THE MOLECULAR GENETICS OF HUMAN COLLAGEN GENES AND OSTEOGENESIS IMPERFECTA

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Karen Molyneux BSc (Liverpool)
Department of Genetics
University of Leicester

April 1991

UMI Number: U527804

UMI U527804

This thesis is dedicated to Moz.

ABSTRACT

The Molecular Genetics of Human Collagen Genes and Osteogenesis
Imperfecta

Karen Molyneux

The collagens are the most abundant proteins in the human body.
Their function is predominantly structural. The mature collagen protein is
composed of three polypeptides which form triple helical structures. The
genes coding for the collagens are large, complex and semi-dispersed
throughout the human genome. They are expressed in a tissue specific,
developmentally regulated manner. Analysis of the fibrillar collagen protein
and gene sequences has revealed a high degree of conservation.

A possible variant in the cDNA sequence coding for α1(III) was
investigated. The variant sequence was phenotypically silent and coded for
an amino acid in the carboxyl end of the helical domain. The variant base
was shown to be an artefact introduced during cDNA synthesis.

Many mutations causing the disease osteogenesis imperfecta (OI)
have been described in the literature. Most are point mutations causing
single amino acid substitutions. The mutation causing a Sillence type III OI
phenotype was determined. It was found to be an inframe 3bp deletion
which resulted in the loss of the last amino acid coded for by exon 19 of
COL1A2, a valine. The mutation, like many of those in the type I collagen
genes which cause OI, coincided in its position with the start of
overmodification of the patients type I collagen molecules.

15 collagen types have been at least partially characterized to date.
In an attempt to characterize new collagen cDNA sequences a library was
produced which was intended to be enriched with collagen coding
sequences. The production of such a library relied on the ability of a
polymerase chain reaction primer to selectively amplify cDNA sequences
coding for the helical domains of collagens. The library contained collagen
sequences, but not enough to make screening the library for new collagen
sequences by random sequencing of clones a viable proposition. This work
continues.

ACKNOWLEDGEMENTS

ABBREVIATIONS

DNA	deoxyribonucleic acid
cDNA	complementary deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
DNase	deoxyribonuclease
RNase	ribonuclease
ddNTP	2'3'-dideoxynucleoside triphosphates
dNTP	2'-deoxyribonucleoside triphosphate
NTP	nucleoside triphospate
PCR	polymerase chain reaction
SDS	sodium lauryl sulphate
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside
amp	ampicillin
IPTG	isopropyl-b-D-thiogalactopyranoside
DEPC	diethylpyrocarbonate
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane-sulphonic acid
PIPES	piperazine-N-N'-bis[2-ethane-sulphonic acid]
DTT	dithiothreitol
BSA	bovine serum albumin
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
MOPS	3-(N-morpholino) propane sulphonic acid
TEMED	N,N,N',N',-tetramethylethylenediamine
EDTA	diaminoethanetetra-acetic acid
TE	Tris-EDTA
bp	base pair
kb	kilo base pairs
Ci	Curie
cpm	counts per minute

CONTENTS

CHAPTER 1

INTRODUCTION

1.1   EXTRACELLULAR MATRIX OF CONNECTIVE TISSUE

Tissues are made up of cells embedded in extracellular matrix. This matrix is comprised of a variety of polysaccharides and proteins which are assembled as an organised network. Connective tissues are generally made up of more extracellular matrix than cells. The amount of connective tissue found in different parts of the body varies. Connective tissue is a major component of skin and bone and a minor component of the brain and spinal cord. It comes in many shapes and forms, from the hard calcified structure of bones and teeth, to the transparent matrix of the cornea and the ropelike, high tensile strength of tendons. The extracellular matrix was once thought to be inert but has since been shown to play a role in regulating the behaviour of the cells it is in contact with.

1.2   COLLAGENS

Major components of connective tissue extracellular matrix are the collagens. Fifteen types of collagen have been at least partially characterized, there may be more. Each collagen type displays a different tissue distribution, and almost certainly a different function, in the body. These functions are determined by the supramolecular structures formed by the collagens.

The collagens are a family of proteins, the members of which all contain one or more areas of triple helical domain. The molecules aggregate to form supramolecular structures which are predominantly used in the body for structural support but they also play active roles in developmental processes, cell attachment, chemotaxis (Postlethwaite *et al.*, 1978) and the binding of antigen-antibody complexes. Collagens are by far the most prevalent proteins in the human body, constituting 25% of total protein. They have been found in all multicellular organisms including sponges (Exposito and Garrone, 1990), sea urchins (D'Alessio *et al.*, 1989) and *C. elegans* (Cox *et al.*, 1989).

1

Each collagen molecule is made up of three polypeptides (α-chains) which may or may not be identical. Each of the three so called α-chains are coiled into a left handed helix with three amino acids per turn. The twisted α-chains are then twisted around each other into a right handed superhelix to form a long thin structure. The triple helix was once thought to resemble a rigid rod structure however, data published in Nestler *et al.* (1983) suggests it may be more appropriately characterized as a semi-flexible rod.

The primary structure of the helical domains of the collagen molecules is a series of Gly-X-Y repeats, where X is proline and Y is a hydroxyproline, around 30% of the time. A glycine at every third position is essential for the formation of the collagen helix. Glycine is the smallest amino acid and fits into the centre of the helix. The relatively high concentration of proline and hydroxyproline stabilizes the helical structure.

With one exception, members of the collagen family are genetically and structurally distinct proteins. They are the products of single locus genes, which are semi-dispersed throughout the human genome.

The collagens have been divided into three groups on the basis of their gross physical features (Miller and Gay, 1987).
GROUP I contains those collagens whose molecular weight is equal to or greater than 95,000 and which possess an uninterrupted helical domain of approximately 300nm in length.
GROUP II collagens have a molecular weight greater than or equal to 95,000 but do not possess an uninterrupted helical domain as long as 300nm.
GROUP III collagens have a molecular weight less than 95,000.

1.3   GROUP I COLLAGENS

The collagens included in this group are types I, II, III, V and XI. These collagens are also called fibrillar collagens because of a tendency of type I, II and III to aggregate into well defined fibre structures. Members of this group make up most of the collagen in the human body. It is probably due to this fact, and the fact that inherited connective tissue diseases such as osteogenesis imperfecta and Ehlers-Danlos syndrome have been found to be caused by mutations in the genes coding for members of this group, that most is known about these collagens.

1.3.1 TYPE I

Type I collagen is found ubiquitously in the human body. It most frequently occurs as a heterotrimer consisting of two αl(I) chains and one α2(I) chains. Homotrimers of αl(I) have been found to be synthesized by cells in culture (Mayne *et al.*, 1975), some tumours (Moro and Smith, 1977) and normal tissues (Wohllebe and Carmichael, 1978; Uitto, 1979a). The helical regions of αl(I) and α2(I) consist of 338 Gly-X-Y repeats (1014 residues).

Type I collagen molecules generally aggregate to form well-structured cross-striated fibres approximately 800-1600Å in diameter. The fibres are found in tissues such as bone, tendon, dentin, dermis and the uterine wall. The distribution of this collagen suggests it has a role to play as a supporting element in tissues.

1.3.2 TYPE II

This collagen is synthesized as a homotrimer of αl(II). It is found in fewer tissues than type I collagen. A rich source of type II collagen is hyaline cartilage, but the collagen is also found in other cartilage structures and in the vitreous body of the eye. Type II collagen molecules form fibres with a much smaller diameter than type I molecules. However, like type I collagen, type II collagen is synthesized as a precursor molecule. These so called procollagen molecules are processed extracellularly and, in doing so, lose one third of their molecular weight. The final collagen molecules are 96% helix. Type II collagen is, so far, unique amongst the fibrillar collagens in that it displays alternative splicing. A 69 amino acid cysteine-rich domain of the amino terminal propeptide is coded for by some species of αl(II) mRNA and not in others (Ryan and Sandell, 1990). The sequence is coded for by exon 2 of COL2A1. The two mRNAs are found in RNA extracted from fetal and juvenile chondrocytes but at different ratios. The alternative splicing is found in a region of the pro αl(II) termed COL1. Paglia *et al.* (1981) observed that that this same region when cleaved from type I collagen was capable of inhibiting αl(II) translation. This work leads to the question of whether this COL1 sequence has a role to play in translational control. αl(II) is a fibrillar collagen and as such is under constraints of molecular size. It is important to note that this alternative splicing does not alter the length or sequence of the mature αl(II) molecule, only that of the proαl(II) molecule.

1.3.3 TYPE III

The type III collagen molecules are homotrimers $[\alpha1(III)]_3$. They have a similar tissue distribution to type I collagen. Type III collagen fibres are found in the more flexible tissues in which type I collagen is found ie. dermis, vessel walls and uterine wall. Fibres derived from type III collagen appear as a fine network compared with type I fibres.

There is data to suggest that type I and type III collagen are coordinately expressed (Miskulin *et al.*, 1986). In fact some fibres have been found to contain both type I and type III collagen (Fleischmajer *et al.*, 1990).

Type III collagen is synthesized as a precursor molecule with extra sequence at the carboxyl and amino ends of the molecule but, unlike types I and II collagen type III collagen   often retains the extra sequence at the amino terminus.

1.3.4 TYPE V

Type V collagen is a minor component of smooth muscle, blood vessels, bone and cornea. There are three type V collagen α-chains, $\alpha1$, $\alpha2$, and $\alpha3$. This collagen occurs as a homotrimer $[\alpha1(V)]_3$ (Haralson *et al.*, 1980) and as two heterotrimers, $[\alpha1(V)]_2\alpha2(V)$ and $\alpha1(V)\alpha2(V)\alpha3(V)$ (Niyibizi *et al.*, 1984). Different cell types may synthesize different trimers. The helical domain of type V collagen is of a similar length to those of the other group I collagens but the procollagen molecule is much larger. This must be due to a much larger amino terminal sequence because the carboxyl propeptide sequence is conserved (Myers *et al.*, 1985). The processed molecule is a similar size to the other group I molecules, therefore the extra sequence is lost during processing.

1.3.5 TYPE XI

Type XI collagen is made up of three α-chains $\alpha1(XI)$, $\alpha2(XI)$ and $\alpha3(XI)$. $\alpha3(XI)$ is a product of the gene coding for $\alpha1(II)$ (COL2A1). The two α-chains differ only in their post-translational modification. The $\alpha3(XI)$ has an

4

elevated level of hydroxylysine linked carbohydrate residues compared with α1(II) (Furuto and Miller, 1983).

There is evidence to suggest that type XI collagen may interact with type II collagen fibrils to control their lateral growth. A similar role has been suggested for type V collagen with respect to type I fibres (Birk *et al.*, 1988). Niyibizi *et al.* (1989) suggested that collagen molecules can be formed using α1(XI), α1(V) and α2(V). COL11A1 gene expression has been detected in numerous tissues (Yoshioka and Ramirez, 1990).

1.4   THE STRUCTURAL SIMILARITIES OF THE GROUP I COLLAGENS

Group I collagens are synthesized by cells as precursor molecules (procollagens). The procollagen molecules are made up of a number of domains:-
1.   an amino propeptide
2.   an amino telopeptide
3.   an uninterrupted helical domain
4.   a carboxyl telopeptide
5.   a carboxyl propeptide

The amino propeptide is 139 amino acids long in pro α1(I) and consists of a globular domain, a short helical domain and a short non-helical, cysteine-rich, domain leading to the amino telopeptide. The amino propeptide regions in the other fibrillar collagens differ slightly from this ie. α2(I) lacks the cysteine-rich globular domain (Tate *et al.*, 1983). The major helical domains of these collagens display a high degree of homology (Miller, 1985) but the most conserved sequences are to be found in the carboxyl propeptide domain (Myers *et al.*, 1985). This may be because the carboxyl propeptide has a role to play in the choice of α-chains and the alignment of the α-chains to form a procollagen molecule.

The telopeptide domains contain the least conserved sequences among the fibrillar collagens. This is surprising since these regions have been found to interact with conserved sequences within the helix during fibrillogenesis (Helseth and Veis, 1981).

Sequence data from types V and IX collagen show them to have reduced homology to the other members of group I but they are still very similar.

The helical domains of the fibrillar collagens are of similar length and as has been mentioned before show sequence homology. The presence of proline and hydroxyproline in the X and Y positions stabilizing the helical structure. The other amino acids in the X and Y positions of the helix establish the correct array of charged and hydrophobic interaction sites, these are required for the correct aggregation of the molecules in fibre formation.

The fibrillar collagens associate laterally in a defined way until a certain diameter is reached. Fibre formation is an entropy-driven process that has the characteristics of crystallization (Kadler *et al.*, 1988). It is initiated by the cleavage of the carboxyl propeptide cleavage site (Kadler *et al.*, 1987). The carboxyl propeptide is thought to be removed before the amino propeptide in type I and III collagen. There does not seem to be an order in which they are removed in type II collagen (Uitto *et al.*, 1979b). The molecules which are laid down in the type I collagen fibre are staggered relative to each other by a distance of 67nm (designated D). A whole mature collagen molecule is approximately 300nm long (4.4D). This staggered conformation allows a gap between the ends of each molecule in the fibre of 40nm. Studies of the fibre formation help to explain why changes in the shape or size of the fibrillar collagen molecules result in disease phenotypes. The collagen molecules are so tightly packed in the fibre that there is little room for any change in its component molecules.

The fibre structure is stabilized by the formation of covalent intra and intermolecular cross-links through specific lysine and hydroxylysine residues in strictly conserved positions within the collagen molecules.

It was originally thought that each type of collagen fibre was made up of one type of collagen. However, the situation has been shown to be more complex. Supramolecular structures have been found to consist of type I and III collagen (Henkel and Glanville, 1982; Fleischmajer *et al.*, 1990), type I and V collagen (Birk *et al.*, 1988) and type II and XI collagen (Vaughan *et al.*, 1988). It is suggested that types V, III and XI serve to control the width of fibres of type I and II collagen. Fleischmajer *et al.* (1990) observed that type III collagen molecules with their amino propeptides still attached covered the outer surface of mature type I collagen fibres. They suggested that the helical domain of the type III collagens align and interact with the type I molecules in the collagen fibre, while the amino propeptide sticks out from the fibril to either prevent more type I collagen molecules attaching to the fibre or to interact with surrounding fibres.

1.5   GROUP II COLLAGENS

The collagen group II is made up of types IV, VI VII and VIII. These molecules do not form long fibrils as the group I molecules do, they form more flexible structures. This property is probably the result of interrupted helical sequences found in the collagens in this group.

1.5.1   TYPE IV

Type IV collagen can be isolated from basement membranes. Each type IV collagen molecule being made up of a combination of three of any of five α-chains, α1, α2, α3, α4 or α5 (Pihlajaniemi *et al.*, 1990a). The type IV collagen molecules are synthesized as a long helical domain (350nm) in between two globular domains. Unlike the fibrillar collagens, these type IV globular domains are not removed during extracellular processing. The whole molecule is 400nm long.

Type IV molecules aggregate to form a 3-d extracellular structure which has a role to play in cell differentiation, tissue regeneration and a specialized function in kidney glomeruli where it serves as a filtration barrier.

The collagen molecules interact at their amino terminal or '7S' domain. These interactions are stabilized by disulphide and covalent cross-links. At the carboxyl end two molecules join end to end through the cysteine containing non-collagenous domain (Weber *et al.*, 1984). Each type IV collagen helical domain contains within it approximately 20 interruptions in the Gly-X-Y repeat sequence. Many of these interruptions are common between the sequences of the α-chains (Pihlajaniemi *et al.*, 1990a). Since the positions of the helical interruptions coincide with the positions of kinks in the helical structure, as seen by electron microscopy (Hoffman *et al.*, 1984), it has been suggested that the differences in the types and positions of these interruptions may determine the α-chain combinations that make up each type IV collagen molecule.

7

1.5.2   TYPE VI

Type VI collagen is made up of a short helical domain (105nm) flanked by large globular domains. Each type VI collagen is made up from a combination of three α-chains. Results published in |Kielty *et al.* (1990) showed changes in the ratios of mRNA coding for these polypeptides in fibroblasts at different stages of development, thus indicating different species of type VI collagen at those stages of development.

α1(VI) and α2(VI) are not synthesized as precursor molecules (Trueb *et al.*, 1986) nor is the exceptionally long α3(VI). This α-chain has twice the molecular weight of the two other α-chains (Bonaldo *et al.*, 1990). The helical and carboxyl domains in the three α-chains are the same length, the difference in size originates from the amino terminal region. This extra sequence consists of eight repeats of a cell attachment sequence.

Antiparallel type VI monomers form dimers which aggregate laterally to form tetramers which are stabilized by disulphide bonds. The tetramers aggregate to form microfibrils which are found in many connective tissues. The short helical domains of the type VI α-chains (335-336 amnio acids) contain two Gly-X-Y repeat interruptions. It is suggested that these allow coiling of the antiparallel triple helices (Furthmayr *et al.*, 1983). The wide distribution of this molecule and the fact that it contains cell attachment sequences, also found in fibronectin and a cell binding protein called osteopontin (Oldberg *et al.*, 1986), suggests that cell adhesion could be one of the major functions of this collagen.

1.5.3   TYPE VII

Type VII collagen molecules are thought to be the substance of the anchoring fibrils which are thought to aid in the attachment of external epithelial basements to the underlying connective tissue stroma. This collagen is secreted from cells as a precursor molecule. It is a homotrimer that posseses a long helical domain (450nm), which is one and a half times the length of the type I helical domain. The mature molecules are more than 90% helical sequence.

To form the anchoring fibril structure two type VII molecules overlap at their amino ends by 60nm. This overlap is stabilized by the presence of disulphide bonds (Morris *et al.*, 1986). The carboxyl domains of

type VII collagen form unusual structures. The structures are composed of three arms which radiate out from the end of the helix. The arm structures are stabilized by the presence of disulphide bonds at the base of the arms (Lunstrum *et al.*, 1986). The type VII dimers are thought to aggregate laterally to form fibres which are attached to the basement membrane through one set of carboxyl arms and to the connective tissue stroma through the opposite set of carboxyl arms (Burgeson *et al.*, 1990). The long type VII helical domain contains within it interruptions in the Gly-X-Y repeat sequence which allow the fibrils a large amount of flexiblity (Bachinger *et al.*, 1990).

1.6   GROUP III COLLAGENS

1.6.1   TYPE VIII

Bovine corneal Descemet's membrane is particularly enriched with type VIII collagen (Labermeier and Kenney, 1983). It has also been isolated from cultured endothelial cells, although it is only synthesized by proliferating, not confluent, cells.

Observations by Benya and Padillo (1986) indicated that type VIII collagen consists of three α-chains each having a molecular weight of 61,000. Each chain contains a central helical domain in between two globular domains. α1(VIII) collagen has been found to have a similar amino acid sequence to another short chain collagen, α1(X). The eight interruptions in the Gly-X-Y repeats in the two collagens were to be found in the same relative positions. Mann *et al.* (1990) sequenced a second type VIII collagen α-chain, α2(VIII). This α-chain, like α1(VIII), contained a short helical region (460 amino acids) which included eight Gly-X-Y interruptions. Like α1(VIII), this second type VIII α-chain also displayed marked sequence homology to α1(X).

1.6.2   TYPE IX

This collagen is found in cartilage tissue, where it represents 1-10% of total collagen. It is also found in the primary corneal stroma of chick embryos (Svobada *et al.*, 1988). Type IX collagen is made up of three

different α-chains which contain three helical domains interrupted by non-helical regions (van de Rest *et al.*, 1985). The helical domains called COL1, 2 and 3 are 115, 339 and 137 amino acids long respectively. Around these regions are non-collagenous domains designated NC1-4.

The NC3 sequence is 5 amino acids longer in α2(IX) compared with α1(IX). The presence of these extra amino acids corresponds with the presence of a kink in the molecule seen using electron microscopy (Irwin, 1986). After covalently linked complexes of types IX and II collagen were detected (Eyre *et al.*, 1987), Vaughan *et al.* (1988) detected the presence of type IX collagen molecules on the surface of type II collagen fibrils. The presence of a hinge in the molecule within COL3 allows type IX collagen to align with type II collagen along the majority of its length, while allowing the amino terminal globular domain NC4 to project out from the type II collagen fibre. It is thought that the NC4 region projecting out from the type II collagen fibre either controls fibre formation or participates in the formation of a link between type II collagen fibres and other components of cartilage.

The theory that the presence of COL3 and NC4 regions protruding from type II collagen fibres could determine the macromolecular structure of the tissue in which the two molecules are found is supported by evidence published in Svobada *et al.* (1988). mRNA coding for α1(IX) was compared after isolation from chick hyaline cartilage and corneal stroma. The morphology of type II collagen fibres from the two tissues is different. Large aggregates of type II fibres are found in hyaline cartilage but not in corneal stroma. mRNA analysis showed the loss of mRNA coding for COL3 and NC4 in mRNA isolated from corneal stroma. The two RNA species coding for α1(IX) chains with and without NC4 and COL3 were found to be the result of using one or other of two translation start codons found within COL9A1, and the splicing out of exon 7. Such alternative splicing has also been demonstrated in embryonic mouse and human fetal RNA extractions (Muragaki *et al.*, 1990).

1.6.3   TYPE X

Type X collagen is synthesized by chondrocytes only when the cartilage is about to undergo bone formation. This suggests the importance of this collagen in the process of bone formation (Jimenez *et al.*, 1986).

It is a short chain collagen composed of 3 α-chains each having a

molecular weight of 59,000. These α-chains do not undergo extracellular processing (Reginato and Jimenez, 1991). The helical region of type X collagen isolated from chick or bovine chondrocytes contain 8 interruptions of the Gly-X-Y sequence. Mostly Gly-X-Y replaced by Gly-X-Gly or Gly-X-Y-X-Y. These interruptions are found in the same places within the helix of the type X collagen synthesized by chick and bovine chondrocytes (Thomas *et al.*, 1991).

Type X collagen molecules are stabilized by inter and intramolecular bonding as other collagens are but they are also stabilized by disulphide bonds which occur between the collagen and proteins in the surrounding matrix structure (Reginato and Jimenez, 1991). Type X molecules are capable of forming dimers which are stabilized by strong non-covalent interactions. This dimer formation is thought to involve the carboxyl non-collagenous domain (Schmid *et al.* 1990). The supramolecular structure formed by type X collagen molecules occur in two forms, a filamentous mat or molecules associated with type II collagen fibres.

1.6.4   TYPE XII

Type XII collagen was first characterized as a cDNA clone synthesized from mRNA isolated from chick tendon (Gordon *et al.*, 1987). Structurally this collagen displays many of the features of type IX collagen (Dublet and van der Rest, 1987). It was therefore assumed that it would be found to associate with a fibrillar collagen and perform a similar function to type IX collagen. Type I collagen is the major fibrillar found in collagen tendon and so a relationship was predicted between these two collagens. To date, this has not been established.

Type XII is a homotrimer (Dublet *et al* 1989). Gordon *et al.* (1990) suggest an unusual structure for type XII collagen. It consists of a 75nm helical domain which contains within it a kink, which is coded for by a non-collagenous domain in between two helical domains. At the amino terminal, three finger-like structures are formed from the non-helical domain from each α-chain. This non-collagenous domain is six times larger than the collagenous domains in the collagen. It folds back on itself to produce 60nm fingers. Gordon *et al.* (1990) suggest these fingers serve a similar function to the NC4 domain of type IX collagen.

11

1.6.5   TYPE XIII

To date, this collagen has been characterized at the cDNA level only. The mRNA coding for this collagen has been detected in skin, intestine, bone and striated muscle (Sandberg *et al.*, 1989). The molecule consists of three helical domains and 4 non-helical domains.

Alternative splicing of mRNA coding for collagen proteins was first observed in type XIII (Tikka *et al.*, 1988). Unlike the collagens which have since been shown to display alternative splicing, the alternative splicing in type XIII collagen can involve helical sequence (Philajaniemi *et al.*, 1990).

The type XIII collagen protein has not yet been isolated, so the question has not been answered as to how the protein copes with α-chains which have helical domains of different lengths. The biological significance of alternative splicing in type XIII collagen is unknown.

1.6.6   TYPES XIV AND XV

cDNA clones coding for these collagens have been isolated but no sequence data has been published.

It seems that the collagens which are synthesized at the highest level by cells have been isolated and characterized ie. group I collagens. Future isolation of new collagens will be dependent on the isolation of cDNA clones coding for them, instead of the isolation of the proteins. The isolation of the proteins may prove to be too difficult because of the low levels of such proteins synthesized by cells.

1.7   COLLAGEN BIOSYNTHESIS

Collagen biosynthesis involves a number of unusual features. Some of the collagens are synthesized as precursor or procollagen molecules and collagen molecules undergo many post-translational modification steps.

Each collagen polypeptide is synthesized from a mRNA template with a signal peptide at the amino terminal, this polypeptide is called a prepro α-chain. The helical domain(s) of the collagens are sandwiched

between two non-helical domains at the amino and carboxyl termini. Most of these terminal non-helical domains are cleaved from the the fibrillar collagen α-chains in the extracellular matrix. The function of these so called propeptide regions presumably being over once the mature molecule is synthesized.

In the cell the three collagen polypeptides associate by hydrophobic and electrostatic interactions among the carboxyl non-collagenous domains. The conserved carboxyl propeptide domain sequence is thought to determine which pro α-chains combine to form a procollagen molecule. The importance of this domain in the formation of collagen molecules is reflected in the fact that it shows the highest degree of sequence similarity between different types and species. The selection of certain combinations of α-chains to form a collagen molecule in cells where more than one collagen type is being synthesized is not fully understood. The number of cysteine residues at the telopeptide end of the carboxyl propeptide may have a role to play (Weil *et al.*, 1987). These cysteines are involved in the forming of disulphide bonds between correctly aligned pro α-chains. Pro α-chains which form heterotrimers ie. α2(I) and α2(V) have three cysteines for disulphide bonding, while pro α-chains which form homotrimers ie. α1(I) and α1(III), have four.

Once the pro α-chains are correctly aligned the helix forms from the carboxyl end of the molecule through to the amino domain (Engel and Prockop, 1990). The propeptides synthesized as coded for by the mRNA are incapable of folding into triple helical structures which are stable at the body temperature of the organism in which they were synthesized. Stable helix formation requires that about 100 proline residues in each α-chain are hydroxylated (Prockop and Kivirikko, 1984). The enzymes which catalyze these hydroxylation reactions are prolyl-4-hydroxylase and prolyl-3-hydroxylase. Most of the proline residues are hydroxylated to 4-hydroxyproline, a few are hydroxylated to 3-hydroxyproline. Only proline residues in the Y position of the Gly-X-Y repeat are hydroxylated. α1(X), α1(VIII), α2(VIII) and type IV α-chains are hydroxylated more than the fibrillar collagen α-chains (Mann *et al.*, 1990). The increased hydroxylation may increase the stability of the collagens with interrupted collagen domains.

Hydroxylation of proline only takes place while the α-chains are in their native form. After the α-chains have associated through their carboxyl propeptides helix formation is delayed only until enough prolines are hydroxylated. Other post-translational modifications involve the

13

hydroxylation of about 20 lysyl residues. Lysyl hydroxylase catalyses these reactions. Only lysine in the Y position of the Gly-X-Y repeat are hydroxylated. Other enzymes may be required to hydroxylate the few lysine residues in the non-helical domains. Different collagen species have different numbers of lysine residues hydroxylated. The number of hydroxylated lysine residues can also vary in the same collagen in different tissues. Following hydroxylation, lysine residues may be glycosylated to form galactosylhydroxylysine residues or glucosylgalactosylhydroxylysine by the use of a glucosyltransferase or a galactosyltransferase. All these modifications cease as the protein forms into a triple helix.

The pro α-chains contain large numbers of glycine-proline and glycine-hydroxyproline bonds which are isomerized from transpeptide bonds to cis bonds. This step probably involves a specific isomerase.

An asparagine linked oligosaccharide group is added to each carboxyl terminal propeptide domain (Kivirikko and Mylylla, 1982). The purpose of this oligosaccharide is unknown but the sequence to which it is attached is highly conserved among the collagens.

