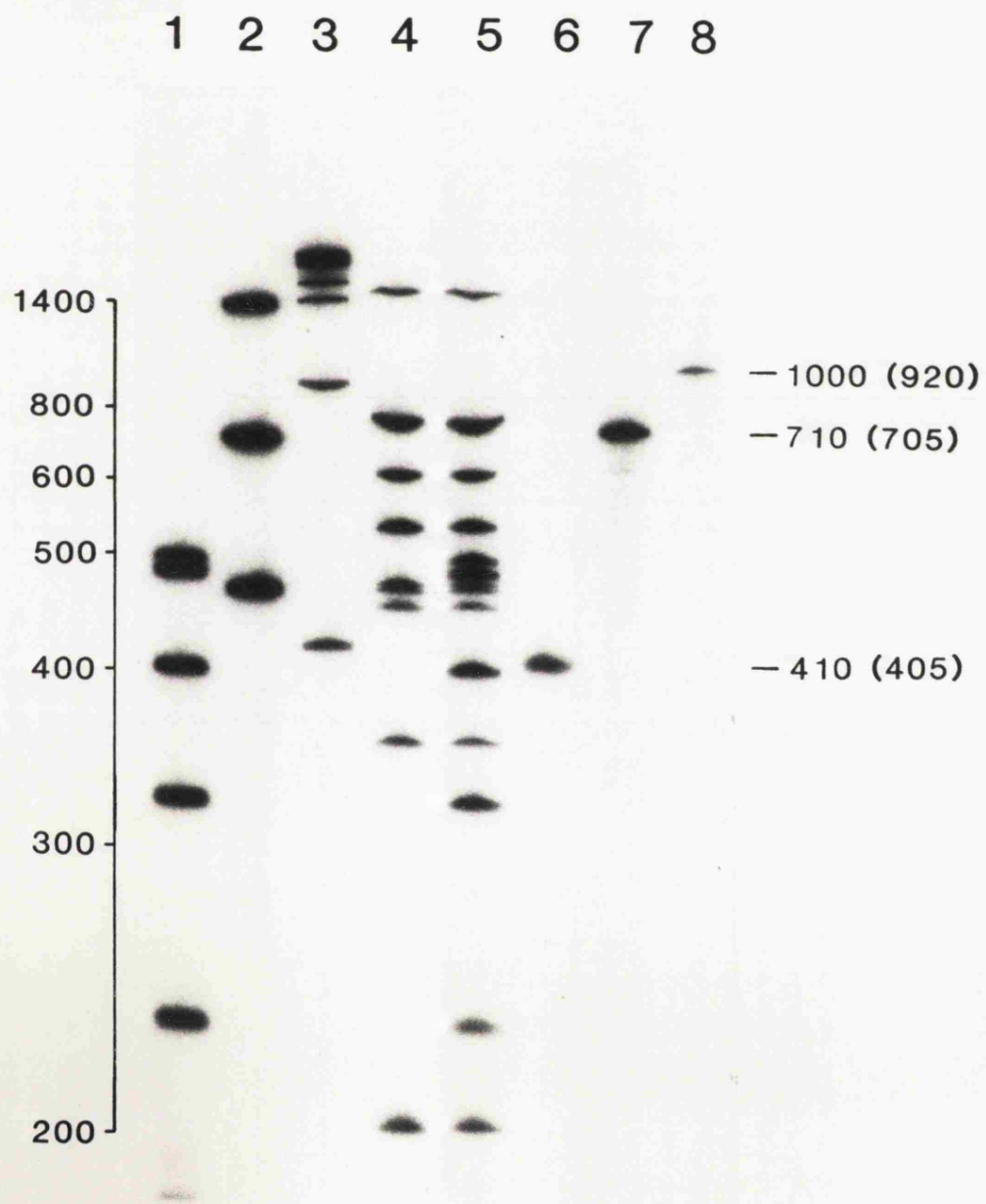
Track 7 contains the transcript generated from clone XE667.

Track 8 contains the transcript generated from clone BX887.

The numbers on the right indicate the sizes of the transcripts as calculated from the size markers. The adjacent numbers in brackets indicate the genuine size of the transcripts.

Autoradiographic exposure was for 2 hours without an intensifying screen.

Sizes are marked in bases.


1 2 3 4 5 6 7 8
1400
800
600
500
400
300
200
— 1000 (920)
— 710 (705)
— 410 (405)

'protected' RNA were conducted as decribed in section 2.17, but using RNase A at a stock concentration of 5µg/ml.

The resulting final (invisible) pellet was dissolved in 20µl TE. 1µl aliquots were added to 7µl distilled water and 4µl loading dye, denatured at 85°C for times of 2 and 5 minutes and then loaded onto a 5% denaturing polyacrylamide gel with the cooling system on. (See Figure 4.4).

As shown in Figure 4.4, the BX887 probe should protect the EX793 RNA in its entirety and the RS355 probe should only protect the 5' terminal 299 bases. The expected protected fragments are seen as discrete, undegraded bands and the short denaturation time of 2 minutes shows that the hybrids are not difficult to denature.

4.4 RNase A protection of human RNA

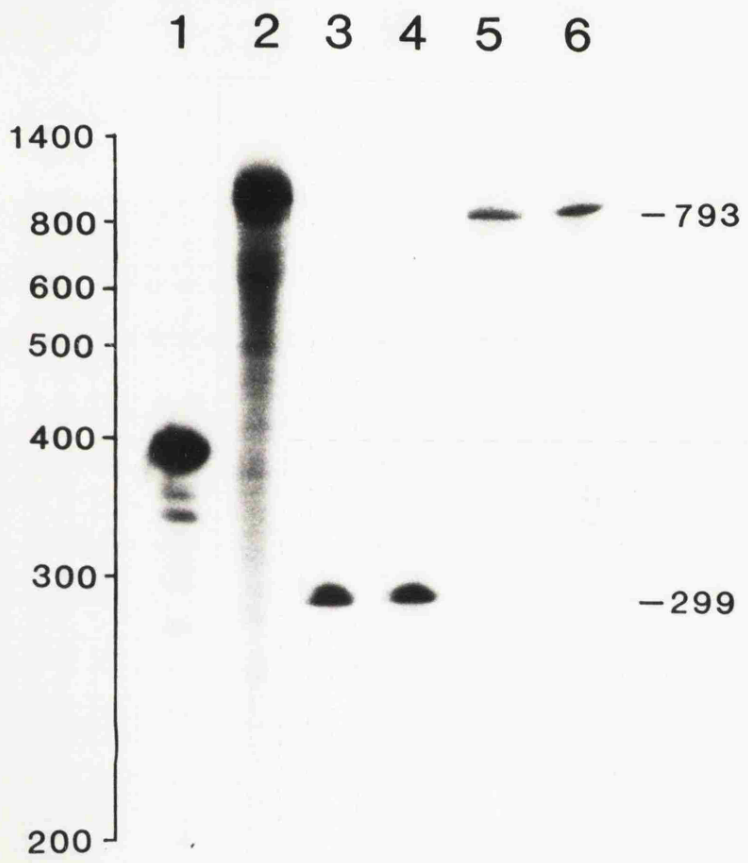
With the success of the above experiment, it was felt that RNase A protection of human RNA should not be technically too difficult. Initially the placental RNA to be 'protected' was tested for purity: 100µg of 2 different RNA preparations were ethanol precipitated and resuspended in 80µl deionised formamide. 20µl 5x hybridisation buffer was added and mixed. The RNA solution was then incubated at 65°C overnight. 4µg samples of incubated and non-incubated RNA were elecrophoresed on a 0.7% agarose gel. Despite some degradation of the 28S and 18S ribosomal RNA's, the bulk of the incubated RNA was in reasonable condition. Considering the poor stability of RNA, the result of this experiment was encouraging. Thus it was considered that the placental RNA samples and the hybridisation solution were of sufficient quality to undergo overnight hybridisation reactions.

A radiolabelled probe was synthesised from XE677 as described in Table 4.1 part (d). The probe was phenol extracted, ethanol

Figure 4.4

RNase A protection of an artificial RNA.

Tracks 1 and 2 contain aliquots of the labelled transcripts (probes) derived from RS355 and BX887 respectively.
Tracks 3 and 4 contain the EX793 sense RNA protected by the probe RS355.
Tracks 5 and 6 contain the EX793 sense RNA protected by the probe BX887.
Probe RS355 protects only the 5' terminal 299 bases of the XE793 sense RNA, whereas probe BX887 protects it in its entirety.
Protected RNA in tracks 3 and 5 was denatured at 85°C for two minutes. Protected RNA in tracks 4 and 6 was denatured at 85°C for five minutes.
Autradiographic exposure was for one hour without an intensifying screen.
Sizes are marked in bases.


1 2 3 4 5 6
1400
800
600
500
400
300
200
−793
−299

precipitated and resuspended in 80µl distilled water. 1 x 10^6 cpm (2µl) was hybridised to 200µg placental RNA. Manipulations were as described in section 2.17, except that the final 'protected RNA' pellet was dissolved in 8µl distilled water. 4µl tracking dye was added to the RNA solution and the whole sample was heated at 85°C for 5 minutes and then electrophoresed on a 5% denaturing polyacrylamide gel along with DNA size markers and an aliquot of the probe (Figure 4.5).

A single band of protected RNA can be seen and is of the expected size. The probe can be seen to be of a slightly larger size due to the presence of about 30 bases of polylinker-derived RNA in the probe, which have been cleaved away from the double stranded RNA hybrid.

The experiment was repeated using the same probe, having been stored for 2 days at -20°C (Figure 4.5). The background, caused by radiolysis, is seen to have increased. Because it will be required to see faint RNase A cleavage products, the background smear needs to be as faint as possible. It was concluded that probes should be used immediately only, and not after storage.

Now that the technique of RNase A protection of human RNA was shown to be working, a brief attempt was made to refine the technique. The stoichiometry of the labelling reaction requires the use of a large amount of the labelled nucleotide (80µCi in the experiment above). As only a proportion of the labelling reaction was used as the probe, it stood to reason that less label could be used by reducing the volume of the labelling reaction. Thus reducing the cost and minimising radiation exposure of the worker.

The labelling reaction was reduced to a 10µl volume (Table 4.1 part e) by making the a 10x NTP mix (100µl = 5µl 100mM GTP, 5µl 100mM ATP, 5µl 100mM UTP, 0.75µl 10mM CTP and 84.25 µl distilled water). This

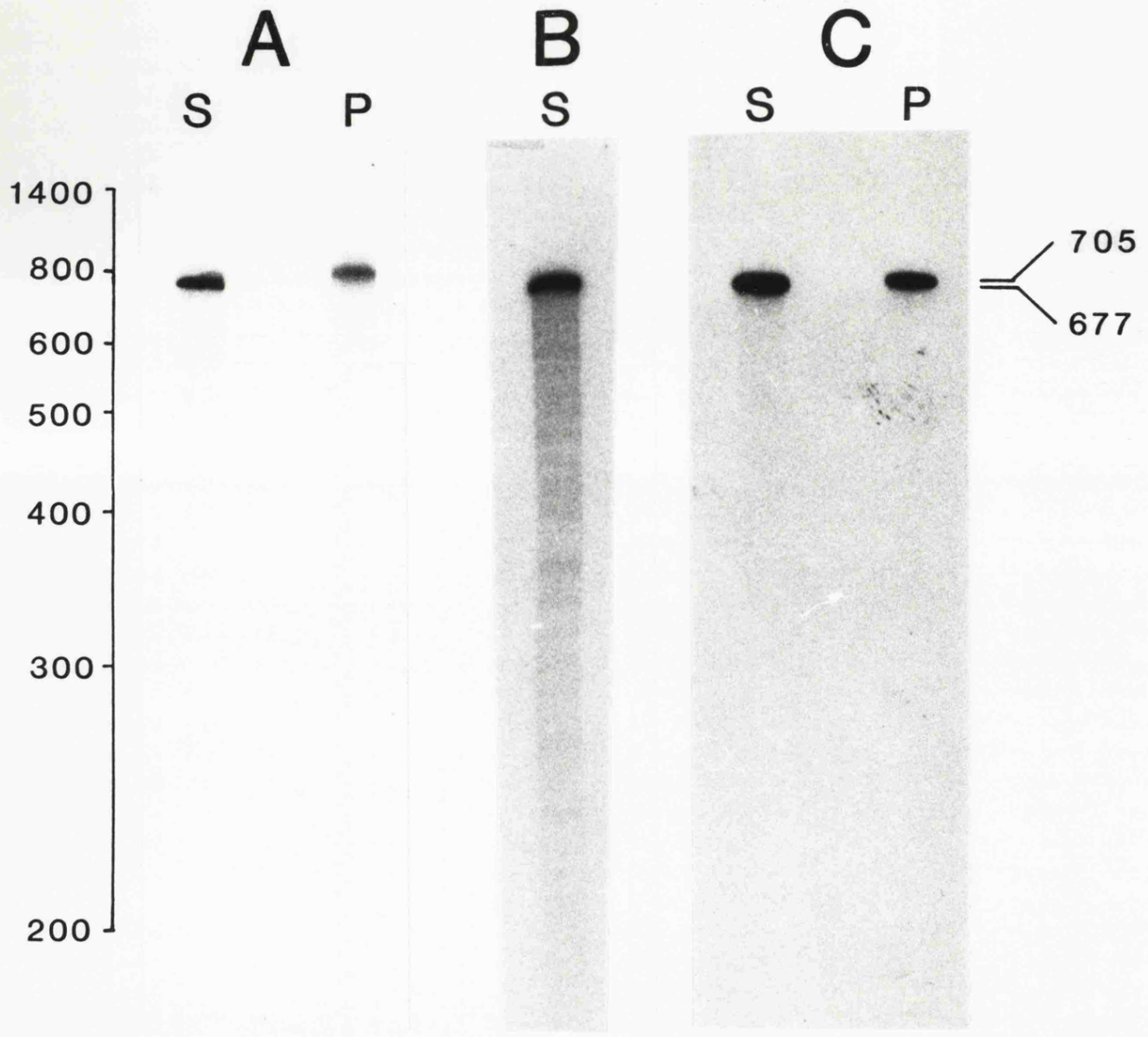
Figure 4.5

RNase A protection of human pro-α1(I) collagen mRNA.

A. Placental pro-α1(I) mRNA protected by probe XE677.
The probe was synthesized in a 20µl volume using 80µCi ^{32}P-CTP.
The hybridisation was conducted with 200µg placental total RNA and 1 x 10^6 cpm probe.
Track S contains the sample, the protected RNA.
Track P contains an aliquot of the probe only.
Autoradiographic exposure was for one day without an intensifying screen.

B. Repeat of RNase protection as in Part A, using the same RNA sample and the same probe, after storage at -20°C for 2 days.
The background smear below the protected RNA band can be seen to have increased significantly.
Autoradiographic exposure was one day without an intensifying screen.

C. Repeat of RNase protection using XE677 probe synthesized in a 10µl volume using 40µCi ^{32}P-CTP.
The hybridisation was conducted with 200µg placental total RNA and 2 x 10^6 cpm probe.
Track S contains the protected RNA sample and track P contains the an aliquot of the probe only.
Autoradiographic exposure was for 6 hours with an intensifying screen.
Sizes are marked in bases.


A
B
C
S
P
S
S
P
1400
800
600
500
400
300
200
705
677

mix was called 10x NTP(c), indicating that it was the mix for labellings where α-^{32}P-CTP is the label to be used. The 10µl labelling reaction was performed with an accordingly reduced amount of isotope; 40µCi, and the amount of probe used for the hybridisation doubled to 2 x 10^6 cpm, see Figure 4.5. Thus the amount of input label could be reduced by half and the quality of the resulting autoradiograph improved by raising the amount of probe used.

4.5 RNase protection of collagen RNA: A problem?

In an experiment to further refine the hybridisation conditions, the BX887 probe was accidentally labelled instead of the XE677 probe which had previously been shown to work well. In this experiment varying amounts of probe and target RNA were tried. However, none of the varying conditions produced any bands on the resulting autoradiograph. RNA incubations with reagents followed by analysis on agarose gels, suggested that the RNA may have been degraded by the formamide in the hybridisation. The experiment was repeated (with BX887 and the same placental RNA sample) with a new batch of formamide for the hybridisation. This experiment did produce a visible band on the autoradiograph, but it was a fuzzy band less than half the size of the expected band (Figure 4.6a). The experiment was repeated again and the same result obtained.

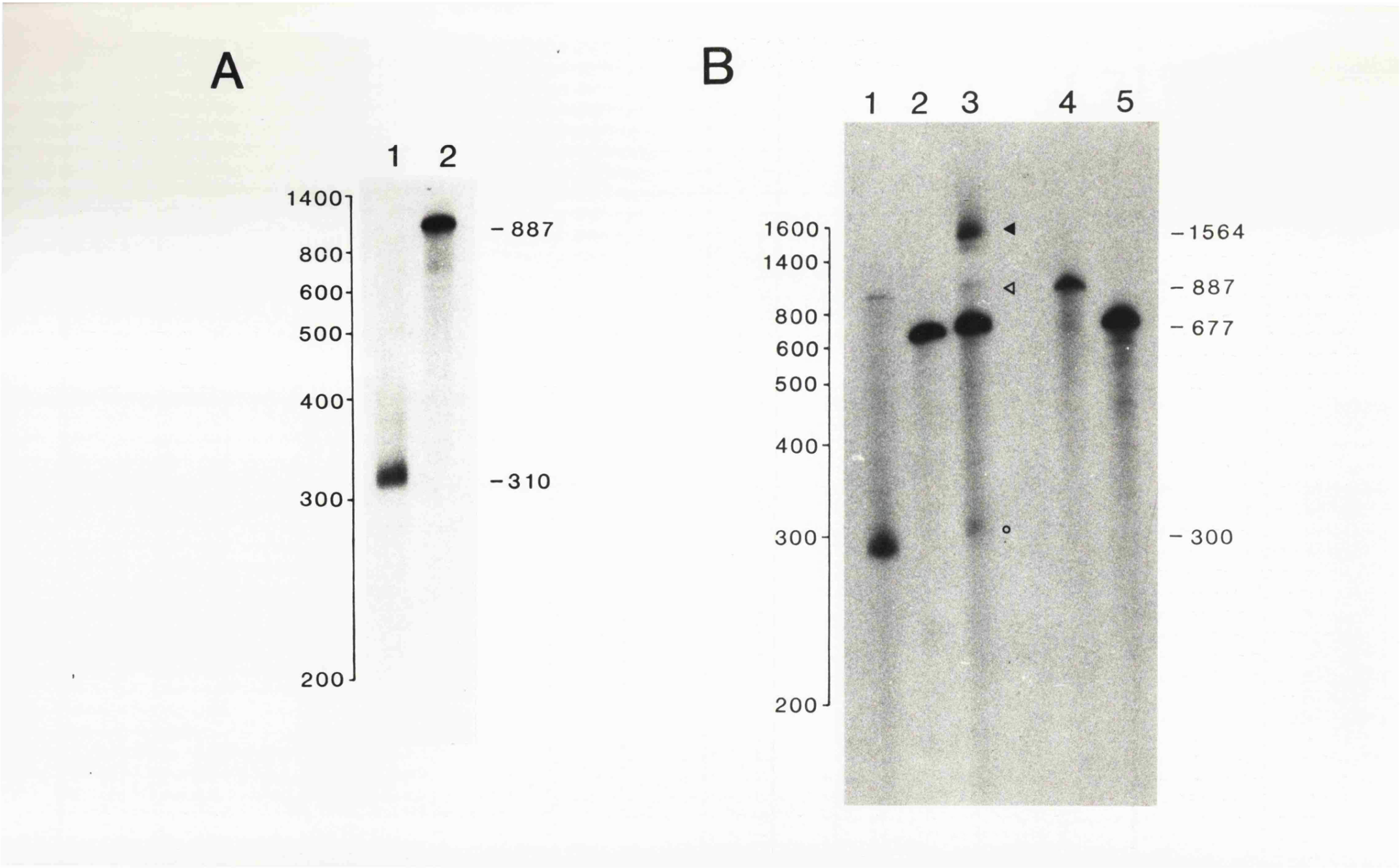
The XE677 probe was then tested to determine if something had gone wrong with the system or whether the fuzzy band effect was specific to BX887. The XE677 probe worked as it previously had done, giving a discrete band of the expected size. Thus either BX887 had been mal-prepared or it was 'misbehaving' at some stage of the procedure. The BX887 clone was finely mapped for restriction endonuclease sites and was determined to be the correct insert in the correct vector.

Figure 4.6

RNase A protection with probe BX887.

A. RNase A protection of human pro-α1(I) collagen mRNA with probe BX887.
Track 1 contains the RNA protected by probe BX887. A band of 887 bases in size was expected. The only band seen is 'fuzzy' in appearance and only approximately 310 bases in size.
Track 2 contains an aliquot of the BX887 probe only. The band is normal in appearance and of the expected size, approximately 920 bases.
Autoradiographic exposure was for one day with an intensifying screen.
Sizes are marked in bases.

B. Dual RNase A protection with probes XE677 and BX887. Tracks 1-3 are RNase protections of a single batch of placental RNA.
Track 1 contains RNA protected by probe BX887. Just visible is the expected-length protected RNA, Marked with the open triangle. At approximately 300 bases is the 'fuzzy' band.
Track 2 contains RNA protected by probe XE677. The expected length band of 677 is seen as normal.
Track 3 contains RNA protected by both probes XE677 and BX877. A band at approximately 1564 bases is observed, marked with the filled-in triangle. This band probably represents mRNA molecules protected by both probes simultaneously (see Figure 4.1). A faint band at 887 bases is present (marked with the open triangle) and the fuzzy band is just visible at around 300 bases (marked with the open circle).
Autoradiographic exposure was for one day with an intensifying screen.
Sizes are marked in bases.


A
1
2
1400
800
600
500
400
300
200
887
310
B
1
2
3
4
5
1600
1400
800
600
500
400
300
200
1564
887
677
300

To prove beyond any doubt that something about the sequence of the BX887 probe was causing it to 'misbehave', the BX887 and XE677 probes were hybridised together to the same RNA sample (Fig 4.6b). The BX887 probe is seen on its own to give the fuzzy band and a faint smear along with a very faint band of the expected size (Figure 4.6b; track 3). When mixed with XE663, the fuzzy band was still present but at reduced intensity. In this track a very large band was present of approximately 1600 bases. This presumably represents RNA molecules protected by the two probes which are adjacent (see Figure 4.1). This has the effect of reducing the intensity of the BX887 fuzzy band on the autoradiograph. The presence of the BX887 fuzzy band and the lack of the expected sized band along with the normal XE677 band shows that the problem is sequence-specific.