Once assembled, the procollagen molecule is released from the cell via the Golgi apparatus. Many collagens are processed extracellularly to remove sequence from the amino and carboxyl termini, though not all. These terminal sequences are cleaved with the aid of specific peptidases which recognise cleavage sites between the telopeptide and propeptide domains.

The mature collagen molecules aggregate in a type specific manner to form supramolecular complexes which are stabilized by intermolecular cross-links. Type I collagens, for examples, undergo a series of complex lysyl and hydroxylysyl oxidative deamination reactions, the resulting aldehydes generated form intermolecular cross-links which increase the tensile strength of the fibres in which they aggregate. The aggregation of the monomers into an ordered structure is a self-driven process and does not require any enzymic catalysis.

It has become clear that the synthesis of collagen trimers and the eventual aggregation of the mature collagen molecules into large supramolecular complexes is determined by the primary structure of the collagen α-chains. Any changes in the amino acid sequence of the constituent α-chains of a collagen may result in the disruption of the collagen molecule synthesis and eventually the synthesis of the collagen multimolecular complex.

1.8   COLLAGEN GENES

Recent reviews have been published by Vuorio and de Crombrugghe (1990) and Sandell and Boyd (1990).

The genes coding for members of the collagen gene family are among the most complex in the human genome. Typically the genes coding for the fibrillar collagens consist of 51-54 exons covering 16-44kb of DNA. A comparison of the gene length (18kb) with mRNA length (4.8kb) of $\alpha 1(I)$ for example clearly shows the majority of collagen gene sequence to be intron. A comparison of the sizes of COL1A1 and COL1A2 (18kb and 35kb respectively) illustrates the fact that the amount of intron sequence varies from gene to gene.

In general collagen genes are semi-dispersed throughout the genome (Law *et al.*, 1990; Heurre *et al.*, 1982; Weil *et al.*, 1988; Solomon *et al.*, 1985; Shows *et al.*, 1989) (Table 1.1). There are exceptions to this rule $\alpha 1(III)$ and $\alpha 2(V)$ (Emanuel *et al.*, 1985a), $\alpha 1(IV)$ and $\alpha 2(IV)$ (Griffin *et al.*, 1987) and $\alpha 1(VI)$ and $\alpha 2(VI)$ (Weil *et al.*, 1988). COL1A3 is located at position 2q31-32.3 (Huerre *et al.*, 1986) syntenic to COL5A2 which has been assigned to 2q14-32. The close proximity of the two genes may indicate a common transcriptional control. This theory is supported by the fact that these collagens share with type I a ubiquitous tissue distribution.

The genes coding for $\alpha 1(IV)$ and $\alpha 2(IV)$ are well characterized. They are closely linked on the long arm of chromosome 13. The genes are arranged in a head to head configuration on 13q34. They are transcribed from opposite strands of DNA; there are between 130 and 41bp between the transcriptional start site of $\alpha 1(IV)$ and the multiple start sites of $\alpha 2(IV)$ (Soininen *et al.*, 1988; Poschl *et al.*, 1988).

COL6A1 and COL6A2 are located on band q223 of chromosome 21 (Weil *et al.*, 1988). The third type VI collagen gene has been assigned to band q37 of chromosome 2. The close chromosomal location of COL6A1 and COL6A2 may facilitate the coordinated expression of the two genes observed by Chu *et al.* (1987). Close proximity is not required for coordinated gene expression of collagen genes. COL1A1 and COL1A2 are located on chromosomes 17 and 7 respectively yet are still coordinately expressed. It is of interest to note that along with COL6A3 the genes coding for $\alpha 1(III)$ and $\alpha 2(V)$ collagens and the extracellular matrix proteins, elastin (Emanuel *et al.*, 1985b) and fibronectin (Henry *et al.*, 1985) are located on 2q. It is known

15

Table 1.1

The molecular species of each collagen is shown with the names denoted to the gene loci which code for the α-chains. The chromosomal location of many of the genes have been determined

Table 1.1

collagen type	constituent chains	gene locus	chromosomal location
1	1(I)	COL1A1	17q21.3- q22
	2(I)	COL1A2	7q21.3- q22
2	1(II)	COL2A1	12q13- q14
3	1(III)	COL3A1	2q24.3- q31
4	1(IV)	COL4A1	13q34
	2(IV)	COL4A2	13q34
	3(IV)	COL4A3	——
	4(IV)	COL4A4	——
	5(IV)	COL4A5	Xq22
5	1(V)	COL5A1	——
	2(V)	COL5A2	2q24.3- q31
	3(V)	COL5A3	——
6	1(VI)	COL6A1	21q223
	2(VI)	COL6A2	21q223
	3(VI)	COL6A3	2q37
7	1(VII)	COL7A1	——
8	1(VIII)	COL8A1	——
9	1(IX)	COL9A1	6q12- q14
	2(IX)	COL9A2	——
	3(IX)	COL9A3	——
10	1(X)	COL10A1	——
11	1(XI)	COL11A1	1p21
	2(XI)	COL11A2	6p212
	3(XI)	COL11A3	12q13- q14
12	1(XII)	COL12A1	——
13	1(XIII)	COL13A1	10q11 -qter

that COL3A1 and COL5A2 are co-expressed (Bornstein and Sage, 1980), It would be of interest to determine if the other extracellular genes in the area are under the same expressional control.

It has been suggested by Solomon *et al.* (1985) and Griffin *et al.* (1987) that selective pressure may encourage the dispersal of collagen genes throughout the genome. This would reduce the chances of unequal crossing over occurring between two genes. The close proximity of some of the collagen genes suggests that these genes are sufficiently different to be located close to each other without risk of unequal crossing over.

1.8.1 THE FIBRILLAR COLLAGEN GENES

Analysis of each exon length of the fibrillar collagen genes has revealed unusual features. All the exons coding for the helical domain are multiples of 9bp in length. Each codes for an integral number of Gly-X-Y repeats. They each begin with the first G of a glycine codon and end with the last base of a Y codon. The most common exon size is 54bp, some exons are 45, 99, 108 or 162bp. At the beginning and end of the helical domain sequence are 'joining' exons which code for the end of the helical sequence, the telopeptide sequence and the beginning of the propeptide sequence. More unusual is the observation that, with one exception, the pattern of the exons coding for the helical domain of the fibrillar collagens are the same. This conservation of exon structure is independent of species. The one exception is the fusion of two 54bp exons to form a 108bp exon in COL1A1 (Chu *et al.*, 1984). Variations in the sizes of the joining exons have been observed. The high degree of exon size conservation between the fibrillar collagens is probably a reflection of the requirement of the collagen proteins to be a predetermined length. Any alteration in the length of the fibrillar collagen would result in an abnormal phenotype.

The highest degree of sequence homology has been observed between the carboxyl propeptide domains of the fibrillar collagens. The last two exons coding for the highly conserved sequence around a carbohydrate attachment site has been found to be 243bp in all the fibrillar collagen genes where sequence is available (Yamada *et al.*, 1983). Equally, the last exon is 144bp in length in all the fibrillar collagen genes. The remaining two exons which complete the coding sequence for the carboxyl region vary in size from 186-244bp.

The exon organization of the amino propeptide region shows most divergence among the fibrillar collagens (Wozney *et al.*, 1981; de Wet *et al.*, 1987; Chu *et al.*, 1984). This divergence of exon structure includes the exons coding for the triple helical domain of the amino propeptide. In α2(I) and α1(III) this helical structure is coded for by by 2 exons, α1(I) has an additional 36bp exon and α1(II) contains two 33bp exons and a 54bp exon.

The variation in the degree of conservation of exon size between the exons coding for the helix and those coding for the amino propeptide suggest different selective pressures acting on the two regions. The conservation of exon size within the helical domain of the fibrillar collagens suggests that the collagen genes have evolved from a 54bp ancestral unit (Yamada *et al.*, 1980). This 54bp unit multiplied enough times to code for a collagen which was the correct size for fibrillar collagen formation. After this, any changes in the length of the collagen molecule were not tolerated. The occurrence of many collagen genes was probably due to the replication of this first collagen gene, followed by divergence of the newly duplicated genes. Since the propeptides are removed from the fibrillar collagens prior to fibre formation then the constraints on their length do not apply.

1.8.2   NON-FIBRILLAR COLLAGEN GENES

The conservation of exon organization seen in the fibrillar collagen genes is not mirrored in the non-fibrillar collagen genes. However, the theory of a common ancestral gene can still be applied. The variations in exon size within these genes can be explained by the presence of interruptions in the Gly-X-Y repeat helical sequence. As a general rule, the more divergent a collagen protein sequence is from the fibrillar collagens, the more its exon organization diverges from the 54bp size conservation seen in the fibrillar collagen genes.

Deviation of exon length from 54bp is often accompanied by split codons. Split codons occur in genes coding for types IX and IV collagen. They occur in a non-random fashion, they are generally the result of the loss or addition of the first G of a glycine codon. Collagens that tolerate variations in helical domain sequence ie. types IV and VI, are also able to tolerate variations in exon size and the presence of split codons.

The most unusual gene organization among the collagen genes is that found in the gene coding for type X collagen. This gene consists of only

three exons. Exon 3 contains sequence which codes for the entire triple helical region, the carboxyl terminal domain and 3' untranslated sequence(Lu Valle *et al.*, 1988; Thomas *et al.*, 1991).

Recent investigations into the exon size distribution of COL6A2 revealed a lack of 54bp exons (Hayman *et al.*, 1990). The most common exon size was 63bp, which led to speculation that this gene may have evolved separately from the other collagens.

1.8.3   INVERTEBRATE COLLAGEN GENES

The 54bp exon pattern also extends to invertebrates such as the sponges (Exposito and Garrone, 1990) and sea urchins (D'Alessio *et al.*, 1989). Therefore it is suggested that the establishment of the exon pattern of the fibrillar collagen genes preceded vertebrate-invertebrate radiation.

1.9   <u>CONTROL OF COLLAGEN GENE EXPRESSION</u>

A recent review of this subject has been published in Ramirez and Di Liberto (1990).

The control of the expression collagen genes is complex and multifactorial. Miskulin *et al.* (1986) observed that the genes coding for types I and III collagen are coordinately expressed by fibroblast tissue culture cells. The three genes coding for the polypeptides that make up these two collagens are located on different chromosomes. Other results indicated that type III collagen gene expression is linked to COL5A2 expression (Bornstein and Sage, 1980). This information is mirrored in the protein data which indicates that complexes of collagen molecules are formed involving types I, III and V collagen. An extra complication which must be accounted for by collagen gene expression is the ability of cells to produce collagen types consisting of more than one combination of α-chains.

As with all genes, collagen gene expression is controlled by a combination of trans-acting and cis-acting elements. The trans-acting elements are specific nuclear proteins which are capable of forming complexes with the DNA sequence cis-acting elements.

In tissue culture, several cytokines and products of oncogenes have been shown to alter collagen gene expression (Bornstein and Sage, 1989).

The study of cis-acting elements involved in collagen gene expression has been made easier by the use of transgenic mice and chimeric genes (Lovell-Badge *et al.*, 1987). Using such methods, the mouse COL1A2 gene has been extensively studied. This gene has three promoter sequences within the 315bp of sequence upstream to the start site of transcription. This gene also possesses an enhancer sequence within its first intron. The enhancer sequence in combination with the promoter sequences confers high and tissue specific expression of the gene (Rossi and de Crombrugghe, 1987a).

The study of the control of COL1A1 expression has been aided by the MOV13 mouse model (Schnieke *et al.*, 1983). The insertion of mouse leukemia retroviral sequence in the first intron of mouse COL1A1 resulted in a block of the transcription of that gene. A mouse, heterozygous for the insertion, displays a type I osteogenesis imperfecta phenotype; a homozygous mouse displays a type II osteogenesis imperfecta phenotype. The phenotype was found to be the result of loss of enhancer sequence from the intron. Negative DNA elements have been detected in the first intron of COL1A1 (Bornstein and McKay, 1988).

A tissue specific enhancer has been detected in the first intron of COL2A1 (Horton *et al.*, 1987). This enhancer sequence is involved in developmentally controlled expression of the gene.

The control of COL4A1 and COL4A2 is unusual because of the unique arrangement of the two genes. The two genes are arranged in a head to head orientation on opposite strands of DNA 130bp apart. In the centre of this 130bp is a possible SP1 binding site (Poschl *et al.*, 1988). SP1 is a general transcription factor. This promoter sequence serves both genes. Transcription of the genes is enhanced by a sequence in the first intron of COL4A1 (Burbelo *et al.*, 1988). This intron sequence contains two possible cis-acting elements. A negative regulatory element is present in the third intron of COL4A2. Preliminary evidence suggests that the methylation of collagen IV regulatory sequences has a role to play in the expression of COL4A1 and COL4A2 (Burbelo *et al.*, 1990).

Rather than controlling the collagen synthesized by a cell by on/off swtiches, collagen gene expression can be altered by alternative splicing. Alternative splicing has been detected in mRNA coding for $\alpha 1(II)$, $\alpha 2(VI)$, $\alpha 1(XIII)$ and $\alpha 1(IX)$. The alternative splicing displayed by $\alpha 1(IX)$ transcripts is accompanied by the use of an alternative transcription start site. The use of the transcription start sites and splicing out of an exon is controlled in a

tissue specific manner (Nishimura *et al.*, 1989). The significance of alternative splicing of mRNA coding for α1(XIII) and α2(VI) is not known. The alternative splicing of mRNA coding for the α1(II) propeptide domain may have a role to play in the control of translation (Ryan and Sandell, 1990). It has been established that fragments of the amino propeptide of type I collagen inhibited α1(II) collagen synthesis in a cell free translation system (Paglia *et al.*, 1981; Hörlein *et al.*, 1981). A more detailed explanation is given in section 1.3.2.

The synthesis of α2(I) mRNA is developmentally controlled in chondrocytes (Bennett and Adams, 1990). Mature chondrocytes do not synthesize any type I collagen protein, although they express the genes COL1A1 and COL1A2. The α1(I) transcripts remain unprocessed in the nucleus. However, α2(I) transcripts are processed and released from the nucleus (Bennett and Adams, 1987). The transcripts are synthesized using an alternative transcription start site and a promoter found in the second intron of COL1A2. This results in the first two exons of COL1A2 being replaced by a 96bp exon and the presence of 4 open reading frames, the largest coding for a polypeptide which would be 71 amino acids long. It has not been determined whether this alternative transcript is translated.

The control of type VI collagen transcription is unusual amongst the collagens. Koller *et al.*, (1991) discovered that the COL6A2 promoter displayed many of the features of the housekeeping gene promoters, and was completely different from those of the collagen genes thus far characterized. This information, along with the protein data which suggests that type VI collagen has a role to play in cell attachment and may be expressed by all cells, lends credence to the speculation that COL6A2 may be looked upon as a housekeeping gene.

Our understanding of the mechanisms that control collagen biosynthesis are still in their infancy. Such control mechanisms were predicted to be complicated because collagen synthesis is dependent on the type and the developmental stage of the cell in which the collagen is synthesized.

1.10   INHERITED CONNECTIVE TISSUE DISEASES

There are many inherited connective tissue diseases of which only a few are discussed here. Although not the most common of the connective tissue diseases, osteogenesis imperfecta (OI), Ehlers-Danlos syndrome (EDS) and Marfan syndrome are the most studied. Other more common but phenotypically milder diseases such as osteoarthritis, osteoporosis and aortic aneurysms are discussed later. The important message that has come out of the results of research into the molecular causes of OI, EDS and Marfan syndrome over the past few years, is that an understanding of the molecular causes of the rare diseases can help in the search for the causes of the more common diseases.

1.11   OSTEOGENESIS IMPERFECTA

1.11.1   SIGNS AND SYMPTOMS

Osteogenesis imperfecta, more commonly known as brittle bone disease, presents itself clinically as a heterogeneous array of symptoms. The symptoms range from bone fragility leading to death *in utero,* to an increased tendency to succumb to bone fractures. The disease is generally classified by clinicians using the four Sillence categories (Sillence *et al.*, 1979).

Type I - Mild form, inherited autosomal dominantly, normal stature, little or no bone deformity, blue sclerae, hearing loss in some cases.

Type II - Lethal form, autosomal dominantly inherited from new mutations, rarely autosomal recessive, death *in utero* or a few hours after birth due to bone fragility, X-ray examination shows beaded ribs formed as a consequence of breakages and inefficient repair *in utero.*

Type III - Progressively deforming form, autosomal dominant and rarely autosomal recessive, sclerae may be blue, dentinogenesis imperfecta, hearing loss, very short stature.

Type IV - Mild to moderate form autosomal dominant, normal sclerae, mild - moderate bone deformity, variable short stature, dentinogenesis imperfecta,

occasionally hearing loss.

The disease occurs in 1 in every 20,000 live births in the U.K. with similar frequencies in other populations. The variability in symptoms sometimes make it difficult to categorize the disease. Some workers have published data to show that the mild form of the disease, type I, often goes undiagnosed or is misdiagnosed as child abuse (Paterson and McAllison, 1989).

1.11.2 DOMINANTLY INHERITED POINT MUTATIONS WHICH CAUSE OSTEOGENESIS IMPERFECTA

Most cases of OI are the result of a mutation in one of the genes coding for the type I collagen polypeptides (COL1A1 or COL1A2). Sykes *et al.* (1990) found consistent linkage between dominantly inherited OI and COL1A1 and COL1A2.

The majority of mutations causing OI are dominantly inherited point mutations. Prockop *et al.* (1989) suggested that the wide variety of point mutations that cause OI is due to the inability of the fibrillar collagens to tolerate changes in their primary structure because all the amino acids which make up the fibrillar collagen polypeptides are essential to their function.

Some trends relating OI phenotypes and the type and position of type I collagen mutations have been proposed. Sykes *et al.* (1990) found that of 8 Sillence OI type IV pedigrees analysed all segregated with COL1A2, all other OI types segregated with either COL1A1 or COL1A2. More specifically, Byers (1990) suggested that a gradient effect is seen when looking at the amino acid substitutions causing OI. Most OI phenotypes are caused by a point mutation substituting a glycine for another amino acid in the type I collagen helical region. The gradient suggested in Byers (1990) is such that, a glycine substitution at the carboxyl end of the helix results in a more severe OI phenotype than a mutation at the amino end of the helix. This conclusion needs to be qualified to allow for the type of substitution, for example an aspartate replacing a glycine has been found to result in a lethal phenotype wherever the mutation occurs in the helix. Substitutions of a cysteine for a glycine most clearly shows this gradient effect.

22

1.11.3 OVERMODIFICATION OF TYPE I COLLAGEN AND OSTEOGENESIS IMPERFECTA

The gradient effect proposed by Byers (1990) may be caused by overmodification of the type I collagen molecules as a result of the presence of a mutation in the helical domain. Bonadio and Byers (1985) found that type I collagen molecules tend to be overmodified from the position of the mutation to the amino end. This, they suggest, is because the substitution of a glycine for a more bulky amino acid slows down the helix formation of the molecule. Glycine being the smallest amino acid fits into the centre of the helix, larger amino acids are not able to do so easily. Helix formation proceeds at a normal rate until the mutation is reached then, as the larger amino acid is accommodated in the helix, helix formation slows down allowing the type I α-chains to be overmodified amino terminal to the mutation. The overmodified collagen molecules display decreased thermal stability, are inefficiently secretion from and degraded by the cell. Incorporation of these unstable molecules into the collagen fibres results in abnormal fibre formation. The greater the proportion of helix which is overmodified the greater the thermal instability of the type I molecule and sometimes the more severe the phenotype caused by the mutation.

1.11.4 COOPERATIVE BLOCK UNFOLDING

An additional factor which may have a role to play in determining whether a glycine substitution results in a lethal phenotype or not, is the positioning of the mutation within the α-helix with respect to the micro-unfolding model proposed by Privalo (1982). Privalo suggested that the collagen helix does not unfold uniformly as it reaches its melting temperature, but that short segments of helix or 'cooperative blocks' unfold at a lower temperature than the rest of the helix. This theory was later confirmed by observations by Kadler et al. (1988). These data, along with the knowledge that areas of the helix rich in proline and especially hydroxyproline are the most stable, led Westerhausen et al. (1990) to compare the thermal stability of type I collagen molecules containing the same type of mutation in different positions along the helix. The mutations were both glycine to serine substitutions, one of the substitutions was at

position 631, the other was at position 598 in the α1(I) helical domain (the first glycine of the domain being numbered 1). Their results suggest that there are areas of thermal sensitivity to be found along the length of the type I collagen helix and areas of thermal stability. If a mutation disrupts a stable area in the helix a decrease in the thermal stability of the whole molecule will be the result, leading presumably to a severe phenotype. However, if the mutation is situated in an area of low thermal stability the overall thermal stability of the molecule would not be much altered, leading to a less severe phenotype.

1.11.5   AUTOSOMAL RECESSIVE MUTATIONS

Most cases of OI are caused by mutations which are inherited in an autosomal dominant manner. There are two documented exceptions (Pihlajaniemi *et al.*, 1984; de Wet *et al.*, 1983), both are the result of defects in COL1A2.

One (de Wet *et al.*, 1983) is a compound heterozygote. A sporadic mutation resulting in the loss of approximately 20 amino acids from α2(I) and a null allele. It is not known if the 20 amino acid deletion would have been enough to cause the lethal phenotype alone, but the null allele inherited from the phenotypically normal father obviously was not.

The case described in Pihlajaniemi *et al.* (1984) was a patient with the progressively deforming form of the disease (type III). The mutation causing the disease was inherited in an autosomal recessive manner from phenotypically normal, consanguineous parents.It was a 4bp frameshift deletion in the last exon of COL1A2. The mutation caused a shift in the termination codon with the result that the mRNA species carrying the mutation was the same length as the normal mRNA. The body was able to tolerate half its α2(I) polypeptides carrying this mutation but not all of it.

1.11.6 GERMLINE MOSAICISM

As has been mentioned before, most cases of osteogenesis imperfecta (OI) are inherited in an autosomal dominant manner. Occasionally the disease recurs in a family, generally more than one offspring being affected from the same parent. The affected offspring have been found to possess the same 'new' mutation in their type I collagen, while their clinically normal parents may synthesize only normal collagen. Such cases have been found to be due to a higher than average occurrence of germline mosaicism associated with the disease, 6-7% of OI cases (Byers *et al.*, 1988). A number of well characterized cases of OI germline mosaicism have been described in the literature (Constantinou *et al.*, 1990a; Cohn *et al.*, 1990; Wallis *et al.*, 1990a). As with most cases of OI, the mutations causing the OI phenotype were found in one or other of the genes coding for the type I collagen polypeptides (COL1A1 or COL1A2). Germline mosaicism has been found to originate from either the mother or father and is often accompanied by somatic mosaicism. (Cohn *et al.*, 1990). The knowledge that mutated collagen α-chains occur in some of the somatic cells of phenotypically normal parents, leads to the speculation that the occurence of an OI phenotype is dependent on the dosage of mutated type I molecules in the individual. Such a dosage effect was seen in the case described in Constantinou *et al.* (1990a). In this case a proband with type II OI (lethal form) was born to a mother with type I OI (mild form). Both were found to carry the same heterozygous mutation, a glycine to cysteine substitution at position 904 in α1(I). The proband synthesized mutated α1(I) polypeptides in all its cells; the mother was mosaic for the synthesis of abnormal α1(I) polypeptides. DNA isolated from the mother's leucocytes contained approximately half the mutated alleles present in the probands fibroblasts.

1.11.7 INSERTION OR DELETION MUTATIONS IN THE TYPE I HELICAL DOMAIN

Not surprisingly most deletions or insertion mutations occurring in the helical domain of type I collagen molecules result in lethal phenotypes (Chu *et al.*, 1983; Byers *et al.*, 1988; Willing *et al.*, 1988; Hawkins, 1989). The length of type I collagen molecules is crucial to the correct formation of fibres. As is always the case with OI there are exceptions to this rule

(Kuivaniemi et al., 1988b; Superti-Furga et al, 1990). Superti-Furga et al. (1990) described a case in which a 39bp deletion in intron 21 of COL1A2 caused the splicing out of exon 21. The mutant chain containing type I collagen molecules produced by the patient were very unstable. It was thought that the mild phenotype (type I OI) displayed by the patient is due to the very unstable type I collagen being degraded by the cell without becoming incorporated into the extracellular matrix. Mild OI phenotypes (type I and IV) are often associated with inhibition of expression of COL1A1 or COL1A2 (Barsh et al., 1982; Sykes et al., 1977). It seems that it is better to have reduced expression of a type I α-chain than to have a defective polypeptide. Although, as has been mentioned before, this depends on the final level of type I collagen incorporated into the extracellular matrix.

The mutation described by Kuivaniemi et al. (1988) was inherited in an autosomal dominant manner and showed variable phenotypic expression in a family. The phenotypes were caused by a 19bp deletion which caused splicing out of exon 11 of the mRNA coding for α2(I) collagen. The authors suggested that the variable phenotypic expression was the result of differing levels of mis-splicing of the α2(I) mRNA in each member of the family.

1.11.8 MUTATIONS IN THE CARBOXYL PROPEPTIDE REGION OF THE TYPE I COLLAGENS

Not all the type I collagen mutations causing OI phenotypes are found in the helical domain of the polypeptides, though most are (Prockop et al., 1989). A few mutations in the carboxyl propeptide region have been described in the literature (Willing et al., 1990; Bateman et al., 1989; Pihlajaniemi et al., 1984). These mutations include frameshift deletions and a 1bp frameshift insertion. The mutation studied by Bateman et al., (1989) caused a disruption in the amino acid sequence of the carboxyl propeptide domain, which led to the inhibition of incorporation of the mutated α-chain into the mature type I collagen molecules. The few mutated α-chains which managed to be included in the mature collagen molecules were overmodified along the whole length of the helix. They displayed decreased thermal stability and were degraded by the cell. As a result only normal type I collagen molecules were found in the extracellular martrix. A consequence of the lack of type I collagen synthesis and intracellular degradation of

abnormal type I collagen molecules was a decrease in type I collagen content of the extracellular matrix. In the case described in Willing *et al.* (1990) half the pro α1(I) synthesized by the patient was very unstable. The unstable polypeptide was caused by a 5bp frameshift deletion at the 3' end of one of the COL1A1 alleles. The product of this mutated allele was a pro α1(I) polypeptide which was 84 amino acids longer than normal. This longer pro α1(I) polypeptide was so unstable it was not available for incorporation into type I collagen molecules, this caused a reduction in the type I collagen content of the extracellular matrix to 50%. The patient suffered from type I OI. The proband described by Bateman *et al.* (1989) had only 20% of the normal type I collagen content in its extracellular matrix and was diagnosed as having type II OI. The mutation causing this reduction in the type I collagen content was a 1bp insertion in exon 49 of COL1A1. The polypeptide coded for by this mutated gene had a truncated carboxyl terminal propeptide. Some of these truncated molecules were incorporated into the α-helical structures before they were degraded by the cell, therefore normal aswell as abnormal pro α1(I) polypeptides were degraded reducing the type I collagen content of the cell to 20%. Thus the less type I collagen in the extracellular matrix, the more severe the phenotype. Characterization of mutations in the different domains of the type I collagen molecule help to increase our understanding of the function of the domains in the mature molecule.

1.11.9 PROTEIN SUICIDE

If the degradation of imperfect type I procollagen molecules was 100% efficient then the severity of an OI phenotype would be dependent on the amount of collagen incorporated into the extracellular matrix. That would depend on the gene in which the mutation was found; COL1A1 or COL1A2. Upon synthesis of defective type I procollagen molecules the cells producing them undergo a phenomenon called 'protein suicide' (Williams and Prockop, 1983). It seems that the resulting thermal instability caused by overmodification of the procollagen molecules encourages the cell's endogenous proteases to degrade the defective procollagen molecules. Each type I procollagen molecule is made up of two pro α1(I) chains and one pro α2(I) chain. It is irrelevant whether the procollagen molecule contains one or two abnormal α-chains, the molecule will be overmodified from the

mutation to the amino end of the molecule and therefore will be thermally unstable and open to endogenous protease digestion. Normal α-chains will be digested along with mutated ones if they are part of an abnormal procollagen molecule. This phenomenon explains why mutations affecting pro α1(I) tend to result in a more severe phenotype than mutations affecting pro α2(I). A mutation affecting pro α2(I) would lead to 50% of the type I procollagen molecules in a cell being overmodified, a mutation affecting pro α1(I) would result in 75% of the type I procollagen molecules being overmodified and therefore open to degradation by the cell. The less the type I collagen in the extracellular matrix the more severe phenotype. Given the fact that not all abnormal type I collagen molecules are degraded by the cell, it should be noted that the more abnormal collagen molecules that are incorporated into the extracellular matrix, the more severe the phenotype.

1.11.10 KINKED TYPE I COLLAGEN MOLECULES

A reduced type I collagen content in the extracellular matrix is not the only means by which an OI phenotype is caused. Disruption of the type I collagen fibre structure is the more common mode of action because the cell does not degrade all the abnormal procollagen molecules. Type I collagen fibre structure is compact and highly ordered. The fibres are made up of many type I molecules packed together therefore, a few abnormal molecules can disrupt the whole fibre structure. Stacey *et al.* (1988) published data to prove that only 10% of the collagen molecules in a fibre need be abnormal to result in a lethal OI phenotype. One way that a small number of mutated α-chain can result in an abnormal collagen fibre is explained in Vogel *et al.* (1988). The authors demonstrate that a point mutation causing the substitution of cysteine for a glycine at position 748 of α1(I) caused a kink to form in the procollagen molecule at the site of the substitution. Type I collagen molecules with kinks in them, if incorporated into collagen fibres, would totally disrupt the structure of the fibres and may, as in this case, result in a lethal OI phenotype. A more common occurrence would be that the overmodified molecules are incorporated into the collagen fibril and disrupt its thermal stability.

28

1.11.11 TYPE I COLLAGEN AND OSTEOGENESIS IMPERFECTA

Although there are no real guidelines to follow when searching for the mutation causing an OI phenotype there are few general rules that help:-

1. Mutations causing OI phenotypes are to be found in COL1A1 or COL1A2.

2. Mutations resulting in reduced expression of the type I collagen genes are less disruptive than mutations which cause the synthesis of a defective type I collagen polypeptide.

3. Mutations in COL1A1 tend to result in more severe phenotypes than those in COL1A2.

4. The general position of a mutation corresponds to the start of post-translational overmodification.

5. Mutations at the carboxyl terminal end of the type I collagen molecules tend to be more severe than those toward the amino-terminl end of the molecule.

Many exceptions to these rules can be cited therefore, the characterization of the mutations causing OI phenotypes is not straight forward.

1.12   EHLERS-DANLOS SYNDROME

1.12.1   SIGNS AND SYMPTOMS

Ehlers-Danlos syndrome (EDS) occurs in 1 in 100,000 births. It is characterized by joint hypermobility and hyperextensive skin. The syndrome may be characterized by these symptoms but the clinical defects displayed by patients suffering from the disease are diverse. So diverse in fact that in some cases it is difficult to differentiate between Ehlers-Danlos syndrome and osteogenesis imperfecta (Sippola *et al.*, 1984).

There are at least 10 types of EDS with a number of subtypes (Beighton *et al.*, 1988). All the different types display the cardinal defects

desicribed above. For most types the biochemical defect is unknown.

EDS TYPES I, II AND III

These three types are inherited autosomal dominantly. They are characterized by hyperextensible and fragile skin, easy bruising and joint hypermobility to a greater or lesser degree.

EDS TYPE IV

This is the most severe and the most rare form of the syndrome. The symptoms displayed by patients with this form of the disease are loose and fragile skin, easy bruisability, 'cigarette paper' scars, arterial ruptures and mitral valve prolapse (Hartsfield and Kousseff, 1990). It is due to the occurrence of the last two symptoms that the life expectancy of patients suffering from type IV EDS is reduced. Survival beyond the 5th decade is rare.

Biochemical studies of tissues or cultured cells from several type IV EDS patients demonstrated a deficiency of type III collagen brought about by defects in the rate of synthesis, (Pope *et al.*, 1975) secretion or stability of type III procollagen molecules (Utani *et al.*, 1990).

EDS TYPE V

Undercrosslinking of collagen molecules has been suggested as the cause of this type of EDS (Differante *et al.*, 1975). This type is linked to the X chromosome (Beighton and Curtis, 1985).

EDS TYPE VI

Type VI EDS is differentiated from the other types by the fact that the eyes of the patient are affected ie. retinal detachment (Beighton, 1970). The patient also suffers from scoliosis. This condition has been demonstrated to be due to underactivity of lysyl hydroxylase (Pinnell *et al.*, 1972). This enzyme hydroxylates lysine residues in the collagen α-chains before helix formation. The hydroxylated lysine residues are involved in the cross-linking of collagen molecules. Type VI is inherited in an autosomal recessive manner.

EDS TYPE VII

The EDS VII phenotype has been demonstrated to be the clinical consequence of retention of the amino propeptide domain of type I collagen. It has been found to be dominantly inherited. The occurrence of the disease provides evidence that the size of type I collagen molecules is critical to the formation of normal type I collagen fibres.

EDS TYPE VIII

The symptoms of this form of the disease include periodontitis, gingival recession and tooth loss.

EDS TYPE IX

The condition which was once known as EDS IX has been re-categorised as a disorder of copper transport (Beighton *et al.*, 1988).

EDS TYPE X

This type of this disorder is inherited in an autosomal recessive manner. It is thought to be caused by an abnormality of the extracellular matrix protein fibronectin.

1.12.2   THE MOLECULAR DEFECTS CAUSING SOME EDS PHENOTYPES

Most is known about the molecular causes of EDS IV and EDS VII. This is probably because these forms of the disease are caused by mutations in collagen genes for which genomic and cDNA clones are freely available.

EDS TYPE IV

Tsipouras *et al.* (1986a) and Nicholls *et al.* (1988) demonstrated that the EDS IV phenotype is linked to COL1A3. Later Superti-Furga *et al.* (1988) characterized a 3.3kb deletion in the triple helical domain of the gene coding for type III collagen (COL1A3) in a patient with EDS IV. The protein product of this mutated gene was shorter than normal. It was incorporated into procollagen molecules which displayed decreased thermal stability and were less efficiently secreted by cells. A single base mutation resulting in an amino acid substitution (glycine to serine) has been demonstrated by Tromp *et al.* (1989) to cause an EDS IV phenotype.

Three splicing mutations have been found along the length of COL1A3 in individuals suffering from EDS IV (Kuivaniemi *et al.*, 1990). One of these splicing mutations caused a phenotype best described as an overlap between familial arterial aneurysms and EDS IV (Kontusaari *et al.* 1990). This case begs the question as to whether the more common connective tissue disorders such as familial arterial aneurysms are the result of collagen gene mutations.

EDS TYPE VII

There are three EDS VII classifications (Beighton *et al.*, 1988);-

VIIA - the result of abnormal pro-α1(I).

VIIB - the result of abnormal pro-α2(I).

VIIC - deficiency in procollagen N-proteinase activity.

The EDS VII phenotype is caused by the retention of the amino terminal propeptide of type I procollagen. Although it is possible that this may be brought about because of defective N-proteinase, such a case has not been described in the literature. This subtype, if is occurs, would be inherited in an autosomal recessive manner, unlike subtypes VIIA or VIIB. To date, only EDS VII cases which are caused by mutations which alter the N-protease cleavage site of pro α1(I) or pro α2(I) have been characterized (Eyre *et al.*, 1985; Cole *et al.*, 1986; Wirtz *et al.*, 1987; Weil *et al.*, 1990). Such mutations caused the splicing out of exon 6 of COL1A1 or COL1A2 during mRNA synthesis. Exon 6 is the 'joining' exon in both genes. It contains the sequences coding for the amino telopeptide and short segments of the amino propeptide and the helix (Chu *et al.* 1984). Contained within this exon is the N-proteinase cleavage site.

Dombrowski and Prockop (1985) suggested that the N-propeptide protease cleavage site conformation consists of three hairpin loops stabilized by lateral interactions. The loss of one of the recognition sites from one of the α-chains would destroy this conformation and result in the inhibition of cleavage. This theory was strengthened by the observations of Vogel *et al.* (1988), who found that a mutation in the helical domain of type I collagen caused a kink to form in the helix. As a result of this kink formation the α-chains were out of normal phase amino terminal to the mutation. This phase shift resulted in the disruption of the N-propeptide cleavage site and the retention of the N-propeptide.

Wirtz *et al.* (1990) used electron microscopy to show retention of the N-propeptide by type I procollagen extracted from a patient with EDS VII.

An 18 amino acid deletion (exon 6) in proα2(I) caused the loss of the N-propeptide protease cleavage site and thus N-propeptide retention. Their results indicate that the α1(I) propeptide remains attached to the α2(I) popeptide through non-covalent interactions, despite being cleaved from the related α1(I) chains. Thus cleavage of α-chains in type I procollagen molecules at the N-propeptide protease cleavage site is dependent on the presence of hairpin loops in each α-chain, but not necessarily any lateral interactions between the three loops. The retention of the amino propeptide domain described in Vogel *et al.* (1988) was probably due to the loss of the hairpin loop structure in one or two of the α1(I) chains brought about by the phase shift in the helix formation.

For most types of EDS the molecular cause is unknown. The quest for the molecular causes of many EDS phenotypes is probably complicated by the fact that each phenotype may be caused by mutations in different genes. This is unlike osteogenesis imperfecta which, in most cases, is caused by a mutation in one of the type I collagen genes (COL1A1 or COL1A2).

1.13   MARFAN SYNDROME

1.13.1   SIGNS AND SYMPTOMS

Marfan syndrome is a relatively common connective tissue disease. It occurs in approximately 4-6/100,000 people (Pyeritz and McKusick, 1979). It is inherited in an autosomal manner and is expressed as a wide variety of symptoms. Unlike OI and EDS diagnosis of Marfan syndrome is typically made not at birth but later in life (Tsipouras, 1990). The symptoms include abnormalities of the eye, aorta and skeleton. Marfan syndrome patients are typically tall with chest and spine deformaties, their limbs are very long compared with their torso. Significant variation of symptoms between families suffering from Marfan syndrome and variations in clinical expression between affected members of the same family are seen. The expected life span of a patient with Marfan syndrome is often reduced because disease of the aorta leads to heart failure or aortic dissection.

1.13.2   THE CHROMOSOMAL LOCATION OF THE MARFAN GENE

The involvement of the skeleton and the cardiovascular system in the disease symptoms led to the assumption that a mutated collagen gene may be the cause of the disease. Genetic linkage studies have excluded many collagen gene loci (Tsipouras *et al.*, 1986; Dalgleish *et al.*, 1987; Ogilvie *et al.*, 1987; Francomano *et al.*, 1988).

An international consortium was set up to collect linkage data and generate an 'exclusion map' (Blanton *et al.*, 1990). Eventually the gene locus involved in causing Marfan syndrome was localized to chromosome 15 (Kainulainen *et al.*, 1990). The position of the locus has been since found to be more centromeric than that suggested by Kainulainen and coworkers. Dietz *et al.* (1991) found it to be tightly linked to markers at 15q15-q21.1. The nature of the protein(s) involved in causing Marfan syndrome are as yet unknown.

1.12.3   COLLAGEN GENES AND MARFAN SYNDROME

Despite the lack of linkage data to be found between Marfan syndrome and the genes coding for the fibrillar collagens, one case has been reported in which a mutation in COL1A2 is indicated as the cause of Marfan symptoms (Byers *et al.*, 1981; Phillips *et al.*, 1990). The mutation was a single base change resulting in an amino acid substitution of arginine[618] by a glutamine. The substitution is in the Y position of a Gly-X-Y repeat. This arginine is highly conserved among the fibrillar collagens (Kuivaniemi *et al.*, 1988a; Bernard *et al.*, 1983; Su *et al.*, 1989; Janeczko and Ramirez, 1989; Weil *et al.*, 1987) and therefore might perform some as yet unknown important function in correct type I collagen formation. This case is generally considered an exception to the rule. Another such exception has yet to be fully characterized Godfrey *et al.* (1990a).

1.13.4   MICROFIBRIL STRUCTURE AND MARFAN SYNDROME

Recent studies have pointed towards a defective component of elastin-associated microfibrillar fibres as the cause of Marfan syndrome (Godfrey *et al.* 1990a). Indirect immunofluoresent studies using monoclonal

antibodies against fibrillin were used. Fibrilin is a major constituent of microfibrils (Maddox *et al.*, 1989). Abnormal microfibrillar structure was observed in skin and fibroblast samples isolated from patients suffering from Marfan syndrome. No such abnormalities were seen in samples isolated from normal individuals. An unique Marfan case vindicated this approach as a route to follow in the search for the defective protein in Marfan patients (Godfrey *et al.*, 1990b). The patient was found to display Marfan symptoms unilaterally. Godfrey and co-workers used indirect immunofluorescence to show that microfibril fibres from the Marfan affected side of the patients body were abnormal (reduced aggregation), while the fibres from the normal side of the body were not. Work is under way to determine if the gene coding for fibrillin is to be found on chromosome 15.

Obviously there is still much more to do until the molecular cause(s) of Marfan syndrome are elucidated. It has taken a long time to get to this stage in the search for the Marfan gene, the rest of the search should be much quicker.

1.14   OTHER CONNECTIVE TISSUE DISEASES

Stickler syndrome is a dominantly inherited condition characterized by vitreo-retinal degeneration, epiphyseal dysplasia and premature degenerative joint disease. Linkage has been suggested between this syndrome and the type II collagen gene (Francomana *et al.*, 1987; Knowlton *et al.*, 1989; Priestley *et al.*, 1990). However Weaver *et al.*, (1989) published results to suggest that mutations in COL2A1 are not the only cause of this disorder.

Type II collagen has also been found to be involved in the occurrence of some forms of spondyloepiphyseal dysplasia (Anderson *et al.*, 1990; Tiller *et al.*, 1990) and primary osteoarthritis associated with a mild chondrodysplasia (Ala-Kokko *et al.*, 1990; Knowlton, 1990). The mutation described in Ala-Kokko *et al.* (1990) was shown by Eyre *et al.* (1991) to be expressed in the patient's cartilage.

Biochemical and immunocytochemical studies have suggested the possible involvement of the COL7A1 gene with the recessive dystrophic phenotype of epidermolysis bullosa (Leigh *et al.*, 1988).

Spotila *et al.* (1991) have found a mutation that substitutes a serine for a glycine in $\alpha2(I)$ at position 661 in a patient with post-

menopausal osteoporosis.

It remains to be determined whether all the common connective tissue disorders are the result of mutations in the collagen genes, and if they are, whether it would be easy to screen for them in large populations. Screening for susceptibility to aortic aneurysms for example, could result in preventative measures that would increase life expectancy.

1.15   DETECTION OF SINGLE BASE MUTATIONS

Reviews of the methods described here are to be found within Cotton (1989) and Rossiter and Caskey (1990).

Before attempting to screen DNA or mRNA for small mutations the sequence should be first screened for gross defects. This can be easily carried out by Southern or northern blotting and hybridization analysis.

Ultimately all mutations must be characterized by sequence analysis. It seems foolhardy to attempt to sequence large areas of DNA without first gaining an idea as to the general location of the mutation. Although this has been attempted by Labhard and Hollister (1990). A better and less time consuming approach is to use one of the screening methods described below to determine the location of the mutation, then to sequence that area of DNA to characterize the mutation.

Many of the screening methods rely on the formation of heteroduplexes of DNA/mRNA fragments. One strand of the heteroduplex should contain the wild type sequence, the other strand the mutant sequence. Two perfectly complementary strands of DNA will anneal to each other along their whole length without interruption. If the two strands of DNA differ at one or more positions they will anneal except at the position(s) at which they differ. Any mismatched bases are open to modification and/or cleavage by enzymes or chemicals. Heteroduplexes can form between two DNA fragments, two RNA sequences or a DNA and an RNA sequence.

1.15.1   RNASE A PROTECTION ANALYSIS

This method was developed by Myers et al. (1985b) and Winter et al. (1985), later it was modified by Gibbs and Caskey (1987). It relies on the ability of pancreatic ribonuclease (RNase A) to recognize and cleave

mismatched RNA bases in a heteroduplex. It requires the synthesis of RNA probes labelled with ^{32}P to a high specific activity. An antisense, wild-type sequence, RNA probe is hybridized to mutant mRNA. After hybridization, the heteroduplex is digested with RNase A. The product(s) of the digestion are electrophoresed through a polyacrylamide gel. The presence of a mismatch base pair between the probe and the mRNA is indicated by cleavage bands on an autoradiograph of the gel. Lack of cleavage of the probe because it is completely protected by the mRNA is seen on an autoradiograph of the gel as a single band, its size corresponding to that of the probe trimmed of any superfluous sequence. The RNA probe can also be protected by DNA sequence in a DNA/RNA hybrid.

This method has been applied to the detection of point mutations causing numerous clinical conditions (Grange *et al.*, 1990; Veres *et al.*, 1987; Dunn *et al.*, 1988; Forrester *et al.*, 1987). There are problems with this method, a major one being that the chemicals and equipment must be rendered RNase free. There is also a limitation on the length of sequence that can be screened at any one time. The probes used in this method are limited in length to between 400bp and 1kb. Probes larger than 1kb result in a high background signal on the final autoradiograph. Genovese *et al.* (1989) adapted the method to increase the length of sequence that can be screened at a time. They were able to screen up to 1.8kb of sequence at one time. They denatured non-radiolabelled wildtype cDNAs sequence and hybridized it to mutant mRNA. This hybrid was subjected to digestion by RNase A as usual then electrophoresed through a denaturing agarose gel. The gel was then northern blotted. The resulting filter was hybridized with an antisense cDNA probe. If there was no mismatch within the original mRNA/cDNA hybrid then no cleavage products would be seen on the autoradiograph of the filter. The presence of a mutation in the mRNA would result in the presence of cleavage bands on the autoradiograph.

The enzyme RNase A does not digest all the mismatches in a heteroduplex. The ability of RNase A to digest a mismatched base is dependent on the base and the sequence by which the base is surrounded (Cotton, 1989). The success rate of the method has been estimated to be between 60% (Cotton, 1989) and 30% (Myers *et al.*, 1985).

1.15.2 CHEMICAL CLEAVAGE

This method was developed by Cotton *et al.* (1988). As with RNase A protection the method relies on the availability of mismatched bases within a heteroduplex. The mismatched bases in this method are modified before they are cleaved. A heteroduplex is formed between a radioactive, wild type sequence probe DNA and DNA or mRNA isolated from an individual with a suspected mutation.

The heteroduplex is treated with hydroxylamine or osmium tetroxide. Hydroxylamine reacts with mismatched C bases and osmium tetroxide reacts with mismatched T bases. After modification these bases are cleaved by the presence of an alkali such as piperidine. The presence of cleavage products is determined by denaturing polyacrylamide gel electrophoresis and autoradiography. Like RNase A mapping, this method identifies not only the presence of a mismatch in the hybrid but also gives an indication as to the position of the mismatch. Due to the fact that only C and T bases can be modified, and subsequently cleaved, both strands of DNA need to be analysed or mRNA and its cDNA must be screened.

A further drawback of this method is high background signal. Cotton and Campbell (1989) determined the success rate of this method to be 100%. However, Theophilus *et al.* (1989) failed to detect all the base changes they were investigating because of high background. This method is time consuming and has the disadvantage of involving the use of highly toxic chemicals.

1.15.3 CARBODIIMIDE

Carbodiimide is used as the reactive chemical in two mutation detection methods. Both methods rely on the ability of carbodiimide to interact in a non-covalent manner with mismatched C or T bases.

The first method was described in Novack *et al.* (1986). This method exploits the fact that DNA containing a base modified with carbodiimide migrates more slowly through a gel matrix than DNA which has not been modified. This method does not indicate the location of a mutation.

The second method employing carbodiimide also involves the modification of mismatched C and T bases within a heteroduplex. After modification, the heteroduplex is used as the template for primer extension

with *Taq* DNA polymerase. The conditions under which the primer extension takes place are such that extension is prevented by the presence of a carbodiimide-modified base (Ganguly and Prockop, 1990). Unlike the first carbodiidmide method this method gives an indication of the location of the mutation within the heteroduplex.

Although the method is claimed to be 100% effect at detecting point mutations occasionally the results are distorted by internal priming of the DNA or non-specific modification of bases (Ganguly and Prockop, 1990).

1.15.4 DENATURING POLYACRYLAMIDE GEL ELECTROPHORESIS

Duplex DNA, when electrophoresed through a denaturing gradient gel, will do so as a duplex until a point is reached in the gel where the denaturant concentration is such that the DNA duplex starts to melt. At this point the migration of the duplex through the gel is retarded. The duplex melts as two domains, high and low melting point. The concentration of denaturant that at which a duplex begins to melt is dependent on the length and the sequence of the DNA fragments in the duplex (Lerman *et al.*, 1984). A mismatched pair in a low melting region of a duplex destabilizes it and as a result reduces the concentration of denaturant at which the migration of the duplex slows down. Duplexes differing by a single base pair can be separated using denaturing gradient gel electrophoresis.

The presence of a mismatched base pair in a high melting domain of a duplex would not alter the moblilty of the duplex through the gel because the low melting domain would melt first. To increase the sensitivity of this method a GC-rich sequence or 'clamp' is attached to one end of the duplex. The presence of this high melting domain effectively makes the whole of the sequence under investigation the low melting domain (Myers *et al.*, 1985c,d). A mismatch anywhere along this domain would reduce the concentration of denaturant at which the duplex melts and thus allow detection of a mutation within this region.

A technique has been developed along the same lines as this by Diaz-Ruiz and Kaper (1979). A thermal gradient is used instead of a denaturing gradient. However, the setting up of a thermal gradient requires expensive equipment.

Neither method enables the location of the mutation to be determined.

1.15.5 POLYMERASE CHAIN REACTION SINGLE STRAND CONFORMATIONAL POLYMORPHISMS

This method involves the use of the polymerase chain reaction as one of the steps (Orita *et al.*, 1989). A fragment of DNA (100-400bp in length) is amplified using the PCR. Included in the PCR is ^{32}P-labelled dCTP or a ^{32}P-labelled primer. The product of the PCR is denatured then electrophoresed through a non-denaturing polyacrylamide gel. The single stranded PCR product adopts a specific comformation as it is electrophoresed through the gel. The conformation that the DNA takes on is stabilized by weak intramolecular bonds. The mobility of these single stranded DNA fragments through the polyacrylamide gel is dependent on the length and the sequence of the DNA. This method has been shown to be able to distinguish between fragments which differ by one base substitution (Orita *et al.*, 1989).

This method has the advantages that it is quick, easy and large areas of DNA can be screened at one time. Clinical applications of this method to detect disease causing mutations have been described in Ainsworth *et al.* (1991), Cawthorn *et al.* (1990) and Dean *et al.* (1990).

1.15.6 HYDROLINK GEL ELECTROPHORESIS

Potentially the simplest method for the detection of single base mutations yet published involves the use of Hydrolink gels (Keen *et al.*, 1991). Hydrolink gels are high resolution gels which are prepared using a gel solution which is obtained in the UK from Hoefer Scientific. These gels are said to be able to separate heteroduplexes containing a single base mismatch from duplexes which are completely homologous. Keen and coworkers used the PCR to amplify fragments of DNA which contained a point mutation. After amplification the DNA was denatured once then reannealed to encourage heteroduplex formation. The DNA was electrophoresed through a Hydrolink gel.

This method has been tested on a limited number of base substitutions (Keen *et al.*, 1991). Further analysis needs to be carried out to determine how sensitive the method is.

This method may be enhanced by the incorporation of nucleotide analogues which alter electrophoretic mobility of heteroduplexes (Kornberg and Lavik, 1989).

The fact that so many single base substitution detection methods can be picked from the literature is a clear indication that none of the methods is perfect. The ideal mutation detection method would satisfy the following criteria:-

1. It would detect all possible base changes regardless of the base composition of the surrounding DNA

2. It would give an indication of the location of the mutation to make characterization of the mutation easier.

3. It would be quick, user-friendly and safe so that eventually mass screening can be performed.

The method that best satisfies these criteria at the moment is the use of Hydrolink gels. However, it has not been thoroughly tested for its ability to detect all possible mutations.

1.16   THE POLYMERASE CHAIN REACTION

Comprehensive reviews of the polymerase chain reaction are published in Gibbs (1990) and Erlich (1989).

It would not be exaggeration to say that the polymerase chain reaction (PCR) has revolutionized the field of molecular biology. It has become a part of many of the procedures used in DNA manipulation and new adaptations of the method are being published constantly. A full review of the PCR and its uses would require more space than is available here. With this in mind I intend to outline the way in which the method works, its components and a few of its uses.

The polymerase chain reaction enables the selective synthesis of predetermined fragments of DNA to quantities of a few micrograms using a few copies of the DNA sequence as starting material. The only requirement of the method is that some of the sequence of the DNA to be isolated must be known. This sequence data is required so that primers can be synthesized to prime DNA synthesis.

41

1.16.1 AN OUTLINE OF HOW THE METHOD WORKS

During the PCR the template DNA is denatured by increasing the temperature of the reaction mix to something in the region of 95°C. The temperature of the reaction mix is then decreased to allow the primers, which are present in a molar excess, to anneal to their complementary sequences in the template DNA. The primers are designed such that one has a sequence which is complementary to the sense template DNA strand and the other has a sequence which is complementary to the antisense template DNA strand. The positions of the primers define the ends of the PCR products.

Having annealed to the template DNA, the primers prime DNA synthesis. At first the enzyme used to catalyse this DNA synthesis was *E.coli* DNA polymerase Klenow fragment (Mullis and Faloona, 1987). This enzyme is heat labile, so DNA synthesis took place at 37°C and the enzyme had to be replenished after each denaturation step; a labour intensive and expensive process. DNA synthesis and therefore the primer annealing at 37°C resulted in many non-specific priming products. A great advance in the specifcity and ease with which this method can be performed was made when a thermostable DNA polymerase isolated from an organism called *Thermus aquaticus* was included in the method (Saiki *et al.*, 1988). This enzyme, *Taq* polymerase, performs the same function as *E.coli* DNA polymerase but it is stable at temperatures up to 95°C (the denaturing temperature) and its optimum working temperature is 70°C. Using this enzyme to catalyse DNA synthesis meant that the annealing temperature of a PCR could be increased to 70°C. Increased specificity of the annealing temperature resulted in increased specificity of the method so that the fragments of DNA could be amplified to the exclusion of all else. The polymerase chain reaction involves repeated cycles of temperature changes which correlate with the denaturation of the template, annealing of the primers to the template and DNA synthesis. Each cycle of PCR results in the synthesis of more DNA which is used as template for further cycles. Theoretically, the number of molecules of a specific DNA fragment will increase by 2^n after n cycles of PCR.

The isolation of *Taq* polymerase has also allowed the automation of this method. Thermocyling machines have been produced which alter the temperature of the reaction mix to allow template denaturation, primer annealing and DNA elongation. The most recently described machine

enables amplification of a 536 bp DNA fragment to take place in less than one hour (Wittner and Garling, 1991).

1.16.2 TEMPLATE MATERIAL

After the isolation of *Taq* polymerase and the production of thermocycling machines, thoughts turned to the multitude of template preparations that can be used as part of the method. Boiled cells (Gussaw and Clarkson 1989), paraffin embedded tissue samples (Impraim *et al.*, 1987) and whole blood (Mercier *et al.*, 1990) have been successfully used to provide template for PCRs. Refinement of the method to increase its sensitivity has lead to lower and lower amounts of DNA template material being required for PCR amplification. This eventually led to Li *et al.* (1988) publishing results to show that it is possible to amplify DNA from a single sperm. Chelly *et al.* (1988) amplified dystrophin mRNA via a cDNA intermediate from lymphoblastoid cells which were thought not to express the mRNA. This and later results (Chelly *et al.*, 1989) led to the speculation that all cells express all their genes, at least at a very low basal level.