The BX887 probe covers the region of the α1(I) mRNA corresponding to the carboxyl-terminal end of the helical region. The XE677 probe covers the region of the mRNA corresponding to the C-telopeptide and the C-propeptide. It was feared that the problem encountered with the BX887 probe was one which would be encountered with any RNA probe correspondong to the helical domain of the α1(I) chain or indeed any other collagen α-chain.

To investigate this, 3 new clones were constructed; EP934, SE737 and TT693, (see Table 5.1 and Figure 5.1). EP934 and SE737 correspond to the middle region of the helical domain and TT693 spans the region of the mRNA corresponding to the end of the helical domain and the C-telopeptide. When used as probes for RNase A protection, all 3 gave results similar to BX887 (Figure 4.7; tracks 1-3). Each gave visible expected-length bands, however that of TT693 was more intense than the other probes (Figure 4.7; track 3). Interestingly, the probe 'signature' remains, i.e. the largest probe: EP934 gave the largest

Figure 4.7

RNase A protection with 3 helical-region derived probes.

Tracks 4, 5 and 6 contain aliquots of the probes EP934, SE737 and TT693.

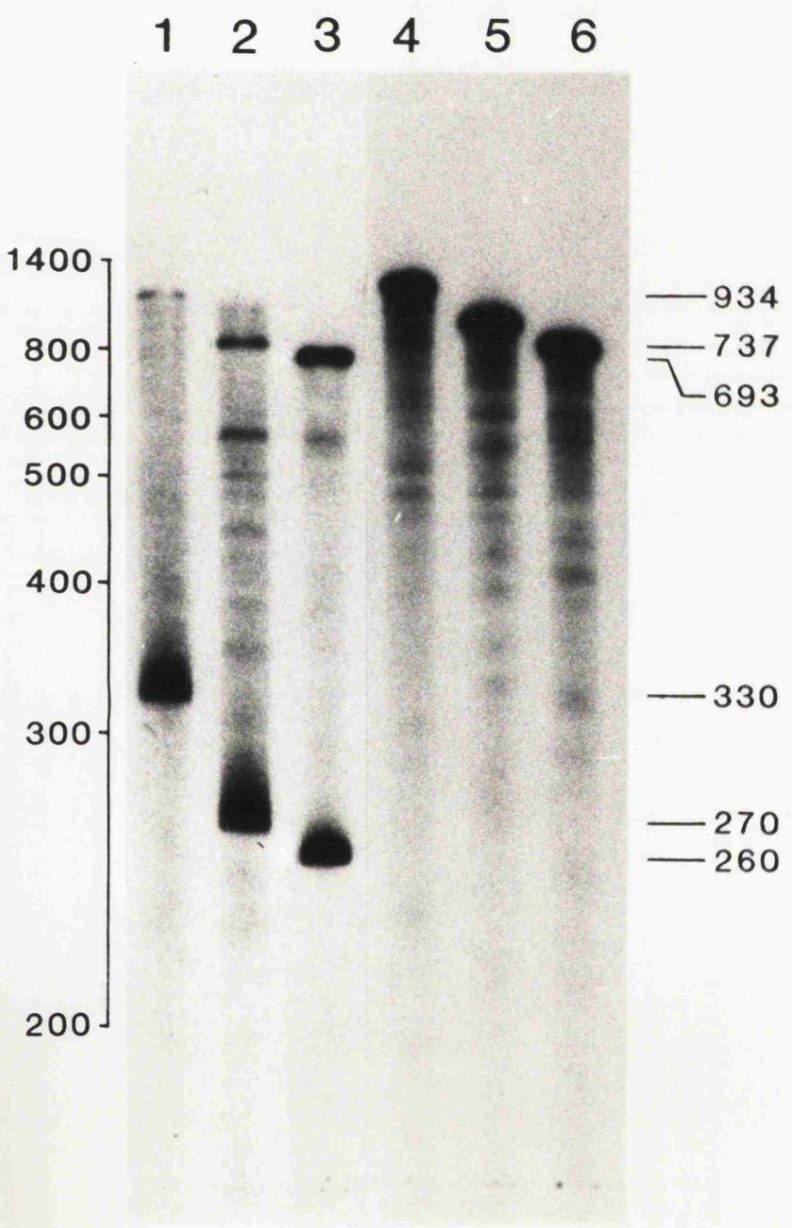
Tracks 1, 2 and 3 contain RNase A protections of human placental RNA with probes EP934, SE737 and TT693 respectively.

Labelling reactions were performed as described in Table 4.1e.

2×10^6 cpm probe was hybridised to 170µg total RNA. RNase A digestion and electrophoresis was as described in section 4.3.

Autoradiographic exposure was for one day with an intensifying screen.

Sizes are marked in bases.


1 2 3 4 5 6
1400
800
600
500
400
300
200
934
737
693
330
270
260

fuzzy band, and the smallest probe: TT693, gave the smallest fuzzy band. The size ratio for the probe sizes is 1:0.79:0.74 and the size ratio for the fuzzy bands was 1:0.81: 0.78.

It seemed likely that these fuzzy bands were caused by secondary structures within the probes themselves or between probes and the target RNAs. In an experiment to determine whether the probe was preferentially hybridising itself and being carried through by inadequate washing of the mAP paper or hybridising partly or completely to the target RNA, the TT693 probe was hybridised to 160µg Aspergillus nidulans RNA (donated by Dr. Raj Beri, ICI Joint Laboratory, University of Leicester) and 160µg human placental RNA (Figure 4.8). No band of any size was seen in the Aspergillus track, even on a long exposure. With the human RNA, the TT693 probe again gave a fuzzy band and a band of the expected size. This showed that the probe must be binding, at least to some extent to the target RNA.

The next experiment was to alter the hybridisation stringency by altering the temperature. Hybridisations using the probe TT693 were performed at 50°C, 60°C and 70°C (Figure 4.8). Each sample still produced the fuzzy band and at similar intensities with respect to the upper bands. Thus raising or lowering the stringency of hybridisation does not alter the intensity of the fuzzy band, therefore, the event causing the fuzzy band had to be occurring after hybridisation.

It had previously been imagined that the fuzzy bands were caused by the formation of some secondary structure prior to RNase A digestion and that the band was small in size because the probe had partially protected itself.

A DEC VAX computer was used to analyse the potential for α1(I) RNA secondary structures with RNAFOLD program on the University of Wisconsin Genetics Computer Group (UWGCG) package. This yielded no

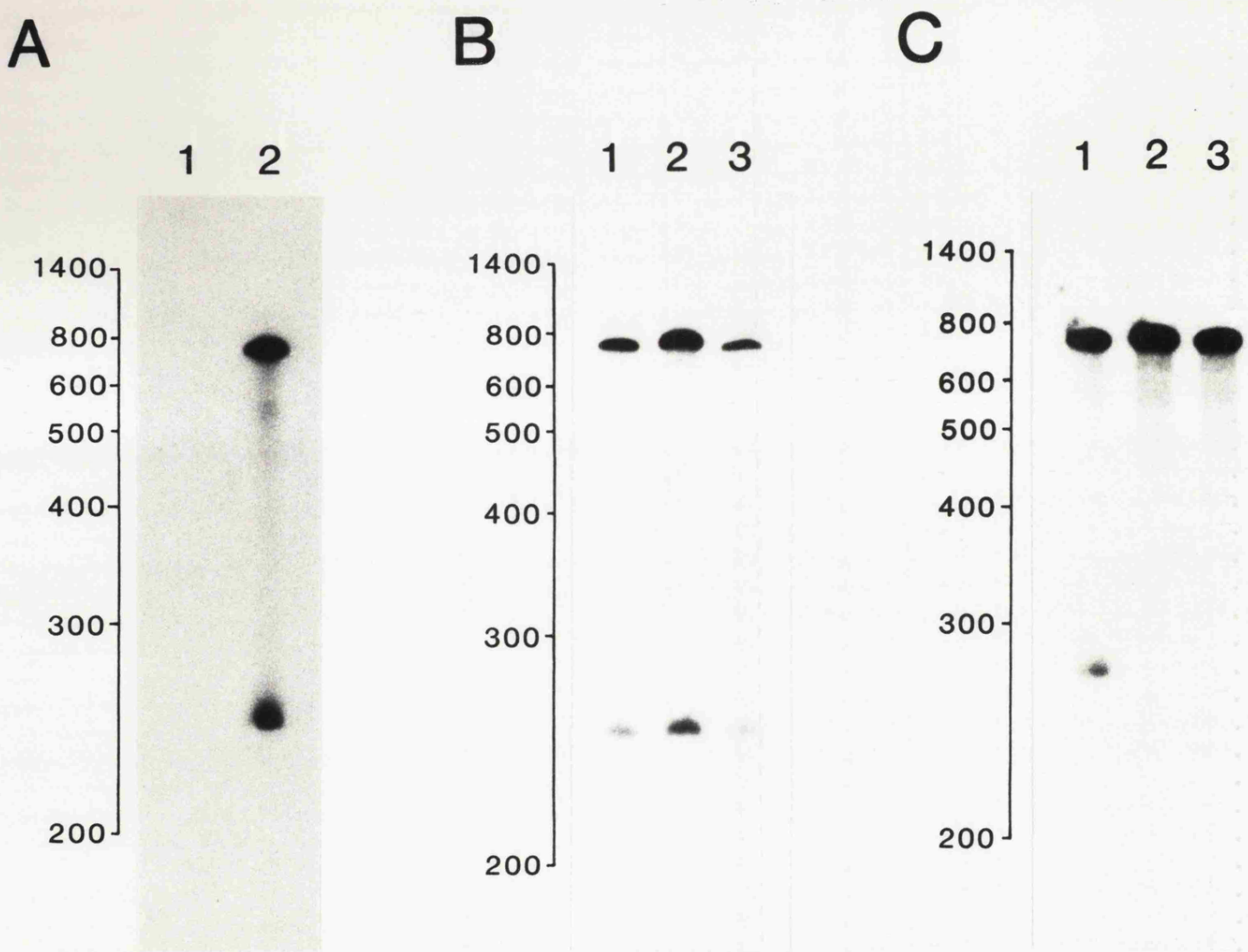
Figure 4.8

Optimising RNase A protection conditions for the helical-region probes TT693.

A. RNase A protection of A. nidulans and human RNA with probe TT693.
Track 1 contains RNase A 'protection' of A. nidulans RNA with probe TT693.
Track 2 contains RNase A protection of human placental RNA with probe TT693.
In both cases, 160µg total RNA was hybridised to 2 x 10^6 cpm probe.
Autoradiographic exposure was for one day with an intensifying screen.

B. RNase A protection of human placental RNA with probe TT693 using differing hybridisation conditions.
Tracks 1, 2 and 3 contain protected RNA from hybridisations with temperatures of 50°, 60°, and 70°C respectively.
In all cases 160µg total human placental RNA was hybridised to 2 x 10^6 probe.
Autoradiographic exposure was for one day with an intensifying screen.

C. RNase A protection of human placental RNA with probe TT693 employing varying final denaturation conditions.
Track 1 contains protected RNA finally resuspended and denatured as before (i.e. resuspended in 8µl H_2O and 4µl tracking dye and heated at 85°C for 5 minutes).
Track 2 contains protected RNA finally resuspended in tracking dye only and heated at 85°C for 7 minutes.
Track 3 contains protected RNA finally resuspended in tracking dye only and heated at 95°C for 7 minutes.
The gel was electrophoresed at a hand-warm temperature (approximately 45°C) instead of the usual cold temperature.
Autoradiographic exposure was for one day with an intensifying screen. Sizes are marked in bases.


A
1 2
1400
800
600
500
400
300
200
B
1 2 3
1400
800
600
500
400
300
200
C
1 2 3
1400
800
600
500
400
300
200

significant foldback potentials for the probe RNAs used.

However, the hybridisation temperature experiment suggested that any probe-probe interaction was not occurring at the hybridisation stage. It therefore seemed possible that any secondary structure could be produced after the RNase A cleavage step and may simply be a problem of insufficient denaturation prior to electrophoresis. To increase the likelihood of obtaining sufficient denaturation conditions, an experiment was performed with the TT693 probe which involved treating one sample as normal, dissolving the final 'protected' RNA pellet of another directly in a new tracking dye, essentially $^{1}/_{3}^{rd}$ old dye (sequencing dye) $^{2}/_{3}^{rds}$ formamide (15ml deionised formamide, 5mg xylene cyanol, 5mg bromophenol blue, 0.1ml 0.5M EDTA pH 8.0, 0.1% SDS) and heated for 7 minutes at 85°C before loading instead of the usual 5 minutes. Another sample was dissolved directly in the new dye and heated at 95°C for 7 minutes.

Instead of the gel being cooled to allow faster electrophoresis and reduce 'smiling' of the bands, the gel was pre-run to warm-up and then warm water circulated to keep the gel at a constant hand-warm temperature (Figure 4.8). The sample handled 'traditionally' was improved; the fuzzy band was reduced in intensity and the upper band increased, presumably due to the increased temperature of the gel. The two samples directly resuspended in the new dyes showed no sign of the fuzzy band and looked fine. Thus the problem was due to insufficient denaturation before loading and could be cured by dissolving the 'protected' RNA directly in tracking dye (essentially 100% formamide plus dye), heating at 85°C for 7 minutes and running the gel at a warm temperarure.

4.6 RNase A mapping of a known OI point mutation

Having solved the problem of helical region derived probes, it was then necessary to use the technique to detect a point mutation in the α1(I) collagen gene. Without being able to detect such a known mutation, when screening for unknown mutations it would never be known if the RNase A cleavage conditions were near enough to the optimum to be able to detect any base substitutions at all.

The positive control chosen was the type II OI cell-line (A.K.) from which Cohn et al., (1986) had characterised the mutation as being a T for G substitution at the first base position of codon 988 of the helical domain. In the attempt to detect this mutation, two probes were used that spanned the mutation: BX887, and SS534 (see Figure 4.9). The methodology was as above including RNase A treatment at 1.7μg/ml for 30 minutes. Visible cleavage products could be seen on the autoradiograph, but were difficult to distinguish from the background of degradation products. The experiment was repeated but with increasing the RNase A concentration to 3.5μg/ml and incubating for 45 minutes. This made the cleavage products more easily visible, but not dramatically so. The experiment was repeated again, this time increasing the RNase A concentration by a factor of 10 to 35μg/ml and incubating for 30 minutes. This dramatically improved the image and made the cleavage products visible at a glance (Figure 4.9). The SS534 probe shows the mutation most clearly; the background is very low and the expected bands of 304 and 230 bases are exactly what is seen. The BX887 cleavage products are not so obvious. The upper cleavage product of 694 bases lies close to a degradation product seen in both the control sample and the mutant sample. The lower product of 193 bases is perhaps more obvious, but is seen as a doublet. It is likely that this is due to 'nibbling' of the cleaved ends of the heteroduplex by the RNase A. This

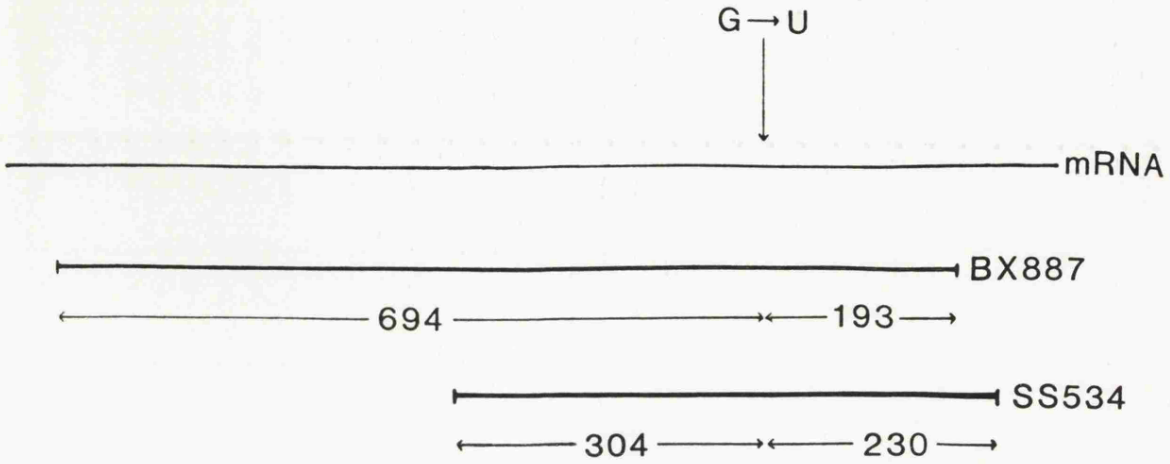
Figure 4.9

RNase A cleavage of a known point mutation.

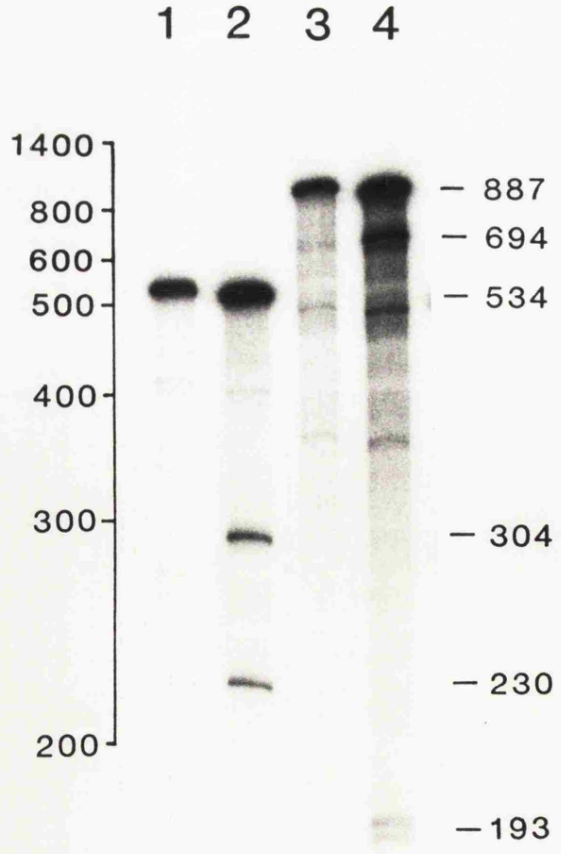
A. Map showing the positions of the A.K. mutation (Cohn et al., 1986) relative to the probes BX887 and SS534. The mutation is a G to T substitution in the codon encoding amino acid 988 of the α1(I) gene.
The numbers show the expected sizes of any RNase A cleavage products.

B. Autoradiograph showing RNase A cleavage of the A.K. α1(I) collagen mRNA.
Tracks 1 and 2 contain normal placental RNA and A.K. RNA respectively, protected by probe SS534.
Tracks 3 and 4 contain normal placental RNA and A.K. RNA respectively, protected by probe BX887.
160μg placental total RNA and 50μg A.K. fibroblast total RNA were used for the hybridisations.
The tracks containing the A.K. protected RNA (2 and 4) show the RNase A cleavage fragments predicted.
Autoradiographic exposure was for three days without an intensifying screen.
Sizes are marked in bases.

A


G → U
mRNA
BX887
694
193
SS534
304
230

B


1 2 3 4
1400
800
600
500
400
300
200
887
694
534
304
230
193

may be a common occurrence in all cleavage products, but is more visible here because of the small size of the fragment allowing the small size variation to be resolvable. Laser scanning densitometry of the Figure 4.9 autoradiograph gave cleavage efficiencies of 38% for SS534 and 58% for BX887. These differences in cleavage efficiency may represent the inaccuracy of the analysis or a difference in cleavage conditions between the two samples.

The success of this experiment proved that RNase A mapping was a viable approach to the detection and mapping of point mutations in OI.

4.7 Discussion

Because of fears of the technical difficulty of RNase A mapping, it was decided to build the technique up in steps and be sure each step was working before going onto the next.

The experiments to produce unlabelled and then labelled in vitro synthesised RNA transcripts met with no problems. The successful protection of a synthetic RNA showed that there were no fundamental problems with the conditions used. Following the successful protection of α1(I) mRNA with probe XE677, the labelling and hybridisation conditions were modified, reducing the amount of input label and increasing the amount of probe in the hybridisation. By making the labelling reagents up as a single solution, the volume of the reaction, and therefore the amount of input label, could be dropped still further. The successful protection of α1(I) mRNA with the XE677 probe, before the problem with helical region derived probes, was very informative in that it demonsrated that the problem was sequence-specific and therefore was a subtle problem and not a fundamental flaw in the system.

Because the problem was sequence-specific it was thought that it

was due to secondary structures in the RNA:probe complex or purely in the probe. Computer analysis of secondary structure potential revealed no significant differences between the XE677 probe and the helical region derived probes, suggesting that the probe was not hybridising to itself. The problem was determined to be due to insufficient denaturation of the RNA:RNA heteroduplex prior to electrophoresis. The GC content of the BX887 probe was seen to be 67% and that of the probe XE677 to be 60%. Because the GC content of BX887 is higher than XE677, it will require stronger denaturation conditions. But the difference between 60% and 67% does not seem a striking one. It is surprising that a GC content difference of only 7% should have had such a dramatic effect upon the behaviour of the two heteroduplexes.

It was necessary, prior to using the technique as a screening method for mutations in the α1(I) collagen gene, to optimise RNase A cleavage conditions. To do this it was either necessary to obtain cleavage of a known point mutation in a type I collagen mRNA or obtain cleavage of an *in vitro* mutagenised synthetic RNA diluted down to 50% of the α1(I) mRNA level in a control RNA sample. Fortunately successful cleavage of the A.K. mutant RNA rendered the generation of *in vitro* mutagenised point mutations for the latter approach unnecessary.

The mutation in this individual was a G to T transversion in the coding strand of the DNA. Thus in the RNA:RNA hybridisation the probe/target mismatch would be C/U. In a study of RNase A cleavage between different types of probe/target mismatches in RNA:DNA heteroduplexes (Myers *et al.*, 1985d), three different C/U mismatches were analysed, each giving only a 5% efficiency of cleavage. However, in a study of mutations in the Influenza virus by RNase A cleavage (of RNA:RNA heteroduplexes) (Lopez-Galindez *et al.*, 1988), 3 different C/U mismatches analysed were cleaved to a 'significant extent'. From this

and the rest of the data presented in these two studies, it would seem that a C/U lies roughly half-way between the easiest type and the most difficult type of mismatch to detect. Thus as a positive control, this C/U mismatch is more informative than, for example, a C/C mismatch which has given efficient cleavage in all examples studied, (Myers et al., 1985d, Lopez-Galindez et al., 1988).

The RNase A concentration to obtain cleavage of the AK mutation was increased to a level that gave significant cleavage (Fig.4.9). Because of a lack of further controls, these cleavage conditions were adopted for the mapping of mutations in the α1(I) gene in a collection of type II OI cell-lines (See Chapter 5).