Amplification of mRNA via a cDNA intermediate also turned out to be relatively straight forward. Crude cell extracts may be used as template material for such PCRs (Ferre and Garduno, 1989). Adaptation of the PCR to amplify mRNA requires only that a single strand cDNA synthesis step precedes the PCR. The second cDNA strand is synthesized by the polymerase chain reaction. Erlich (1989) published a method that enables cDNA synthesis and the PCR to take place in one tube.

Amplification of mRNA via a cDNA intermediate allows easy analysis of RNA sequences. This is especially useful when screening very large genes, such as members of the collagen gene family (Labhard and Hollister, 1990) and the dystrophin gene (Roberts *et al.*, 1990), for mutations which cause disease phenotypes. It is a reflection of the versatility of the method that most of the single base mutation detection methods described previously can involve PCR amplification of DNA or mRNA.

The amplification of mRNA via a cDNA intermediate using the PCR allows easy analysis of mRNA splicing to take place. Kuivaniemi *et al.* (1990) used PCR to show that the same sequence change in the splice site consensus sequence of different introns within a gene results in mis-splicing consequences,

1.16.3   QUANTITATIVE PCR

PCR amplification of RNA has been used in the detection of retroviruses such as the humanpapilloma virus (Young *et al.*, 1988) and HIV (Kemp *et al.*,1989). The amount of viral RNA available for amplification can be quantified by including in the PCR a known amount of control template and a second set of primers. The control PCR and the experimental PCR take place at the same time in the same reaction mix. Quantification of the template material can only take place after a limited number of temperature cycles. As a general rule a PCR runs for 30 cycles of temperature changes, unless a relatively high concentration of plasmid DNA is used as the template material. During the reaction the concentration of DNA increases exponentially until it reaches a certain level and then it will increase no more. To determine the quantity of template material the reaction must be terminated while the DNA concentration is increasing exponentially. Only during this time will the amount of control PCR product and the amount of 'experimental' PCR product be proportionally related. Why the rate of DNA amplification slows down after a certain number of cycles is unclear. It is not that the PCR substrates, the primers and deoxyribonucleotides are depleted. They are known to be present in excess from the beginning to end of the reaction. What is more likely is that the *Taq* polymerase activity is decreased because of the many temperature increases to 95°C. Although *Taq* polymerase can withstand temperatures of up to 95°C its optimum temperature is 70°C. The other reason for the DNA concentration levelling off is that after many temperature cycles the template concentration in the reaction is so high, it denatures then anneals to itself in preference to the primers. This suggestion is supported by the observation that high template starting concentrations inhibit PCR amplification.

1.16.4   BREAKING NEW GROUND WITH THE PCR

The PCR has been adapted to amplify not only regions of DNA for which the sequence has already been determined but also some regions which are new. The amplification of previously uncharacterized regions of DNA or mRNA is normally dependent on one of the primers having a

44

sequence homologous to the template material. DNA synthesis is extended from known primer sequence to the poly (A^+) tail or a similar sequence artificially attached to the end of the template material (Ohara *et al.*, 1989). Such methods are normally used for the amplification of mRNA sequence. Amplification of DNA which has not previously been sequenced was achieved by Triglia *et al.* (1988). First the DNA was digested with a restriction endonuclease which digested the DNA so that fragments of DNA are produced which contain the known sequence and neighbouring uncharacterized sequence. The digested DNA fragments were circularized by ligating at a low DNA concentration and then used as template for PCR amplification. The primers used in the amplification reaction anneal to either strand of the circularized DNA at positions which are back to back to each other within the characterized region.

DNA or mRNA for which no sequencing data has been published can be amplified using degenerate primers. These primers are synthesized with part of their sequence not predetermined G, A, T or Cs but among the population of primers the positions are occupied by an equal number of each of the bases. Such primers are used when the primer sequence is determined by protein sequence, they allow for codon degeneracy (Lee *et al.*, 1988). They are also used when the sequence of the gene under investigation is known for one species but not for the species from which the template DNA has been isolated.

Instead of just isolating DNA fragments the PCR has also been used in the construction of new DNA sequences. Adaptations of the PCR have been used for site-directed mutagenesis and the production of chimeric genes (Vallette *et al.*, 1989; Helmsley *et al.*, 1989).

1.16.5   SEQUENCING PCR PRODUCTS

The most frequent use of the PCR is amplification of fragments of DNA prior to further analysis. Often changes in the sizes of DNA fragments seen by gel electrophoresis provide enough information, for example in the amplification of DNA for use in genetic fingerprinting (Jeffreys *et al.*, 1988). However, the amplified DNA is often sequenced. Sequencing amplified DNA revealed one of the drawbacks of the method, it has an inherently high base misincorporation rate (Dunning *et al.*, 1988; Tindall and Kunkel, 1988). The high error rate is a result of the enzyme not possessing a 3' to 5' exonuclease

activity (Innis *et al.*, 1989). This high error rate has always to be considered when sequencing PCR products, especially if the PCR product is cloned then sequenced. Isolation and sequencing of single clones results in misincorporated bases being included in the sequencing ladder. To overcome this problem many clones must be sequenced to differentiate between PCR errors and real sequence changes. A quicker route to overcome this problem is to directly sequence the PCR product. Direct sequencing involves the sequencing of all the PCR product DNA on the same sequencing ladder. All the bases in the PCR fragment are about equally likely to be misincorporated, so the variant bases are spread throughout the sequence and only the most commonly occurring sequence is seen on the ladder. This sequence will be the real sequence of the amplified DNA unless the concentration of the starting material is very low and the error occurs early on in the amplification reaction. In this case the misincorporated base will be present in the majority of DNA fragments and would be seen on the sequencing ladder. Direct sequencing also has the advantage that it is quicker than subcloning the amplified fragment, transforming it into bacteria then isolating the plasmid before sequencing.

Another problem encountered when working with PCR amplified DNA is that of cloning the DNA. If restriction endonuclease recognition sites are incorporated into the 5' end of the primers with a sufficient number of bases 5' to them, the PCR fragments can be digested with the restriction endonuclease before ligating to vector DNA (Scharf *et al.*, 1986). Attempts to blunt-end ligate PCR amplified DNA have resulted in limited success. It seems that *Taq* polymerase adds a single base (often an adenine) to the end of each DNA strand it synthesizes (Clarke *et al.*, 1988). Measures can be taken to overcome this problem, these include the removal of the extra base by treating the DNA with the Klenow fragment of *E.coli* DNA polymerase (Helmsley *et al.*, 1989), or ligating the DNA to a vector with a 5' thymidine base overhang (Holton and Graham, 1991; Marchuk *et al.*, 1991). Other more elaborate methods are described in Stocker (1990), Shuldiner *et al.* (1990) and Alsanidis and Jong (1990).

Described above are minor problems encountered when working with PCR, these are far outweighed by the ability of PCR to make areas of DNA or RNA easily accessible from many different nucleic acid sources. The ease with which this method can be successfully applied to many DNA manipulation techniques means that it will probably find use in more procedures in the future.

46

1.16.6   ALTERNATIVES TO THE POLYMERASE CHAIN REACTION

All except one of the alternative PCR methods overcome the need for denaturing and annealing steps in the reactions. The system requiring the denaturation step involves the annealing of a RNA probe to the DNA template then DNA amplification involving a bacterial enzyme Qβ-Replicase (Lorneli *et al.*, 1989). The other methods are very similar in design to that published by Compton (1991). This method which is called nucleic acid sequence-based amplification (NASBA) involves the annealing of the first primer to a RNA template. DNA is synthesized from the primer using the RNA as template and reverse transcriptase. The RNA strand of the hybrid is then digested by RNase H. The second strand is synthesized using the second primer and reverse transcriptase. The double stranded product contains within it a T7 RNA polymerase promoter. Its sequence is incorporated into the first primer. This promoter activates RNA synthesis by the third enzyme, T7 RNA polymerase. The RNA strand is digested by RNase H then the single stranded cDNA is available for reverse transcriptase activity to synthesize the complementary DNA strand. Systems such as this are said to have an advantage over conventional PCR in that they do not require the use of a thermocycling machine. None of these methods are used on a routine basis by many laboratories. Conventional PCR is by far the most popular method.

1.17   THE AIMS OF THIS PROJECT

To add to the accumulating information concerning the collagen gene family and the effects of mutations in these genes on phenotype.

Analysis of the sequences of the genes coding for the fibrillar collagens has revealed an inability to tolerate sequence changes. Most of the changes in the amino acid sequence of the fibrillar collagens result in a disease phenotype. A possible silent amino acid variant was investigated.

The collagen amino acid sequence changes that cause the connective tissue disease osteogenesis imperfecta are many and diverse. α2(I) collagen mRNA, isolated from patients with non-lethal osteogenesis imperfecta, was screened for the presence of mutations using RNase A mapping.

Fifteen collagen types have been at least partially characterized to date. A novel method of producing a cDNA library enriched with collagen coding sequences was attempted.

CHAPTER TWO

MATERIALS AND METHODS

2.1   TISSUE CULTURE CELL LINES

Fibroblast cell lines GM1093, GM1436A, GM2573, GM2645 and GM2695 were obtained from the National Institute of General Medical Sciences Human Genetic Mutant Cell Repository, Institute of Medical Research, New Jersey, USA.
The other cell lines were gifts from:-

CF 1-4	Dr D. Shaw
	University of Wales,
	Cardiff,
	U.K.
LW, KO, SC,	Professor P.H. Byers
R, PJ, JR,	University of Washington,
PIR and JB	USA.
Transformed	Dr N.J. Royle
lymphoblasts	University of Leicester.

2.2   PLASMIDS AND BACTERIAL STRAINS

2.2.1   RECOMBINANT CLONES

The $\alpha2$(I) clone, HP2010 (Kuivaniemi *et al.*, 1988) was a gift from Dr D.J. Prockop, Thomas Jefferson University, Philadelphia, USA.

The $\alpha2$(V) clone Hf 511 (Weil *et al.*, 1987) and the $\alpha1$(I) clone HF404 (Chu *et al.*, 1982) were gifts from Dr F. Ramirez, SUNY-Health Science Center at Brooklyn, USA.

The type VI collagen clones P18 ($\alpha1$(VI)), P1 ($\alpha2$(VI)) and P24 ($\alpha3$(VI)) (Chu *et al.*, 1987) were donated by Dr M-L Chu, Thomas Jefferson University, Philadelphia, USA.

49

The α1(III) clone p(III)-55 was provided by Dr R. Dalgleish. It is a combination of the two clones p(III)-33 and p(III)-21 (Miskulin *et al.*, 1983).

2.2.2   NON-RECOMBINANT PLASMIDS

The plasmids pBluescript SK⁻/KS⁻ (Short *et al.* (1988) Stratagene Cloning Systems, San Diego, USA) and pTZR19R and pTZ19R (Pharmacia, Mead *et al.* 1986) were used in the construction of recombinant plasmids.

pUC13 (Messing, 1983) and M13mp18 (Yanis-Perron *et al.*, 1985) were used in the construction of DNA markers.

2.2.3   BACTERIAL STRAINS

E. coli strains JM83 *rec*A and XL-1 blue were used.

E. coli JM83 *rec*A was produced by Mark Matfield (1983) and was a gift from Professor W. Brammar (ICI Joint lab., Leicester University) The genotype of this strain was *ara*, δ(*lac-pro*), *Str*A, *thi*1, f80d*laci*q, *ZM*15, *rpsL*, *rec*A.

E.coli XL-1 blue was purchased from Stratagene, its genotype is, *end*A1, *hsd*R17, (rk-, mk+), *sup*E44, *thi-1*, λ-, *rec*A1, *gyr*A96, *rel*A1, (*lac-*), [F', *pro*AB, *lac*IqZVM15, Tn10, (*tet*ʳ)]

2.3   ENZYMES, ANTIBIOTICS AND CHEMICALS

The following were obtained from:-

Sigma
RNase A, Proteinase K, ampicillin, chloramphenicol, tetracycline, isoproyl-β-D-thiogalacto-pyranoside (IPTG), polyethylene glycol (PEG) grade 6000, herring sperm DNA, diethylpyrocarbonate (DEPC), N-2-hydroxyethylpiperazine-N-2-ethane-sulphonic acid (HEPES), piperazine-N-N'-bis[2-ethane-sulphonic acid] (PIPES), Orange G spermidine-HCl, β-mercaptoethanol and dithiothreitol,.

Pharmacia

DNase I, T7 RNA polymerase, DNA polymerase I (Klenow fragment) Moloney Murine Leukemia Virus reverse transcriptase, T7 DNA polymerase, Ficoll 400, dideoxyribonucleotides (ddNTPs), deoxyribonucleotides (dNTPs), ribonucleotides (NTPs), dextran sulphate and bovine serum albumin (BSA)

Fisons

Phenol, polyvinylpyrrolidine, 3-(N-morpholino)propane sulphonic acid (MOPS), formamide, formaldehyde, and 2-amino-2-(hydroxymethyl)propane-1,3-diol (Tris)

BDH

Urea, guanidinium thiocyanate and caesium chloride.

Bio-Rad

N,N,N',N'-tetramethylethylenediamine (TEMED) and ammonium persulphate.

Anglian

Taq DNA polymerase and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal).

Amersham

mAP paper, [α-^{32}P] dCTP, [α-^{32}P] CTP, [α-^{35}S] dATP and Hybond-N filters

University of Leicester, Biochemistry Department

Primers for the polymerase chain reactions and sequencing reactions were synthesized by J. Keyte.

The origin of many other reagents are stated through out the text, all were of analytical grade, where available.

2.4   MEDIA

2.4.1   FOR BACTERIA

Luria broth was 10g Difco Bacto tryptone, 5g Difco Bacto yeast extract and 5g NaCl per litre of distilled water. Luria agar was the liquid media solidified by the addition of 15g of Bactoagar per litre. Ampicillin

51

(50μg/ml) and tetracycline (12.5μg/ml were made up in as stocks in 50% ethanol and stored at -20°C.

2.4.2 FOR TISSUE CULTURE

Dulbecco's Modification of Eagles Minimum essential medium (DMEM, Gibco, (Morton, 1970) was used to culture all fibroblast cells, it contained 4.5g glucose per litre and was stored at 4°C. RPMI medium (Gibco) was used for the culture of lymphoblasts. DMEM or RPMI was supplemented with penicillin (100units/ml, Sigma), streptomycin (100μg/ml, Sigma) and 10% foetal calf serum (Gibco). After sterility testing the serum was stored at -20°C as aliquots. Phosphate buffered saline (Gibco) was obtained as a 10 x solution which was diluted with 'tissue culture clean' water before use. Trypsin-EDTA (0.5g trypsin and 0.2g EDTA per litre) was obtained from Gibco and was stored as aliquots at -20°C.

2.5 GENERAL NUCLEIC ACID MANIPULATION TECHNIQUES

2.5.1 PHENOL EXTRACTION

An equal volume of phenol:chloroform:isoamylalcohol (25:24:1) was added to the DNA solution. The mixture was vortexed for two minutes then centrifuged at 13,000rpm for three minutes in a MSE Microcentaur centrifuge. The top, aqueous layer was removed without disturbing the interface. Chloroform extraction was performed in the same way using chloroform:isoamylalcohol 24:1.

2.5.2 ETHANOL PRECIPITATION

One tenth volume of 3M sodium acetate pH5.2 or one half volume of 7.5M ammonium acetate were added to the DNA solution. After mixing, two and one half volumes of ethanol were added. The solution was mixed well before incubation at -80°C for 15 minutes or -20°C for 30 minutes. The DNA precipitate was collected by spinning at 13,000 rpm in a MSE Microcentaur microcentrifuge or at 10,000rpm in a RC5B Sorvall centrifuge,

using a HB4 rotor.

2.5.3 DETERMINATION OF DNA AND RNA CONCENTRATIONS

DNA or RNA concentrations were determined by measurement of absorbance of U.V. light at 260nm in a Hitachi U-2000 spectrophotometer. A reading of 0.02 OD units (0.03 units for RNA) corresponded to a DNA concentration of 1µg/ml Alternatively samples were compared with known concentrations of plasmid DNA on agarose gels.

2.5.4 DNA DIGESTION BY RESTRICTION ENDONUCLEASES

DNA was digested as described in the manufactures protocol, using the buffer provided. Digest reactions were usually incubated at 37°C for 1 hour. Spermidine-HCl was added to inefficient restriction endonuclease digest reactions, to increase their efficiency (Bouche, 1981; Pingoud, 1985). At the end of the reaction time, the digest was spun to collect the mix at the bottom of the tube. A one tenth volume of Orange G dye (0.1% w/v Orange G, 20% w/v Ficoll 400, 100mM Na_2EDTA) was added to the digest mix before loading onto an agarose or polyacrylamide gel.

2.5.5 PRECAUTIONS TAKEN WHEN WORKING WITH RNA

Due to the presence of RNases on glassware, tips, tubes and in solutions, all equipment and solutions to be used with RNA were treated with 0.1% diethylpyrocarbonate (DEPC) overnight. The DEPC was degraded by autoclaving. Care was taken when using RNase free solutions or equipment to keep then 'clean'.

2.6 <u>ELECTROPHORESIS OF DNA</u>

0.8% agarose gels were used to separate large DNA fragments. Fragments shorter than 600bp in length were separated with the use of 4% Nusieve agarose gels. DNA fragments which were shorter than 600bp in

length and of a similar size were separated using a polyacrylamide gels.

2.6.1 SEPARATION OF DNA FRAGMENTS USING AGAROSE GELS

Agarose gels were made using Seakem HGT agarose, Seaplaque LGT agarose or Nusieve GTG agarose all from ICN Biomedicals Ltd., High Wycombe, Bucks. Usually the agarose was dissolved in E buffer (200mM Tris base, 10mM EDTA adjusted to pH7.7 with glacial acetic acid.) however separation of low molecular weight fragments was enhanced by dissolving the agarose in TBE buffer (90mM Tris base, 90mM boric acid and 2.4mM Na_2EDTA pH 8.3). Which ever buffer was used, it contained ethidium bromide at 0.02mg/ml. Molecular weight markers were usually λ/Hind III and ϕX174/Hae III. Marker DNA samples were heated at 65°C for 5 minutes before loading onto the gel, to disassociate the cohesive ends of the λ. Before 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 at 15-45 volts overnight, until the dye had reached the bottom of the gel.

2.6.2 RNA GELS

Samples of RNA isolates were electrophoresed through 0.8% agarose gel. The samples were mixed with 0.1 volumes of 10 x Orange G loading dye before loading into the wells, alongside λ/Hind III and ϕX174/Hae III markers. The gel was run at not more than 70 volts. As with DNA gels, the RNA gels and buffer contained 0.02mg/ml ethidium bromide.

2.6.3 NON-DENATURING POLYACRYLAMIDE GELS

A stock of 40% bisacrylamide was prepared by dissolving 38g acrylamide (Serva) and 2g N,N'-methylbisacrylamide (Uniscience) in a final volume of 100ml of dH_2O. The solution was deionised by mixing with 5g of Amberlite resin (Serva) for 1 hour. The resin was removed by filtration and the bisacrylamide stock stored at 4°C.

Enough bisacrylamide stock was used to give the required percentage gel. This depended on the size of the fragments to be separated.

The bisacrylamide was mixed with 2ml of 10 x TBE (0.9M Tris base, 0.9M boric acid, 24mM Ma$_2$EDTA, pH8.3) and dH$_2$O to 40ml. 140μl of 10% ammonium persulphate and 30μl of TEMED were mixed with the gel, which was immediately poured into a 1mm thick Bio-Rad Protean II gelmould 6mm wide, 2cm deep well-formers were inserted into the gel. When the gel had polymerized the combs were removed and the gel was inserted into the gel running apparatus. The upper and lower buffer reservoirs were filled with 0.5 x TBE .

0.1 volume of loading dye containing bromophenol blue (0.1%) and xylene cyanol (0.1%) in 50% w/w glycerol was added to each sample before loading into the wells.

The gel was electrophoresed at 25mA with cold water circulating through the cooling tank. When the dyes in the loading buffer indicated the DNA had run far enough the gel mould was dismantled. The gel was removed from the mould and rinsed in 0.5 x TBE plus 0.02mg/ml ethidium bromide for 20 minutes to stain the DNA. It was rinsed in distilled water for 5 minutes to reduce the background staining. The DNA was visualised using a U.V. transilluminator.

2.7   PHOTOGRAPHING GELS

Ethidium bromide stained gels were visualised using a Chromato U.V. transilluminator Model TM-4 (U.V.P. Inc. California, USA).

A Polaroid MP-4 land camera was used with Kodak Tmax 100 professional film. The film was developed using Kodak LX 24 X-ray developer and fixed in Kodak FX-40 X-ray fixer.

2.8   RECOVERY OF DNA FROM AN AGAROSE GEL

This method was adapted from the one described by Yang et al. (1979). DNA was passed through an agarose gel as described previously. A sheet of dialysis membrane was cut to give a single layer 1cm wider than the band of DNA on the gel and 0.5cm longer than the depth of the gel. The dialysis membrane was boiled in TE, pH7.6. The DNA in the gel was visualised using a long wave U.V. wand (Ultra-Violet Products Inc. California, USA). Exposure time of the DNA to the U.V. light was kept to a minimum to

reduce DNA damage. With the current switched off, the gel was cut just below the required DNA band, the cut extended to 1.5cm beyond each side of the DNA band and down to the gel plate. Using flat ended forceps, the dialysis membrane was inserted vertically into the cut until it rested on the gel plate. The projecting part of the membrane was folded so that it rested on the gel surface pointing towards the well. The DNA was run onto the membrane at 200V. When the DNA was fully loaded the cut was extended so that the gel was cut in half. With the current still on, the upper part of the gel was gently prised away from the lower part to open up a gap of 1-2mm, such that the face of the membrane was not touching the gel. The gel was left to run for a few minutes to ensure complete loading, this was checked using the U.V. wand. Without switching the current off the membrane was securely grasped with the forceps and in one single movement transferred to a 1.5ml Eppendorf tube. This was done as rapidly as possible. A corner of the membrane was caught in the Eppendorf lid before spinning at 13,000rpm in a Microcentaur microcentrifuge for 1 min. The membrane was rinsed with 30μl distilled water and spun again. Once the DNA had been removed from the membrane, this was checked using the U.V. wand, the membrane was discarded. The DNA solution was phenol extracted, to remove any agarose, and ethanol precipitated before it was used for further analysis.

2.9   LIGATION OF DNA FRAGMENTS

From a method described in Crouse *et al.* (1983).
500ng of insert DNA were digested with appropriate restriction endonucleases and passed through a 0.6% low gelling temperature agarose gel alongside 500ng of digested vector DNA. If the ends of the vector DNA were the same, it was treated with calf intestinal alkaline phosphatase as described in section 2.10.

The bands on the gel corresponding to the digested vector and insert DNA were excised while being visualised on a U.V. transilluminator. The gel slices were transferred to preweighed Eppendorf tubes and weighed. The volume of gel excised with each slice was estimated by assuming that the gel had the density of water, 1g/ml. The gel slices were heated to 65°C and appropriate volumes of each were used in a ligation reaction containing 50ng of vector DNA, an equimolar amount of insert DNA, 30mM Tris-HCl, pH7.6, 10mM MgCl$_2$, 5mM DTT, 1mM dATP and 1unit of T4 DNA ligase. The

final volume of the ligation reaction was 100µl, it was incubated at 15°C overnight.

2.10   ALKALINE PHOSPHATASE TREATMENT OF DNA

To prevent self-ligation of vector DNA after restriction endonuclease digestion it was treated with alkaline phosphatase (calf intestinal, Boehringer Mannheim GmbH, West Germany). After restriction endonuclease digestion the DNA digest mix was heated at 65°C for 5 minutes to inactivate the endonuclease. 0.01 unit of calf intestinal alkaline phosphatase was added to the digest mix which was then incubated at 37°C for 30 minutes.

The DNA was electrophoresed on a 0.6% low gelling temperature agarose gel and prepared for ligation as described in section 2.9.

2.11   TRANSFORMATION OF DNA INTO E. COLI

DNA was transformed into E. coli using the method described in Hutchinson and Halverson (1980) using calcium chloride or by electroporation (Dower et al. 1988). Electroporation was used when high efficiency was required but the calcium chloride method was more reliable.

2.11.1   PREPARATION OF COMPETENT CELLS FOR CALCIUM CHLORIDE TRANSFORMATION

Adapted from a method published in Hutchinson and Halverson (1980).

An 8ml culture of E.coli JM83 recA or XL-1 blue was grown overnight in Luria broth at 37°C, in a shaking water bath. 1ml of this culture was used to inoculate a fresh 50ml of Luria broth. The cells were allowed to grow at 37°C until the optical density of the culture at 560nm was between 0.36 and 0.44. The culture was transferred to sterile containers and spun in a centrifuge at 3,000rpm for 5 minutes. The supernates were discarded and the pellets resuspended in 20ml 5mM calcium chloride, which was ice cold and had been filter sterilized. The cells were incubated on ice for 15 minutes

before being spun again at 3,000rpm for 5 minutes The supernates were again discarded and the pellets resuspended in 5ml 50mM calcium chloride, 10% glycerol solution, which was ice cold and sterile. 200µl aliquots of this final bacterial suspension were dispensed into sterile Eppendorf tubes resting on an ethanol/dry ice bath. The samples were stored at -80°C before use. They were kept for up to one month at -80°C.

2.11.2   CALCIUM CHLORIDE TRANSFORMATION OF DNA INTO E.COLI

After being thawed on ice competent cells were incubated, on ice, in the presence of 10ng of the DNA to be transformed in 10mM $MgCl_2$, 10mM Tris-HCl, pH7.4 and 10mM $CaCl_2$. The cells were heat-shocked at 37°C for 90 seconds, then left at room temperature for 10 minutes. 1ml of Luria broth was added to each sample before incubation at 37°C for 1 hour. Aliquots of each sample were spread onto Luria agar plates supplemented with ampicillin (100µg/ml) to select for plasmid containing bacteria and X-gal (50µg/ml) and IPTG (50µg/ml) to select for recombinant plasmids. Non-recombinant plasmids are able to complement the chromosomal lacZ deletion and thus produce β-galactosidase in the presence of inducer (IPTG), giving rise to a blue colour on cleavage of the chromogenic substrate X-gal. Recombinant plasmids, which contain an insert interrupting lacZ, will give rise to white colonies. The plates were incubated at 37°C overnight.

2.11.3   PREPARATION OF COMPETENT CELLS FOR TRANSFORMATION BY ELECTROPORATION

This method was used when a high transformation rate was required, it is explained in Dower et al. (1988).

E. coli XL-1 blue was grown in 10ml of Luria broth overnight. 2ml of this culture were used to inoculate 100ml of fresh pre-warmed Luria broth. This culture was grown at 37°C in a shaking water bath until the cell density at 600nm was about 0.55. The cells were pelleted by centrifugation at 4,000 for 15 minutes at 4°C, in a Sorval RC5B centrifuge using a HB4 rotor. The cells were rinsed, first in 100 ml then 50ml of sterile distilled water (4°C), then in 4ml of sterile 10% glycerol, 4°C, the final pellet was resuspended in 2ml of 10% glycerol (sterile, 4°C). This process served to both

58

concentrate the cells into a dense suspension and to wash away the large amounts of electrolytes in the culture medium. While being rinsed the cells were kept at 4°C. If not for immediate use, the cells were frozen as 40μl aliquots on a dry ice/ethanol bath and stored at -80°C.

2.11.4   TRANSFORMATION BY ELECTROPORATION

A Bio-Rad pulse generator, pulse controller and sample carriage were used. The transforming DNA was cleaned by phenol extraction and ethanol precipitation, the final pellet was dissolved in 10μl of distilled water. A wide range of DNA concentrations can be used with this method (Dower *et al.*, 1988). 1μl of the DNA solution was mixed with the 40μl of competent cells, which were still on ice. The cell suspension was transferred to a chilled electroporation cuvette. The cuvette containing the cells was inserted into the sample carriage and the short electrical pulse delivered. Immediately after this the cuvette was removed and 1ml of SOC medium (2% Bactotryptone, 0.5% yeast extract, 10mM NaCl, 0.25mM KCl, 20mM MgCl$_2$, 20mM MgSO$_4$ and 20mM glucose) was added to the cells. The cells were transferred to a glass tube and shaken at 37°C for 1 hour to allow the cells to recover and start to express their antibiotic resistance genes. Aliquots of the cells were spread onto Luria agar plates containing antibiotic, X-gal and IPTG.

2.12   TRANSFER OF NUCLEIC ACID TO A NYLON FILTER THEN HYBRIDIZATION OF A RADIOACTIVE PROBE TO THE IMMOBILIZED DNA.

2.12.1   TRANSFER OF DNA FROM AN AGAROSE GEL TO AN NYLON FILTER BY SOUTHERN BLOTTING

The filter used was Hybond-N from Amersham. The blotting method is based on the method originally described by Southern (1975).

The gel was photographed. After being removed from its plastic support the gel was washed for 7 minutes in 0.25M HCl (250ml). This partially depurinates the DNA within the gel matrix producing smaller DNA fragments. The gel was rinsed briefly in distilled water, then washed for 30

minutes in denaturing solution (0.5M NaOH, 1.5M NaCl). After rinsing again in distilled water the gel was washed in 250ml neutralizing solution (3M NaCl, 0.5M Tris-HCl pH7.4). The gel was transferred to the blotting apparatus, which consisted of a raised glass plate covered with 3MM chromatography paper (Whatman) soaked in 20 x SSC (3M NaCl, 0.3M Na_3citrate). The gel was placed on top of this. The 3MM paper around the edge of the gel was covered with cling film. A sheet of nylon filter and two pieces of 3MM paper were cut to the size of the gel, they were prewetted in 3 x SSC (0.45M NaCl, 45mM Na_3citrate). The nylon membrane was placed on top of the gel in such away that no bubbles were formed between the gel and the membrane. The two sheets of 3MM paper were placed on top of the membrane, again avoiding the formation of bubbles between the layers. A pile of paper towels were placed on top of this assembly and on top of this a glass plate and a weight.

The blot was allowed to proceed for at least two hours with frequent changes of the paper towels. The blot was dismantled and the filter rinsed in 3 x SSC before being allowed to dry at room temperature for one hour. When the filter had dried it was wrapped in Saran Wrap cling film. The DNA was fixed to it by exposure to a U.V. transilluminator for 45 seconds.

2.12.2 PREPARING A RADIOLABELLED PROBE

Adapted from the method described in Feinberg and Vogelstein (1984).

The DNA fragment to be labelled was separated from vector by digestion with a restriction endonuclease and electrophoresis through a 0.6% low gelling temperature agarose gel in 1 x E buffer containing 0.02mg/ml ethidium bromide. The band on the gel corresponding to the required DNA fragment was excised from the gel and placed into an Eppendorf tube. The gel slice was heated to 65°C for 5 minutes and a volume of the gel was removed which contained 10ng of DNA (the density of the gel slice was said to be 1g/ml). The gel aliquot was placed into a separate Eppendorf tube, the remaining gel slice was stored at 20°C for further use. The 10ng of DNA were made up to 9.8μl with distilled water and boiled for 5 minutes. After cooling to 37°C the following were mixed with the probe template:- 3μl OLB (240mM Tris-HCl pH8.0, 25mM $MgCl_2$, 49.7mM β-mercaptoethanol, 4.8mM each of dATP, dTTP and dGTP (prepared as stock

solutions at 0.1M, dissolved in 3mM Tris-HCl pH7.0), 1mM HEPES (prepared as a 2mM stock solution titrated to pH6.6 with NaOH) and hexadeoxyribonucleotides at 30 O.D units/ml (prepared as a stock, resuspended in 3mM Tris-HCl and 0.2M EDTA)),

0.6μl BSA (10mg/ml),

1μl [α-^{32}P] dCTP (3000 Ci/mMol, 10μCi/μl),

0.6μl Klenow fragment (1unit/μl).

This labelling reaction was mixed and either left at room temperature overnight or incubated at 37°C for two hours.

2.12.3 DETERMINATION OF THE PERCENTAGE INCORPORATION OF [α-^{32}P] dCTP INTO A PROBE

The probe labelling reaction was stopped by the addition of 85μl of stop solution (20mM NaCl, 20mM Tris-HCl pH7.5, 2mM EDTA, 0.25% SDS). After mixing, 1μl of this probe solution was mixed with 11μl of distilled water. This dilute solution was used to determine the percentage of radioactive nucleotide incorporated into the probe. Two 5μl aliquots of the diluted probe were spotted onto two separate 1cm^2 pieces of DE81 paper (Whatman DEAE ion exchange chromatography paper). One square was left at room temperature, the other was washed 6 times in 0.5M disodium hydrogen phosphate for 5 minutes each time, twice in distilled water for 1 minute and twice for 1 minute in ethanol, all on a revolving platform. These washes removed any unincorporated nucleotides. After drying, the radioactivity retained on the two squares of DE81 paper were compared by measuring Cerenkov radiation in a Packard Tricarb scintillation counter. If the radioactive nucleotide incorporation was 50% or above the probe was considered of a high enough specific activity to be used in hybridization experiments.

2.12.4 HYBRIDIZATION OF A RADIOLABELLED DNA PROBE TO DNA FIXED TO NYLON FILTER

This is an adaption of a method by Reed and Mann (1985). The prehybridization solution and the hybridization solution were essentially the same, 0.5% dried skimmed milk (Cadbury's Marvel), 1.5 x SSPE (0.27M NaCl, 15mM sodium phosphate pH7.7, 1.5mM Na_2EDTA), 1% SDS and 6% polyethylene glycol 6000. The dried milk was completely dissolved in distilled water before the other components were added. The filter was added to 20ml of prehybridization solution and incubated at 65°C in a shaking water bath for 1 hour. Hybridization took place in a fresh 20ml of solution, prewarmed at 65°C, with the addition of the labelled probe which had been boiled for 5 minutes and placed on ice for a few minutes before being added to the hybridization solution. Hybridization took place at 65°C overnight in a shaking water bath. Hybridization for 4 hours was sufficient if the concentration of DNA on the filter was high.

2.12.5 WASHING NYLON FILTERS AFTER HYBRIDIZATION

After hybridization the filter was carefully removed from the hybridization solution, and rinsed four times in 4 x 100ml aliquots of 3 x SSC (0.45M NaCl, 45mM Na_3citrate), 0.1% SDS. Then twice for 10 minutes each in two more 100ml aliquots of 3 x SSC, 0.1% SDS. The filter was then washed in four aliquots of 0.5 x SSC, 0.1% SDS for 15 minutes each time. All the washes took place at 65°C. If the filter required further washing it was washed in the more stringent wash solution, 0.1 x SSC, 0.1% SDS, at 65°C.

2.12.6 TRANSFER OF RNA FROM AN AGAROSE GEL TO A NYLON FILTER BY NORTHERN BLOTTING

3μg aliquots of total cellular RNA (up to 3μl) were denatured by glyoxalation. They were mixed with 7μl of GFM buffer (1.1M deionised glyoxal, 78% formamide and 0.6 x MOPS) and incubated at 55°C for 15 minutes. λ/Hind III markers were treated in the same way. If the RNA was to be visualised, 1μl of 10mg/ml ethidium bromide was added to each sample. After incubation 1μl of 10 x L buffer (40% deionised formamide, 49% glyoxal,

10% 10 x MOPS buffer (0.2M MOPS, 50mM sodium acetate, 1mM EDTA), 0.1% of bromophenol blue and xylene cyanol.) was added to each sample.

The samples were immediately loaded onto a neutral gel in 1 x MOPS buffer. The gel was run at about 100V, but did not exceed 50-60mA, to prevent the gel overheating.

Transfer apparatus was set up as for a Southern blot. The gel was transferred directly from the gel tank to the blotting apparatus, no further denaturation steps were needed. After transfer, the filter was quickly rinsed in 10 x SSC, placed on 3MM paper (Whatman) and dried at 37°C. The RNA was bound to the filter by U.V. irradiation for 45 seconds. The filter was baked at 80°C for 1 hour to reverse the glyoxalation. The filter was stored at 4°C until required.

2.12.7 HYBRIDIZATION OF A RADIOLABELLED PROBE TO A NORTHERN BLOT

The probe was labelled as described in section 2.12.2. Prehybridization and hybridization were as described in Denhardt (1966). The filter was prehybridized for 2 hours at 42°C in 20ml prehybridization buffer (50% v/v deionised formamide, 0.47 x Denhardt's solution (100 x solution = 10g BSA pentex fraction V, 10g Ficoll and 10g polyvinyl pyrolidone made up to 500ml with deionised water), 4.7 x SSPE, 0.1% sodium lauryl sulphate (SDS), 0.18 mg/ml heat-denatured herring sperm DNA, 10% dextran sulphate). The dextran sulphate was dissolved in the SSPE then the other components were added, with the exception of the salmon sperm DNA. The prehybridization buffer was preheated to 42°C. The salmon sperm DNA was boiled for 10 minutes, placed on ice for a few minutes then centrifuged briefly before being mixed in with the rest of the buffer immediately before using.

Dextran sulphate solution is very viscous, so the bubbles in the solution were removed before the solution was used to avoid a high background on the filter.
Hybridization buffer was prepared in the same way as the prehybridization buffer. The probe was boiled for 10 minutes before it was added to the rest of the buffer, immediately before use.

Hybridization took place overnight at 42°C in a shaking water bath.

The filter was washed in 100ml aliquots of 1 x SSC, 0.1% (w/v)

SDS, twice at room temp for 15 minutes each and twice at 50°C for 15 minutes each.

2.12.8   AUTORADIOGRAPHY

After washing, the wet filter was placed on cling film (Saran) covered cardboard and the whole thing was covered with cling film. The filter was autoradiographed either at room temperature without an intensifying screen or at ⁻80°C with an intensifying screen (Kodak X-omatic, regular). The X-ray films were developed using Kodak LX-24 X-ray developer and fixed using Kodak FX-40 X-ray fixer.

2.12.9   STRIPPING A PROBE FROM A FILTER

100ml of 0.1% SDS were heated to 95°C, the filter was added to this and shaken until the SDS solution cooled to room temperature. To check the stripping had removed the probe, the filter was autoradiographed overnight at ⁻80°C with an intensifying screen.

2.13   ISOLATION OF PLASMID DNA

2.13.1   LARGE SCALE PREPARATION OF PLASMID DNA

Adapted from a method by Ish-Horowitz and Burke (1981).

A 10ml overnight culture of *E. coli* containing the required clone was grown in Luria broth plus antibiotics. 8ml of this culture were used to inoculate 400ml of prewarmed Luria broth, plus antibiotics, in a 2L flask. When the optical density of this culture, at 600nm, was 1.5, 80mg of chloramphenicol were added. The culture was incubated overnight at 37°C.

The bacteria were harvested by centrifuging at 7,000rpm for 10 minutes at room temperature in a GS3 rotor and an RC5B Sorvall centrifuge. The supernate was discarded and the bacterial pellet suspended in 12ml solution I (50mM glucose, 25mM Tris-HCl pH8.0, 10mM EDTA). This was left at room temperature for 10 minutes then 24ml of freshly prepared solution II were added (0.3N NaOH, 1% SDS,), after mixing gently

the solution was placed on ice for 5 minutes. 12ml 5M potassium acetate pH4.8 (4°C) were added to the mixture, this was swirled gently and returned to the ice for 15 minutes.

The mixture was split into two 50ml polypropylene centrifuge tubes and spun at 10,000rpm for 10 minutes at 4°C in a HB4 rotor and an RC5B Sorvall centrifuge. The supernatant was filtered through polyallomer wool (SUPA synthetic filter wool, SUPA Aquatic Supplies Ltd, Chesterfield.) into a measuring cylinder. The volume was noted and 0.6 volumes of propan-2-ol were added. This was mixed and left at room temperature for 20 minutes before centrifuging at 10,000rpm in a HB4 rotor as before. Each pellet was washed with 5ml 70% ethanol and centrifuged again. The pellets were dissolved in 6ml TE (10mM Tris-HCl, 1mM Na_2EDTA, pH7.6) then pooled in a 15ml Corex tube (Dupont instruments). The pH of the solution was determined, the solution was neutralized with Tris base as necessary. After accurately measuring the volume of the nucleic acid solution 1g caesium chloride and 0.1ml of ethidium bromide (10mg/ml) were added for each ml of solution. After thorough mixing the solution was centrifuged at 10,000rpm for 10 minutes at room temperature. The pellet and pellicle were discarded

The density of the solution was measured using a field refractometer (Bellingham and Stanley Ltd.) The refractive index should have been 1.3874; 1.57g/ml, if not the density was adjusted by adding caesium chloride or TE. The solution was loaded into a quick seal tube (Beckman Instruments Inc. 5/8" x 3") and topped up with CsCl solution, 1.57g/ml density. The tube was sealed with a tube sealer (Beckman Instruments Inc.) and placed in a Beckman 75Ti rotor with a balance tube. An adaptor was placed on top the top of the tube. The gradient was spun overnight at 65,000, 20°C in a Sorvall OTD 65B ultracentrifuge.

The plasmid band was transferred to a 15ml Corex tube using a syringe and a 18g needle. The ethidium bromide was extracted from the DNA solution with equal volumes of CsCl saturated propan-2-ol, the extraction was repeated until the upper propan-2-ol layer was colourless, the extraction procedure was repeated once more. The DNA was separated from the CsCl in the solution by the addition of 2 volumes of distilled water and 6 volumes of ethanol (-20°C). This mixture was placed in iced water for 10 minutes before being spun at 10,000rpm for 10 minutes at 4°C. The resulting pellet was washed in 70% ethanol before being dissolved in 50µl TE.

2.13.2 MINI PREPARATION OF PLASMID DNA

From a method by Serghini *et al.* (1989).

A 5ml culture was grown overnight at 37°C. 1.5ml of this culture were transferred to an |Eppendorf tube and centrifuged in an MSE Microcentaur microcentrifuge at 13,000rpm for 30 seconds. The supernate was removed and the pellet resuspended in 50μl TNE (10mM Tris-HCl, 100mM NaCl, 1mM Na$_2$EDTA pH8.0). This cell suspension was phenol extracted, 13μl 7.5M ammonium acetate and 130μl ethanol were mixed with the aqueous phase which was subsequently left on ice for 10 minutes. The mixture was spun at 13,000rpm for 10 minutes to pellet the precipitate. The supernatant was removed and the pellet washed in 80% ethanol before being dissolved in 25μl TE.

2.5μl aliquots of the 'miniprep' DNA solutions were digested with restriction endonucleases to release the insert and to check the integrity of the clone.

2.13.3 QUICK PLASMID PREPARATION

This is an adaptation of the method described in section 2.13.1.

A 6ml culture was grown overnight in selecting media. 3ml of this culture were used to inoculate 100ml of fresh, prewarmed broth containing antibiotic to select for the plasmid containing bacteria. The cells were grown to an optical density at 600nm of 1.5. 20mg of chloramphenicol were added to the culture to increase the copy number of the plasmid. The culture was then incubated at 37°C overnight in an orbital shaker. The culture was divided between two 50ml centrifuge tubes and harvested by spinning at 7,000rpm for 10 minutes at room temperature. Each pellet was resuspended in 3ml solution I (see large scale prep.) and left at room temperature for 5 minutes. 6ml of solution II (see large scale prep) were added to each tube. The cell suspensions were mixed gently and placed on ice for 5 minutes. 4.5ml of 5M potassium acetate pH4.8, 4°C, were added to each tube to precipitate the chromosomal DNA and most of the proteins. After mixing, the tubes were placed on ice for 5 minutes before being spun at 10,000rpm for 10 minutes at 4°C. The supernatant was filtered through polyallomer wool (SUPA synthetic filter wool, SUPA Aquatic Supplies Ltd. Chesterfield) into 30ml Corex tubes. 0.6 volumes of propan-2-ol were added to the filtered supernatant which was mixed, then left to stand at room temperature for 10

min. The precipitate was collected by spinning at 10,000rpm for 10 minutes at 4°C. Each pellet was washed with 70% ethanol and spun again as described above. The pellets were dissolved in 1.5ml TE. 750μl 7.5M ammonium acetate were added to the DNA solution to precipitate any residual protein, after allowing to stand for 10 minutes at room temperature the solution was spun at 10,000rpm for 10 minutes at room temperature. The supernatant was transferred to a 15ml Corex tube and 2.5 volumes of ethanol were added. The solution was left at room temperature for 5 minutes then the DNA was recovered by spinning at 10,000rpm for 10 minutes at 4°C. The pellet was washed with 70% ethanol and spun as before. The final pellet was resuspended in 200μl TE. Because of the large amounts of RNA present when DNA was isolated in this way RNase A at a final concentration of 0.5μg/ml was added to all restriction digest reactions.

2.14   MANIPULATION OF TISSUE CULTURE CELLS

2.14.1   TISSUE CULTURE GROWTH CONDITIONS

Fibroblast cell lines were grown as monolayers at 37°C and 5% CO_2 in DMEM supplemented with 10% foetal calf serum, penicillin and streptomycin.

All culture manipulations were carried out in a class II microbiology safety cabinet (Medical Air Technology Ltd.). Glassware to be used with cells was rinsed extensively in distilled water before drying and sterilization at 200°C for 5 hours.

All solutions were made with double distilled and deionised water and were filter sterilized using 0.2μm Acrodisc filter units (Gelman Sciences Inc., USA).

2.14.2   DIVIDING TISSUE CULTURE CELLS

Cell cultures were divided when they had reached confluency. Volumes of media etc. varied with size of flask to be divided.

The medium was removed from the flask and the cells rinsed with Dulbecco's phosphate buffered saline. After the saline was removed, Trypsin-EDTA solution was added. The cells were incubated in the presence of this

67

solution at 37°C until they 'rounded up' and came away from the flask surface. The cell suspension was then passed up and down a sterile Pasteur pipette to separate the clumps of cells. Once the clumps of cells were separated, an equal volume of DMEM supplemented with 10% foetal calf serum was added to the cell suspension to inactivate the trypsin. The cell suspension was divided between three flasks each the same size as the original one and each containing fresh media. The cells were incubated in a Gallenkamp incubator set at 37°C and 5% CO_2.

2.14.3 FREEZING TISSUE CULTURE CELL STOCKS

The volumes quoted here are for one $125cm^2$ tissue culture flask of confluent cells.

The medium was removed and the cells rinsed with Dulbecco's phosphate buffered saline. After the saline was removed 2ml trypsin-EDTA (0.5g trypsin and 0.2g EDTA/litre) were added to the cells. The cells were incubated at 37°C until they just started to "round up" and a few detached from the flask surface, as seen under a phase contrast microscope (Olympus CK2). The trypsin was removed, taking care to remove all of it but not any cells. 2ml of DMEM supplemented with 10% foetal calf serum and 10% dimethyl sulphoxide (tissue culture grade, Sigma) were added to the cells which were suspended in the media by pipetting. A phase contrast microscope was used to check that all the cells were suspended in the medium. 1ml aliquots of the suspension were transferred to cryotubes, these were placed in a polystyrene box containing a plastic 500ml beaker full of liquid nitrogen. The lid was placed on the polystyrene box and the cells were left for 20 minutes to freeze. When frozen, the cells were transferred to a 25 litre dewar containing liquid nitrogen. The cells were kept in liquid nitrogen until such time as they were needed.

2.14.4 RECOVERING TISSUE CULTURE CELLS FROM FROZEN STOCKS

The frozen cell suspension was removed from the liquid nitrogen and defrosted in a 37°C water bath. When the suspension was liquid it was transferred to a $50cm^2$ tissue culture flask. 9ml of DMEM plus 10% foetal

calf serum were added to the cell suspension. The first 1ml was added dropwise, the rest was added 1ml at a time. After each addition the flask was gently agitated. In this way the osmotic shock suffered by the cells was minimised. After about 6 hours the cells were checked under a phase contrast microscope to see if they had attached to the flask surface. The medium was replaced with fresh medium and the cells were left to grow to confluency.

2.14.5   ISOLATION OF RNA FROM TISSUE CULTURE CELLS

Adapted from a method by Chomozynski and Sacchi (1987).

The volumes and method described here are for two 175cm^2 tissue culture flasks of confluent fibroblast cells. The cells were allowed to grow to confluency then left to grow for a further 3 days. It had been previously shown that fibroblasts synthesize proportionally more collagen mRNA at confluency than when actively growing (Miskulin et al., 1986).

The media was removed form the flasks using an aspirator (Charles Austin Pumps Ltd.). The cells were washed with Dulbecco's phosphate buffered saline. They were removed from the flask surfaces by the addition of 2 x 2ml solution D (4M guanidinium thiocyanate, 25mM sodium citrate, pH7.0, 0.5% sodium N-lauroyl-sarcosine and 0.1M β-mercaptoethanol) to each flask. The cells from the two flasks were pooled in a 30ml Corex tube (Dupont Instruments) to create a final volume of 8ml. The cells were mascerated for 10 seconds using a 'Polytron' blender (Kinematica GmbH, Switzerland). To the solution were added 0.8ml 2M sodium acetate pH4.0, 8ml water saturated phenol and 1.6ml chloroform:isoamylalcohol (49:1). This mixture was shaken for 10 seconds then left on ice for 15 minutes before being centrifuged at 9,000rpm for 20 minutes at 4°C. The aqueous phase was removed to a fresh siliconized 30ml Corex tube. 10ml of propan-2-ol were mixed with the aqueous phase before incubating at ⁻20°C for 2 hours. The RNA precipitate was collected by centrifuging at 9,000rpm for 20 min at 4°C. The pellet was redissolved in 600μl solution D and transferred to a DEPC-treated 1.5ml Eppendorf tube. 600μl of propan-2-ol were added to this before incubating at ⁻20°C for 45 minutes. The RNA was collected as a pellet by spinning at 13,000 rpm for 20 minutes at 4°C. The pellet was washed with 70% ethanol (made with DEPC-treated water) then dissolved in 80μl DEPC-treated water. A 1.5μl aliquot was electrophoresed through a

0.8% agarose gel to check the quality of the RNA. The quantity of the RNA was determined by U.V. absorbance at 260nm. The rest of the RNA was stored at ⁻80°C until required.

RNA was isolated from lymphoblasts using this method. The cells from 6 x 10ml cultures were used. They were pooled and collected by centrifugation at 1,100rpm for 10 minutes in an MSE Centaur centrifuge. After being washed with Dulbecco's phosphate buffered saline, the cells were suspended in 8ml solution D. The method was continued as described above.

2.14.6   ISOLATION OF DNA FROM TISSUE CULTURE CELLS

DNA was isolated from a $175cm^2$ flask of confluent fibroblast cells.

The media was removed using an aspirator and the cells were washed with Dulbecco's phosphate buffered saline. 5ml of trypsin-EDTA solution were added to the cells. The cells were incubated at 37°C until such time as the cells 'rounded up' and came away from the flask surface, as seen using a phase contrast microscope. The cell suspension was removed to sterile, disposable, capped centrifuge tube and 5ml of medium containing 10% foetal calf serum were added. The cell suspension was pelleted at 1,100rpm for 10 minutes in an MSE Centaur centrifuge. The pellet was resuspended in 1.4 ml 1 x SSC and transferred to a sterile Eppendorf tube, which was centrifuged at 6,500rpm for two minutes in a MSE Microcentaur microcentrifuge. The supernatant was removed and the pellet resuspended in 200μl white cell suspension media (0.3M sodium acetate, 20mM Tris-HCl pH7.5 and 1mM Na_2EDTA), 15μl 10% SDS were added. After mixing for 1 minute, the solution was phenol extracted, back extracted by the addition of 180μl TE pH7.6, phenol extracted again, then chloroform extracted. 20μl of 20 x SSC and 1μl RNase A (10mg/ml, boiled) were added to 200μl of DNA solution, this was incubated at 37°C for 1hour. 20μl of 10% SDS and 2μl proteinase K (1mg/ml) were added to the solution which was then incubated at 50°C for 1hour. The solution was phenol extracted and back extracted with 100μl TE. These two steps were repeated then the solution was extracted with chloroform:isoamylalcohol. Two volumes of ethanol were added to the final aqueous phase, this was mixed then spun at 13,000rpm for 1 minute. The resulting pellet was washed in 70% ethanol and resupended in 200μl TE.

Where possible all the steps in the preparation were carried out under 'PCR-

clean' conditions.

The quality of the DNA was determined by electrophoresing a 1μl aliquot through 0.8% agarose gel. The quantity of the DNA was calculated from its U.V. absorbance as described in section 2.5.3.

2.15 AMPLIFICATION OF NUCLEIC ACID SEQUENCE USING THE POLYMERASE CHAIN REACTION

2.15.1 AMPLIFICATION OF GENOMIC DNA USING THE POLYMERASE CHAIN REACTION

This method was adapted from a paper by Kogan *et al.* (1987).

Due to the extreme sensitivity of the method great care was taken to make sure all tubes, tips and reagents were free from contaminating DNA. If RNA was used as the template, the tips, tubes and reagents were rendered RNase free before they were used.

Reaction mixes were set up containing 1μg or less of template DNA or total cellular RNA, 16.6mM ammonium sulphate. 67mM Tris-HCl (pH8.8 at 25°C), 6.7mM $MgCl_2$, 100mM β-mercaptoethanol, 6.7μM EDTA and 170ng/μl BSA (nuclease free). 1.5mM each of dATP, dGTP, dTTP, and dCTP were added with 50pmoles of each primer and 3 units *Taq* polymerase (Amersham). After mixing thoroughly, the mixture was overlaid with paraffin oil to prevent evaporation. The reaction was carried out in a Programmable Dri-Block PHC-1 (Techne) using temperature cycles of 95°C (denaturation), 55-65°C (annealing) and 70°C (elongation), thirty cycles were run as a rule.

2.15.2 POLYMERASE CHAIN REACTION AMPLIFICATION OF mRNA USING REVERSE TRANSCRIPTASE TO SYNTHESIZE THE cDNA

This method is adapted from one described by H. Erlich (1989)

There are two steps to the method, the first being the reverse transcriptase step.

To a final volume of 20μl of 1 x reaction buffer (50mM KCl, 20mM Tris-HCl pH8.4 at room temperature, 2.5mM $MgCl_2$ and 0.1mg/ml BSA (nuclease free)) were added:-

2μl 10 x dNTPs (10mM each of dATP, dGTP, dTTP and dCTP)

1μl RNasin (Promega, 40 units/μl)

50pmoles antisense primer

1μg RNA template (total cellular, heated at 90°C for 5 minutes then cooled on ice)

1.5μl Moloney Murine Leukemia virus reverse transcriptase (Pharmacia, 15 units/μl)

1.5μl DTT (100mM).

The reaction mix was left at room temperature for 10 minutes then incubated at 42°C for 60 minutes. The reaction mix was collected at the bottom of the centrifuge tube by spinning in a Microcentaur microcentrifuge for 30 seconds then 80μl of 1 x buffer were added to the tube with 50pmoles of the sense primer and 3 units of *Taq* polymerase (Amersham). After mixing well 50μl of liquid paraffin (Fisons) were layered ontop of the reaction mix and the tube was placed in a Programmable Dri-block PHC-1 (Techne). 30 cycles of 95°C for 1 minute, 55°C for 1 minute and 70°C for 1-10 minutes were run through.

2.15.3   ANALYSIS OF AMPLIFIED DNA

The paraffin oil was removed by ether extraction and an aliquot of the reaction electrophoresed through an agarose or acrylamide gel.

Before PCR amplified DNA was digested with a restriction endonuclease it was phenol extracted and ethanol precipitated. The digest reactions were made up as described in section 2.5.4 except, 0.1 volume of 40mM spermidine-HCl was added 10 minutes into the reaction time. The DNA was digested for one hour.