CHAPER 5

SCREENING FOR MUTATIONS IN TYPE II OI BY RNase A MAPPING

5.1 Introduction

With a tight time limit imposed on this project, it was decided to restrict the search for mutations in OI to the lethal (type II) variants. For reasons described in section 1.4.3, it was considered that for any given type II OI individual, the mutation was most likely to lie in the α1(I) collagen gene. In addition, because of the hypothesised 'gradient effect', a lethal mutation was considered most likely to lie in the 3' end of the helical-encoding region of the gene.

Thus the type of OI for which the site of the mutation could best be predicted was type II. The site of the mutation most likely being towards the 3' end of the helical-encoding region of the α1(I) collagen gene. A bonus of this 'educated guess' was that the longest continuous length of type I pro-collagen cDNA available at the time, covered this region. One disadvantage of this strategy was the paucity of type II OI cell-lines, which are necessary as the source of RNA. This being due to their low incidence combined with the rarity of post-mortem skin biopsies taken except in large hospitals where clinicians often realise the potential usefulness of such samples.

A total of 11 fibroblast cell-lines from type II OI babies were obtained, with initial passage numbers of between 4 and 10. These were each grown for several passages and the total cellular RNA harvested. RNA from these cell lines was checked for sizable deletions or insertions by Northern blotting, using the α1(I) cDNA clone Hf404 as a radiolabelled probe. All the α1(I) mRNAs appeared to be of the same size and therefore any rearrangements in the coding sequence of COL1A1

were likely to be fine rather than gross.

5.2 Subcloning of overlapping cDNA's

As a result of the high specific activity of the probes used for RNase A mapping, some degradation by radiolysis will occur between the labelling and the subsequent electrophoresis of the 'protected' probe fragment. This, and the probable presence of 'RNase sensitive sites' in the probe/RNA heteroduplexes, is what causes the background seen on the autoradiographs. The larger the probe used the greater this background becomes (R. Gibbs, personal communication). As a consequence of this and the resolving range of denaturing polyacrylamide gels, the upper limit to the size of the probes used in this study was 1000 bases. A lower limit for the size of the probe also exists: probes smaller than around 350 bases may dissociate from their target RNA during the elution of the probe/mRNA heteroduplex from the 'mAP' paper (R. Gibbs, personal communication).

In this study it was intended to cover the entire region of the pro-α1(I)mRNA covered by the cDNA clones Hf404 and Hf677 (Chu et al., 1982). These two overlapping clones correspond to approximately two-thirds of the helical region and the entire C-terminus (see Figure 5.1).

It was felt that the subclones used for this study should overlap as any cleavage site approximately 50 bases or less from the end of the probe, would produce a large fragment indistinguishable from the probe and a small fragment which would be run off the gel and probably too faint to see anyway.

Initially six overlapping subclones were constructed for the generation of pro-α1(I) antisense RNA probes (see Figure 5.1 and Table 5.1). To enable the construction of subclones which span the overlap of

TABLE 5.1

In Vitro Transcription Template Details

Clone	Origin	Insert ends	Vector	Vector ends	Linearised by
EP934	Hf404	*EcoRI/PvuII*	pTZ19	*EcoRI/SmaI*	*EcoRI*
AB720	Hf404	*EcoRI/BamHI*	pTZ19	*EcoRI/BamHI*	*AvaI*
SE723	Hf404	*SstI/EcoRI*	pTZ18	*SstI/EcoRI*	*HindIII*
BX887	JR1081	*BamHI/XhoI*	pTZ19	*BamHI/SalI*	*BamHI*
TT693	JR1081	*ThaI*	pTZ18	*SmaI*	*BamHI*
BE574	Hf677	*XhoI/EcoRI* (partial)	pTZ18	*SalI/EcoRI*	*BalI*
EH473	Hf404	*EcoRI/HinfI*	pTZ19	*EcoRI/HincII*	*EcoRI*
XP794	EP934	*EcoRI/PvuII*	pTZ19	*EcoRI/SmaI*	*XhoI*
RS355	JR1081	*RsaI/Sau3A*	pTZ19	*SmaI/BamHI*	*EcoRI*
SS534	JR1081	*Sau3A*	pTZ18	*BamHI*	*HindIII*
XE677	Hf677	*XhoI/EcoRI*	pTZ18	*SalI/EcoRI*	*HindIII*
EX793	Hf677	*EcoRI/XhoI*	pTZ18	*EcoRI/SalI*	*HindIII*
EB1536	Hf404	*EcoRI/BamHI*	pTZ19	*EcoRI/BamHI*	*EcoRI*

Figure 5.1

RNA probe map.

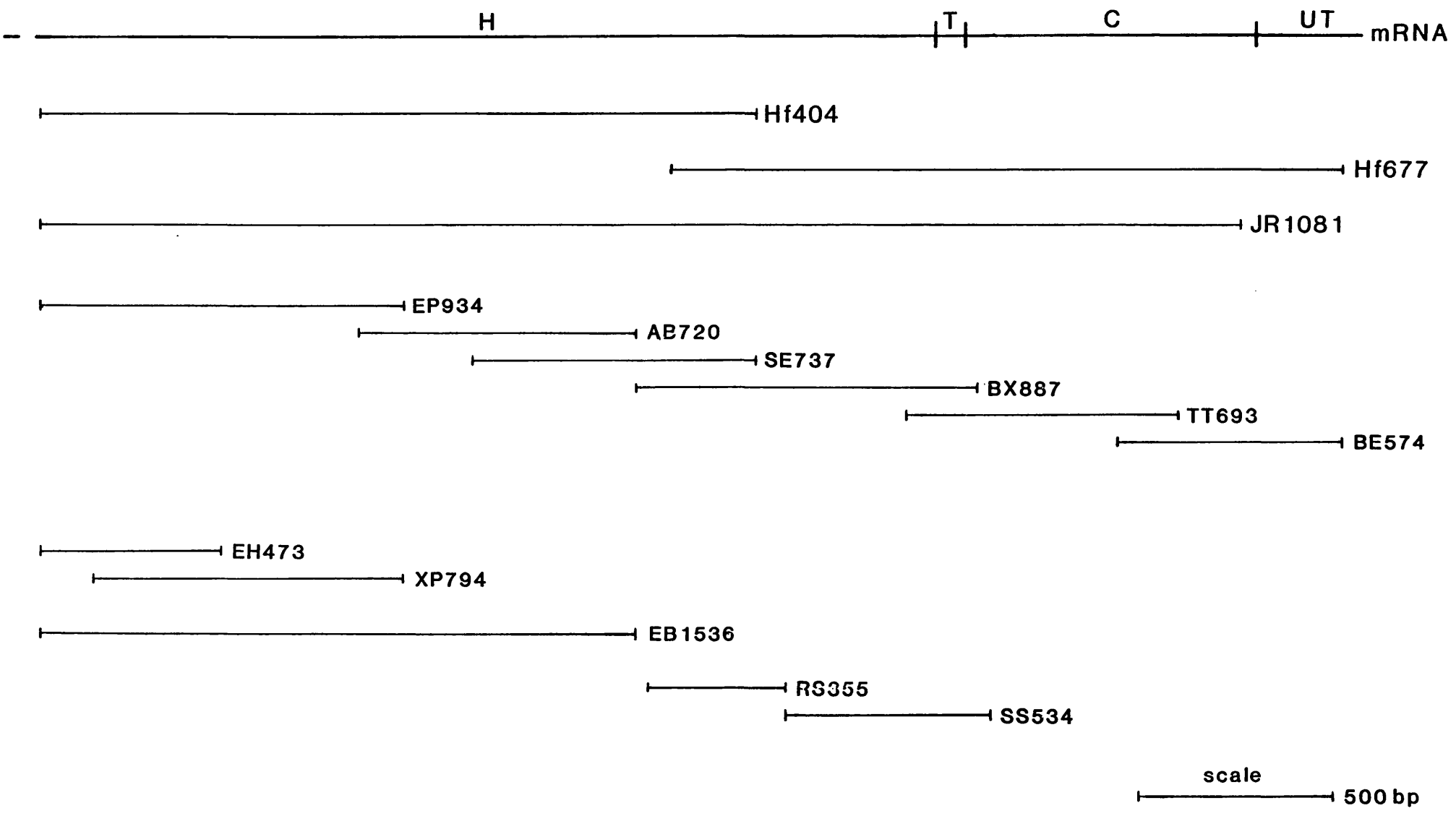
Diagram showing the relative positions of all the antisense RNA probes described in this chapter.

Marked on the mRNA are regions corresponding to protein domains: H = helical domain; T = Carboxyl telopeptide; C = Carboxyl propeptide; UT = 3' untranslated region.

Hf404, Hf677 and JR1081 are cDNA clones from which the probes were derived.

Probes EP934, AB720, SE737, BX887, TT693 and BE574 were designed to span the entire region of the α1(I) mRNA covered by the cDNA clones Hf404 and Hf677.

Probes EH473,XP794, EB1536, RS355 and SS534 were used to investigate putative abnormal RNase A cleavage patterns obtained with the above probes.


H
T
C
UT
mRNA
Hf404
Hf677
JR1081
EP934
AB720
SE737
BX887
TT693
BE574
EH473
XP794
EB1536
RS355
SS534
scale
500 bp

Hf404 and Hf677, the two clones were spliced together using a unique NcoI site in the overlap region and subcloned into the EcoRI site of pUC13. This clone was called pJR1081.1. Because of an internal EcoRI in the Hf677 cDNA, the 3' approximately 250bp were lost, making the JR1081 insert 3090bp. The insert was then subcloned into the EcoRI site of pBR328. This subclone was called pJR1081.2. The transfer of the insert to pBR328 gave a larger size difference between insert and vector. This facilitated the isolation of the insert, enabling the isolation and subcloning of fragments using 4-cutter type restriction endonucleases without the confusion and complication of co-migrating vector DNA.

5.3 RNase A mapping of mutant cell-lines

The factor limiting the number of samples that could be handled at once, was the mRNA purification step. The time taken just to apply each sample to the 'mAP' paper was usually 15-20 minutes. The pipetting action adopted in this step becomes very uncomfortable when the position has been held for more than an hour, hence the maximum number of samples that could tolerably be handled in one session was considered to be four; and the comfortable maximum three.

So, in general, for each probe labelling reaction, aliquots of the probe were hybridised to 3 different mutant RNA samples. The first 3 samples to be screened were GM6260, GM5747 and GM2328 (see section 2.1), (hereafter these samples are called 6260, 5747 and 2328). These RNAs were hybridised to the TT693 probe, covering the helical domain / C-propeptide boundary and the 4 helical domain probes: BX887, SE737, AB720 and EP934 . Probes TT693, BX887 and SE737 gave no unusual bands (Figure 5.2). All autoradiographs were given long exposures to bring out any faint bands. Thus the background with all probes seems high, but the background pattern does not differ between samples.

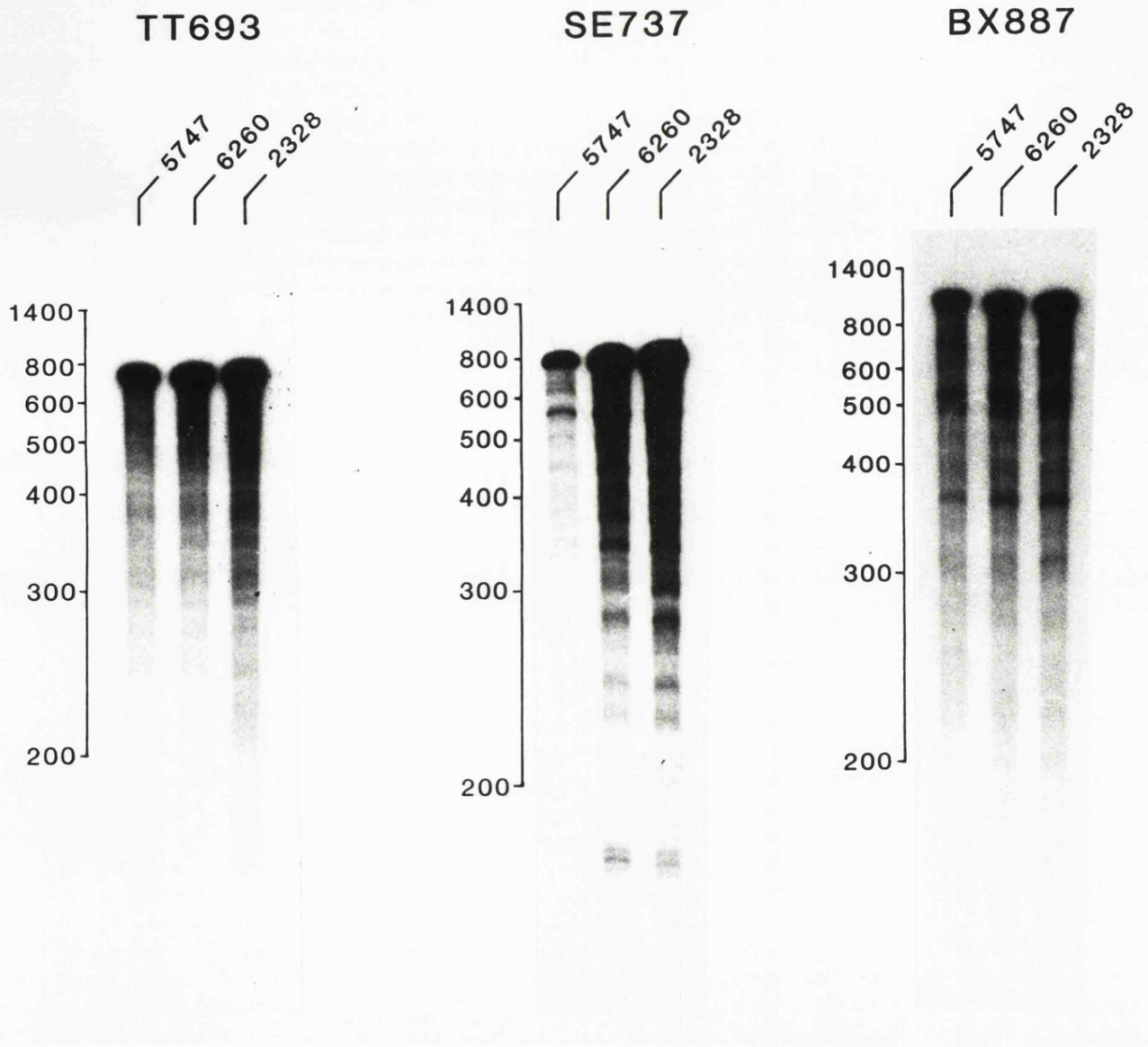
Figure 5.2

RNase A protection of type II OI samples with three different probes.

Autoradiographs of RNase A protection gels with RNA samples 5747, 6260 and 2328, hybridised to the antisense RNA probes TT693, SE737 and BX887. All procedures were as described in section 2.17.

Autoradiographic exposures were for one day at -80°C.

Sizes are marked in bases.


TT693
SE737
BX887
5747
6260
2328
1400
800
600
500
400
300
200

However with probe EP934, faint bands at approximately 340 and 619 bases are seen in sample 6260 and not 2328 (in this hybridisation the quality of the 5747 RNA sample was too poor for comparason) (Figure 5.3). If these two bands, being as faint as they are, had not approximately added up to the size of the probe (934 bases), they would have been ignored. As they did add up, they could represent very inefficient cleavage of a mismatch. As any mismatch could lie in either of two possible positions, a new probe was constructed called XP794 (see Figure 5.1 and Table 5.1). The XP794 probe covers both orientations: if the mismatch is in the right-hand position, the expected bands would be approximately 619 and 175 bases; in the left-hand position, approximately 454 and 340 bases (Figure 5.3).

A faint band <u>is</u> present at approximately 175 bases (sized to 160 bases, Figure 5.3), but no band is visible around 619 bases. In addition, two bands are present in the 2328 sample which are absent from 6260. These bands are approximately 155 and 195 bases in size and therefore make little sense.

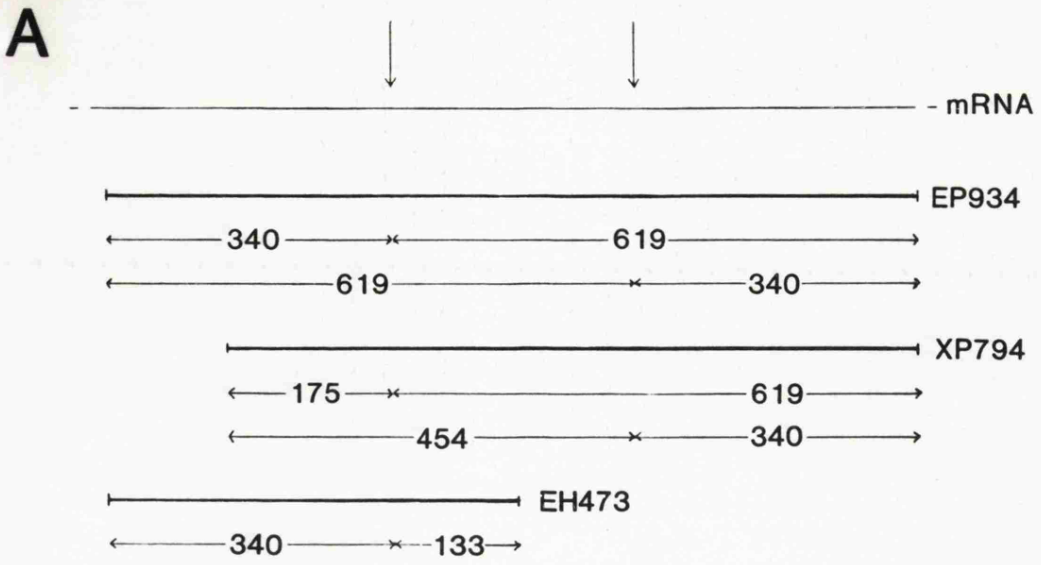
If the band at 175 bases in 6260 was the result of a mismatch cleavage, the upper band of approximately 619 bases may be masked by a co-migrating degradation product. A new probe, EH473, (see Figure 5.1 and Table 5.1) which overlaps the possible cleavage site should, if the mismatch is real, confirm it as being so. In the event of a genuine mismatch, the EH470 probe should reveal a band of approximately 340 bases. No such band is obvious in the 6260 sample (Figure 5.3). However there is a band in this region which is present in both samples, but is perhaps comparatively more intense in 6260 than in 2328. Despite the possibility remaining that a mismatch could be present which is cleaved at very low efficiency, the cleavage was not convincing enough to justify the risk of pursuing the putative variant and finding no

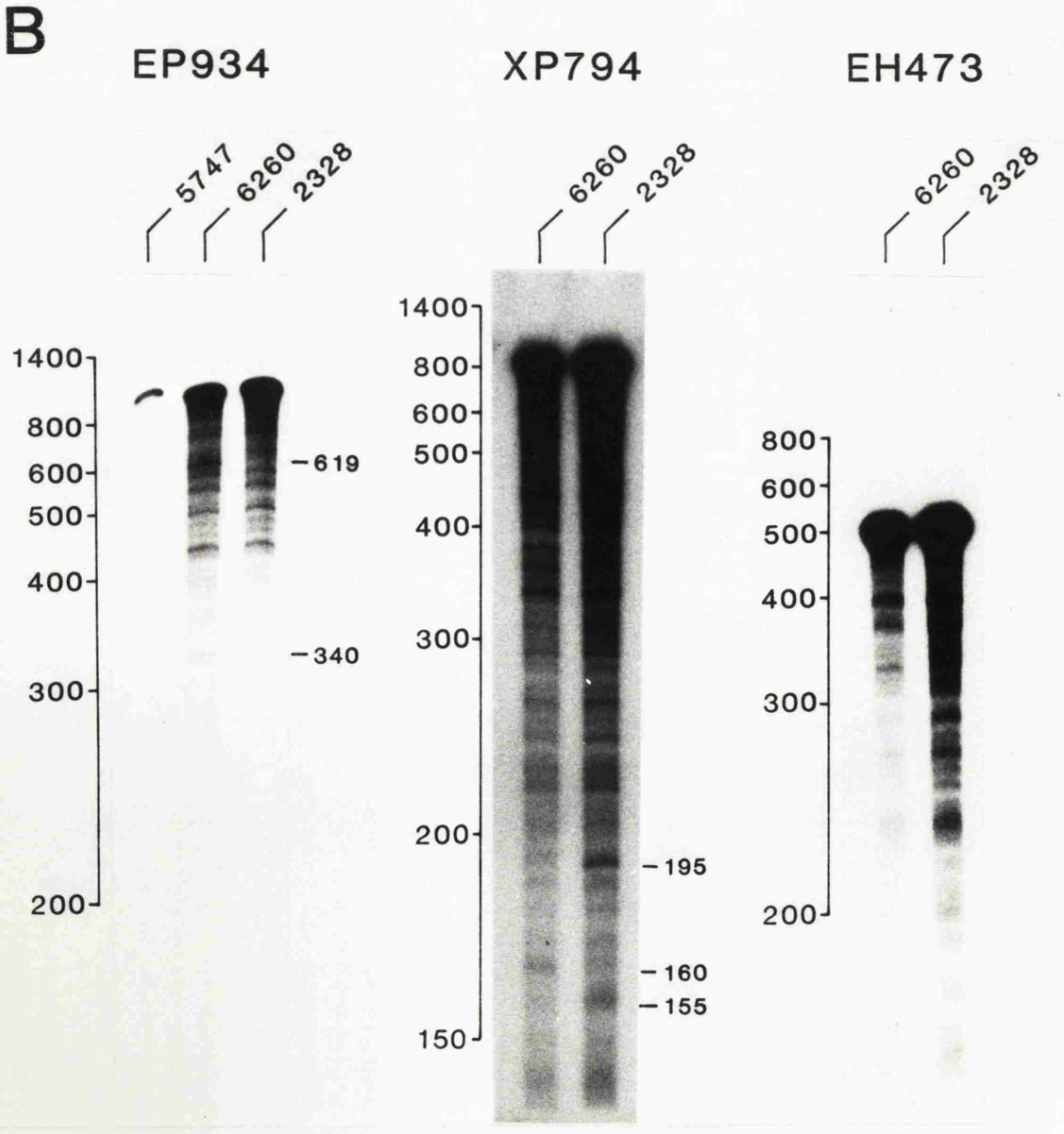
Figure 5.3

Investigation of an abnormal RNase A cleavage pattern obtained with probe EP934.

A. Diagram showing the relative positions of antisense RNA probes EP934, XP794 and EH473.
The vertical arrows indicate the two possible cleavage sites as identified by probe EP934 with the RNA sample 6260. The numbers indicate the sizes, in bases, of cleavage products expected with the different probes depending upon the orientation of the putative cleavage site.

B. Autoradiographs of RNase A protection gels with RNA samples 5747, 6260 and 2328 hybridised to antisense RNA probes EP934, XP794 and EH473.
Probe EP934 produced bands of approximately 619 and 340 bases with sample 6260, but not samples 5747 and 2328.
Probe XP794 produced a band of approximately 160 bases in sample 6260 and bands of approximately 195 and 155 bases in sample 2328.
Probe EH473 produced no visible bands that are not present in both samples 6260 and 2328.
Autoradiographic exposures were for four days at room temperature for probes EP934 and XP794 and overnight at room temperature for probe EH473.
Sizes are marked in bases.


A
mRNA
EP934
340
619
619
340
XP794
175
619
454
340
EH473
340
133


B
EP934
XP794
EH473
5747
6260
2328
6260
2328
6260
2328
1400
800
600
500
400
300
200
619
340
1400
800
600
500
400
300
200
150
195
160
155
800
600
500
400
300
200

abnormal DNA sequence.

The remaining helical-region probe to be used on these three cell-lines was AB720 (see Figure 5.4). This probe produced, like the others, a background smear with faint discrete bands present, but the pattern identical in all samples. But this probe however, in the background smear produced 2 bands of unusually high intensity in all samples. Because of their intensity these bands were sized, giving approximate sizes of 240 and 480 bases, which add up to the size of the probe (720 bases). As these bands were present in all samples, it was thought that they could be due to one of 4 things:

1. Co-incidentally sized degradation products.
2. Base polymorphism.
3. cDNA Cloning artefact.
4. A site particularly sensitive to RNase A.

To investigate whether a particular site was being cleaved, a new probe, EB1536 (see Figure 5.1 and Table 5.1), was used to determine the orientation. In the 'left-hand' position this probe should give bands of approximately 1056 and 480 bases and in the 'right-hand' position should give bands of 1296 and 240 bases (Figure 5.4). On the 'long-run', no band of 420 bases could be seen, but a band of approximately 1300 bases was present (Figure 5.4). On the 'short-run', a band of approximately 245 bases was present (Figure 5.4). On this run the upper band was less obvious being near the resolution limit of the gel. Thus the EB1536 probe, backs up the possibility, but not unequivocally so, that a cleavage site may be present in the 'right-hand' position.

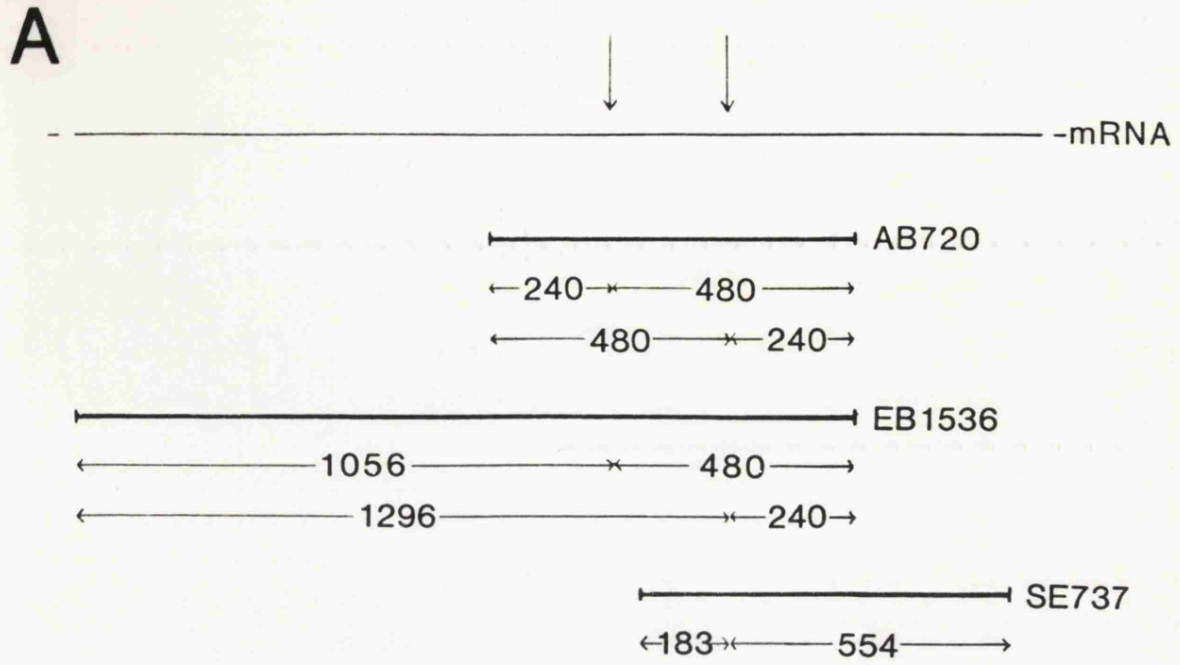
The 'right-hand' position was overlapped by the previously used SE737 probe (Figure 5.2). A prominent band had been noticed on this autoradiograph at the time, but with no obvious 'partner' band, this

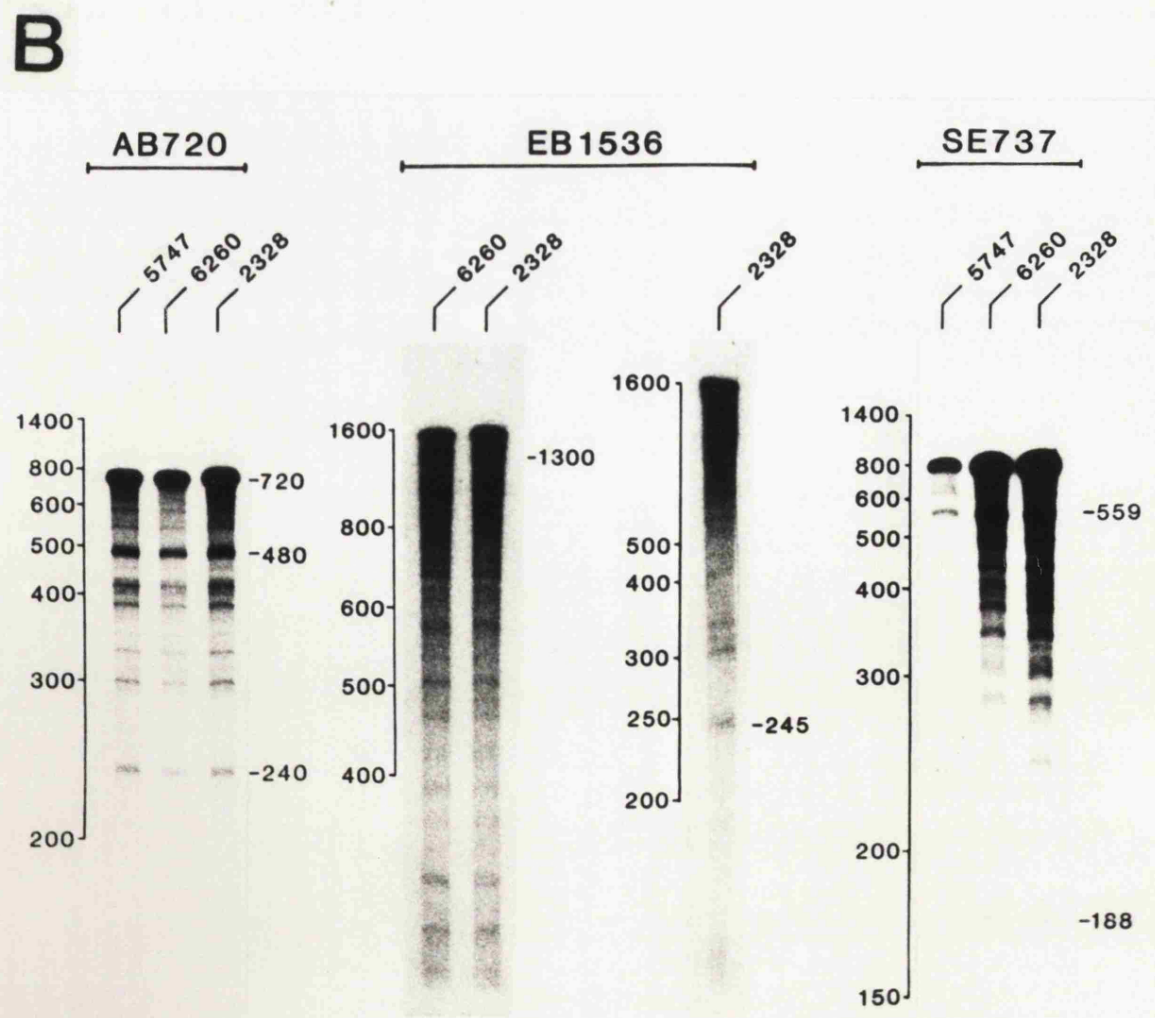
Figure 5.4

Investigation of an abnormal RNase A cleavage pattern obtained with probe AB720.

A. Diagram showing the relative positions of antisense RNA probes AB720, EB1536 and SE737.
The vertical arrows indicate the two possible positions of a cleavage site as identified by probe AB720. The numbers indicate the sizes, in bases, of cleavage products expected with the different probes depending upon the orientation of the cleavage site.

B. Autoradiographs of RNase A protection gels with RNA samples 5747, 6260 and 2328 with antisense RNA probes AB720, EB1536 and SE737.
Probe AB720 produces additional bands of approximately 480 and 240 bases in all three samples.
Probe EB1536 produces additional bands of approximately 1300 and 245 bases, mapping the putative cleavage site to the right-hand position (see above).
Probe SE737 confirms the putative cleavage site as being in the right hand position by producing bands of approximately 559 and 188 bases.
Autoradiographic exposures were for one day at room temperature.
Sizes are marked in bases.


A
-mRNA
AB720
240
480
480
240
EB1536
1056
480
1296
240
SE737
183
554


B
AB720
EB1536
SE737
5747
6260
2328
6260
2328
2328
5747
6260
2328
1400
800
600
500
400
300
200
-720
-480
-240
1600
800
600
500
400
-1300
1600
500
400
300
250
200
-245
1400
800
600
500
400
300
200
150
-559
-188

was passed by. For a cleavage site in the 'right-hand' position, probe SE737 should have given bands of approximately 554 and 183 bases. The prominent band on this autoradiograph was sized to 559 bases. A possible heterogeneous 'partner' band was sized to 188 bases (Figure 5.4). With the evidence of these three probes together, it would appear that cleavage is occurring by one cause or another at this site.

Although interesting, it was decided to only investigate this phenomenon further, if after screening the remainder of the type II OI cell-lines, no mutation had been found.

5.4 RNase A mapping of a mutation in type II OI

In the next step of the strategy, RNA samples from the type II OI cell-lines F.W., P.E., F.B. and M.G. (see section 2.1) were to be screened for mutations. The first probe to be used was BX887, as it was considered the most likely probe to reveal a mutation. In one of the samples, F.B., this probe revealed a variant with clarity beyond any doubt at all (Figure 5.5). The upper band was sized to 545 bases and the lower to 335 bases, thus the 2 fragments add up to the size of the probe (887 bases). The lower band is seen as a doublet, but this was not a source of concern as this phenomenon had been seen before in the positive control cell-line A.K. (see section 4.5)

The experiment was repeated using the F.B. RNA and a control sample, and at the same time the SS534 probe was used in an attempt to orientate the position of the variant. The BX887 probe repeated the result, but the SS534 probe failed to detect it (Figure 5.6). If the variant lay in the 'right-hand' position, the SS534 probe should have been cleaved into fragments of approximately 372 and 162 bases .

To positively check that the variant did lie in the 'left-hand' position, probe RS355 was used. This probe should then be cleaved into

Figure 5.5

A clearly abnormal RNase cleavage pattern.

Autoradiograph of an RNase A protection gel with type II OI RNA samples F.W., P.E., F.B. and M.G. hybridised to antisense RNA probe BX887. RNA sample F.B. produced cleavage products of approximately 545 and 335 bases not seen with the other samples. Autoradiographic exposure was for 5 hours at -80°C.
Sizes are marked in bases.

BX887

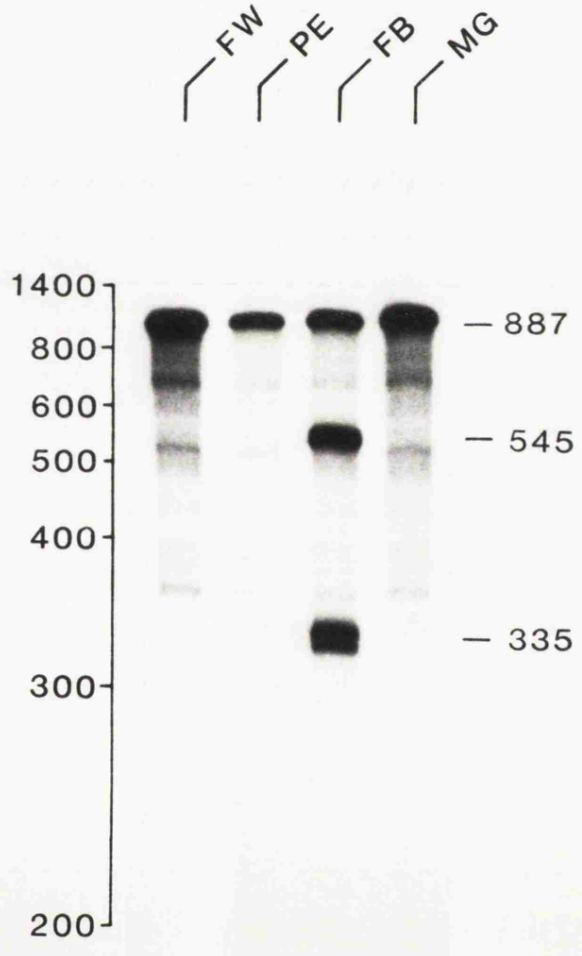
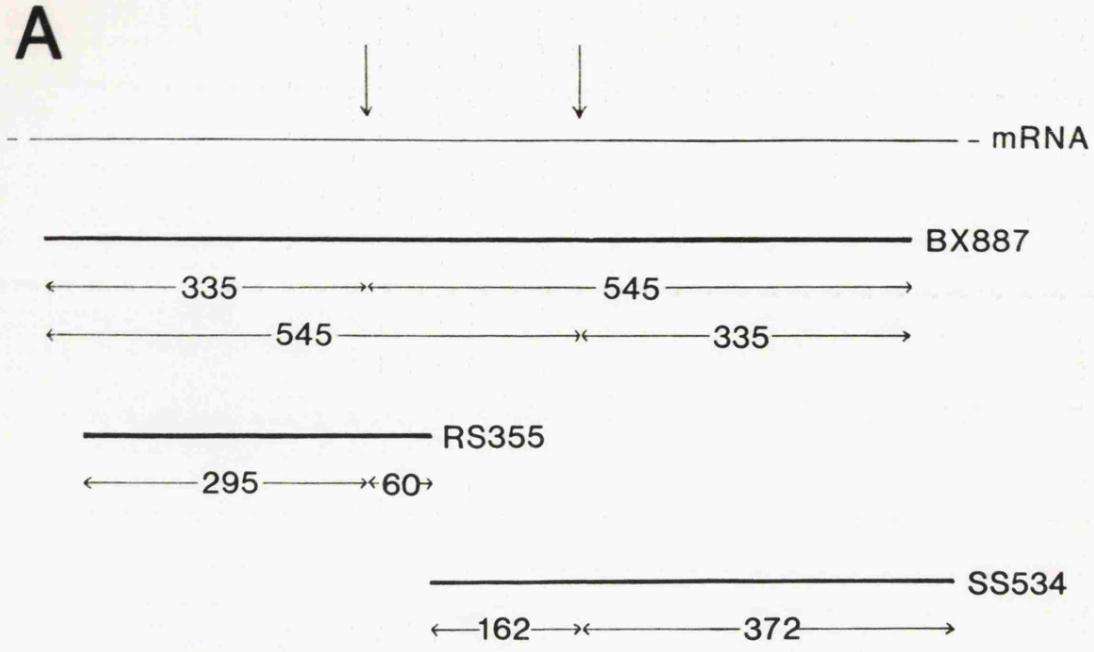


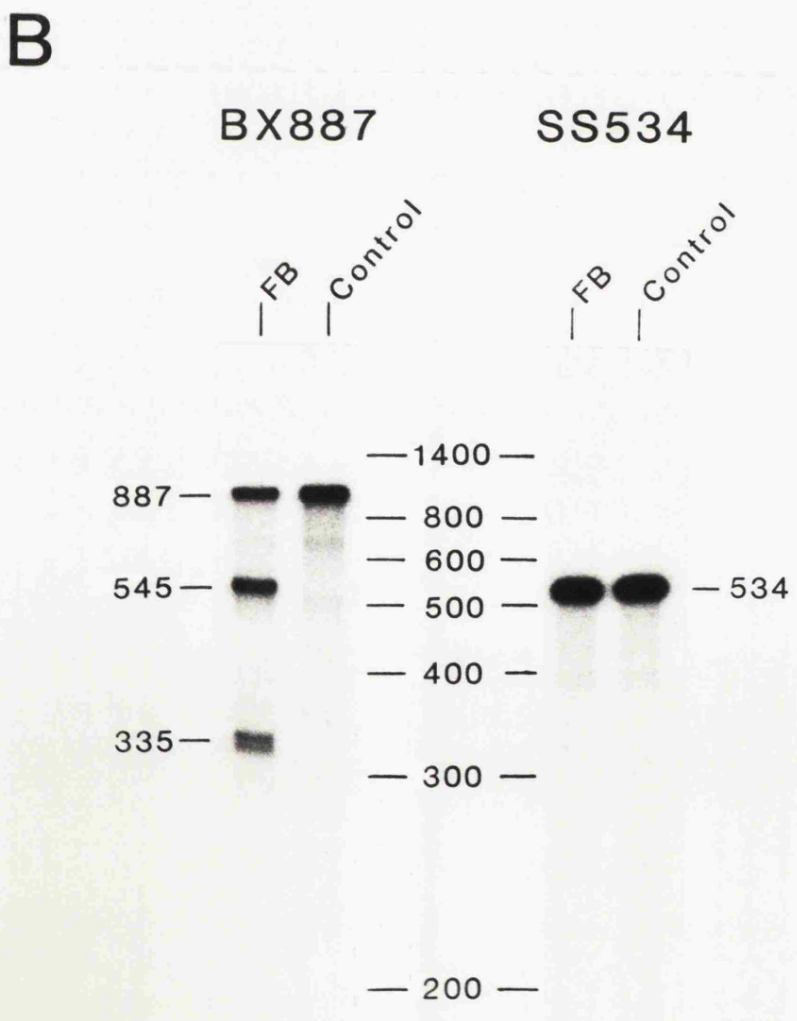
Figure 5.6

RNase A mapping of a putative mutation in sample F.B.

A. Diagram showing the relative positions of antisense RNA probes BX887, RS355 and SS534. The vertical arrows indicate the possible positions of the RNase A cleavage site identified by probe BX887 in RNA sample F.B.
The numbers indicate the sizes of the cleavage products observed with probe BX887 and expected with probes RS355 and SS534 if the cleavage site is in the left-hand or right-hand orientation respectively.

B. Autoradiographs of RNase A protection gels with the RNA sample F.B. and a control RNA sample hybridised to the antiense RNA probes BX887 and SS534.
As before, sample F.B. produced additional cleavage bands of approximately 545 and 335 bases, whereas probe SS534 produced the full-length protected band only. This result provisionally maps the cleavage site to the left-hand position.
Autoradiographic exposure was for one day at room temperature.
Sizes are marked in bases.


A
mRNA
BX887
335
545
545
335
RS355
295
60
SS534
162
372


B
BX887
SS534
FB
Control
FB
Control
1400
887
800
600
545
500
534
400
335
300
200

fragments of approximately 295 and 60 bases (Figure 5.6). As this probe was small and the smaller expected cleavage product only about 60 bases long and therefore likely to be difficult to see, the F.B. RNA and a control sample when mapped with this probe were run on a 40cm long sequencing type gel (using the usual RNase A mapping gel-mix). This length gel was used in an attempt to finely map the position of the variant by calculating the map distance between the wild-type probe band and the upper cleavage product. In order to obtain as accurrate a map as possible, a DNA sequence was run alongside the RNA samples as a size ladder. Two sets of samples were electrophoresed, one for a much longer period than the other (see Figure 5.7 for long run and Figure 5.8 for short run).

In Figure 5.7 the probe band is seen to be heterogeneous and the upper cleavage product seen as a cluster of at least 6 bands varying in size by 25 bases. The middle of this cluster was calculated, using the DNA sequence, as being approximately 65 bases from the lower edge of the wild-type band. The heterogeneous appearance of the bands was entirely unexpected and, at the time, unexplainable. The presence however, of a band cluster approximately 65 bases smaller than the probe was consistent with what was expected for a positive mapping of the variant in the 'left-hand' orientation. The bands on the 'short run' (Figure 5.8) look less heterogeneous indicating that the unusual heterogeneity of both the wild-type and the upper cleavage product seen in Figure 5.7, was possibly a consequence of the long distance electrophoresis rather than a consequence of unusual cleavage at this particular site. It was thus considered that there was sufficient data in favour of the presence of a variant in the F.B. cell-line to justify characterising the region of cDNA or genomic DNA corresponding to the position of the variant.

Figure 5.7

Positive mapping of a putative mutation in sample F.B.

'The long-run.'

Autoradiograph of a 40cm-long 6% polyacrylamide RNase protection gel with the RNA sample F.B. and a control RNA sample hybridised to antisense RNA probe RS355.
These samples were electrophoresed for $5^{1}/_{2}$ hours.
On the left is an aliquot of the probe.
On the right is a sequencing ladder (Marked S.L.) which serves as an elaborate size marker.
The F.B. RNA sample produced a heterogeous cleavage product consisting of at least six discrete bands varying in size by approximately 25 bases. The middle of the cleavage product is approximately 65 bases smaller than the fully protected probe band.
The cleavage product represents the upper cleavage product, predicted as being 295 bases in length (Figure 5.6). The lower cleavage product was run off the gel.
Autoradiographic exposure was for one and a half days at room temperature.
Size differences are marked in bases.