2.16   REAMPLIFICATION OF DNA

An aliquot of PCR product was electrophoresed through an agarose gel to check that the PCR had been successful, and to separate excess primers from the amplified DNA. Once clear of the primers, the band on the gel corresponding to the amplified fragment was excised and placed in a 1.5ml Eppendorf tube. 100μl of sterile water were added to the gel slice, this was frozen to -20°C and thawed twice, to help elute the DNA. 2μl of the DNA solution were used as the template for a PCR, 15 cycles were usually

sufficient.

2.17   RNASE A PROTECTION ANALYSIS

The methods described in the following section are from Hawkins (1989).

2.17.1   END LABELLING MARKERS

500ng of pUC13 were digested with the restriction endonuclease *Hpa*II at 37°C for 1 hour. 500ng of M13mp18 were digested with *Msp*I at 37°C for 1 hour. To each of these digest reactions, 1μl [α-^{32}P] dCTP (3000Ci/mmol, Amersham), 0.5μl 4M NaCl, 0.5 units Klenow and 2.5μl H$_2$O were added. The reactions were allowed to stand at room temperature for 15 minutes. This allows the Klenow to fill in the first space in the sticky ends of the DNA fragments with [α-^{32}P] dCTP. The reaction mixes were phenol extracted and back extracted with water. The remaining protein was precipitated out of solution by adding 0.5 volume 7.5M ammonium acetate and leaving at room temperature for 10 minutes. The precipitated protein was pelleted by spinning at 13,000rpm for 15 minutes. The supernates were transferred to fresh Eppendorf tubes, where the DNA was precipitated by the addition 2.5 volumes of ethanol and incubating at ⁻80°C for 15 minutes. The precipitate was collected as described previously. The supernates were discarded and the DNA pellets were resuspended in 100μl aliquots of sterile water. The markers were stored at ⁻20°C until needed.

Approximately 40 counts of each marker were mixed with 4μl of loading dye (10mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue in deionised formamide) then made up to 12μl with sterile water before being loaded onto an RNase A protection gel

2.17.2 DENATURING POYACRYLAMIDE GEL FOR RNASE A PROTECTION EXPERIMENTS

A 5% polyacrylamide gel was used:-

16.8g urea (Serva Feinbiochemica GmbH and Co. Heidelberg Germany.)

2ml 10 x TBE

8ml formamide

5ml 40% bisacrylamide

The urea was dissolved in the gel mix by swirling in a 65°C water bath. The gel mix was made up to 40ml with distilled water, then allowed to cool to room temperature. 240µl of 10% ammonium persulphate were added with 60µl TEMED, the solution was mixed by swirling gently, drawn up a 50ml syringe then poured into a Bio-Rad Protean II gel mould. The gel was allowed to polymerize for 1 hour.

2.17.3 PROBE TEMPLATE PREPARATION

Fragments of the α2(I) cDNA clone Hp2010 were subcloned into the phagemid vectors pTZ18R or pTZ19R. The orientation of the subcloning was such that an antisense RNA probe could be synthesised from the T7 polymerase promoter. Large scale preparations of these subclones were prepared as described in section 2.13.1. Approximately 50µg of DNA were linearized with a restriction endonuclease to give 5' overhang ends. The DNA solution was phenol extracted twice and back extracted with distilled water twice. The final aqueous layer was transferred to a DEPC-treated Eppendorf tube and all steps from then on were RNase free. The DNA solution was chloroform extracted, the aqueous layer was saved and to it was added sodium acetate pH5.2 at a final concentration of 0.4M and two volumes of ethanol. The DNA was precipitated at ⁻80°C for 15 minutes. The precipitate was collected by centrifuging at 13,000rpm for 20 minutes. The pellet was washed with 70% ethanol (RNase free), dried under vacuum then resuspended in 10µl RNase free TE.

The final concentration of the template solution was estimated by diluting 1µl with 19µl TE and running known volumes of this solution on a 0.8% agarose gel with known amounts of pUC13 digested with EcoRI.

Having determined the template concentration, the stock solution

was diluted with RNase free TE to give a final concentration of 0.4µg/µl.

2.17.4 SYNTHESIS OF THE ANTISENSE RNA PROBE

Probe labelling mixes were set up:-

10 x NTP(c)	1µl
100mM DTT(Sigma)	1µl
5 x buffer 1µl	
2mg/ml BSA	0.5µl
RNasin	0.8µl (35 units)
0.5 x T7 RNA polymerase	0.7µl (10 units)
[α-^{32}P] CTP (800Ci/mMol)	40µCi
template	0.6µl
H$_2$O	to 10µl

10 x NTP(c) consisted of 5mM each of GTP, ATP and UTP and 75µM CTP.
5 x buffer was 200mM Tris-HCl pH8.25, 30mM MgCl$_2$, 10mM spermidine and 250mM NaCl.
The T7 RNA polymerase (20 units/µl) was diluted two fold in 1 x manufacturers assay buffer just before adding to the reaction mix.

All the solutions were RNase-free. Care was taken when preparing the reaction mix not to contaminate it with RNase. All the solutions, with the exceptions of RNasin and T7 RNA polymerase, were equilibrated to room temperature before use. At low temperatures the spermidine in the buffer is liable to precipitate the DNA template out of solution.

The mixture was mixed by pipetting then incubated at 37°C for 2 hours, when 1µl (6units) of RNase-free DNase I was mixed into the reaction. The reaction was incubated at 37°C for a further 15 minutes.

After centrifuging briefly, 40µl of DEPC-treated water were added to the reaction mix. This solution was phenol extracted and, after removing the aqueous layer, the phenol was back extracted with 50µl of DEPC-treated water. The aqueous layers were pooled. The RNA was precipitated by the addition of 50µl 7.5M ammonium acetate and 500µl of ethanol, incubating at ⁻80°C for 15 minutes and centrifuging at 13,000rpm for 15 minutes. The

supernate was carefully removed to an Eppendorf tube using a pipette. The tube was spun again for a few seconds and the rest of the ethanol was removed. No pellet was visible at this stage. To check that enough [α-^{32}P] CTP had been incorporated into the probe, the counts in the pellet and the supernatant were compared using a Geiger counter (series-900 Mini-Monitor). Enough [α-^{32}P] CTP was incorporated into the probe if there were approximately 5 times more counts in the 'pellet' than in the supernatant.

The 'pellet' was dissolved in 80µl of DEPC-treated water and left on ice until required. Approximately 2 x 10^6 cpm of the probe were added to 50µg of total cellular RNA. The sodium chloride concentration was adjusted to 0.1M by the addition of 2.5M NaCl. The RNA was precipitated out of solution by the addition of 2.8 volumes of ethanol. The mixture was chilled on ice for 5 minutes and the precipitate collected by spinning at 13,000rpm. The ethanol was removed as before.

The pellet, which was at this stage quite large, was dissolved in 160µl of deionised formamide. This was done by making sure that the pellet was floating in the formamide then leaving at room temperature for 30 minutes. When the pellet had dissolved 40µl of 5 x hybridization buffer (1.8M NaCl, 0.2M PIPES pH6.7) were added. This mixture was heated to 85°C for 5 minutes to denature the RNA, then incubated at 60°C overnight to allow the probe to hybridize to the cellular RNA.

2.17.5   mRNA PURIFICATION

This step is dependent on the properties of messenger affinity paper (mAP, Amersham). For each hybridization reaction to be purified, a piece of mAP was cut approximately 4mm x 4mm using a clean scalpel blade. Care was taken not to touch the paper or in anyway contaminate it with RNase. The piece of mAP was soaked in 0.5M NaCl for 5 minutes before being transferred to a piece of 3MM paper (Whatman, cut to 5cm x 5cm, autoclaved and dried). 100µl of 0.5M NaCl were mixed with the hybridization mix which was then spotted onto the mAP one drop at a time. Care was taken to make sure that the RNA solution was absorbed through the mAP and not just the surrounding 3MM paper. After loading the whole of the hybridization solution, the mAP was washed in 3 aliquots of 0.5M NaCl (DEPC-treated) for 15 minutes each. The mAP was removed from the salt

solution and blotted dry by allowing the edge of it to touch a tissue. The mAP was transferred to a fresh Eppendorf tube to which 180μl of DEPC-treated water were added, this was heated to 65°C for 10 minutes, chilled on ice for 10 minutes, then the mAP removed. The mRNA/probe hybrid was now in solution.

2.17.6   RNASE CLEAVAGE

The RNA solution was left to warm to room temperature then 20μl of 2.5M NaCl solution were added with 100μl RNase A solution. The RNase A solution (100μg/ml in 3 x R buffer (20mM NaCl, 10mM LiCl, in 0.3 x TE pH7.5) had been boiled for 10 minutes and stored at 4°C. The digest reaction was left at room temperature for 30 minutes, then stopped by the addition of 20μl 10% SDS and 10μl proteinase K (10mg/ml) and incubating at 37°C for 20 minutes. 10μg of calf liver tRNA were added to the solution which was subsequently phenol extracted. The aqueous layer was transferred to a fresh Eppendorf tube and ethanol precipitated by the addition of 750μl of ethanol and chilling on ice for 5 minutes. The precipitate was pelleted by spinning at 13,000rpm for 15 minutes. The RNA was rinsed with 200μl of ethanol and collected as before. After carefully removing all of the ethanol, the RNA was resuspended in 12μl of loading dye (0.1% xylene cyanol, 0.1% bromophenol blue, 10mM EDTA made up to volume with deionised formamide).

2.17.7   RUNNING THE GEL

The gel was fitted into the Bio-rad Protean II gel tank, 0.5 x TBE was poured into the lower and upper buffer chambers and the wells were thoroughly rinsed with buffer to remove any excess urea. The gel was prerun for 30 minutes at 25mA. A thermostatic circulator (LKB) was used to maintain the temperature of the gel at about 35°C.
Aliquots of the markers and the RNA probe (about 20-30 counts of each, as measured using a Geiger counter) were made up to a volume of 8μl with DEPC-treated water. 4μl of loading dye were added to each sample. The wells in the gel were rinsed again with buffer then the markers, probe, and RNase reaction mixes were loaded into the wells. The gel was run at 25mA and 30°C until the bromophenol blue dye band reached the bottom of the gel.

The gel mould was taken apart and the gel was fixed in 10% acetic acid, 10% methanol for 15 minutes, transferred to a piece of 3MM paper and dried under vacuum for 20 minutes in a Bio-Rad Model 583 gel drier. The gel was autoradiographed overnight with an intensifying screen.

2.18   DNA SEQUENCING

2.18.1   ISOLATION OF PLASMID FOR SEQUENCING

Plasmid was isolated using a Qiagen plasmid pack 100 (Diagen) according to the manufactures instructions. The method required the overnight growth of a 100ml culture of *E. coli* containing the plasmid, in Luria broth supplemented with antibiotic to select for the plasmid. The final pellet was dissolved in 100µl of sterile water.

2.18.2   ALKALI-DENATURATION OF PLASMID DNA PRIOR TO SEQUENCING

DNA was alkali-denatured and spin-dialysed using a method described in Murphy, and Ward, (1989).

Approximately 15µg of plasmid DNA were made up to 20µl with sterile water, 5µl of 1M NaOH, 1mM Na_2EDTA were mixed with the DNA which was then incubated at 37°C for 15 minutes. Salt was removed from the DNA solution by passing it down a spin-dialysis column. The DNA recovered from the column was used immediately as a template for sequencing reactions.

2.18.3   PREPARATION OF A SPIN-DIALYSIS COLUMN

Sepharose CL-6B (Pharmacia) was equilibrated in T0.1E (10mM Tris-HCl, pH8.0, 0.1mM Na_2EDTA), the buffer volume was adjusted to produce a packed gel:buffer ratio of 2:1. 25µl of a slurry of 200 micron glass beads (Jencons Ballotini beads N^O11) in water were placed into a 0.5ml Eppendorf with its base partially pierced with a 21G needle. 300µl of the Sepharose slurry were placed on top of the glass beads. The tube was placed

inside a 1.5ml Eppendorf tube, which was in turn placed inside a 15ml disposable tube and spun at 200g for 4 minutes (1000rpm) in a Sorvall Ominspin centrifuge. The 0.5ml tube containing the column was transferred to another 1.5ml Eppendorf tube and used within one hour, before the gel matrix was allowed to dry.

2.18.4   SEQUENCING REACTIONS

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

8.5µl (6.4µg) of alkali-denatured template, described above, were mixed with 10ng of sequencing primer in 40mM Tris-HCl (pH7.5), 20mM $MgCl_2$ and 50mM NaCl. This was heated to 65°C for 2 minutes then allowed to cool to 30°C in a beaker of water over 30 minutes. To this template-primer mix was added 1µl 100mM DTT, 2µl labelling mix (0.5µM each of dGTP, dCTP, dTTP), 1µl [α-^{35}S]dATP (9410Ci/mMol, 10µCi/µl) and 2 units of T7 DNA polymerase. The mixture was mixed thoroughly, avoiding the formation of bubbles and incubated at room temperature for 3 minutes. 3.5µl aliquots of the labelling mix were added to 2.5µl aliquots of termination mixes (each containing 80µM dGTP, dATP, dCTP and dTTP plus 8µM of one or other of the following ddGTP, ddATP, ddCTP, or ddTTP and 50mM NaCl, prewarmed at 37°C). The solutions were incubated at 37°C for 5 minutes before 4µl of stop solution (95% deionised formamide, 20mM Na_2EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) were added. The mixes were spun for a few seconds in a MSE Microcentaur microcentrifuge and stored at ⁻ 20°C until required.

Described above is the method used to sequence denatured plasmid DNA. The method was adapted to allow direct sequencing of PCR product. 5pmoles of PCR amplified DNA was boiled for 3 minutes in the presence of an equimolar amount of sequencing primer, 40mM Tris-HCl pH7.5, 20mM $MgCl_2$ and 50mM NaCl. The boiled solution was snap frozen in dry ice. This template-primer mix was then treated as described above.

2.18.5 SEQUENCING GEL

A 6% acrylamide gel was used containing 8.3M urea and 0.5 x TBE (45mM Tris base, 45mM boric acid, 1.2µM Na_2EDTA pH8.3). The gel mix was filtered by suction filtering through Whatman N^o1 filters using a Buchner funnel. 350µl of freshly prepared 10% ammonium persulphate and 24µl TEMED were mixed with 75ml of the gel mix. The gel was poured into a 40cm long sequencing gel mould (Bethesda Research Laboratories, BRL) and left to polymerize for one hour. The gel was mounted into a gel tank (BRL) and 0.5 x TBE was poured into the upper and lower buffer reservoirs. 0.5cm well-forming combs were inserted into the top of the gel mould until they made contact with the gel, the wells were rinsed with buffer to remove excess urea. The gel was prerun at 2000 volts until the temperature of the gel was about 55°C.

4µl aliquots of the sequencing reactions were denatured by boiling for 2 minutes then loaded immediately into the wells. The gel was run at 2000 volts until it was judged, using the dyes in the stop solution as a guide, that the gel had run far enough. On a 6% acrylamide gel bromophenol blue runs at approx. 26 bases and xylene cyanol runs at 106 bases.

At the end of the run, the gel was fixed in 10% methanol, 10% acetic acid for 20 minutes. The gel was transferred to a piece of Whatman 3MM paper and dried in a Bio-Rad gel drier for 60 minutes. Gels were autoradiographed for 12 hours or 1 week as necessary.

2.19   SYNTHESIS OF cDNA TO BE USED AS TEMPLATE FOR PCR

2.19.1   SINGLE STRAND cDNA SYNTHESIS

The method was derived from D'Alessio *et al.* (1987).

To a DEPC-treated, screw top Eppendorf on ice was added, 1µg poly(A)$^+$ RNA (a gift from Dr S.Kirk, isolated from chorionic membrane), 25µl oligo $d(T)_{12-18}$ (Pharmacia) and 4µl DEPC-treated water. The RNA was denatured by heating to 65°C for 5 minutes then cooled on ice for 5 minutes. 10µl of 1st strand buffer (250mM Tris-HCl pH8.3, 37.5mM KCl, 2.5mM $MgCl_2$ and 50mM DTT) were mixed in with the RNA solution. 2.5µl of a mix containing 10mM of each dGTP, dATP, dTTP and dCTP were added with 2.5µl of Moloney Murine Leukemia Virus reverse transcriptase (BRL,

200units/μl). The tube was kept on ice, 10μl were removed and 1μl of [α-^{32}P] dCTP (3000Ci/mmol) was added to this aliquot. Both tubes were incubated at 37°C for 1 hour. The reaction was stopped in each tube by the addition of 0.08 volume of 0.2M Na$_2$EDTA. After mixing gently the tubes were incubated at 65°C for 10 minutes. The solutions were phenol extracted, back extracted then ethanol precipitated by the addition of 0.1 volume 7.5M ammonium acetate and 2.5 volumes of ethanol. After chilling the tubes were spun in a Microcentaur microcentrifuge for 10 minutes to pellet the nucleic acid. The pellets were washed with 70% ethanol then 100% ethanol, before being vacuum dried. The non-radioactive pellet was stored at ⁻20°C until required. The radioactive pellet was dissolved in 100μl of water. Approximately 50 counts per second (as measured using a Geiger counter) were denatured by glyoxalation. The DNA was glyoxalated by adding 4μl DMSO, 0.8μl 100mM phosphate buffer (100mM disodium hydrogen phosphate, 100mM dihydrogen sodium phosphate pH7) and 1.5μl deionised glyoxal and heating at 50°C for 1 hour. To check that the cDNA synthesis had been successful, the glyoxalated cDNA was run overnight through a 1% phosphate gel alongside glyoxalated markers. The markers were prepared as described in section 2.19.2. The gel was dried in a Biorad gel drier and autoradiographed overnight.

2.19.2 PREPARATION OF [α-^{32}P] dCTP END-LABELLED λ/HINDIII RESTRICTION FRAGMENTS

A modification of the method published in Anon. (1984).

The following were mixed in an Eppendorf tube:-

3μl 10 x T/M/C (60mM Tris-HCl pH7.5, 60mM MgCl$_2$, 200mM NaCl)

3μl 10mM DTT

15.5μl dH$_2$O

4μl λ/Hind III (250μg/μl)

1μl of solution containing 0.5mM each dATP, dTTP and dGTP.

1μl [α-^{32}P] dCTP (300Ci/mmol)

0.5 units Klenow fragment (Pharmacia)

This reaction mix was incubated at room temperature for 15 minutes. It was phenol extracted and the resulting phenol layer was back extracted with an equal volume of dH$_2$O. The two aqueous layers were pooled. The protein in the aqueous solution was precipitated by the addition of 0.5 volume of 7.5M

81

ammonium acetate and leaving at room temperature for 5 minutes. The precipitate was collected by spinning in a Microcentaur microcentrifuge. The supernatant was saved and the nucleic acid ethanol precipitated, the pellet was washed with 80% ethanol and the pellet dissolved in 100μl dH$_2$O. Approximately 40 count per second, measured with a Geiger counter, were glyoxalated as described in section 2.19.1 before being electrophoresed through a 1% phosphate gel.

2.20 PLASMID LIBRARY SCREENING

2.20.1 PREPARING THE BLOTS

An adaptation of the method by Buluwela *et al.* (1989) was used.

White colonies were picked from plates containing X-gal, IPTG and antibiotic to select for plasmid containing bacteria. Each colony was used to inoculate a separate well in a microtitre plate containing 100μl of sterile LUB supplemented with 0.1 volume of 10 x HMFM (36mM K$_2$HPO$_4$, 13mM KH$_2$PO$_4$, 20mM Na$_3$citrate, 10mM MgSO$_4$ and 44% w/v glycerol). The colonies were picked with the aid of sterile Gilson pipette tips. Approximately 1,000 colonies were picked in this way, then left at 37°C for 1 hour. The plates were stored at ‾80°C until required.

The colonies were replicated onto Nylon filters (Hybond-N) using a 96 pronged "hedgehog". Dry nylon filters were used. The filters were cut to leave a margin of about 0.5cm on all sides. The hedgehog was flamed, cooled, dipped into the microtitre wells, agitated gently then lifted out and pressed gently but firmly onto the surface of the filters, making sure that all the prongs made contact with the filter, and lifted vertically off. This procedure was replicated for each microtitre plate. The filters were laid onto Luria agar plates containing antibiotic. The colonies were left to grow overnight at 37°C, after which each colony appeared as a bacterial "blob" at the infected position. The coordinates of any clones which did not grow were noted. The nylon filters were lifted off the plates and placed onto blotting paper soaked in 2 x SSC, 5% SDS then left for 2 minutes. The blotting paper with the filters on it was "cooked" in a microwave at full power (650 Watts) for 2.5 minutes. This treatment fixed the clone DNA to the filters. The filters were rewetted with 5 x SSC, 0.1% SDS and the bacterial debris was wiped from the surface with a gloved finger. After drying at room temperature and

wrapping in cling film (Saran Wrap), the filters were stored at 4°C.

2.20.2 PROBE PREPARATION

This was carried out essentially as described in sections 2.12.2 and 2.12.3.

2.20.3 HYBRIDIZATION OF A RADOIACTIVE PROBE TO LIBRARY FILTERS

This was carried out as described for Southern blot hybridization, section 2.12.4. The stringency of hybridization was altered by changing the temperature.

2.20.3 WASHING FILTERS AFTER HYBRIDIZATION

The filters were washed at the same temperature as the hybridization with 3 x SSC, 0.1% SDS. More stringent washes were used as required.

2.21 COMPUTER PROGRAMMES

Analysis of DNA sequences and RNA secondary structure were performed using the Genetics Computer Group program suite (Kanehisa *et al.*, 1984) on the DEC VAX computer at the University of Leicester.

The word processing was carried out using Microsoft Word version 5.

CHAPTER THREE

A POSSIBLE POLYMORPHISM IN THE α1(III) COLLAGEN CODING SEQUENCE

3.1 INTRODUCTION

At the time of carrying out this work no silent mutations had been characterized in collagen coding sequence DNA. The lack of published silent mutations probably did not prove that none existed. The emphasis is to look for mutations in individuals who suffer from a disease linked to a collagen disease, not to look for sequence changes in individuals with a normal phenotype. In the search for mutations causing disease phenotypes no phenotypically silent mutations had been found, not even mutations which changed the nucleic acid sequence but not the polypeptide sequence. It was argued that mutations found in individuals with abnormal phenotypes caused the phenotype and were not silent because protein data supported the DNA/RNA evidence. As a rule, collagen genes do not tolerate mutations very well (Prockop *et al.*, 1989) probably because collagen proteins are complex, well ordered aggregates of many collagen molecules. A slight change in the collagen molecule structure can disrupt the whole protein.

This work was initiated by comparing the two α1(III) cDNA clone sequences published by Miskulin *et al.* (1983) and Mankoo and Dalgleish (1988), p(III)-21 and p(III)-33 respectively. These clones were both synthesized from mRNA isolated from the same tissue culture cell line, Human Foetal Lung fibroblasts (HFL-1, ATCC CCL135). The clones covered sequence at the 3' end of the α1(III) mRNA and overlapped (Miskulin *et al.*, 1983). Examination of the published sequences showed them to differ in this overlapping region by one base. A thymidine was found in p(III)-33 where a cytosine was found in p(III)-21 (Mankoo and Dalgleish, 1988). This base difference, if real, would result in an amino acid difference between the polypeptide sequences coded for by the clones:

p(III)-33 - serine.

p(III)-21 - proline.

The amino acid substitution would be at position 1011 in the α-helix. The first glycine of the α-helix being numbered 1. Glycine substitutions in this area have been found to result in overmodification of the type I collagen 5' to the substitution, a decrease in the thermal stability

84

of the type I molecule and an osteogenesis imperfecta phenotype (Bonadio and Byers, 1985). However as far as was known this cell line was normal in its collagen production. It was postulated that because the substitution did not involve a glycine in the α-helix and involved type III collagen rather than type I, it could be silent or even recessive, only producing a diseased phenotype in the homozygous state. The only recessive osteogenesis imperfecta phenotype characterized was caused by 4bp frameshift deletion in exon 1 of COL1A2 (Pihlajaniemi et al., 1984). Having never encountered a silent mutation at that time (1988) the consequences of the variant were difficult to predict.

The variant would result in the production of a restriction fragment length polymorphism (RFLP). The sequence in p(III)-21 [CCAGG] could be cut by the restriction endonucleases EcoRII and BstNI. The sequence in p(III)-33 could not be cut using these enzymes. The presence/absence of this RFLP in the DNA isolated from the cell line HFL-1 may have been detected by restriction digestion, Southern blotting and hybridization however, the sizes of the DNA fragments could not be predicted because the sequence of intron 4 was not known. A restriction map had been published covering this area in Chu et al. (1985) but the enzymes EcoRV and BstNI were not included. With the arrival the PCR it was considered easier to use the PCR and direct sequencing instead of RFLP analysis to determine the presence or absence of the base change.

Since the two clones are derived from the same cell line the base difference in their sequence could have arisen due to:-

1. The cell line from which the clones were derived being heterozygous for this position in its genome.

2. The production of an artefact in one of the clones during cDNA synthesis.

The polymerase chain reaction was used to determine which scenario best explains the presence of the base difference. A 900bp fragment of the collagen gene, COL1A3, was amplified using the PCR so as to include the 'variant' base difference. The amplified fragment was sequenced directly. If the base difference between the two clones was the result of the production of an artefact during cDNA synthesis, the sequence obtained by directly sequencing the PCR product would match that of either p(III)-33 or p(III)-21. If the cell line from which the clones were synthesized was heterozygous for this position, then both the cytosine and thymidine would be seen at the same position on the sequencing gel. The presence of both bases at the same

position on the sequencing ladder would be brought about because, during the PCR and direct sequencing reactions, both alleles contained within the genome are amplified and sequenced simultaneously. The sequences of both alleles are superimposed on the same sequencing ladder.

3.2 THE AMPLIFICATION OF A FRAGMENT OF COL1A3 FROM HUMAN FOETAL LUNG CELLS

The polymerase chain reaction was carried out in a similar way to that described in section 2.15.1. Due to the fact that this work was accomplished before the manufacture of thermocycling machines it was carried out with the aid of three temperature baths. The baths were set at 95°C (PEG 4000), 55°C (water) and 70°C (water). 35 cycles consisting of one minute at each temperature were performed. The template for the PCR was 1μg of genomic DNA isolated from the HFL-1 cell line. The sequences and positions of the primers are shown in Figure 3.1. The primers were designed so the resulting PCR fragment could be sequenced using the primer PCR-2. The size of the PCR fragment was predicted from data in Chu *et al.* (1985) to be approximately 900bp-1kb.

Confirmation that the 900bp fragment was the only product of the PCR was determined by electrophoresing one tenth volume of the reaction product through an agarose gel (Figure 3.2). The remaining 90μl of reaction mix were ether extracted to remove the paraffin, ethanol precipitated and redissolved in 15μl of TE. The whole of this DNA solution was gel purified and reamplified as described in section 2.16. Five reamplification reactions were pooled, ethanol precipitated, dissolved in TE and electrophoresed through an agarose gel. The 1kb fragment was recovered from the gel as described in section 2.8. The concentration of the final solution was determined. 10μg of amplified 900bp fragment were isolated.

3.3 PARTIAL SEQUENCING OF THE 900bp FRAGMENT

Using the direct sequencing method described in section 2.18.4, the fragment was sequenced using the primer PCR-2 to prime the sequencing reactions.

The results of the sequencing reactions (Figure 3.3) clearly show

Figure 3.1

The sequences of the primers PCR-1 and PCR-2.

The positions of the primers are also shown. Information regarding the positions of the primers was from Chu *et al.* (1985b). It should be pointed out that the numbering of the exons in this figure is from that of Chu *et al.* (1985b). The exons are numbered from the 3' end of the gene, normally exons are numbered from the 5' end of the gene. Exon 4 in the figure represents the carboxyl joining exon. It codes for helical sequence, telopeptide and some propeptide sequence.

Position X represents the position of the possible variant base. It is 75bp from the 5' end of the PCR product.