RS355

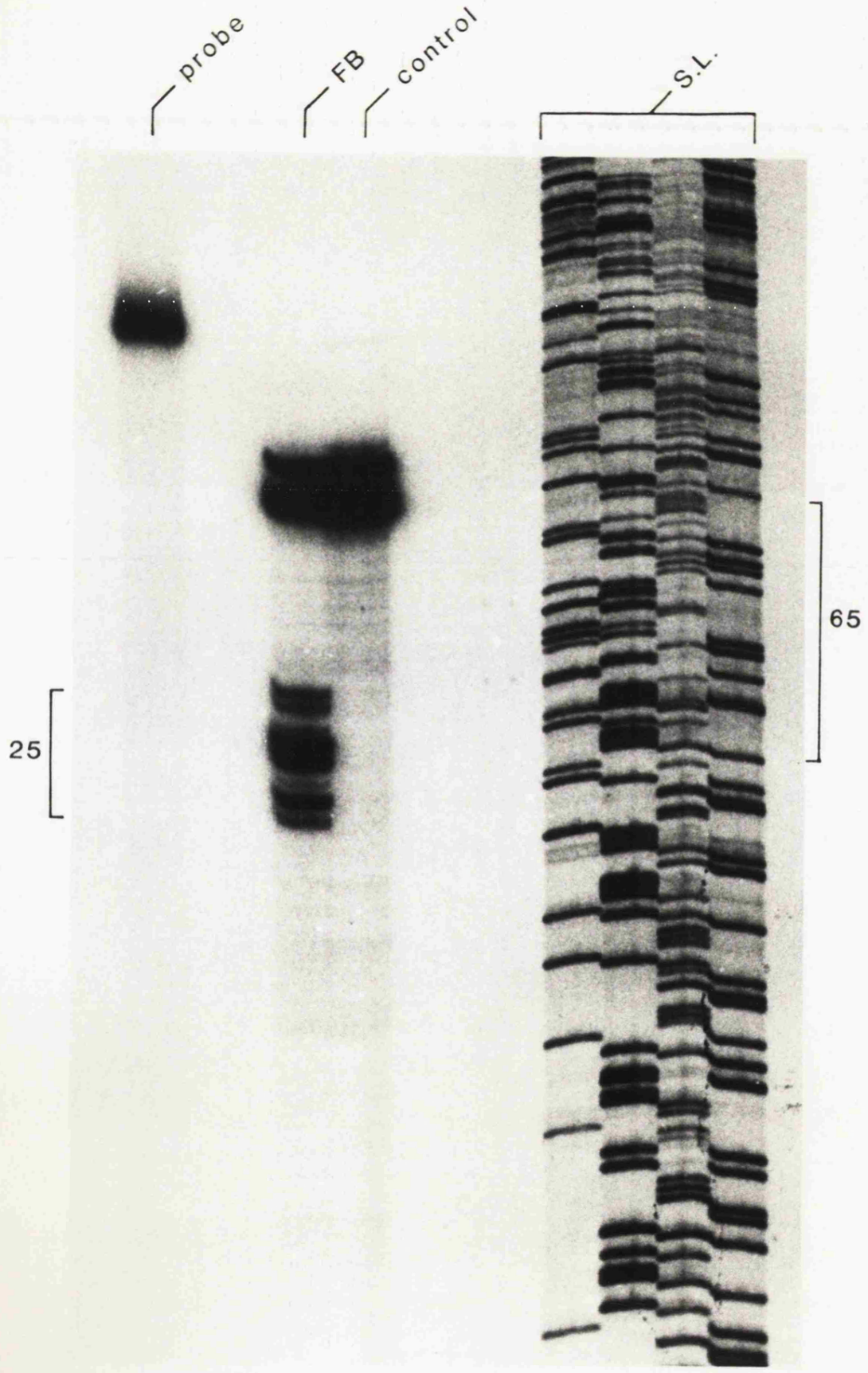
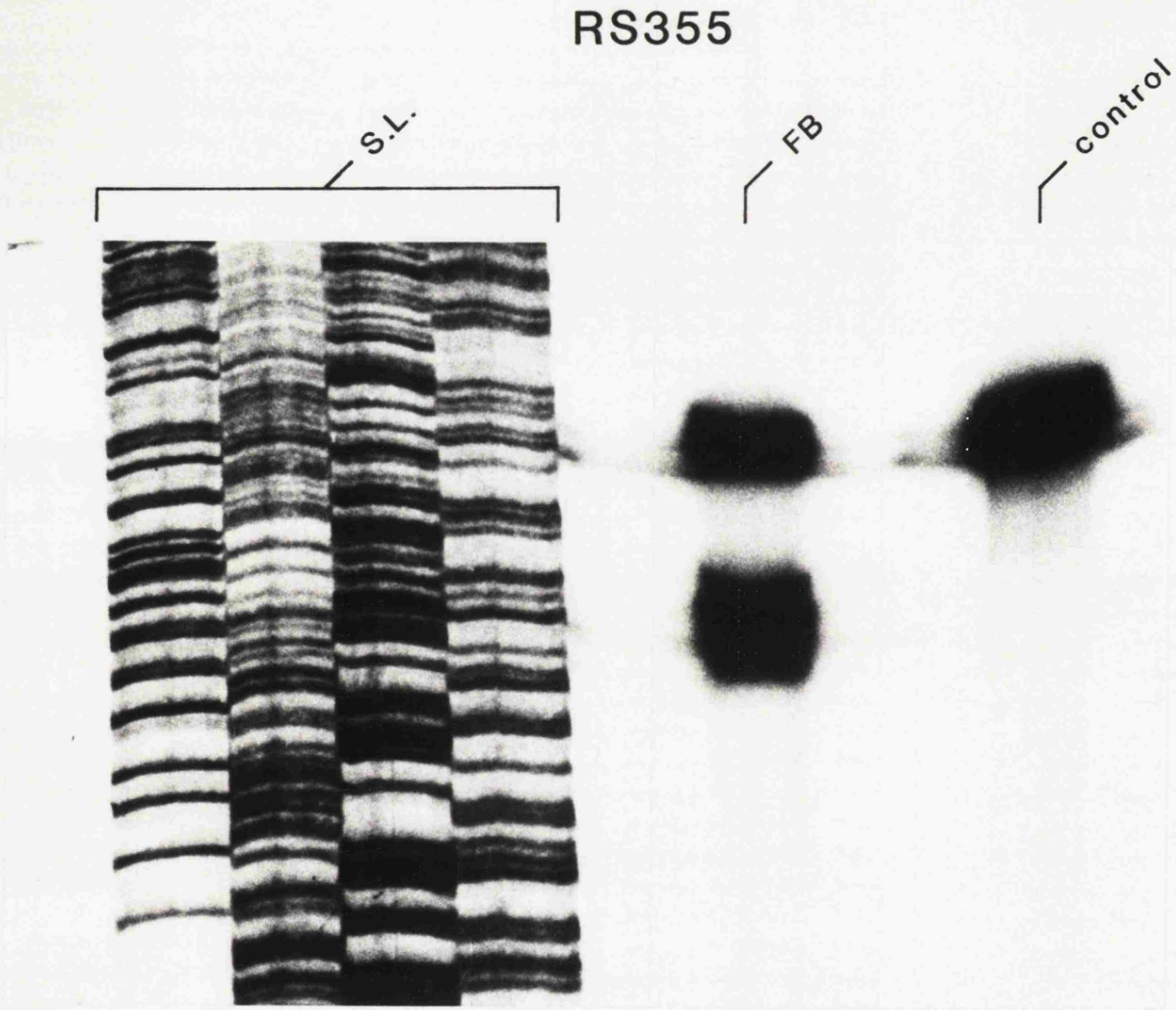


Figure 5.8

Positive mapping of a putative mutation in RNA sample F.B.

'The short run.'

Autoradiograph of a 40cm-long 6% polyacrylamide RNase A protection gel with the RNA sample F.B. and a control RNA sample hybridised to antisense RNA probe RS355.
These samples were identical to those in Figure 5.7 and electrophoresed on the same gel, but only for two hours.
Autoradiographic exposure was for one and a half days.


RS355
S.L.
FB
control

At the DNA level, the cleavage of probe RS355 maps the variant to a position approximately 65bp 5' to the Sau3AI site at the 3', or 'right-hand' end, of the RS355 subclone. From the pro-α1(I) cDNA sequence in the EMBL computer database, this Sau3AI site was found to correspond to amino acid number 887 of the helical domain. Hence the approximate position of the variant was at amino acid 868. Using the genomic structure of COL1A2 (Chu et al., 1982) to deduce the intron/exon junction positions on the pro-α1(I) cDNA sequence, amino acid 887 was mapped to the 3' end of exon 43, a 108bp exon. This meant that the region of the variant lies towards the 5' end of exon 43 of COL1A1 (see Figure 5.9).

5.5 Discussion

The pro-α1(I) cDNA clones Hf404 and Hf677 were used to generate RNA probes to screen for mutations in the α1(I) gene in type II OI. Only the pro-α1(I) gene and type II OI's were chosen in this study of OI, to maximise the chances of finding a mutation for the minimum amount of work.

With the only experience of this technique being the positive control, described in section 4.5, it was not known how obvious or well-disguised a mutation might be. It was not known at what point it became unwise to scrutinise the autoradiograph backgrounds any more closely. RNase A may cleave a particular mismatch at a very low efficiency, so the cleavage products from some mutants will surely be very faint.

Because sizeable deletions or insertions would have been demonstrated by Northern blotting and because most mutations in OI would appear to be point mutations (see section 1.4.3), it was considered that any mutations found would be point mutations or small

Figure 5.9

DNA sequence of COL1A1 exon 43.

DNA sequence of exon 43 showing the corresponding protein sequence and amino acid numbers in the α1(I) chain helical domain.

Marked is the Sau3AI site which defines the 3' end of the RS355 subclone.

Marked by the vertical arrow is the approximate position of the RNase A cleavage site of F.B.

EXON 43

```
   GGTGACCGTGGTGAGACCGGCCCCGCTGGACCCCCTGGTGCTCCTGGTGCTCCTGGTGCC
 1 ---------+---------+---------+---------+---------+---------+ 60
   CCACTGGCACCACTCTGGCCGGGGCGACCTGGGGGACCACGAGGACCACGAGGACCACGG

   GlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaProGlyAlaProGlyAla

    6   7   8   9 860 1   2   3   4   5   6   7   8   9 870 1   2   3   4   5
```

```
                                        S
                                        a
                                        u
                                        3
                                        A
   CCTGGCCCCGTTGGCCCTGCTGGCAAGAGTGGTGATCGTGGTGAGACT
61 ---------+---------+---------+---------+-------- 108
   GGACCGGGGCAACCGGGACGACCGTTCTCACCACTAGCACCACTCTGA

   ProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThr -

    6   7   8   9 880 1   2   3   4   5   6   7   8   9 890 1
```

deletions or insertions. Therefore any RNase cleavage products from a probe/mRNA mismatch should approximately add up to the wild-type size. Like other probes, EP934, produced many faint bands that were visible in all samples. But two faint bands however were present in 6260 but not in 2328 that added up to the wild-type band size. It was felt that these bands could represent products of a very inefficient mismatch cleavage.

This possible variant was investigated with two new overlapping probes. One of these probes gave some evidence in support of a mismatch, but overall the results were inconclusive. It may well be that a true variant had been revealed, but that the faint cleavage products with probes XP794 and EH473 were hidden in the background smear of degradation products. There thus becomes a point at which a possible mutation, as identified by faint cleavage products, becomes too difficult to verify. Only after a great deal more experience of this technique will the point at which cleavage products are too faint to be likely to be verifiable become apparent.

Probe XP794 showed a phenomenon not seen with any other probe. This probe, 794 bases long, produced bands of 200 and 160 bases in 2328 and not in 6260. With no other probe were 'extra' bands seen that did not add up to the size of the wild-type band. Reasons for this remain obscure.

Apart from a role in identifying and mapping mutations, RNase A mapping has the potential to identify base polymorphisms. These may be as a consequence of the degeneracy of the genetic code, with the base substitutions lying in the 'wobble' position of the codon or may be base substitutions leading to amino acid polymorphisms. The discovery of any polymorphisms in the pro-α1(I) coding sequence would be useful not only as linkage markers, but as tools to analyse the relative

expression of the two COL1A1 genes in individuals with normal COL1A1 genes. In addition, RNA polymorphisms could be used to investigate the presence or absence of germline imprinting on the COL1A1 gene (for review see Solter, 1988).

The three probes AB720, EB1536 and SE723, together indicate that there is either a dissimilarity between the cDNA clone Hf404 or that a region of double-stranded RNA is particularly sensitive to attack by RNase A. The former rationale would either be due to a base polymorphism or due to a cloning artefact in Hf404 by way of a substituted base. If a small region had been deleted or inserted, the cleavage of the probes would most likely have been more pronounced. Due to a lack of time, the cause of this apparent cleavage could not be pursued. In order to determine the basis of the apparent cleavage many more samples, than the just three so far, should be analysed. The cell-line from which Hf404 was derived should be obtained and its RNA mapped with these probes. If these approaches failed to identify the basis of the apparent cleavage, then the possibility of a particularly RNase A sensitive site would remain as the only explanation.

If the apparent cleavage was due to a cloning artefact than the mRNA derived from both alleles of the three OIs would produce mismatches when hybridised to the probe; that is to say the mismatch would be 'homozygous'. If this was the case, then cleavage would have to be very inefficient as compared with the A.K. single base mutation which is heterozygous. The similar relative intensities of the apparent cleavage products in the three samples analysed would indicate that they are all of the same genotype, that is to say all heterozygous or all homozygous.

Because of the lack of mutations found in the $\alpha 1(I)$ and $\alpha 2(I)$ C-propeptides in OI (see section 1.4.3), the probe covering this

region, BE574, was not intended to be used until all others had drawn a blank.

The next batch of samples were first tested with BX887, the probe considered most likely to reveal a mutation in type II OI. The F.B. sample yielded the result required to verify the power of RNase A cleavage to detect and map mutations in OI. In this sample, in stark contrast to the the bands seen in the previous batch of samples, cleavage was complete (assuming the variant to be heterozygous). Rather than feeling that valuable time had been wasted investigating the faint bands seen in the previous batch of samples, it was felt that the degree of cleavage here was almost too good to be true for a point mutation. That being the only type of mutation anticipated, as at the time, no small deletions or insertions had been found in the helical domain of collagen genes.

The existing subclones RS355 and SS534 were used to orientate the variant to the 'left-hand' position. The appearance of the lower cleavage product as a doublet with the A.K. sample and now the F.B. sample with the probe BX887 was assumed to be due to 'nibbling' by the RNase A around the cleavage site. However the cleavage product seen with probe RS355, with a cluster of 6 bands, indicated that perhaps something more specific was occurring at the cleavage site. But no sense was made of the clue that there were 3 pairs of bands, 8 or 9 bases different in size, until after the variant had been sequenced (see section 6.5).

From information derived from the structure of COL1A2, exon 43 of COL1A1 was deduced to encode amino acids 856-891. Probe RS355 mapped the mutation to approximately amino acid position 868, a glycine residue! It was felt though, that this mapping was not accurate enough to name a particular amino acid. It was felt however that the mapping

was probably accurate to ±2 (Gly-X-Y) units from this site. That is to say in the region of amino acids 862-875.

The variant was now further characterised.

CHAPTER 6

CHARACTERISATION OF A MUTATION IN OI

6.1 DNA amplification

Traditional characterisation of the F.B. mutation would have necessitated cloning, either by cDNA or genomic means from large quantities of RNA or DNA, followed by subcloning prior to sequencing. The recent development of the polymerase chain reaction (PCR) (Mullis et al., 1986), enables rapid analysis of mutations using very small amounts of starting material.

The ideal strategy for the analysis of mutations by PCR is to directly sequence the PCR products. However because the mutation anticipated in this study was a heterozygous point mutation, it was intended to sequence PCR-derived clones first. In direct sequencing, a heterozygous point mutation will be seen as faint bands in two tracks and therefore less obvious than the mutant allele on a sequence of cloned DNA. In addition, until recently the quality of results from direct sequencing was not as good as those from cloned sequencing. However, because of the infidelity of the polymerase chain reaction, in which the rate of misincorporations over 30 cycles may be as high as 1 in 300 (Sharf et al, 1986 and Dunning et al, 1988), it was intended to confirm that the change seen in the clone was genuine by direct sequencing.

To facilitate cloning, the PCR primers were designed around clonable enzyme sites in the region of interest. From the published sequence of the α1(I) cDNA clone Hf677 (Chu et al., 1982), it was seen that the two exons flanking the mutant exon contained unique NarI sites 153bp apart. Because only the cDNA sequence was available, it had to be

verified that there were no additional NarI sites lying in the introns between exons 42 and 44.

From the genomic map corresponding to α1(I)CB6 (Cohn et al, 1986), it was estimated that intron 42 is approximately 60bp and intron 43 is approximately 400bp in size. This would make the genomic fragment contained by the NarI sites in exons 42 and 44 approximately 650bp in size.

When the 153bp NarI-NarI fragment derived from the cDNA clone Hf677 was used as a radiolabelled probe and hybridised with NarI digested DNA of CG103 (Barsh et al., 1982), a cosmid clone containing most of the COL1A1 gene, a band calculated to be approximately 700bp was obtained.

Thus it appeared that there were no additional NarI sites in introns 42 and 43, or if there were, they would be close enough to the known sites to not affect the strategy.

Various possible PCR primers were then designed to flank or cover the NarI sites. Five possible 20-mer oligonucleotide primers were designed and used to screen the GenBank and EMBL computer data bases. It was felt that this search was necessary, because, due to the lack of published intronic sequence, the primers had to lie in helical domain exons. Thus the primers could potentially hybridise to other parts of the same gene or to other collagen genes. Any cross-hybridisation would bring about the possible production of aberrant PCR products and would act as a drain on the primers.

Two primers were chosen that were thought to have the least potential to cross-hybridise. The two primers (left and right) were made by John Keyte (Department of Biochemistry, Leicester University).

The strategy for amplifying the mutation is shown in Figure 6.1.

A 10μl aliquot of a 50μl amplification of 1μg of F.B. DNA was

Figure 6.1

Amplification of the F.B. mutant region.

The figure represents the mutant region of COL1A1.

The boxes represent the exons with the number of each one within the boxes.

The numbers below the boxes indicate the size of the exon, in base pairs.

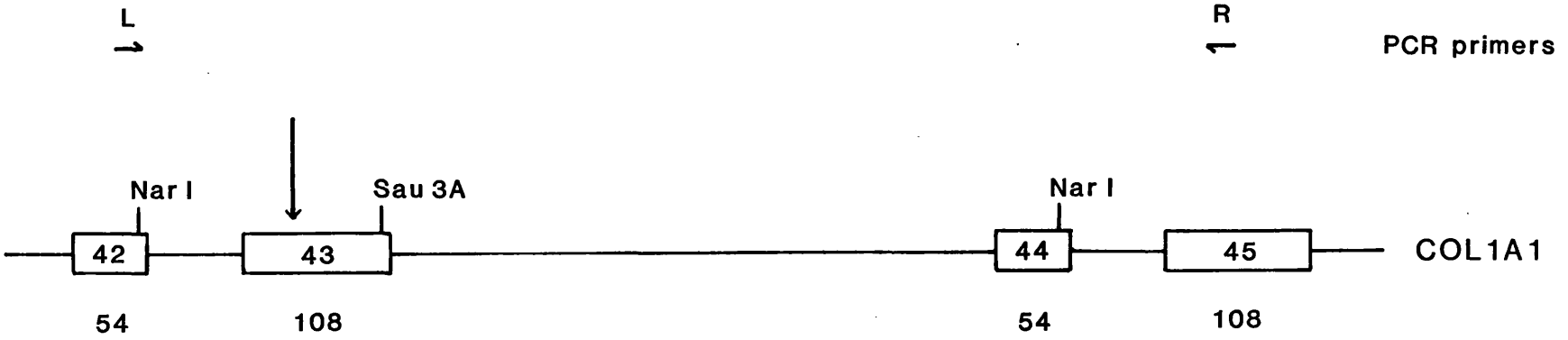
The two NarI sites present in this segement of DNA are marked.

The Sau3AI site of exon 43 is also marked.

The vertical arrow indicates the approximate position of the mutation.

The position of the PCR primers are marked above the gene. Both primers are 20-mers. The left hand primer (L) has the sequence: $5'$CCGAAGGTTCCCCTGGACGA$3'$ and the right hand primer (R) has the sequence: $5'$CGCCCTGTTCGCCTGTCTCA$3'$.

The fragment amplified was approximately 850bp in size.


L
R
PCR primers
Nar I
Sau 3A
Nar I
42
43
44
45
COL1A1
54
108
54
108

visualised on a 1% agarose gel. A single band of approximately 850bp was seen, with very little background smear at all. That this was the correct fragment was confirmed by Southern blotting, using the 153bp NarI-NarI cDNA fragment as the radiolabelled probe. Thus the correct fragment of DNA had been amplified with minimal non-specific co-amplification.

A further 15 µl of the amplification reaction was electrophoresed on a 1% agarose gel and the amplified fragment excised. The gel slice was placed in a microfuge tube and 100µl distilled water added. The tube was frozen and thawed twice, bringing about a little seepage of DNA from the frozen gel slice into the water. 13µl of this DNA solution was reamplified for 20 cycles. The re-amplified material was phenol extracted, ethanol precipitated and dissolved in 10µl TE.

6.2 Cloning and sequencing of the mutant locus

2.5µl of the PCR-amplified material (approximately 250ng) was digested with NarI and ligated into alkaline phosphatase treated ClaI digested Bluescript SK^- plasmid. The ligated DNA was then transformed into freshly prepared E.coli JM83recA.

The mutation had been mapped to a region approximately 100bp from the 5' NarI site and hence approximately 600bp from the 3' NarI site. So cloning in only one orientation would make the region of the mutation amenable to sequencing. Due to a lack of suitable restriction endonuclease sites in the sequence as orientation markers, 14 white colonies were picked for sequencing, giving a 98% chance ($1-(1-P)^{14}$, where P=0.25) that both alleles would be sequenced in the correct orientation.

DNA to be sequenced was prepared as described in section 2.19 and sequencing was conducted using the Bluescript KS 17-mer primer.

Of the 14 clones sequenced, 11 were of the desired orientation. Of these 11 clones, 5 were seen to have a 9bp deletion from the region mapped by RNase A cleavage. The sequence obtained from the two different alleles is shown in figure 6.2.

The sequence is seen to have a repetitive nature. The repeat length is 9bp. Two full tandem repeats are present, and a third repeat unit of 8bp; one base short of a full repeat, is also present. Thus the 'window' for a 9bp deletion could lie anywhere in a region of 26bp, making exact analysis of the deletion boundary impossible. The repetitive region in the normal allele covers 9 amino acids; P-G-A-P-G-A-P-G-A. So wherever the deletion occurred, even if it split up codons at each deletion junction, one amino acid triplet will have been lost and the other two will remain.

6.3 Analysis of mutation inheritance

In order to confirm that the deletion was not a 'silent' variant, it was necessary to confirm that neither of the proband's parents also carried the deletion. It was also necessary to prove that the deletion was not a PCR artifact.

It had been expected that any mutation would be a base substitution, and it was intended to use direct sequencing of the PCR product to confirm the observation. However, for this deletion, it was anticipated that a simpler technique could be adopted. The 'polymerase chain reaction' was carried out on DNA from F.B., his parents and controls. The DNA was cleaned by phenol extraction and ethanol precipitation and dissolved in 17µl distilled water. 1µl samples were then digested with various restriction endonucleases in single and double digests that might be expected to produce fragments of a size in which a 9bp difference should be detectable on an acrylamide gel. The

Figure 6.2

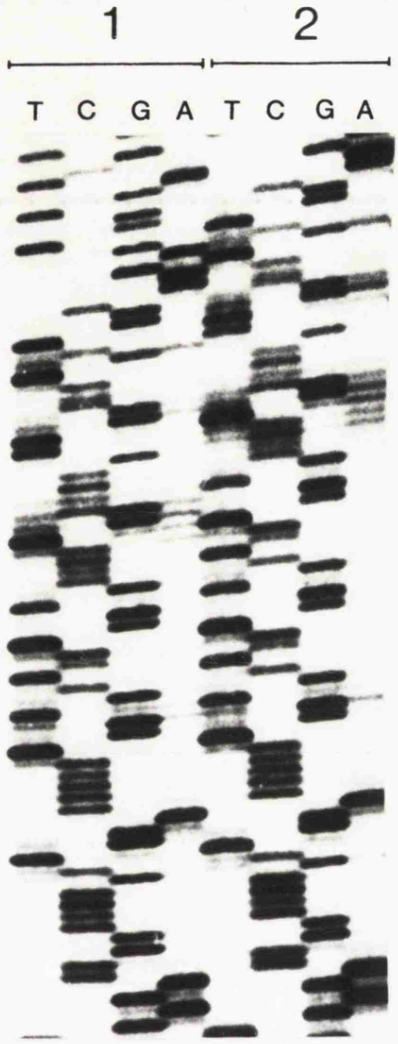
DNA sequence of the F.B. mutant allele.

A. DNA sequence of the mutant and wild-type alleles.
1 and 2 signify the mutant and normal alleles respectively.
9bp from the repetitive region in the middle of the picture can be seen to have been deleted in the mutant allele.
Autoradiographic exposure was for two days.

B. Mixed DNA sequence of the mutant and wild-type alleles.
1 and 2 signify the mutant and normal alleles respectively.
The ddTTP, ddCTP etc. tracks of the mutant and wild-type alleles have been electrophoresed side by side. This autoradiograph clearly shows the position of the deletion.
Autoradiographic exposure was for one day.

C. DNA and protein sequences of the wild-type and mutant F.B. COL1A1 alleles.

A



B

C

GGACCCCCTGGTGCTCCTGGTGCTCCTGGTGCCCCTGGC
Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly
865

Wild type allele

GGACCCCCTGGTGCTCCTGGTGCCCCTGGC
Gly Pro Pro Gly Ala Pro Gly Ala Pro Gly
865

Mutant allele

samples were then electrophoresed through a 12% acrylamide gel. The presence of unknown restriction endonuclease cut sites in the introns of the PCR material, made analysis of most digests tried, too difficult.

The best result was from a BstEII/MboI double digest (see Figure 6.3): The fragment containing the repetitive region was clearly seen as a 92bp band in the control samples and the parents of F.B.. F.B.'s DNA however, had this band but at half the intensity of the controls. In addition, another faint band of similar intensity was visible of approximately 83bp in size .

Thus the deletion could be visualised on an acrylamide gel. The deletion was proved to not be a PCR artifact, as half the population of amplified DNA molecules could be seen to carry the deletion. In addition, the mutation was shown to be de novo as neither parent had the deletion, strongly suggesting that the mutation is the disease cause and not a silent inherited variant.

DNA from the proband's parents was tested for parentage by hybridising a Southern blot of the family's DNA to a cocktail of three mini-satellite locus-specific probes: pλg3 (Wong et al., 1986), λMS1 and λMS32 (Wong et al., 1987) (see Figure 6.4). These probes detect loci with heterozygosities each in excess of 97%. All five bands visible in the F.B. track can be seen to have been inherited from either parent. If each of the three loci is assumed to have a heterozygosity of 97%, the probabilities of incorrect paternity and maternity are 0.006 and 0.03 respectively. This result shows that the parents are the true parents of F.B. and that no mix up of cell-lines or DNA samples has occurred in all the manipulations carried out.

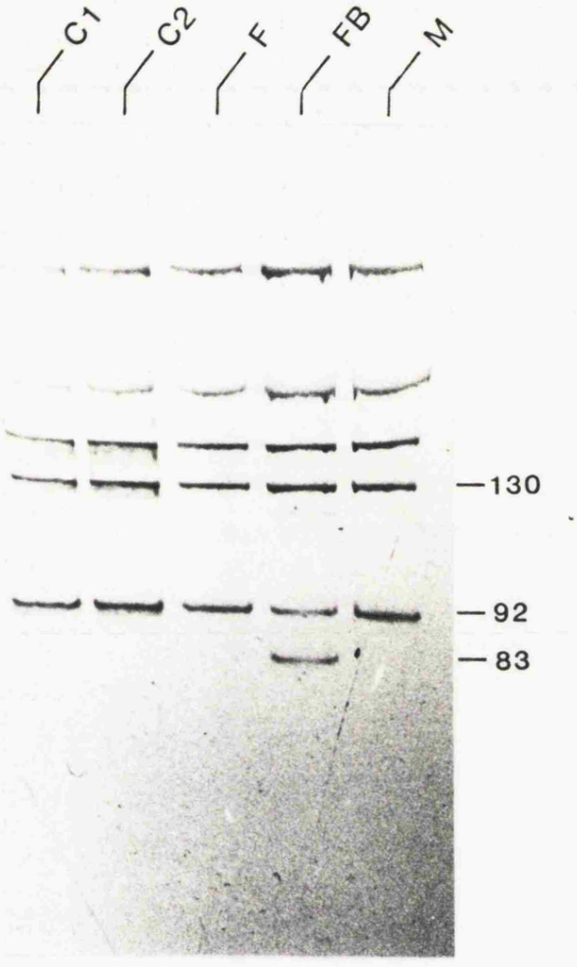
Figure 6.3

Analysis of mutation inheritance.

A. Negative image of an ethidium bromide stained 12% acrylamide gel.
1µg of genomic DNA from two control samples (C1 and C2), the father of F.B. (F), F.B. and the mother of F.B. (M) was amplified for 30 cycles as in section 2.18. $^{1}/_{17}{}^{th}$ of the recovered DNA was double digested by MboI and BstEII.
All samples have a band at 92bp, the fragment spanning the F.B. mutation. In F.B. the 92bp band is half as intense as that of both of his parents and the controls. F.B. has an additional band of 83bp not seen in the other samples.
This gel clearly shows that the F.B. deletion was de novo.

B. Restriction map of exon 43.
B = BstEII, M = MboI.
The BstEII/MboI fragment of COL1A1 exon 43 spans the F.B. mutant region which is marked as a filled-in box. In the normal allele this fragment is 92bp in size, and 83bp in the mutant allele.

A



B

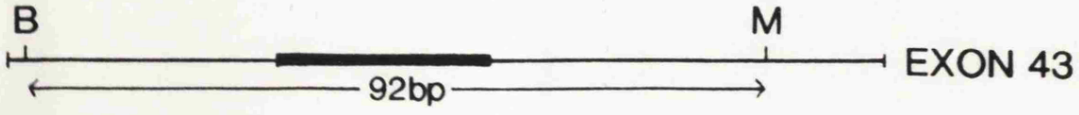
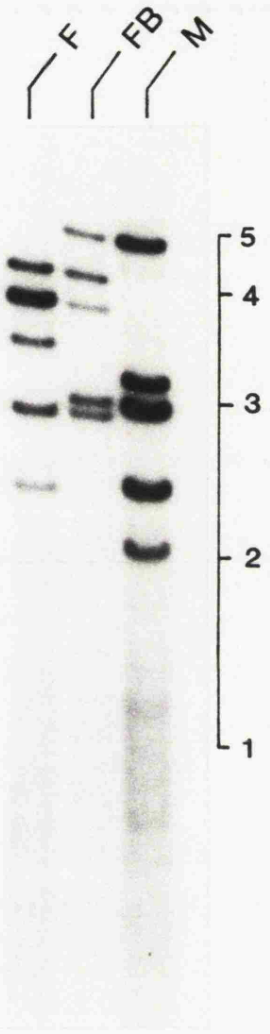


Figure 6.4

Parentage test Southern blot.

Southern blot of 0.8% agarose gel containing approximately 5µg HinfI digested DNA from the father of F.B. (F), F.B. and the mother of F.B. (M).
The locus-specific minisatellites pλg3, (Wong et al., 1986), λMS1 and λMS32 (Wong et al., 1987) were used simultaneously as the radiolabelled probes.
Hybridisation was carried out using the Marvel mix method (section 2.12.1) and post-washed down to 0.1x SSC at 65°C.
Sizes are marked in kb.
Autoradiographic exposure was for one day without an intensifying screen.


F
FB
M
5
4
3
2
1

6.4 A HaeIII RFLP in COL1A1

One of the digests used to try to detect the deletion on acrylamide gels was a HaeIII digest. An array of 14 bands was seen on the gels, but the pattern given by the control samples was markedly different to that of F.B. (Figure 6.5). In the digest of F.B. DNA and the DNA of his parents, an extra band 'c' is seen. The bands 'h' and 'n' seen in the controls, are at half the intensity in the parents, and absent from F.B..

As the two HaeIII sites that flank the deletion lie in exon 43 and are hence not separated by any intronic DNA, we can expect a size of 45bp for the HaeIII fragment covering the normal allele and 36bp for the mutant allele.

In figure 6.5, band 'k' (sized to 48bp) is seen to be about half as intense in F.B. than in the controls. This is therefore the 45bp HaeIII fragment from the normal allele. The band 'm' (sized to 36bp), which is not present in the parents and controls, is about half as intense as band 'n' seen in the controls. This is therefore the fragment containing the deletion.

With this understood, it seems likely that the two bands 'h' and 'n' (sized to 35bp and 63bp respectively) seen in the control DNAs, which are absent from F.B., are allelic to band 'c' (sized to 96bp) seen only in the DNA from F.B. and his parents. Thus it appears that F.B. lacks a HaeIII site that was present in all of 15 control DNAs analysed.

Despite not being present in any of the (British) control DNAs analysed the reality of this variant is confirmed by its inheritance to F.B. from his parents (who are of Swiss origin). Both parents of F.B. can be seen to be heterozygous for the variant. All three bands corresponding to the variant in the parents can be seen to be at

Figure 6.5

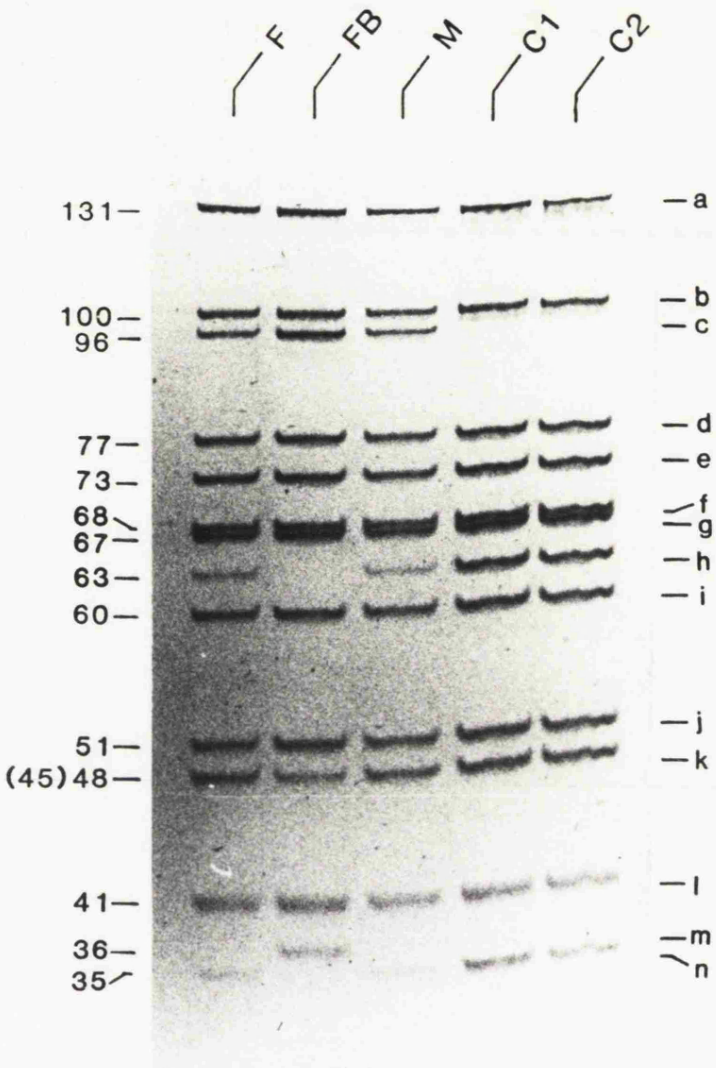
A HaeIII RFLP in COL1A1.

A. Negative image of an ethidium bromide stained 12% acrylamide gel.
1µg of genomic DNA from two control samples (C1 and C2), the father of F.B. (F), F.B. and the mother of F.B. (M) was amplified for 30 cycles as in section 2.18. $^{1}/_{17}{}^{th}$ of the recovered DNA was digested by HaeIII.
The F.B. deletion is shown by bands k and m . Band k, calculated as being 48bp in size represents the 45bp normal HaeIII fragment of exon 43 (see part B). Band m represents the 36bp HaeIII fragment of the mutant allele.

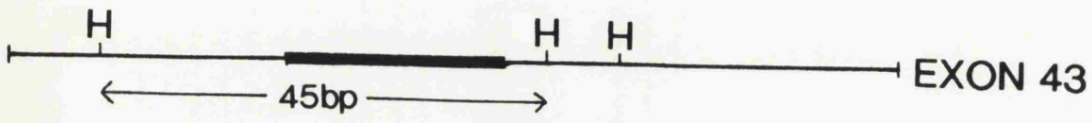
In addition to the F.B. deletion, the restriction fragment pattern of F.B. and his parents is different to the control samples. F.B. and his parents have band c, a band absent from the controls. F.B. does not have bands h and n, which are present in all other samples. In the parents of F.B., the bands c, h and n are approximately half as intense with respect to their other bands. In F.B., band c is approximately at the same intensity as his other bands. Bands h and n are therefore allelic to band c; F.B. being homozygous for the absence of the HaeIII site and his parents heterozygous.

B. Restriction map of exon 43.
H = HaeIII.
The largest HaeIII fragment of COL1A1 exon 43 spans the F.B. mutant region which is marked as a filled-in box. In the normal allele this fragment is 45bp in size and 36bp in the mutant allele.

A


F
FB
M
C1
C2
131
100
96
77
73
68
67
63
60
51
(45)48
41
36
35
a
b
c
d
e
f
g
h
i
j
k
l
m
n

B


H
H
H
45bp
EXON 43

reduced intensities as compared with the other bands. Therefore the variant is genuine, but is present at best at a moderately low frequency in the Anglo-Saxon population.

6.5 Discussion

With the F.B. mutation accurately mapped by RNase A mapping, characterisation proved to be very rapid. Thus the techniques of RNase A mapping and PCR amplification can be seen to complement each other well, and coupled together, constitute a very powerful method of detecting and characterising mutations. However, the fact that the F.B. mutation was found to be a small deletion highlights the fact that RNase A mapping will favour the detection of deletions and insertions over point mutations.

As the deletion arose in a repetitive region, the precise definition of the mutation is unknown, there being 18 possible 'windows' for the deletion in the 26bp repetitive region. The deletion boundaries could have occurred within codons without the possibility of producing an amino acid substitution as a result of the deletion. Thus wherever the deletion occurred, the result would be the loss of one Gly-Ala-Pro unit. Possible mechanisms for such a deletion include unequal crossing-over, unequal sister-chromatid exchange, DNA slippage at replication and gene conversion.

In retrospect, the descision to choose cloned sequencing rather than direct sequencing was inadvertantly a good one. To simulate a direct sequence of the F.B. mutation, the two alleles were mixed *i.e.* each ddNTP track contained *both* alleles. This gave a result that was not obviously abnormal, but impossible to read from the point of the mutation onwards. Only after several attempts at the direct sequencing would it probably have become apparent that a deletion or insertion was

responsible. Then cloned sequencing would have to have been initiated. Thus it is probably advisable, that when sequencing PCR-amplified mutations detected by RNase A mapping, that cloned sequencing should be performed and not direct sequencing.

In the light of the sequencing data, the cause of the heterogeneous cleavage products obtained with probe RS355 (see Figure 5.7) can be speculated. In the formation of the heteroduplex of the probe and the mutant RNA, a 'bubble' may be formed with three repeats on the probe strand and two repeats on the mRNA strand. Three conformations of interaction between the strands are possible, each involving loops in the probe strand. If the bubble 'zips-up' from one end of the repetitive region only, a loop at the other end of the repetitive region will be formed producing either the upper or the lower doublet-band depending upon the direction of the 'zipping-up'. If the bubble 'zipped-up' from each end simultaneously, the middle repeat will loop out and the middle doublet-band will be seen on the gel. From the intensity of the middle doublet, it would appear that the simultaneous bi-directional 'zipping-up' is the preferred event. The presence of each product as a doublet may represent partial cleavage of the loop by RNase A, or nibbling from the ends of the heteroduplex.

That the deletion was a de novo mutation was clearly shown in Figure 6.3. Because neither of the parents carried the deletion, it was necessary to verify parentage before any conclusions could be made about the mutation. That the deletion was de novo strongly suggests that it is the disease-cause. However, post-translational modification data from the F.B. cells (performed by Dr. A. Superti-Furga, Universitats-Kinderklinik, Zurich), showed the hydroxylation of the four lysine residues between the deletion and the carboxyl end of the helical domain. The finding that overmodification occurred along the

entire length of the triple helix (1014 amino acid residues) is clearly at odds with the finding of a deletion at approximately amino acid position 870.

Amino acid substitutions of glycine residues are believed to cause overmodification amino terminal to the substitution site by delaying triple helix propagation, thereby exposing nascent α-chains to the modification enzymes for longer than normal. In contrast, an α-chain with a deletion that leaves the Gly-X-Y motif intact would be expected to be readily incorporated into the triple helix causing little, if any, overmodification. However, a case of lethal OI has been described in which a multi-exon deletion in the α2(I) gene resulted in the overmodification of the type I collagen chains amino-terminal to the deletion junction (Willing et al., 1988), thus suggesting that the triple helical structure may not be normal or possibly too unstable to prevent overmodification.

In the case of F.B., overmodification has not only occurred amino-terminal to the deletion but carboxyl-terminal too. One can speculate that the amino acid residues in the X and Y positions of the Gly-X-Y motif play an important role in interchain interactions, and mis-alignment of the correct amino acids in these positions could have a strong enough effect as to delay helix formation and thus cause over-modification. It may be that nascent α-chains amino-terminal to the entwined α-chains may not be as free of interchain interactions as have been proposed in the models of collagen triple helix formation. If such interchain interactions were disrupted by a $(Gly\text{-}X\text{-}Y)_n$ deletion, then triple helix formation could be effected prior to the helix formation-front reaching the deletion junction, thus causing overmodification beyond the region of the mutation.

In addition to the overmodification of procollagen molecules, it

is possible that such a mutation could cause a disease phenotype by interrupting the charge pattern and effecting the interchain hydrogen-bonding of collagen molecules.

As yet no phenotypically silent variants have been found in collagen gene exons, so their incidence cannot even be guessed. The incidence of OI is approximately 1 in 20,000. If we were to assume that the rate of phenotypically observable mutations in the type I collagen genes is 1 in 20,000 per generation, then the probability that in this proband a second, independent, de novo mutation was the cause of the disease phenotype would be very low indeed. The possibility remains that F.B. could be a compound heterozygote, that is to say, he may have inherited a normally silent variant from one parent which, in the presence of the Gly-X-Y deletion, leads to the lethal phenotype.

A case of a similar mutation; the deletion of amino acids 874-876 from the α1(I) chain of a proband with type II OI has been observed previously (P. Byers, personal communication to A. Superti-Furga). But in that case, determination that the deletion was not an inherited variant was not possible as the parents were not available.

The post-translational modification data from F.B. indicates that the deletion is having an entirely unexpected effect upon the protein. That the deletion of a single Gly-X-Y unit from COL1A1 causes lethal OI is at the least, surprising and at the most, unbelievable. By an altered charge pattern the deletion may be expected to cause mild OI, or by altering the N-propeptidase cleavage site, cause type VII EDS.

With the F.B. deletion being such an 'unlikely' mutation in type II OI, it is of some importance that its role in the disease pathogenesis is determined.

One way to do this is to take advantage of the existence of

Mov-13 cells. These are fibroblasts derived from mice with a retrovirus inserted in the first intron of COL1A1 (Schnieke et al., 1983; Harbers et al., 1984). Fibroblasts homozygous for the retroviral insertion fail to synthesize any α1(I) chains and therefore no type I collagen. It has been shown by Stacey et al., (1987), that the type I collagen-deficient phenotype of these cells can be rescued by the transfection of a vector containing the human α1(I) full-length cDNA, following immortalisation. Stacey et al., (1988), have shown that Mov-13 cells will produce mutant type I collagen following transfection with a mutated mouse genomic COL1A1 clone.

To determine whether the F.B. deletion causes the observed overmodification pattern, the human α1(I) full-length cDNA could be in vitro mutagenised to produce the F.B. deletion. When transfected into immortalised Mov-13 cells, mutant type I collagen would be produced. If this type I collagen showed the same overmodification pattern as F.B., then it would be safe to say that the 9bp deletion was indeed the cause of the type II OI phenotype. To take the experiment to the extreme, transgenic mice could be produced containing the 9bp deletion, and if they had the type II OI phenotype, the deletion must be the disease cause.

The HaeIII dimorphism found in the study of the F.B. mutation was a coincidental bonus, but also perhaps is an indication of the potential power of the polymerase chain reaction. It is likely that the introns of collagen genes contain many dimorphic sites for 4-cutter restriction endonucleases. Segments of the genes may be amplified, end labelled and partially digested with a 4-cutter enzyme, electrophoresed and autoradiographed. Any differences in the banding pattern between individuals would indicate an RFLP. Any new RFLPs in collagen genes would improve the polymorphism information content (PIC value) of the

genes and thus may be useful as linkage markers.

The 9bp deletion described here is unlike any published mutation in OI. All other deletions involved the loss of complete exons and the deletion junctions have always been in introns. The case of the homozygous 4bp deletion in exon 51 of COL1A2 (Pihlajaniemi et al., 1984) was different, in that the helical domain was normal, but that the initial assembly of the three α-chains was disrupted.

If the 9bp deletion is the sole cause of the type II OI phenotype, then it is a novel mutation in more than a purely academic respect. In no mutation has overmodification of the protein occurred carboxyl-terminal to the mutation. That this should occur at all is surprising enough, but that it should be caused by a 9bp deletion is shocking. If the deletion is the sole disease-cause, then a major re-think on the molecular pathogenesis of OI may have to be initiated.