Figure 3.1

Exon 5 (108 bp)

Intron 4 (approx. 750 bp)

Exon 4 (97 bp)

PCR-1

PCR-2

approx. 900bp

75 bp
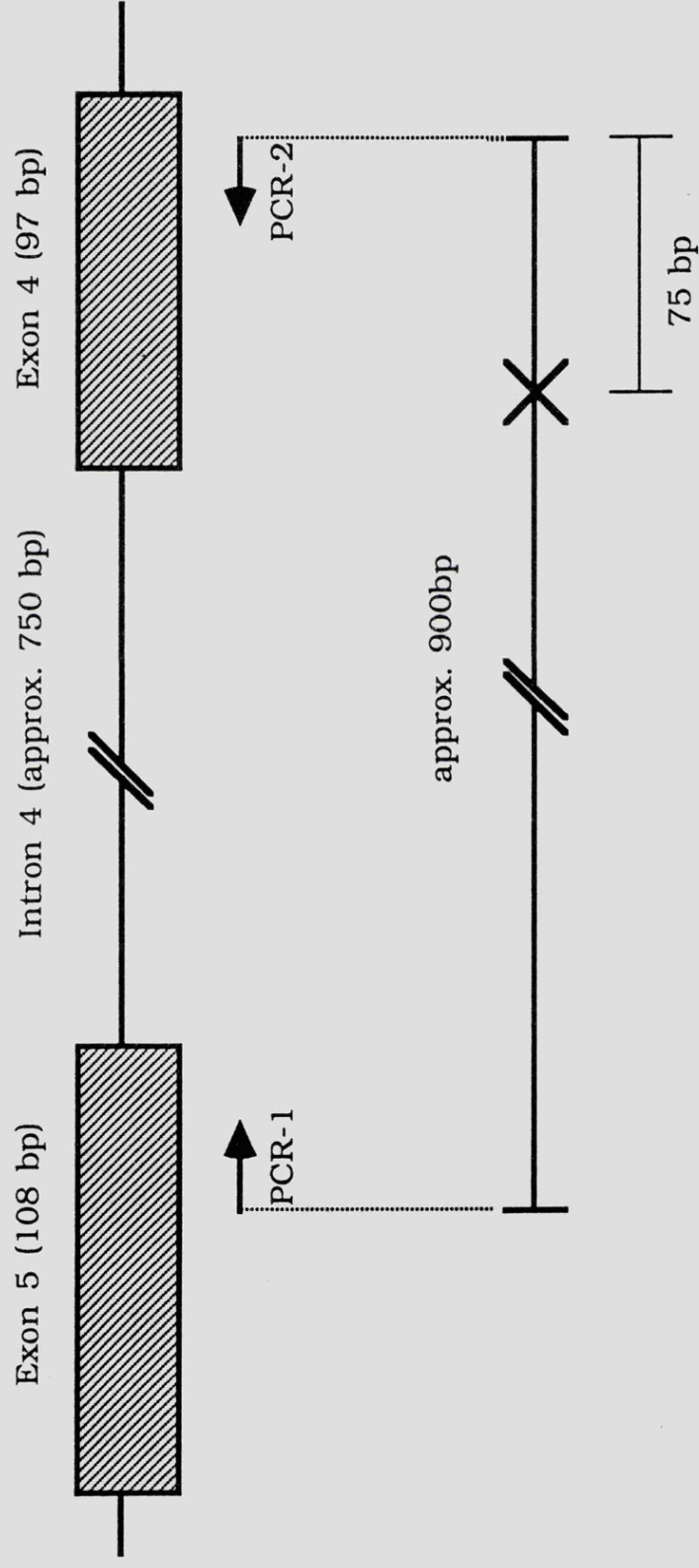
5'

3'

PCR-1 - TAGGGTCGTTACCGTCGCCG

PCR-2 - AGGGTAACCTGGTGGTCCCG

Figure 3.2

Amplification of HFL-1 DNA using primers PCR-1 and PCR-2 results in a single 900bp fragment.

The photograph shows the product of PCR amplification using genomic DNA from the fibroblast cell line HFL-1 as a template and the primers PCR-1 and PCR-2. The PCR product was electrophoresed through a 0.8% agarose gel and stained with ethidium bromide.

The size of the PCR product was predicted from information in Chu *et al.* (1985) to be approximately 900bp-1kb.

Lane 1 - λ/*HindII* and φX174/*HaeIII* markers

Lane 2 - Amplified DNA from the cell line HFL-1

FIGURE 3.2
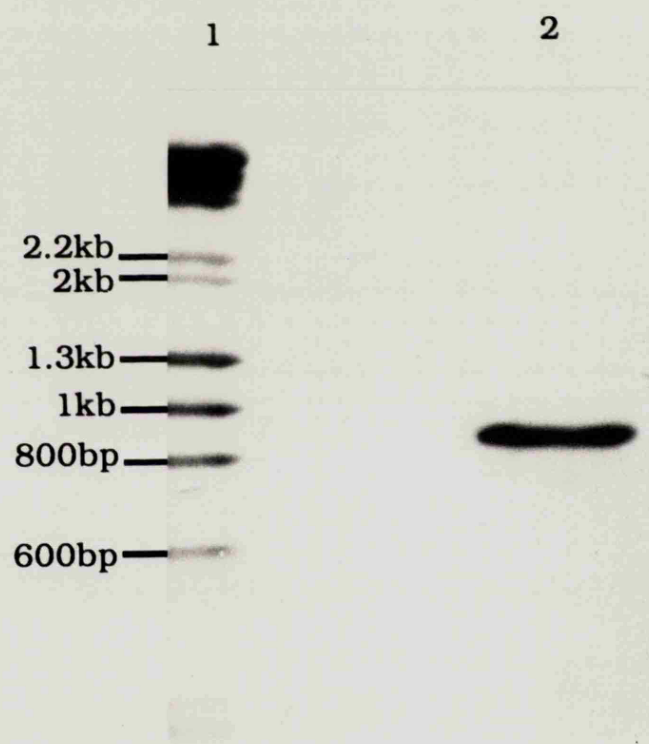
Figure 3.3

The genome of HFL-1 does not contain a variant within the region analysed.

      The sequencing autoradiograph shows the sequence of part of the two alleles of COL1A3 of the fibroblast cell line HFL-1.

      The sequencing ladder does not reveal any heterozygosity.

Figure 3.3
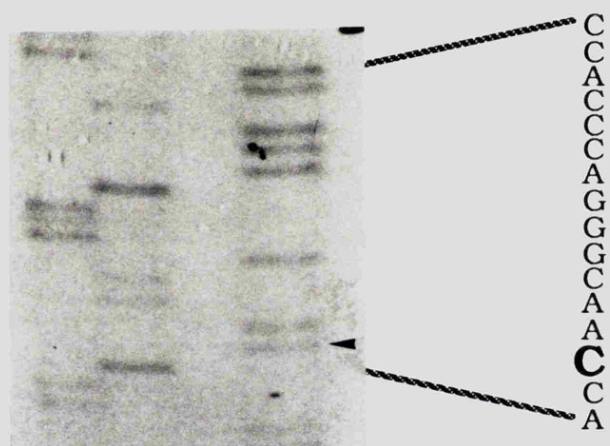
that the cell line from which the two clones were derived was not heterozygous for the base which was different between between the two clones. A cytosine is seen at this position on the sequencing ladder with no sign of the thymidine seen when p(III)-33 was sequenced. PCR amplification and direct sequencing of the same region of COL3A1 from DNA isolated from other phenotypically normal individuals revealed a proline codon coding for amino acid 1011 of the α1(III) helical domain. These results confirm the conclusion that the normal sequence of this region contains a proline codon not a serine codon at position 1011.

3.4   DISCUSSION

From these results it can be concluded that the variant base between the overlapping sequences of the clones p(III)-33 and p(III)-21 was an artefact produced during the cDNA synthesis of one of the clones. Since the direct sequencing results of the HFL-1 cell line and other normal DNA samples agreed with the sequence published for clone p(III)-21 at this position (Chu *et al.*, 1983; Loidl *et al.*, 1984) it can be concluded that the cDNA synthesis artefact is the thymine found in p(III)-33.

This work was carried out because of the lack of published data concerning phenotypically silent amino acid changes found in the coding regions of fibrillar collagen genes. The inability of collagens to tolerate changes in their amino acid sequence is considered to be due the importance of each amino acid in the eventual structure of the collagen fibrils (Prockop *et al.*, 1989).

Since the completion of this work three phenotypically silent variants have been described in the literature. Two are conservative base changes one in exon 6 of COL1A2 (Weil *et al.*, 1990) and the other in exon 33 of COL3A1 (Tromp *et al.*, 1991). A base change which caused an amino acid substitution in COL3A1 was described in Zafarullah *et al.* (1990). The amino acid was an alanine (GCT) at position 530 of the triple helix which was substituted for a threonine (ACT). The change affected an amino acid in the Y position of a Gly-X-Y repeat.

There are few α-helix amino acid substitutions reported in the literature which do not involve a glycine. One was found in the DNA from a family suffering from osteoarthritis associated with mild chondrodysplasia. (Ala-Kokko *et al.*, 1990). The substitution involved an arginine to cysteine

87

substitution at position 519 of the α1(II) helical domain. As researchers look towards the link between collagen genes and more common heritable connective tissue disorders, it will be interesting to determine if mutations which result in amino acid changes within the helix but do not involve a glycine will be found to be the cause of these diseases. Although, Kontusaari *et al.* (1990) found a splice site mutation in the α-helix of α1(III) in a family with a history of aortic aneurysms and EDS IV, and Spotila *et al.* (1990) found a glycine to serine substitution at position 661 in the α-helix of α2(I) in a patient with post-menopausal osteoporosis. As with the link between rare connective diseases and the collagen genes, the link between the more common connective tissue diseases and the collagen genes is complex.

CHAPTER FOUR

RNASE A MAPPING TO IDENTIFY A MUTATION IN A PATIENT WITH NON-LETHAL OSTEOGENSIS IMPERFECTA (OI)

4.1   INTRODUCTION

It has been established that the genes coding for $\alpha1(I)$ and $\alpha2(I)$ collagen (COL1A1 and COL1A2 respectively) are linked to dominantly inherited osteogenesis imperfecta (OI) (Sykes *et al.*, 1990). There are many published cases where mutations in COL1A1 or COL1A2 are shown to be the causes of OI (Prockop *et al.*, 1989). Such cases support the linkage studies and show that the mutations are not to be found in genes near to those coding for the type I collagens but actually in COL1A1 or COL1A2. Rare exceptions to this are published in Aitchison *et al.* (1988), Daw *et al.* (1988) and Wallis *et al.* (1990a). Therefore to identify the mutation causing a non lethal OI phenotype, it is necessary to examine COL1A1 and/or COL1A2. However, because these genes are so large (18kb and 35kb respectively), it is more practical to look for mutations affecting the $\alpha1(I)$ or $\alpha2(I)$ polypeptide chains using the mRNA coding for them. RNase A mapping was considered the best method to use to identify mutations which change the coding sequence of these collagen genes.

RNase mapping (Hawkins, 1989) was employed as the screening method because it was proven to be successful at detecting a control base substitution (Hawkins and Dalgleish, unpublished) and a 9 base pair deletion (Hawkins, 1989; Hawkins *et al.*, manuscript in preparation). At the beginning of the work, it was assumed that RNase cleavage tended to bias towards detecting deletions or insertions. Mismatches resulting from deletions, insertions or rearrangements offer greater potential for RNase A cleavage because of the more extensive single-stranded regions produced within the hybrid. However, the fact that the method was able to detect the control base substitution was encouraging.

It was decided to screen the mRNA coding for the polypeptide $\alpha2(I)$ for mutations from individuals with non-lethal OI because, with the protein suicide theory in mind (Williams and Prockop, 1983), it was expected that a mutation in COL1A2 would result in a less severe phenotype than a mutation in COL1A1. A full length $\alpha2(I)$ collagen cDNA clone was given as a

gift by Dr D.J. Prockop. The clone was used to look for mutations causing the non-lethal osteogenesis imperfecta phenotype in a number of patients. Also available was 75% of the α1(I) cDNA, which could be used in the screening experiments.

Since, to date, no reliable correlation has been identified between the position of a mutation within a collagen gene and an OI phenotype, the whole of the α2(I) mRNA had to be screened in order to determine if a mutation was present. A full-length α2(I) cDNA clone was required to do this because RNase A mapping involves the synthesis of radioactive antisense RNA probes from cDNA templates. To screen the whole of the α2(I) mRNA overlapping cDNA probes covering it were required.

During all of this work the ability of the RNase A mapping method to detect only 60% of point mutations (Cotton, 1989) and the fact that the mutation, which an individual may be carrying, may be in COL1A1 not COL1A2 were taken into consideration.

4.2 CONSTRUCTION OF cDNA TEMPLATES FOR RNA PROBE SYNTHESIS

Using the full-length α2(I) cDNA clone Hp2010 (Kuivaniemi *et al.*, 1988a), overlapping subclones were produced in the phagemid transcription vectors pTZ18R and pTZ19R. The numbering and positioning of the subclones in relation to the full-length α2(I) cDNA clone are shown in Figure 4.1. PTZ18R and pTZ19R were employed because they contain a T7 RNA polymerase promoter. The subcloned fragments were between 1kb and 400bp in length and were inserted into the vector in such a way as to be able to run off, from the cDNA template, an antisense RNA probe *in vitro*. RNA probes which are larger than 1kb degrade easily, causing high background on the final autoradiographs; conversely, probes which were smaller than 400bp did not hybridize sufficiently well to the cellular RNA. The sizes of the subclones and the endonuclease sites into which they were inserted are shown in Table 4.1.

Large scale preparations of the subclones were prepared as described in section 2.13.1. The templates were linearized by digestion with a restriction endonuclease to give a 5' overhang. Templates with a blunt end or 3' overhang restrict probe synthesis and result in a low radioactive nucleotide incorporation. After digestion the templates were extracted with

Figure 4.1

The positions of the α2(I) cDNA subclones.

      A diagram to show the positioning of the α2(I) subclones to be used in RNase A protection analysis, in relation to the cDNA clone Hp2010 (Kuivaniemi *et al.*, 1988)

ATG = Start of transcription.

STH = Start of triple helical domain.

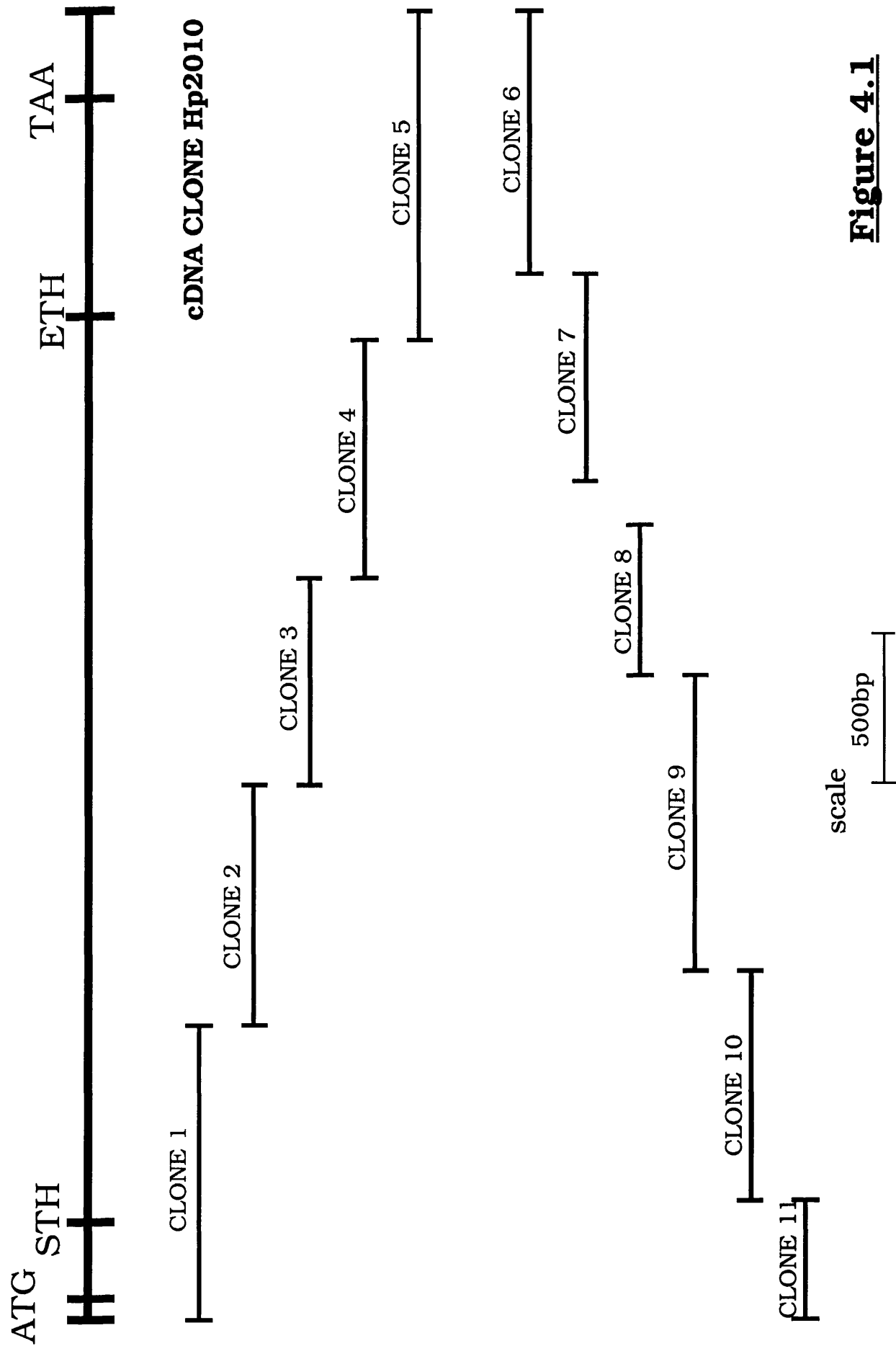ETH = End of triple helical domain.

TAA = end of translation.

Figure 4.1

Table 4.1

The size of the α2(I) subclones. Also displayed are the restriction endonuclease sites to which the inserts were ligated.

Table 4.1

clone number	restriction endonuclease sites	size of insert (to the nearest 10bp)
1	Tha I - Tha I	960
2	Tha I - Tha I	790
3	Tha I - Xho I	860
4	Xho I - Xba I	830
5	Xba I - Hind III	930
6	Hind III - Sst II	730
7	SstI - Pst I	780
8	Hinc II - Hinc II	550
9	Hinc II - Pst I	1030
10	Xho II - Pst I	780
11	Pst I - EcoR I	400

phenol, then chloroform to render them RNase free (section 2.17.3).

4.3 RNA ISOLATION FROM TISSUE CULTURE CELL LINES

Tissue culture fibroblast cell lines derived from patients with non-lethal OI and one with OI type II were propagated. RNA was isolated from them as described in section 2.14.5.

Little was known of the clinical features of the 5 patients from whom the 5 cell lines obtained from ATCC were derived, except that they suffered from non-lethal OI.

The cells from Cardiff were from a phenotypically normal father and his three affected sons (CF1-4). No additional information was known about these cells or their donors.

The eight cell lines donated by P. Byers (University of Washington, USA.) were sent with additional information concerning the initiation of overmodification of the α1(I) polypeptides produced by each line of cells. This information is summarized in Table 4.2. The cyanogen bromide mapping experiments were carried out, as described by Bonadio and Byers (1985), on α1(I) only. The results in Table 4.2 refer to the start of overmodification in α1(I), overmodification would start at an equivalent position in α2(I) (Figure 4.2). With the exceptions of the mutation found by Hawkins (1989) and some whole exon deletions (Sippola, *et al.*, 1984), the start of overmodification has been found to coincide with the position of a mutation in the propeptide. Taking the cyanogen bromide mapping results into consideration, these cell lines were screened only across the area around the start of overmodification.

Each of the RNA preparations were checked for the presence of a gross defect in the mRNA coding for α2(I) collagen by northern blot analysis. None were detected (results not shown).

4.4 RNASE A PROTECTION OF THE α2(I) PROBE 10 BY mRNA FROM CELL LINES LW, GM1436A AND GM2573

The probe transcribed from subclone 10 (Figure 4.1) was prepared as described in section 2.17.4. Approximately 2×10^6 cpm of probe were hybridized to 50μg of total cellular RNA from cell lines GM1436A, GM2573

Table 4.2

The results of cyanogen bromide mapping of α1(I) polypeptides from the fibroblast cell lines from the USA.

Only the α1(I) polypeptides from each cell line were subjected to cyanogen bromide mapping. Overmodification within a type I collagen molecule affects the three a-chains and will start in the same relative position in all three chains. It is more informative to look for the start of overmodification in the α1(I) polypeptide because α1(I) mapping yields 4 cyanogen bromide polypeptides whereas α2(I) cyanogen bromide mapping yields only 2 large polypeptides.

Table 4.2

Cell Line	O.I. Phenotype	Start of overmodification
R	II	CB8
PJ	III	CB7
KO	III/IV	CB8
LW	III	CB8
JR	IV	CB8
JB	III	CB8
PIR	III	Not Localized
SC	IV	CB8

Figure 4.2

Cyanogen bromide maps of the α1(I) and α2(I) polypeptides.

Cyanogen bromide cleaves polypeptides at methionine amino acids. These maps have been deduced by analysing the amino acid sequences of the α-chains

The position of subclone 10 in relation to the cyanogen bromide peptides is shown.
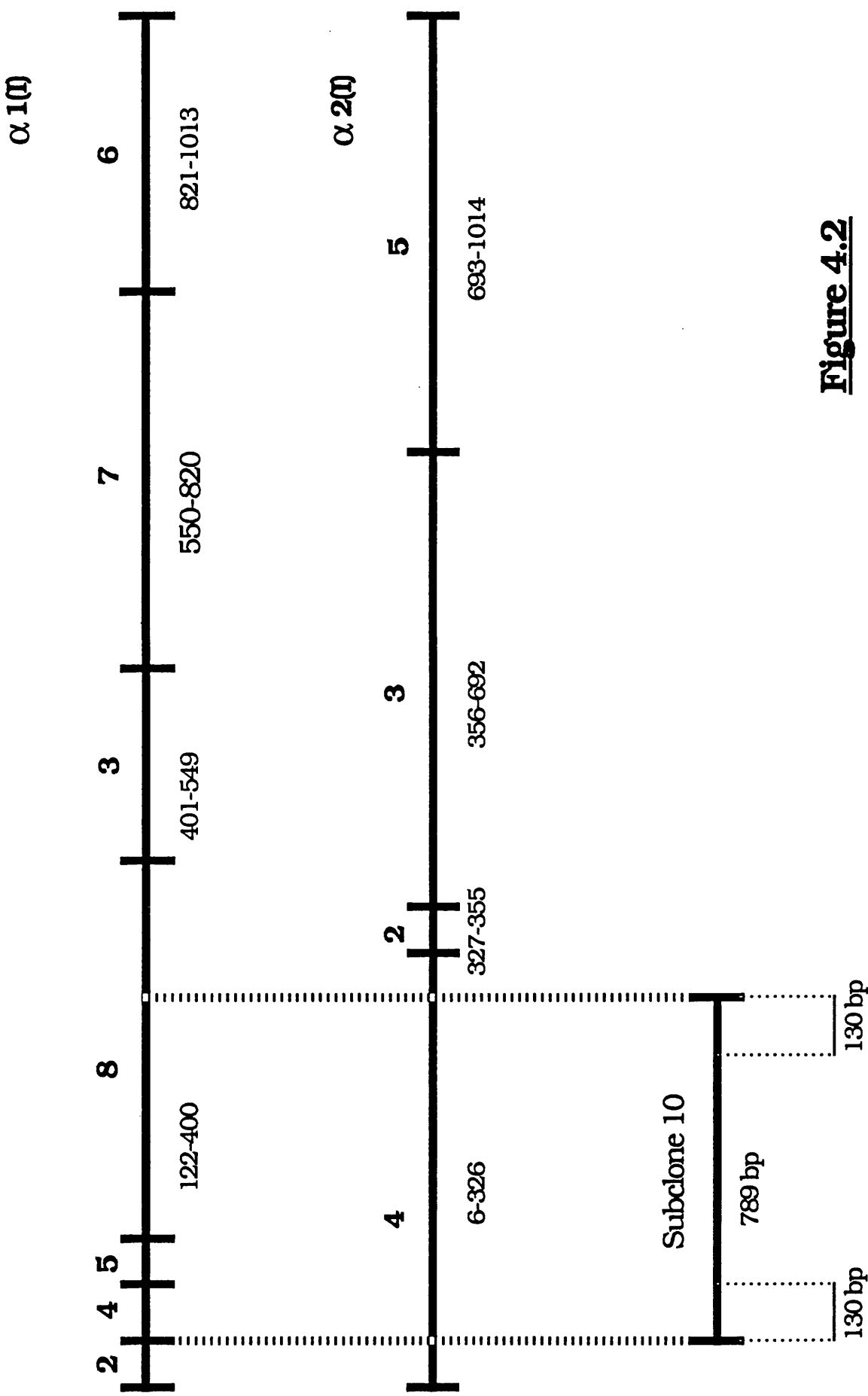
Figure 4.2

and LW.

Cyanogen bromide mapping of the α1(I) collagen chains from cell line LW revealed overmodification originating within α1(I)CB8, (α2(I)CB4, Figure 4.2). Part of the mRNA coding for α2(I)CB4 is covered by subclone 10 (Figure 4.2). The probe/mRNA hybrids were digested with RNase A and treated as described in section 2.17.6. The resulting RNA was electrophoresed through a denaturing polyacrylamide gel alongside marker DNA. The gel was fixed then dried, before being autoradiographed for 48 hours.

The autoradiograph (Figure 4.3) shows complete probe protection when the probe was hybridized with RNA from cell lines GM1436A and GM2573. The probe was cleaved when it was digested with RNase A after hybridization to RNA from cell line LW. The lane in Figure 4.3 containing probe 10 hybridized to RNA from LW has a band corresponding to the fully-protected probe, 780 bases, and two bands corresponding to the cleaved probe, 650 bases and 130 bases. This result indicates that a mismatch may be found in the area covered by the last 130 bases of probe 10. The protein analysis would suggest that the mismatch is covered by the 3' end of the probe, which corresponds to α2(I)CB4 (Figure 4.2).

4.5 RNASE DIGESTION OF SUBCLONE 10/*FokI* HYBRIDIZED TO RNA FROM CELL LINE LW

A RNase mapping experiment was performed using subclone 10 digested with the restriction endonuclease *FokI*. Sequence data from Kuivaniemi *et al.* (1988a) indicated that *FokI* digestion of the subclone 10 would remove 180bp from the 5' end of the probe template (Figure 4.4A). If the mismatch corresponded to the 3' end of clone 10, cleavage products would be seen on the autoradiograph (470 bases and 130 bases). If the mismatch corresponded to the 5' end of subclone 10, no cleavage products would be seen on the autoradiograph. A large scale preparation was made of the new template (sublone 10/*FokI*). This was cleaned to render it RNase free (section 2.17.3) and an antisense RNA probe was synthesized from it. RNase A protection analysis was repeated using this probe and RNA from LW.

Figure 4.4B shows a band running at approximately 600 bases (normal allele) and two cleavage products running at approximately 460 and 140 bases (abnormal allele), indicating that the mismatch was covered by the 130-140 bases at the 3' end of probe 10. This experiment confirmed the

Figure 4.3

A mismatch between the α2(I) probe 10 and mRNA from the cell line LW.

A RNase A protection autoradiograph showing probe 10 fully protected by RNA from cell lines GM1436A and GM2573 and the cleaved probe after hybridization with RNA from cell line LW.