Ironically this novel mutation was not found as an exclusive consequence of a novel technique, but was coincidentally found by a novel technique. It was hoped that RNase A mapping would begin to redress the bias in the detection of collagen gene mutations. However, the first mutation found, the F.B. mutation, could have as easily been found with the well-used technique of S1 mapping.

REFERENCES

Abraham, P.A., Perjda, A.J., Carnes, W.H. and Vitto, J. (1982) Marfan syndrome: Demonstration of abnormal elastin in aorta.
J. Clin. Invest. 70: 1245-1252.

Aitchison, K., Ogilvie, D., Honeyman, M., Thompson, E. and Sykes, B. (1988) Homozygous osteogenesis imperfecta unlinked to collagen I genes.
Hum. Genet. 78: 233-236.

Allenbach, E.S., Boettiger, D., Pacifici, M. and Adams, S. (1985) Control of types I and II collagen and fibronectin gene expression in chondrocytes delineated by viral transformation.
Mol. Cell. Biol. 5: 1002-1008.

Bailey, A.J., Sims, T.S. and Light, N. (1984) Cross-linking in type IV collagen.
Biochem. J. 218: 713-723.

Baldwin, C.T., Constantinou, C.D., Dumars, K.W. and Prockop, D.J. (1989) A single base mutation that converts glycine 907 of the α2(I) chain of type I procollagen to aspartate in a lethal variant of osteogenesis imperfecta.
J. Biol. Chem. 264: 3002-3006.

Barsh, G.S., David, K.E. and Byers, P.H. (1982) Type I osteogenesis imperfecta: A nonfunctional allele for proα1(I) chains of type I procollagen.
Proc. Natl. Acad. Sci. USA 79: 3838-3842.

Barsh, G.S., Roush, S.L., and Gelinas, R.E. (1984) DNA and chromatin structure of the human α1(I) collagen gene.
J. Biol. Chem. 259: 14906-14913.

Barsh, G.S., Roush, C.L., Bonadio, J. and Byers, P.H. (1985) Intron-mediated recombination may cause a deletion in an α1 type I collagen chain in a lethal form of osteogenesis imperfecta.
Proc. Natl. Acad. Sci. USA 82: 2870-2874.

Bateman, J.F., Chan, D., Mascara, T., Rogers, J.G. and Cole, W.G. (1986) Collagen defects in lethal perinatal osteogenesis imperfecta. Biochem. J. 240: 699-708.

Bateman, J.F., Chan, D., Walker, I.D., Rogers, J.G. and Cole, W.G. (1987) Lethal perinatal osteogenesis imperfecta due to the substitution of arginine for glycine at residue 391 of the α1(I) chain of type I collagen.
J. Biol. Chem. 262: 7021-7027.

Bateman, J.F., Lamande, S.R., Dahl, H.M., Chan, D. and Cole, W.G. (1988) Substitution of arginine for glycine in the collagen α1(I) chain in lethal ostoegenesis imperfecta.
J. Biol. Chem. 263: 11627-11630.

Berk, A.J. and Sharp, P.A. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids.
Cell 12: 721-732.

Bernard, M., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B.R. and Ramirez, F. (1988) Cloning and sequencing of pro-α1(XI) collagen cDNA demonstrates that type XI belongs to the fibrillar class of collagens and reveals that expression of the gene is not restricted to cartilagenous tissue.
J. Biol. Chem. 263: 17159-17166

Biggin, M.D., Gibson, T.J. and Hong,E.F. (1983) Buffer gradient gels and 35S label as an aid to Rapid DNA sequence determination.
Proc. Natl. Acad. Sci. USA 80: 3963-3965.

Bornstein, P. and Sage, H. (1980) Structurally distinct collagen types.
Ann. Rev. Biochem. 49: 957-1003.

Bornstein, P., McKay, J., Morishima, J.K., Devarayalu, S. and Gelinas, R.E. (1987) Regulatory elements in the first intron contribute to transcriptional control of the human α1(I) collagen gene.
Proc. Natl. Acad. Sci. USA 84: 8869-8873.

Borntstein, P. and McKay, J. (1988) The first intron of the α1(I) collagen gene contains several transcriptional regulatory elements.
J. Biol. Chem. 263: 1603-1606.

Boucek, R., Noble, N.L., Gunja-Smith, Z. and Butler, W.T. (1981) The Marfan syndrome: a deficiency of chemically stable collagen cross-links.
New. Engl. J. Med. 305: 988-991.

Bouche, J.P. (1981) The effect of spermidine on endonuclease inhibition by agarose contaminates.
Anal. Biochem. 115: 42-45.

Boyd, C.D., Toth-Fejel, SE., Gadi, I.K., Litt, M., Condon, M.R., Kolbe, M., Hagen, I.K., Kurkinen, M., Mackenzie, J.W. and Magenis, E. (1988) The genes coding for human pro α1(IV) collagen and pro α2(IV) collagen are both located at the end of the long arm of chromosome 13.
Am. J. Hum. Genet. 42: 309-314

Breul, S.D., Bradley, K.H., Hance, A.J., Schafer, M.P., Berg, R.A. and Crystal, R.G. (1980) Control of collagen production by human diploid fibroblasts.
J. Biol. Chem. 255: 5250-5260.

Burgeson, R.E., Morris, N.P., Murray, L.W., Duncan, K.G., Keene, D.R., and Sakai, L.Y. (1985) The structure of type VII collagen.
Ann. N. Y. Acad. Sci. 460: 47-57.

Byers, P.H., Siegel, R.C., Peterson, K.E., Rowe, D.W., Holbrook, K.A., Smith, L.T., Chang, Y-H. and Fu, J.C.C. (1981) Marfan syndrome: Abnormal α2 chain in type I collagen.
Proc. Natl. Acad. Sci. USA 78: 7745-7749.

Byers, P.H., Starman, B.J., Cohn, D.H. and Horwitz, A.L. (1988) A novel mutation causes a perinatal lethal form of osteogenesis imperfecta.
J. Biol. Chem. 263: 7855-7861.

Byers, P.H., Tsipouras, P., Bonadio, J.F. and Starman, B.J. (1988) Perinatal lethal osteogenesis imperfecta (OI type II): A biochemically heterogeneous disorder usually due to new mutations in the genes for type I collagen.
Am. J. Hum. Genet. 42: 237-248.

Castagnola, P., Moro, G., Descalzi-Cancedda, F. and Cancedda, R. (1986) Type X collagen synthesis during in vitro development of chick embryo tibial chondrocytes.
J. Cell Biol. 102: 2310-2317.

Chebloune, Y., Trabuchet, G., Poncet, D., Cohen-Solal, M., Faure, C., Verdier, G. and Nigon, V.M. (1984) A new method for detection of small modifications in genomic DNA applied to the δ-β globin gene cluster.
Eur. J. Biochem. 142: 473-480.

Chomczynski, P. and Sacchi, N. (1987) Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.
Anal. Biochem. 162: 156-159.

Chu, M-L., Myers, J.C., Bernard, M.P., Ding, J-F. and Ramirez, F. (1982) Cloning and characterisation of five overlapping cDNAs specific for the human proα1(I) collagen chain.
Nucleic Acids Res. 10: 5925-5934.

Chu, M-L, Williams, C.J., Pepe, G., Hisch, J.L., Prockop, D.J. and Ramirez, F. (1983) Internal deletion in a collagen gene in a perinatal lethal form of osteogenesis imperfecta.
Nature 304: 78-80.

Chu, M-L., Weil, D., de Wet, W., Bernard, M., Sippola, M. and Ramirez, F. (1985a) Isolation of cDNA and genomic clones encoding the human proα1(III) collagen.
J. Biol. Chem. 260: 4357-4363.

Chu, M-L, Garguilo, V., Williams, C.J. and Ramirez, F. (1985b) Multi-exon deletion in an osteogenesis imperfecta variant with increased type III collagen mRNA.
J. Biol. Chem. 260: 691-694.

Cole, W.G., Chan, D., Chambers, G.W., Walker, I.D. and Bateman, J.F. (1986) Deletion of 24 amino acids from the pro-α1(I) chain of type I procollagen in a patient with the Ehlers-Danlos syndrome type VII. J. Biol. Chem. 261: 5496-5503.

Cohn, D.H., Byers, P.H., Steimann, B. and Gelinas, R.E. (1986) Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human proα1(I) collagen gene. Proc. Natl. Acad. Sci. USA 83: 6045-6047.

Cohn, D.H., Apone, S., Eyre, D.R., Starman, B.J., Andreassen, P., Charbonneau, H., Nicholls, A.C., Pope, F.M. and Byers, P.H. (1988) Substitution of cysteine for glycine within the carboxyl-terminal telopeptide of the α1 chain of type I collagen produces mild osteogenesis imperfecta. J. Biol. Chem. 263: 14605-14607.

Conner, B.J., Reyes, A.A., Morin, C., Hakura, K., Teplitz, R.L. and Wallace, R.B. (1983) Detection of sickle-cell β^S-globin allele by hybridisation with synthetic oligonucleotides. Proc. Natl. Acad. Sci. USA 80: 278-282.

Constantinou, C.D., Pack, M., Nielson, K.B. and Prockop, D.J. (1988) Substitution of cysteine for glycine 904 of the α1(I) chain in a lethal variant of osteogenesis imperfecta. The asymptomatic mother is a carrier of a different mutation in type I procollagen. Am. J. Hum. Genet. 43 A180 (abstract).

Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988) Reactivity of cytosine and thymine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc. Natl. Acad. Sci. USA 85: 4397-4401.

Crouse, G.F., Frischauf, A. and Lehrach, H. (1983) An integrated and simplified approach to cloning into plasmid and single-stranded phages. Meth. Enz. 101: 78-89.

D'alessio, M., Weil, D., Prince, J., Bateman, J., Cole, W., Hollister, D. and Ramirez, F. (1988) Differential expression of a splicing mutation in the two type I procollagen genes.
Am. J. Hum. Genet. 43: A181 (Abstract).

Dalgleish, R., Trapnell, B.C., Crystal, R.G. and Tolstoshev, P. (1982) Copy number of a human type I α2 collagen gene.
J. Biol. Chem. 257: 13816-13822.

Dalgleish, R., Woodhouse, M. and Reeders, S. (1985) An RFLP associated with the human type III collagen gene (COL3A1).
Nucleic Acids Res. 13: 4609.

Dalgleish, R., Williams, G. and Hawkins, J.R. (1986) Length polymorphism in the proα2(I) collagen gene: An alternative explanation in a case of Marfan syndrome.
Hum. Genet. 73: 91-92.

Dalgleish, R., Hawkins, J.R. and Keston, M. (1987a) Exclusion of the α2(I) and α1(III) collagen genes as the mutant loci in a Marfan syndrome family.
J. Med. Genet. 24: 148-151.

Dalgleish, R. (1987b) Southern blotting.
In gene cloning and analysis. Ed. G.J. Boulnois. pp 45-60. Blackwell scientific publications. Oxford.

Deak, S.B., Nicholls, A., Pope, F.M. and Prockop, D.J. (1983) The molecular defect in a nonlethal variant of osteogenesis imperfecta.
J. Biol. Chem. 258: 15192-15197.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX.
Nucleic Acids Res. 12: 387-396.

Dickson, L.A., Pihlajaniemi, T., Deak, S., Pope, F.M., Nicholls, A., Prockop, D.J. and Myers, J.C. (1984) Nuclease S1 mapping of a homozygous mutation in the carboxyl-propeptide-coding region of the proα2(I) collagen gene in a patient with osteogenesis imperfecta.
Proc. Natl. Acad. Sci. USA 81: 4524-4528.

Dickson, L.A., de Wet, W., Liberto, M.D., Weil, D. and Ramirez, F. (1985) Analysis of the promoter region and N-propeptide domain of the human pro-α2(I) collagen gene.
Nucleic Acids Res. 13: 3427-3438.

Driesel, A.J., Schumacher, A.M. and Flavell, R.A. (1982) A HindIII restriction site polymorphism in the human collagen α1(I)-like gene of chromosome 7.
Hum. Genet. 62: 175-176.

Duncan, K.G., Fessler, L.I., Bachinger, H.P. and Fessler, J.H. (1983) Procollagen IV. Association to tetramers.
J. Biol. Chem. 262: 17724-17727.

Dunn, J.M., Phillips, R.A., Becker, A.J. and Gallie, B.L. (1988) Identification of germline and somatic mutations affecting the Retinoblastoma gene.
Science 241: 1797-1800.

Dunning, A.M., Talmud, P. and Humphries, S.E. (1988) Errors in the polymerase chain reaction
Nucleic Acids Res. 16: 10393.

Engvall, E., Hessle, H., and Klier, G. (1986) Molecular assembly, secretion and matrix deposition of type VI collagen.
J. Cell Biol. 102: 703-710.

Emanuel, B.S., Cannizzaro, L.A., Seyer, J.M. and Myers, J.C. (1985a) Human α1(III) and α2(V) procollagen genes are located on the long arm of chromosome 2.
Proc. Natl. Acad. Sci. USA 82: 3385-3389.

Emanuel, B.S., Cannizzaro, N.L., Ornstein-Goldstein, N., Indik, Z.K., Yoon, K., May, M., Oliver, L., Boyd, C. and Rosenbloom, J. (1985b) Chromosomal localization of the human elastin gene
Am. J. Hum. Genet. 37: 873-882.

Eyre, D.R., Shapiro, F.D. and Aldridge, J.F. (1985) A heterozygous collagen defect in a variant of the Ehlers-Danlos syndrome type VII.
J. Biol. Chem. 260: 11322-11329.

Eyre, D.R. Upton, M.P., Shapiro, F.D., Wilkinson, R.H. and Vawter, G.F. (1986) Non-expression of cartilage type II collagen in a case of Langer-Saldino achondrogenesis.
Am. J. Hum. Genet. 39: 59-67.

Feinberg, A.P. and Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity.
Anal. Biochem. 137: 266-267.

Fessler, J.H. and Fessler, L.I. (1978) Biosynthesis of procollagen.
Ann. Rev. Biochem. 47: 129-162.

Fessler, L.I., Shigaki, N. and Fessler, J.H. (1985) Isolation of a new procollagen V chain from chick embryo tendon.
J. Biol Chem. 260: 13286-13293.

Fischer, S.G. and Lerman, L.S. (1983) DNA fragments differing by single base substitutions are separated by denaturing gradient gels: Correspondence with melting theory.
Proc. Natl. Acad. Sci. USA 80: 1579-1583.

Focht, R.J. and Adams, S.L. (1984) Tissue specificity of type I collagen gene expression is determined at both transcriptional and post-transcriptional levels.
Mol. Cell. Biol. 4: 1843-1852.

Forrester, K., Almoguera, C., Han, K., Grizzle, W.E. and Perucho, M. (1987) Detection of high incidence of K-ras oncogenes during human colon tumorigenesis.
Nature 327: 298-303.

Francomano, C.A., Liberfarb, R.M., Hirose, T., Maumenee, I.H., Streeten, E.A., Meyers, D.A. and Pyeritz, R.E. (1987) The Stckler syndrome: evidence for close linkage to the structural gene for type II collagen.
Genomics I; 293-296.

Fraser, R.D.B., Macrae, T.P., Miller, A., and Suzuki, E. (1983) Molecular conformation and packaging in collagen fibrils. J. Mol. Biol. 167: 497-521.

Freeman, G.J. and Huang, A.S. (1981) Mapping temperature-sensitive mutants of vesicular stomatis virus by RNA heteroduplex formation. J. Gen. Virol. 57: 103-117.

Furuto, D.K. and Miller, E.J. (1981) Characterization of a unique collagenous fraction from limited pepsin digests of human placental tissue: molecular organization of the native aggregate. Biochemistry 20: 1635-1640.

Furuto, D.K. and Miller, E.J. (1983) Different levels of glycosylation contribute to the heterogeneity of α1(II) collagen chains derived from a transplantable rat chondrosarcoma. Arch. Biochem. 134: 320-324.

Gerstenfeld, L.C., Crawford, D.R., Boedtker, H. and Doty, P. (1984) Expression of type I and III collagen genes during differentiation of chick embryonic chick myoblasts in culture. Mol. Cell. Biol. 4: 1483-1492.

Gibbs, R.A. and Caskey, C.T. (1987) Identification and localization of mutations at the Lesch-Nyhan locus by ribonuclease A cleavage. Science 236: 303-305.

Goldring, M.B. and Krane, S.M. (1987) Modulation by recombinant interleukin I of synthesis of types I and III collagens and associated procollagen mRNA levels in cultured human cells. J. Biol. Chem. 262: 16724-16729.

Gordon, M.K., Gerecke, D.R. and Olsen, B.R. (1987) Type XII collagen: Distinct extracellular matrix component discovered by cDNA cloning. Proc. Natl. Acad. Sci. USA 84: 6040-6044.

Grant, W.T., Wang, G-J. and Balia, G. (1987) Type X collagen synthesis during endochondral ossification in fracture repair. J. Biol. Chem. 262: 9844-9849.

Green, M.R., Maniatis, T. and Melton, D.A. (1983) Human beta-globin pre-mRNA synthesized *in vitro* is accurately spliced in *Xenopus* oocyte nuclei.
Cell 32: 681-694.

Griffin, C.A., Emanuel, B.S., Hansen, J.R., Cavenee, W.K. and Myers, J.C. (1987) Human genes encoding basement membrane α1(IV) and α2(IV) chains map to the distal long arm of chromosome 13.
Proc. Natl. Acad. Sci USA 84: 512-516.

Grobler-Rabie, A.F., Brebner, D.K., Vandemplas, S., Wallis, G., Dalgleish, R., Kaufman, R.E., Mattew, C.G.P. and Boyd, C.D. (1985a) Polymorphism of DNA sequence in the human proα2(I) collagen gene.
J. Med. Genet. 22: 182-186.

Grobler-Rabie, A.F., Wallis, G., Brebner, D.K., Beighton, P., Bester, A.J. and Mattew, C.G. (1985b) Detection of a high frequency *RsaI* polymorphism in the human proα2(I) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta.
EMBO J. 4: 1745-1748.

Halme, T., Savunen, T., Aho, H., Vihersaari, T. and Penttinen, R. (1985) Elastin and collagen in the aortic wall: Changes in the Marfan syndrome and annuloaortic ectasia.
Exp. and Mol. Path. 43: 1-12.

Hanahan, D. and Meselson, M. (1983) Plasmid screening at high density.
Meth. Enz. 100: 333-342.

Harbers, K., Kuehn, M., Delius, H. and Jaenisch, R. (1984) Insertion of retrovirus into the first intron of α1(I) collagen gene leads to lethal mutation in mice.
Proc. Natl. Acad. Sci USA 81: 1505-1508.

Hartung, S., Jaenisch, R. and Breindl, M. (1986) Retrovirus insertion inactivates mouse α1(I) collagen gene by blocking initiation of transcription.
Nature 320: 365-367.

Henderson, A.S., Myers, J.C. and Ramirez, F. (1983) Localization of the human α2(I) collagen gene (COL1A2) to chromosome 7q22.
Cytogenet. Cell Genet. 36: 586-587.

Henke, E., Leader, M., Tajima, S., Pinnell, S. and Kaufman, R. (1985) A 38 base pair insertion in the proα2(I) collagen gene of a patient with Marfan syndrome.
J. Cell. Biochem. 27: 169-174.

Henry, I., Jeanpierre, M., Bernard, M., Weil, D., Grzeschik, K.H., Ramirez, F., Chu, M-L. and Junien, C. (1985) The structural gene for fibronectin (FN) maps to 2p323-qter.
Cytogenet. Cell Genet. 40: 650 (abstract).

Henry, I., Bernheim, A., Bernard, M., van der Rest, M., Kimura, T., Jeanpierre, C., Barichard, F., Berger, R., Olsen, B.R., Ramirez, F. and Junien, C. (1988) Mapping of a human fibrillar collagen gene, pro α1(XI) (COL11A1), to the p21 region of chromosome 1.
Genomics 3: 87-90.

Hofmann, H., Fietzek, P.P. and Kuhn, K.J. (1980) Comparative analysis of the sequences of the three collagen chains α1(I), α2(I) and α1(III).
J. Mol. Biol. 141: 293-314.

Huerre, C., Junien, C., Weil, D., Chu, M-L., Morabito, M., N'Guyen, V.C., Myers, J.C., Foubert, C., Gross, M.S., Prockop, D.J., Boue, A., Kaplan, J.C., de la Chapelle and Ramirez, F. (1982) Human type I procollagen genes are located on different chromosomes.
proc. Natl. Acad. Sci. USA 79: 6627-6630.

Huerre-Jeanpierre, C., Henry, I., Bernard, P., Galano, P., Weil, D., Grzeschik, K.H., Ramirez, F. and Junien, C. (1986) The proα2(V) collagen gene (COL5A2) maps to 2q14-2q32, syntenic to the proα1(III) collagen locus (COL3A1).
Hum. Genet. 73: 64-67.

Hutchinson, K.W. and Halverson, H.O. (1980) Cloning of randomly sheared DNA fragments from a lysogen of Bacillus subtilis. Identification of prophage-containing clones.
Gene 8: 267-278.

Ignotz, R.A., Endo, T. and Massague, J. (1987) Regulation of fibronectin and type I collagen mRNA levels by transforming growth factor-B.
J. Biol. Chem. 262: 6443-6446.

Ish-Horowitz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning.
Nucleic Acids Res. 9: 2989-2998.

Kapoor, R., Bornstein, P. and Sage, H. (1986) Type VIII collagen from bovine Descemet's membrane: structural characterisation of a triple-helical domain.
Biochemistry 25: 3930-3937.

Keene, D.R., Sakai, L.Y., Lunstrum, G., Morris, N.P. (1987) Type VII collagen forms an extended network of anchoring fibrils.
J. Cell Biol. 104: 611-621.

Khillan, J.S., Schmidt, A., Overbeek, P.A., de Crombrugghe, B. and Westphal, H. (1986) Developmental and tissue specific expression directed by the α2 type I collagen promoter in transgenic mice.
Proc. Natl. Acad. Sci. USA 83: 725-729.

Kosher, R.A., Kulyk, W.M. and Gay, S.W. (1986) Collagen gene expression during limb cartilage differentiation.
J. Cell. Biol. 102: 1151-1156.

Kraft, R., Tardiff, J., Krauter, K.S. and Leinwand, L.A. (1988) Using mini-prep plasmid DNA for sequencing double-stranded templates with sequenase.
Biotechniques 6: 544-547.

Kuivaniemi, H., Sabol, C., Tromp, G. Sippola-Thiele, M. and Prockop, D.J. (1988) A 19 base pair deletion in the pro-α2(I) gene of type I procollagen that causes in-frame RNA splicing from exon 10 to exon 12 in a proband with atypical osteogenesis imperfecta and his asymptomatic mother.
J. Biol. Chem. 263: 11407-11413.

Laitinen, O., Uitto, J., Iivanainen, M., Hannuskela, M. and Kivirikko, K.I. (1968) Collagen metabolism of the skin in Marfan's syndrome.
Clin. Chim. Acta. 21: 321-326.

Law, M.L. Tung, L., Morse, H.G., Berger, R., Jones, C., Cheah, K.S.E. and Solomon, E. (1986) The human type II collagen gene (COL2A1) assigned to 12q14.3.
Ann. Hum. Genet. 50: 131-137.

Lee, B., Vissing, H., Ramirez, F., Rogers, D. and Rimoin, D. (1989) Identification of the molecular defect in a family with spondyloepiphyseal dysplasia.
Science 244: 978-980.

Lohler, J., Timpl, R. and Jaenisch, R. (1984) Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoetic and mesenchymal cell death.
Cell 38: 597-607.

Lopez-Galindez, C., Lopez, J.A., Melero, J.A., de la Fuente, L., Conception, M., Ortin, J., Perucho, M. (1988) Analysis of genetic variability and mapping of point mutations in influenza virus by the RNase A mismatch cleavage method.
Proc. Natl. Acad. Sci. USA 85: 3522-3526.

Lovell-Badge, R.H., Bygrave, A., Bradley, A., Robertson, E., Tilly, R. and Cheah K.S.E. (1987) Tissue-specific expression of the human type II collagen gene in mice.
Proc. Natl. Acad. Sci. USA 82: 4050-4054.

Lyons, B.L. and Schwartz, R.I. (1984) Ascrobate stimulation of PAT cells causes an increase in transcription rates and a decrease in degradation rates of procollagen mRNA.
Nucleic Acids Res. 12: 2569-2579.

Mayne, R. and Burgeson, R.E (eds.) (1987) Structure and function of collagen types.
Academic Press Inc.

McLachlen, A.D. (1976) Evidence for gene duplication in collagen.
J. Mol. Biol. 107: 159-174.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering.
Protein Engineering 1: 67-74.

Melton, D.A., Krieg, P.A., Maniatis, T., Zinn, K. and Green, M.R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter.
Nucleic Acids Res. 12: 7035-7056.

Merlino, G.T., McKeon, C., de Crombrugghe, B. and Pastan, I. (1983) Regulation of the expression of genes encoding types I, II, and III collagen during chick embryonic development.
J. Biol. Chem. 258: 10041-10048.

Messing, J. (1983) New M13 vectors for cloning.
Meth. Enz. 101: 20-89.

Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either strand of double-digest restriction endonuclease fragments.
Gene 19: 269-276.

Miller, E.J. (1985) The structure of the fibril-forming collagens.
Ann. N. Y. Acad. Sci. 460: 1-13.

Miller, E.J. and Gay, S. (1987) The collagens: an overview and update.
Meth. Enzym. 144: 3-41.

Miskulin, M., Dalgleish, R., Kluve-Beckerman, B., Rennard, S.I., Tolstoshev, P., Brantly, M. and Crystal, R.G. (1986) Human type III collagen expression is coordinately modulated with type I collagen genes during fibroblast growth.
Biochemistry 25: 1408-1413.

Morris, N.P., Keene, D.R., Glanville, R.W., Bentz, H. and Burgeson, R. (1986) The tissue form of type VII collagen is an antiparallel dimer.
J. Biol. Chem. 262: 11345-11350.

Morton, H.J. (1970) A survey of commercially available tissue culture media.
In Vitro 6: 89-108.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) Specific enzymatic amplification of DNA In Vitro: The polymerase chain reaction.
Cold Spring Harbor Symp. Quant. Biol. LI : 263-273.

Myers, R.M., Fischer, S.G., Maniatis, T. and Lerman, L.S. (1985a) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis.
Nucleic Acids Res. 13: 3111-3129.

Myers, R.M., Fischer, S.G., Lerman, L.S. and Maniatis, T. (1985b) Nearly all single base substitutions in DNA fragments attached to a GC-clamp can be detected by denaturing gradient gel electrophoresis.
Nucleic Acids Res. 13: 3131-3145.

Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985c) Detection of single base substitutions in total genomic DNA.
Nature 313: 495-498.

Myers, R.M., Larin, Z. and Maniatis, T. (1985d) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes.
Science 230: 1242-1246.

Myllyla, R., Koivu, J., Pihlajaniemi, T. and Kivirikko, K.I. (1983) Protein disulphide-isomerase activity in various cells synthesizing collagen.
Eur. J. Biochem. 134: 7-11.

Nemethy, G. and Scheraga, H.A. (1986) Stabilisation of collagen fibrils by hydroxyproline.
Biochemistry 25: 3148-3188.

Nicholls, A.C., Pope, F.M. and Graig, D. (1984). An abnormal collagen α chain containing a cysteine in autosomal dominant osteogenesis imperfecta.
Br. Med. J. 284: 112-113

Nicholls, A.C., De Paepe, A., Narcisi, P., Dalgleish, R., De Keyser, F., Matton, M. and Pope, F.M. (1988) Linkage of a polymorphic marker for the type III collagen gene (COL3A1) to atypical autosomal dominant Ehlers-Danlos syndrome type IV in a large Belgian pedigree.
Hum. Genet. 78: 276-281.

Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. and Olsen, B.R. (1986) The developmentally regulated type X collagen gene contains a long open reading frame without introns.
J. Biol. Chem. 261: 5041-5050.

Noll, W.W. and Collins, M. (1987) Detection of human DNA polymorphisms with a simplified denaturing gel electrophoresis technique.
Proc. Natl. Acad. Sci. USA 84: 3339-3343.

Novack, D.F., Casna, N.J., Fischer, S.G. and Frod, J.P. (1986) Detection of single base-pair mismatches in DNA by chemical modification followed by electrophoresis in 15% polyacrylamide gel.
Proc. Natl. Acad. Sci. USA 83: 586-590.

Ogilvie, D.J., Wordsworth, B.P., Preistly, L.J., Dalgleish, R., Schmidtke, J., Zoll, B. and Sykes, B.C. (1987a) Segregation of all four major fibrillar collagen genes in the Marfan syndrome.
Am. J. Hum. Genet. 41: 1071-1082.

Ogilvie, D.J., Aitchison, K. and Sykes, B.C. (1987b) An RFLP close to the human collagen I gene COL1A1.
Nucleic Acids Res. 15: 4699.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms.
Proc. Natl. Acad. Sci. USA 86: 2766-2770.

Paglia, L., Wilczek, J., de Leon, L.D., Martin, G.R., Horlein, D. and Muller, P. (1979) Inhibition of procollagen cell-free synthesis by amino-terminal extension peptides.
Biochemistry 18: 5030-5034.

Pihlajaniemi, T., Dickson, L.A., Pope, F.M., Korhonen, V.R., Nicholls, A., Prockop, D.J. and Myers, J.C. (1984) Osteogenesis imperfecta: Cloning of a pro-α2(I) collagen gene with a frameshift mutation.
J. Biol. Chem. 259: 12941-12944.

Pihlajaniemi, T., Myllyla, R., Seyer, J., Kurkinen, M. and Prockop, D.J. (1987) Partial characterisation of a low molecular weight human collagen that undergoes alternative splicing.
Proc. Natl. Acad. Sci. USA 84: 940-944.

Pingoud, A. (1985) Spermidine increases the accuracy of type II restriction endonucleases.
Eur. J. Biochem. 147: 105-109.

Pinnel, S.R., Krane, S.M., Kenzora, J.E. and Glimcher, M.J. (1972) A heritable disorder of connective tissue (Hydroxylysine-deficient collagen disease).
N. Engl. J. Med 286: 1013-1020.

Pinnel, S.R. and Murad, S. (1983) Disorders of collagen. In Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S., eds.
The metabolic basis of inherited disease. 5th ed. New York: McGraw-Hill 1425-1449.

Poschl, E., Pollner, R. and Kuhn, K. (1988) The genes for the α1(IV) and α2(IV) chains of human basement collagen type IV are arranged head-to-head and are separated by a bidirectional promoter of unique structure.
EMBO J. 7: 2687-2695.

Priest, R.E. Moinuddin, J.F. and Priest, J.H. (1973) Collagen of Marfan syndrome is abnormally soluble.
Nature 245: 264-266.

Prockop, D.J. (1984) Osteogenesis imperfecta: Phenotypic heterogeneity, protein suicide, short and long collagen.
Am. J. Hum. Genet. 36: 499-505.

Pyeritz, R.E. and McKusick, V.A. (1979) The Marfan syndrome: Diagnosis and management.
New Engl. J. Med. 300: 772-777.

Pyeritz, R.E. and McKusick, V.A. (1981) Basic defects in the Marfan syndrome.
New Engl. J. Med. 305: 1011-1012.

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes.
Nucleic Acids Res. 13: 7207-7221.

van der Rest, M., Rosenberg, L.C., Olsen, B.R. and Poole, A.R. (1986) Chondrocalcin is identical with the C-propeptide of type II collagen.
Biochem. J. 237: 923-925.

van der Rest, M. and Mayne, R. (1988) Type IX collagen proteoglycan from cartilage is covalently cross-linked to type II collagen.
J. Biol. Chem. 263: 1615-1618.

Retief, E., Parker, M.I. and Retief, A.E. (1985) Regional chromosome mapping of human collagen genes alpha 2(I) and alpha 1(I) (COL1A2 and COL1A1)
Hum. Genet. 69: 304-308.

Rossi, P. and de Crombrugghe, B. (1987a) Formation of a type I collagen RNA dimer by intermolecular base-pairing of a conserved sequence around the translation initiation site.
Nucleic Acids Res. 15: 8935-8956.

Rossi, P. and de Crombrugghe, B. (1987b) Identification of a cell-specific transcriptional enhancer in the first intron of the mouse α2 (type I) collagen gene.
Proc. Natl. Acad. Sci. USA 84: 5590-5594.

Rossouw, C.M.S., Vegeer, W.P., Du Plooy, S., Bernard, M.P., Ramirez, F. and de Wet, W. (1987) DNA sequences in the first intron of the human pro-α1(I) collagen gene enhance transcription.
J. Biol. Chem. 262: 15151-15157.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986) Analysis of enzymically ampified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes.
Nature 324: 163-166.

Scharf, S.J., Horn, G.T. and Erlich, H.A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences.
Science 233: 1076-1078.

Schnieke, A., Harbers, K. and Jaenisch, R. (1983) Embryonic lethal mutation in mice induced by retrovirus insertion into the α1(I) collagen gene.
Nature 304: 315-320.

Shenk, T.E., Rhodes, C., Rigby, R.W.J. and Berg, P. (1975) Biochemical method for mapping mutational alterations in DNA with S1 nuclease: The location of deletions and temperature sensitive mutations in simian virus 40.
Proc. Natl. Acad. Sci. USA 72: 989-993.

Sillence, D.O., Senn, A. and Danks, D.M. (1979) Genetic heterogeneity in osteogenesis imperfecta.
J. Med. Genet. 16: 101-116.

Sippola, M., Kaffe, S. and Prockop, D.J. (1984) A heterozygous defect for structurally altered pro-α2(I) chain of type I procollagen in a mild variant of osteogenesis imperfecta.
J. Biol. Chem. 259: 14094-14100.

Smith, F.I., Parvin, J.D. and Palese, P. (1986) Detection of single base substitutions in influenza virus RNA molecules by denaturing gradient gel electrophoresis of RNA-RNA or DNA-RNA heteroduplexes.
Virology 150: 55-64.

Soberon, X., Corvarribias, L. and Bolivar, F. (1980) Construction and characterisation of new cloning vehicles IV: Deletion derivatives of pBR322 and pBR325.
Gene 9: 287-305.

Soininen, R., Huotari, M., Hostikka, S.L., Prockop, D.J. and Tryggvason, K. (1988) The structural genes for α1 and α2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.
J. Biol. Chem. 263: 17217-17220.

Solomon, E., Hiorns, L., Dalgleish, R., Tolstoshev, P., Crystal, R. and Sykes, B. (1983) Regional localisation of the human α2(I) collagen gene on chromosome 7 by molecular hybridisation.
Cytogenet. Cell Genet. 35: 64-66

Solomon, E., Hiorns, L.R., Spurr, N., Kukinen, M., Barlow, D., Hogan, B.L.M. and Dalgleish, R. (1985) Chromosomal assignments of the genes coding for human type II, III and IV collagen: A dispersed gene family.
Proc. Natl. Acad. Sci. USA 82: 3330-3334.

Solter, D. (1988) Differential imprinting and expression of maternal and paternal genomes.
Annu. Rev. Genet. 22: 127-146.

Superti-Furga, A., Gugler, E., Gitzelmann, R. and Steinmann, B. (1988) Ehlers-Danlos syndrome type IV: A multi-exon deletion in one of the two COL3A1 alleles affecting structure, stability, and processing of type III procollagen.
J. Biol. Chem. 263: 6226-6232.

Sykes, B., Oglvie, D., Wordsworth, P., Anderson, J. and Jones, N. (1986) Osteogenesis imperfecta is linked to both type I collagen structural genes.
Lancet 2: 69-72.

Tajima, S., Ting, J.P.Y., Pinnel, S.R. and Kaufman, R.E. (1984) Isolation and characterisation of a human proα2(I) collagen gene segment.
J. Invest. Dermatol. 82: 265-269.

Thompson, E.M., Young, I.D., Hall, C.M. and Pembrey, M.E. (1987) Recurrence risks and prognosis in severe sporadic osteogenesis imperfecta.
J. Med. Genet. 24: 390-405.

Tromp, G. and Prockop, D.J. (1988) Single base mutation in the proα2(I) collagen gene that causes efficient spicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame proα2(I) chain.
Proc. Natl. Acad. Sci. USA 85: 5254-5258

Tromp, G., Kuivaniemi, H., Shikata, H. and Prockop, D.J. (1989) A single base mutation that substitutes serine for glycine 790 of the .α1(III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome IV.
J. Biol. Chem. 264: 1349-1352.

Tsipouras, P., Myers, J.C., Ramirez, F. and Prockop, D.J. (1983) Resriction fragment length polymorphism associated with the proα2(I) gene of type I procollagen: Application to a family with an autosomal dominant form of osteogenesis imperfecta.
J. Clin. Invest. 72: 1262-1267.

Tsipouras, P., Borresen, A-L., Bamforth, S., Harper, P.S. and Berg, K. (1986a) Marfan syndrome: exclusion of genetic linkage to the COL1A2 gene.
Clin. Genet. 30: 428-432.

Tsipouras, P., Byers, P.H., Schwartz, R.C., Chu, M-L., Weil, D., Pepe, G., Cassidy, S.B. and Ramirez, F. (1986b) Ehlers-Danlos syndrome type IV: cosegregation of the phenotype with to a COL3A1 allele of type III procollagen.
Hum. Genet. 74: 41-46.

Veres, G., Gibbs, R.A., Scherer, S.E. and Caskey, C.T. (1987) The molecular basis of the sparse fur mouse mutation.
Science 237: 415-417.

Vogel, B.E., Minor, R.R., Freund, M. and Prockop, D.J. (1987) A point mutation in a type I procollagen gene converts glycine 748 of the α1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta.
J. Biol. Chem. 262: 14737-14744

Vogel, B.E., Doelz, R., Kadler, K.E., Hojima, Y., Engel, J. and Prockop, D.J. (1988) A substitution of cysteine for glycine 748 of the α1 chain produces a kink at this site in the procollagen I molecule and an altered N-proteinase cleavage site over 225nm away.
J. Biol. Chem. 19249-19255.

de Vries, W.N. and de Wet, W.J. (1986) The molecular defect in an autosomal dominant form of osteogenesis imperfecta.
J. Biol. Chem. 261: 9056-9064.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) Hybridisation of synthetic oligodeoxyribonucleotide to ΦX174 DNA : the effect of single base pair mismatch.
Nucleic Acids Res. 6: 3543-3557.

Walsh, M.J., Leleiko, N.S. and Sterling, Jr.,K.M. (1987) Regulation of types I, III and IV procollagen mRNA synthesis in glucocorticoid-mediated intestinal development.
J. Biol. Chem. 262: 10814-10818.

Weil, D., Bernard, M., Gargano, S. and Ramirez, F. (1987) The pro-α2(V) collagen gene is evolutionarily related to the major fibrillar-forming collagens.
Nucleic Acids Res. 15: 181-198.

Weil, D., Mattei, M-G., Passage, E., Van Cong, N., Pribula-Conway, D., Mann, K., Deutzmann, R., Timpl, R. and Chu, M-L. (1988a) Cloning and chromosomal localisation of human genes encoding the three chains of type VI collagen.
Am. J. Hum. Genet. 42: 435-445.

Weil, D., Bernard, M., Combates, N., Wirtz, M.K., Hollister, D.W., Steinmannn, B. and Ramirez, F. (1988b) Identification of a mutation that causes exon skipping during collagen pre-mRNA spicing in an Ehlers-Danlos syndrome variant.
J. Biol. Chem. 263: 8561-8564.

Weil, D., D'Alessio, M., Ramirez, F., de Wet, W., Cole, W.G., Chan, D. and Bateman, J.F. (1989) A base substitution in the exon of a collagen gene causes alternative splicing and generates a structurally abnormal polypeptide in a patient with Ehlers-Danlos syndrome type VII.
EMBO J. 8: 1705-1710.

Weintraub, H. and Groudine, M. (1976) Chromosomal subunits in active genes have an altered conformation.
Science 193: 848-856

Weiss, E.H., Cheah, K.S.E., Grosveld, F.G., Dahl, H.H.M., Solomon, E. and Flavell, R.A. (1982) Isolation and characterization of a human collagen α1(I)-like gene from a cosmid library.
Nucleic Acids Res. 10: 1981-1994.

Wenstrup, R.J., Cohn, D.H., Cohen, T. and Byers, P.H. (1988) Arginine for glycine substitution in the triple-helical domain of the products of one α2(I) collagen allele (COL1A2) produces the osteogenesis imperfecta type IV phenotype.
J. Biol. Chem. 263: 7734-7740.

de Wet, W., Chu, M-L. and Prockop, D.J. (1983a) The mRNAs for the pro-α1(I) and pro-α2(I) chains of type I procollagen are translated at the same rate in normal human fibroblasts and in fibroblasts from two variants of osteogenesis imperfecta with altered steady state ratios of the two mRNAs.
J. Biol. Chem. 258: 14385-14389.

de Wet, W., Pihlajaniemi, T., Myers, J., Kelly, T.E. and Prockop, D.J. (1983b) Synthesis of a shortened pro-α2(I) chain and decreased synthesis of pro-α2(I) chains in a proband with osteogenesis imperfecta.
J. Biol. Chem. 258: 7721-7728.

de Wet, W., Sippola, M., Tromp, G., Prockop, D., Chu, M-L. and Ramirez, F. (1986) Use of R-loop mapping for the assessment of human collagen mutations.
J. Biol. Chem. 261: 3857-3862.

Wiestner, M., Krieg, T., Horlein, D., Glanville, R.W., Fietzek, P. and Muller, P.K. (1979) Inhibiting effect of procollagen peptides on collagen biosynthesis in fibroblast cultures.
J. Biol. Chem. 254: 7016-7023.

Williams, C.J. and Prockop, D.J. (1983) Synthesis and processing of a type I procollagen containing shortened pro-α1(I) chains by fibroblasts from a patient with osteogenesis imperfecta.
J. Biol. Chem. 258: 5915-5921.

Willing, M.C., Cohn, D.H., Starman, B., Holbrook, K.A., Greenberg, C.R. and Byers, P.H. (1988) Heterozygosity for a large deletion in the α2(I) collagen gene has a dramatic effect on type I collagen secretion and produces perinatal lethal osteogenesis imperfecta.
J. Biol. Chem. 263: 8398-8404.

Winter, E., Yamamoto, F., Almoguera, C. and Perucho, M. (1985) A method to detect and characterise point mutations in transcribed genes: Amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells.
Proc. Natl. Acad. Sci. USA 82: 7575-7579.

Wong, Z., Wilson, V., Jeffreys, A.J. and Thein, S.L. (1986) Cloning a selected fragment from a DNA 'fingerprint': isolation of an extremely polymorphic minisatellite.
Nucleic Acids Res. 14: 4605-4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) characterisation of a panel of highly variable minisatellites cloned from human DNA.
Ann. Hum. Genet. 51: 269-288.

Woodbury, D., Benson-Chanda, V. and Ramirez, F. (1989) Amino-terminal propeptide of human pro-α2(V) collagen conforms to the structural criteria of a fibrillar procollagen molecule.
J. Biol. Chem. 264: 2735-2738.

Wozney, J., Hanahan, D., Tate, V., Boedtker, H. and Doty, P. (1981) Structure of the pro-α2(I) collagen gene.
Nature 294: 129-135.

Yamada, Y., Avvedentimento, V.E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. and de Crombrugghe, B. (1980) The collagen gene: evidence of its evolutionary assembly by amplification of a DNA segment containing an exon of 54bp.
Cell 22: 887-892.

Yamada, Y., Kuhn, K. and de Crombrugghe, B. (1983) A conserved nucleotide sequence coding for a segment of the C-propeptide is found at the same location in different collagen genes.
Nucleic Acids Res. 11: 2733-2744.

Zinn, K., DiMaio, D. and Maniatis, T. (1983) Identification of two distinct regulatory regions adjacent to the human β-interferon gene.
Cell 34: 865-879.

ABSTRACT

The Molecular Genetics of Inherited Connective Tissue Disorders.

John Ross Hawkins

The collagens are a family of structural proteins which function as an extracellular framework in eukaryotic organisms. They are characterised by a unique protein conformation which consists of three polypeptide chains in a triple helix. In vertebrates, at least thirteen collagen types encoded by at least twenty non-allelic genes have been identified.

The collagen genes constitute a multi-gene family with a probable common evolutionary origin. Though generally dispersed, some of the genes are known to be clustered on certain chromosomes.

Alterations in collagen genes can result in a heterogeneous group of heritable disorders of the connective tissues. There is accumulating evidence that similar phenotypes are due to similar mutations or location of mutations.

One inherited disease of the connective tissues is the Marfan syndrome (MS). The genetic basis of this disease is still unknown. The four genes of the fibrillar collagens types I, II and III were chosen as candidate genes for a study of linkage in a large MS family. Three of these genes were excluded as being the mutant gene in this family.

The heritable brittle bone disease, osteogenesis imperfecta (OI) is usually caused by mutations in either of the two genes for type I collagen. The technique of RNase A cleavage mapping was used to search for mutations in the α1(I) gene in the lethal type of OI.

The mutation in one individual was detected and mapped to exon 43 of the α1(I) gene (COL1A1). The mutant gene was amplified, cloned and sequenced revealing a 9bp deletion from a repetitive region of one allele. It is surprising that such a mutation could produce OI in its most severe form. Studies to verify that the deletion is the disease-cause remain to be undertaken.