From left to right the lanes of the gel were loaded with :-

^{32}P-labelled pUC13/*HpaII* (markers)

^{32}P-labelled M13mp18/*MspI* (markers)

Probe 10 protected from RNase A digestion by mRNA from cell line GM1436A

Probe 10 protected from RNase A digestion by mRNA from cell line GM2573

Probe 10 hybridized to RNA from the cell line LW and digested with RNase A. (LW)
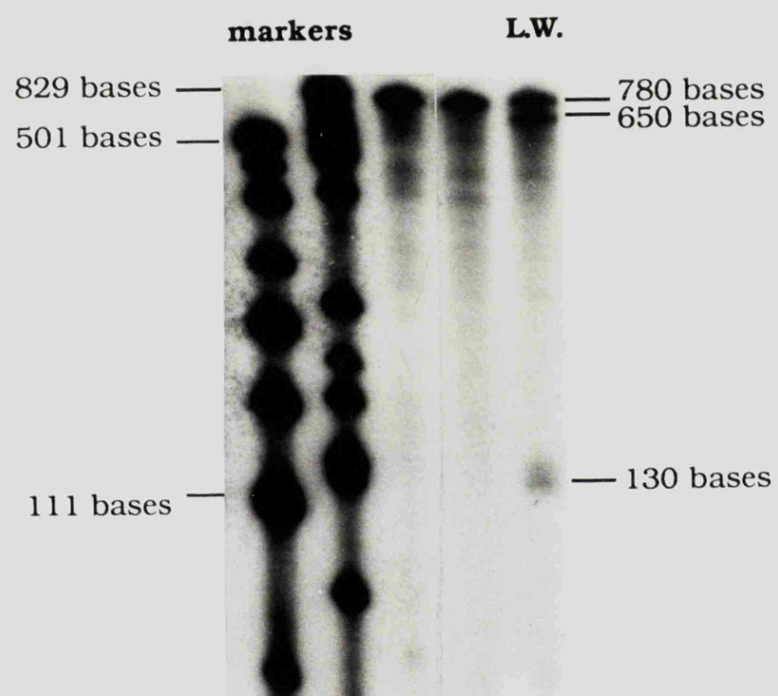
Figure 4.3

Figure 4.4

A mismatch between probe 10/*FokI* and RNA from cell line LW.

A

A map to show the position of the *FokI* site within subclone 10 and the size of the new probe template.

B

The mismatch between probe 10 and mRNA isolated from LW is located at the 3' end of probe 10.

A RNase A protection autoradiograph showing probe 10/*FokI* cleaved by RNase A digestion after hybridization to cellular RNA from LW.

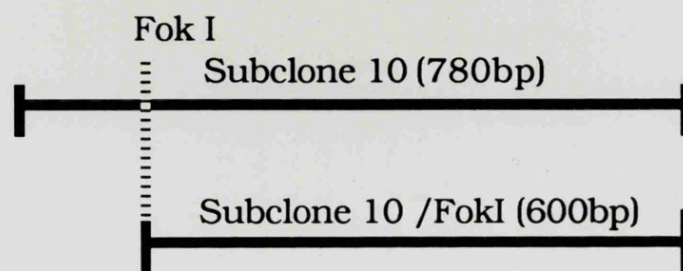From left to right the lanes of the gel have been loaded with:-

^{32}P-labelled pUC13/*HpaII* (markers)
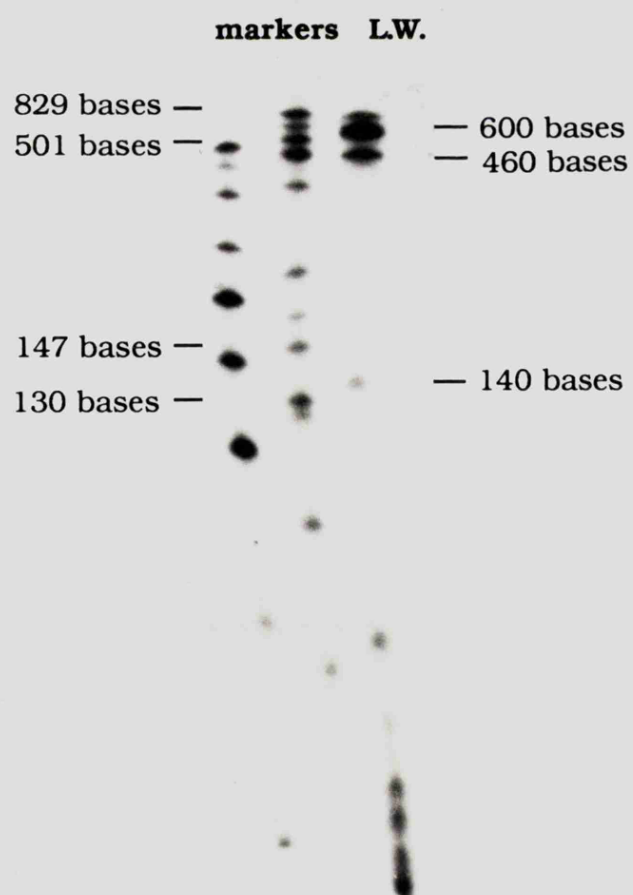
^{32}P-labelled M13mp18/*MspI* (markers)

Probe 10/FokI hybridized with RNA from LW fibroblast and digested with RNase A.

Figure 4.4

A

Fok I

Subclone 10 (780bp)

Subclone 10 /FokI (600bp)

B

markers   L.W.

829 bases —

501 bases —

— 600 bases

— 460 bases

147 bases —

130 bases —

— 140 bases

results obtained previously, suggesting that a mismatch existed between mRNA isolated from the cell line LW and probe 10, therefore the result obtained previously was not an artefact.

To determine the exact positioning of the mismatch in relation to probe 10, RNase A mapping was repeated using probe 10/*FokI* and RNA from LW. The resulting digested hybrid was run through a 40cm long sequencing-like gel containing RNase A gel mix, alongside a sequencing ladder. Using the sequencing ladder as markers, the size of the smaller cleavage product was determined (Figure 4.5A). On the sequencing gel, the smaller cleavage product was seen to be 4 bands running at 107, 111, 115 and 119 bases. Multiple cleavage products, such as these, were seen by Hawkins (1989). These multiple cleavage bands were subsequently found to be due to the presence of a 9bp deletion in the mRNA.

The size of the smaller cleavage products along with sequence data from Kuivaniemi *et al.* (1988a), placed the approximate position of the mismatch at the 3' end of exon 19 or the 5' end of exon 20, at or about amino acid 260 (if the first glycine in the helix of the mature α-chain is numbered 1). (Figure 4.5B)

4.6 PCR AMPLIFICATION OF THE DNA SEQUENCE CONTAINING THE SUSPECTED MUTATION

Having established that the mutation was to be found in COL2A1 at or around the codon for amino acid 260. DNA around this area, from the cell line LW, was amplified using the PCR, with the intention of sequencing the DNA and therefore characterizing the mutation.

PCR primers were designed to enable amplification of this region of DNA. Unfortunately only the cDNA sequence of this section of the COL1A2 was known. The primers were designed to keep the PCR product as short as possible since the size of intron 19 was not accurately known. A small fragment size was also chosen for ease of sequencing. For easier cloning, each primer was designed with a restriction endonuclease recognition site incorporated into it (Figure 4.6). Problems encountered by myself and other workers upon trying to digest PCR product with restriction endonucleases led to careful consideration as to the choice of enzyme recognition site incorporated into the primers. Primer A2B contained an *EcoRI* site. An 8bp 5' extension sequence was added to primer A2B to encourage cleavage of the

Figure 4.5

Fine mapping of the mismatch between probe 10/*FokI* and mRNA from the cell line LW.

A

An autoradiograph showing the sizes of the shorter cleavage products produced after RNase A digestion of a probe 10/*FokI*-LW hybrid.

From left to right:-

Small RNase A cleavage products running at approximately 107, 111 bases, 115 bases and 119 bases.

Sequencing ladder markers.

B

A diagrammatic representation of the approximate position of the mismatch.

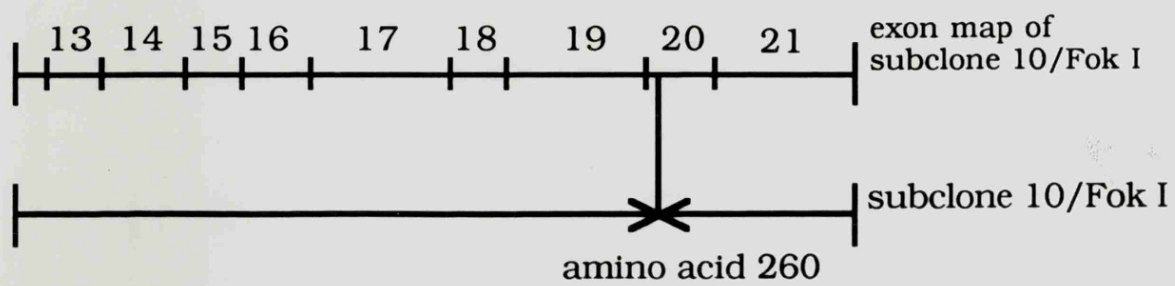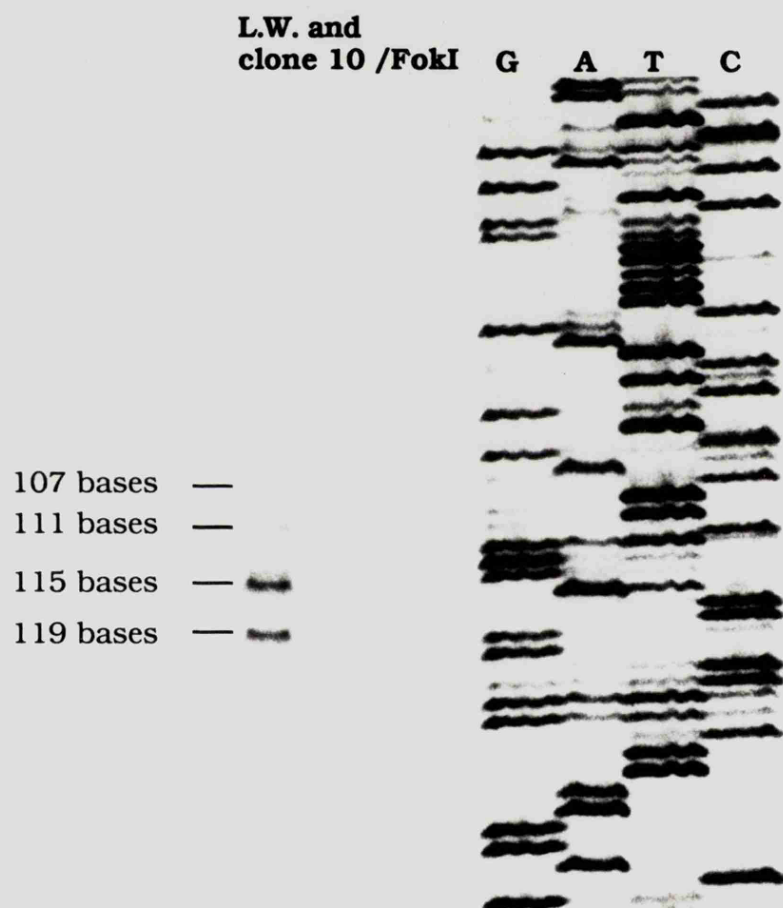X = the mismatch at the 3' end of exon 19 or the 5' end of exon 20.

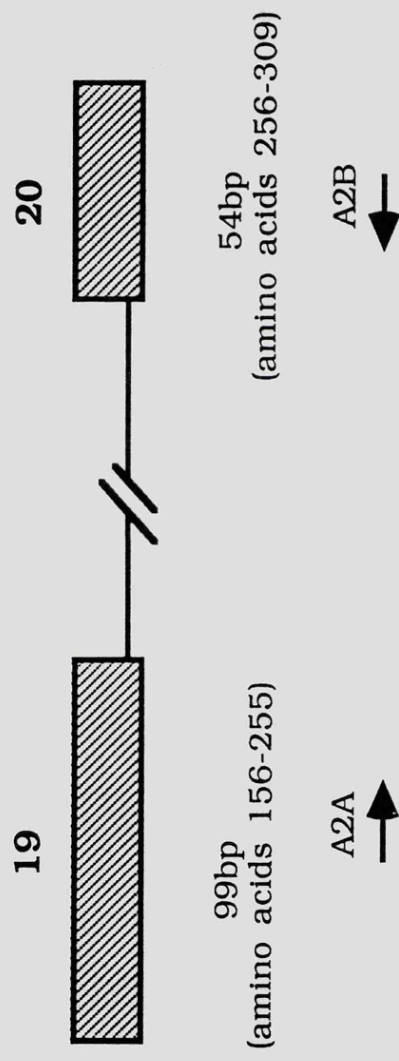# Figure 4.5

Figure 4.6

PCR primers A2A and A2B.

An exon map of part of COL1A2.

The positions of the sequences within COL1A2 which the PCR primers A2A and A2B will anneal to are shown (arrows).

The size of the PCR product could not be predicted because the intron sequence was not known.

Hatched boxes = exons

Lines = introns

19

99bp
(amino acids 156-255)

A2A

20

54bp
(amino acids 256-309)

A2B

5'   3'

SstII

Primer A2A = TCCCTGGACCCCGCGGTATT

EcoRI

Primer A2B = TCACCGGTGAATTCGGGCTCACCCCTTGTTACCGC

Figure 4.6

EcoRI site (Jeffreys *et al.*, 1990). Primer A2A was designed with an *SstII* site within it. This *SstII* site was homologous to that in the COL1A2 sequence and was located as near to the 3' end of the primer as conveniently possible. It has been found that some enzyme recognition sites at the 5' end of PCR primers are not cut efficiently (Jung *et al.*, 1990).

PCR amplification of this area of COL1A2 was performed as described in section 2.15.1, using DNA isolated from the fibroblast cell line LW as template. One tenth volume of PCR product was run through a 0.8% agarose gel. A DNA fragment running at about 600bp was seen in the gel, no contaminating bands were seen.

4.7   CHARACTERIZATION OF THE MUTATION BY SEQUENCE ANALYSIS

Attempts were made to directly sequence the 600bp PCR product using the PCR primers as sequencing primers, as described in section 2.18.4. Unfortunately, this resulted in what looked like primer slippage along the exon sequence. Only what looked like intron sequence could be determined on the sequencing ladder. This result was obtained when sequencing in each direction. It was therefore decided that the PCR fragment should be cloned and then sequenced.

One tenth volume of the PCR product was cleaned by phenol extraction and ethanol precipitated. The DNA was digested with *SstII* and *EcoRI* for 1 hour at 37°C in the presence of spermidine-HCl (4mM). The digested DNA was ligated into pBluescript SK⁻, which was subsequently transformed into *E. coli* XL-1Blue (described in sections 2.9 and 2.11.2 respectively). DNA preparations were made of 5 recombinant clones (as described in section 2.18.1). The clones were sequenced using M13 forward and reverse primers, to obtain the sequence of exon 19 and exon 20 in each clone.

All of the clones contained a normal exon 20 sequence. Three of the clones contained a normal exon 19 sequence, the remaining two clones contained a 3 base pair deletion at the 3' end of exon 19 (Figure 4.7). The loss of these three bases resulted in the inframe deletion of the codon coding for amino acid 255, a valine.

This result agreed with the results obtained from the RNase A mapping experiments, which predicted that the mutation would be found at
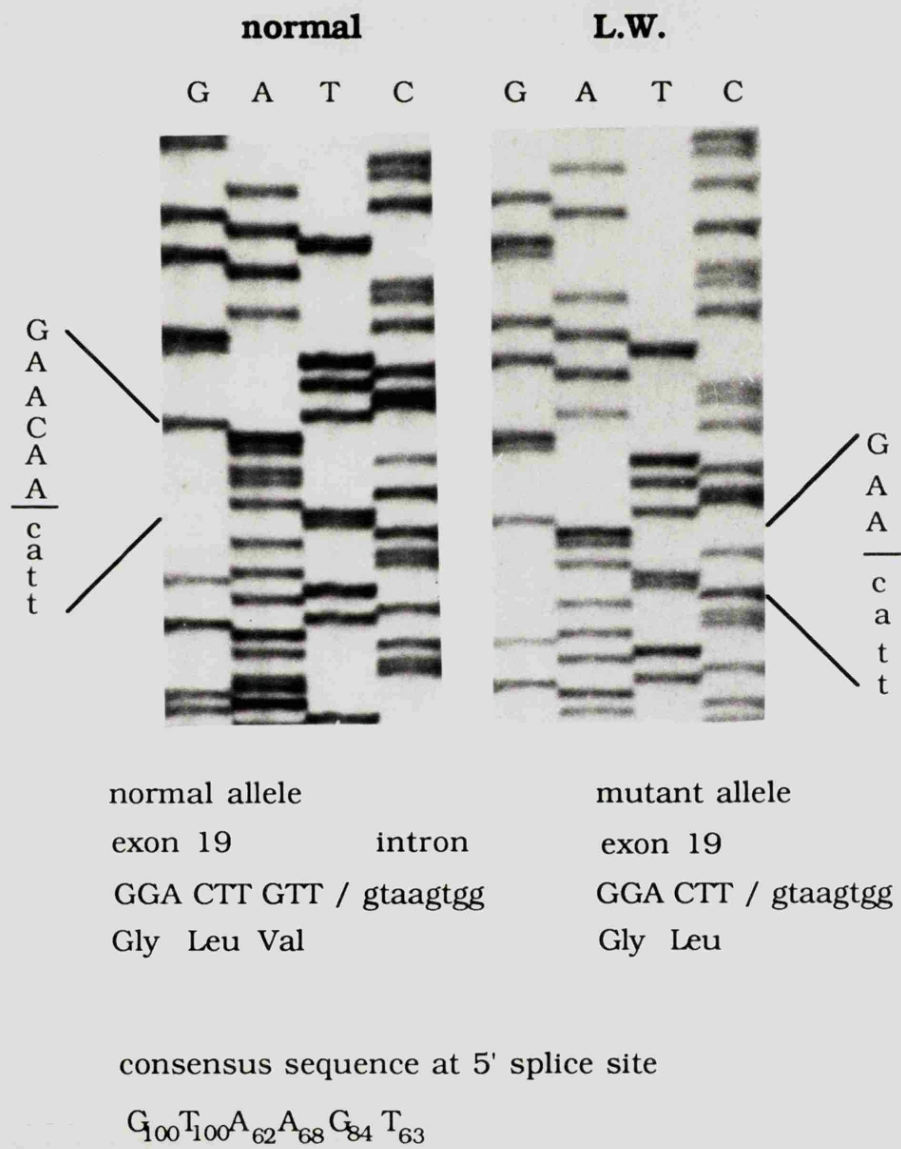
Figure 4.7

The sequence difference between the α2(I) probe 10 and mRNA from the cell line LW is a 3bp deletion.

The sequencing gel autoradiograph shows the sequence of the normal allele and the abnormal allele. The sequences on the autoradiographs are that of antisense DNA. The sequences written below the autoradiograph are those of sense DNA.

One COL1A2 allele from LW is normal in this region, the other has a 3bp deletion. The 3bp which are deleted code for amino acid 255, the last amino acid coded for by exon 19.

Figure 4.7

	normal					L.W.			
	G	A	T	C		G	A	T	C

G
A
A
C
A
A
‾
c
a
t
t

G
A
A
‾
c
a
t
t

normal allele   mutant allele

exon 19   intron   exon 19

GGA CTT GTT / gtaagtgg   GGA CTT / gtaagtgg

Gly Leu Val   Gly Leu

consensus sequence at 5' splice site

$G_{100}T_{100}A_{62}A_{68}G_{84}T_{63}$

or around amino acid 260. The appearance of the four cleavage products running at 107,111, 115 and 119 bases on the sequencing-like gel may have given a clue that the mutation involved more than one base. Just as the multiple bands seen by Hawkins (1989) may have indicated that many bases were involved in that mutation.

It seems that even if the primer slippage had not been encountered during the direct sequencing, then this 3bp deletion would not have been as easily characterized using that method. Direct sequencing involves sequencing both alleles on the same sequencing ladder, if a deletion is present in one allele the ladder would be clear up until the deletion then it would be muddled. Cloning then sequencing was inevitable. The results so far suggested that LW carried a 3bp deletion which caused the inframe deletion of the last codon of exon 19. A codon which codes for a valine. LW is heterozygous for this deletion as indicated by the fact that half the probe in the Figures 4.3 and 4.4 is fully protected.

4.8   IS THE 3bp DELETION INHERITED BY LW FROM EITHER PARENT?

Having determined that the mutation detected by RNase A mapping was the deletion of the last 3 bp of exon 19 of COL1A2, it was of interest to find out if the mutation was inherited from either parent.

LW is the daughter of phenotypically normal, non-consanguineous parents. Genomic DNA was isolated from the lymphocytes of her mother (EW) and her father (PW) by members of P. Byer's laboratory (University of Washington, USA). Using these DNAs and DNA isolated from the LW cell line, parentage was confirmed using single locus specific minisatellite probes (gifts from J. Armour, University of Leicester) and the method described by Jeffreys *et al.* (1980). Three single locus probes were hybridized individually to a Southern blot of 1.5µg of each of the three DNA preparations digested with *Mbo*I. The three probes used were cMS621, cMS600 (Armour *et al.*, 1990) and λMS32 (Wong *et al.*, 1987). Using these probes a positive parentage result of 1 x 10^9 was obtained.

The DNA sequence coding for the 3' end of exon 19 and the 5' end of exon 20 was amplified, using primers A2A and A2B and genomic DNA from LW and her parents as template. The result of all three of the PCRs was a 600bp fragment, as seen when 0.1 volume of PCR product was

electrophoresed through a 0.8% agarose gel.

Direct sequencing of these fragments was not possible for reasons explained in section 4.7. The inheritance of the 3bp deletion was determined by the size difference of the PCR products from LW and her parents. However, DNA fragments of about 600bp cannot be separated from a fragment 3bp shorter, even on a polyacrylamide gel. The PCR products had to be digested into smaller fragments before running through a polyacrylamide gel to determine the presence or absence of the 3bp deletion.

Sequencing of part of intron 19, when looking for the mutation in LW, revealed a *FokI* site at the 5' end of the intron. There are no *FokI* sites in the sequence of exon 19 or 20 of COL1A2. It was predicted that digestion of the PCR products would result in a fragment 81bp in length from the mutant allele and a fragment 84bp in length from the normal allele. Fragments this size can be separated on a 12% polyacrylamide gel. The PCR product from the parents and LW were treated as described in section 2.15.3, digested with *FokI*, then electrophoresed through a 12% polyacrylamide gel for 3 hours. The gel was photographed, see Figure 4.8.

The photograph of the gel shows the presence of two alleles differing in size by 3 base pairs in the lane containing the amplified and digested DNA from LW. Only one band, representing a single allele, is seen in each lane containing amplified and digested DNA from the parents. The bands running above the two allele bands in the lane containing DNA from LW are the result of heteroduplex formation.

It is concluded from this result that the 3bp deletion found in LW's DNA is not inherited from either parent.

The results of this experiment also confirm that the 3bp deletion found in two of the 5 clones sequenced was not a PCR artefact. The presence of such artefacts may be misleading if the artefact is present in a large proportion of clones.

4.9   DOES THIS 3bp DELETION AT THE 3' END OF EXON 19 CAUSE A CHANGE IN THE SPLICE SEQUENCE?

The position of the deletion (the last 3 bases of exon 19) suggests that its presence may cause a change in the splicing pattern of the hnRNA in this area. Some cases of osteogenesis imperfecta have been found to be the result of mis-splicing of a type I collagen hnRNA, brought about by a

Figure 4.8

The 3bp deletion is not inherited by LW from either of her parents

DNA isolated from fibroblast cells from LW is heterozygous for the 3bp deletion, while DNA from lymphoblasts isolated from LW's parents contain only one COL1A2 allele.

Genomic DNA isolated from skin fibroblasts from LW and lymphoblasts from her parents was amplified using the method described in section 2.15.1. The resulting DNA was digested with the resrition enzyme *FokI* before electrophoresis through a 12% acrylamide gel. An 84bp DNA fragment is the result of amplification and digestion of the normal allele, an 81bp DNA fragment represents the mutated allele.
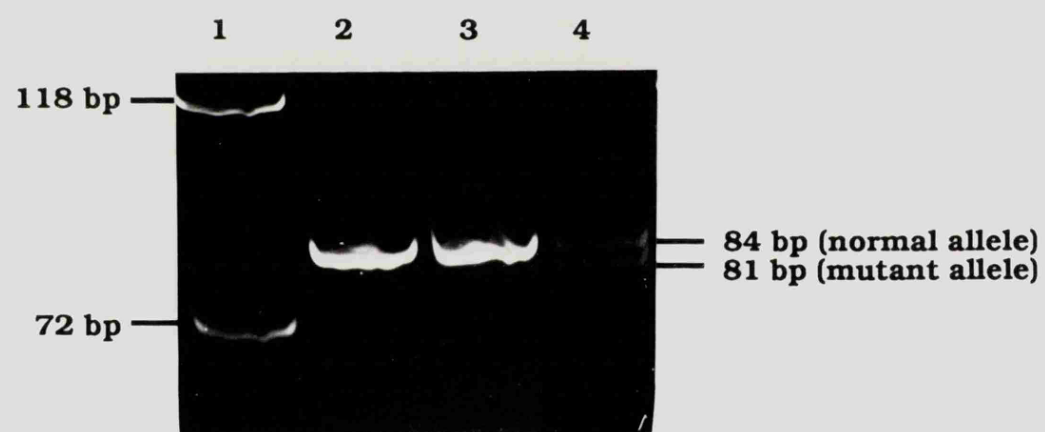
Lane 1 - φX174/*HaeIII*

Lane 2 - DNA from the mother

Lane 3 - DNA from the father

Lane 4 - DNA from LW

Figure 4.8

mutation in the splice site consensus sequence (Bonadio *et al.*, 1990; Tromp and Prockop, 1988).

If mis-splicing of the mRNA coding for α2(I) was occurring in the cells of LW, a change in the splice site consensus sequence at the 3' end of exon 19 would result in the mis-splicing of exon 19.

The results from RNase A mapping, Figure 4.3 and 4.4B, indicate that mis-splicing does not occur because the lengths of the two cleavage products add up to the length of the protected probe. If exon 19 had been mis-spliced, the lengths of the two cleavage products would not add up to the length of the full-length probe, there would some bases missing.

Sequencing data suggests that no abnormal splicing would occur as a result of the deletion. The intron sequence of both the normal and mutated allele remain unchanged by the mutation:-

exon 19/GTAAGT....intron 19.....CCCTAG/exon 20

Despite the presence of the deletion, the last two bases of exon 19 in the mutant and normal allele are the same:-
exon 19
normal allele AGA GGA CTT G<u>TT</u>/intron sequence
exon 19
mutant allele AGA GGA C<u>TT</u>/intron sequence

The 5' splice site consensus sequence (Lear *et al.*, 1990) has been found to be :-
exon   intron
A_{64} G_{73} [G_{100} T_{100} A_{62} A_{68} G_{84} T_{63}]

The numbers to the right of each base represent the frequency, as a percentage, that each nucleotide is found at that position.

To confirm the assumption that the 3bp deletion does not alter the splicing out of intron 19, the area of mRNA around the mutation site was amplified using the PCR. The resulting DNA was electrophoresed through a polyacrylamide gel to separate the two allelic bands which would run 3bp apart, if no splicing changes were present.

4.10 PCR AMPLIFICATION OF mRNA USING REVERSE TRANSCRIPTASE TO SYNTHESIZE THE cDNA.

The method used is described in section 2.15.2 and uses Moloney Murine Leukemia Virus reverse transcriptase to synthesize the cDNA, then *Taq* polymerase in the polymerase chain reaction.

The method was carried out using total cellular RNA from the cell line LW as the template and primers A2A and A2B. 30 PCR cycles were run; 95°C, 1 minute, 55°C, 1 minute, and 70°C, 1 minute. One tenth volume of the PCR product was run through a 12% polyacrylamide gel until the xylene cyanol dye reached the bottom of the gel. After staining with ethidium bromide, the gel was visualized on a U.V. transilluminator and photographed, (Figure 4.9). Two of the bands seen in Figure 4.9 are 135bp and 132bp long, they therefore represent the two alleles, normal and mutated respectively. The 132bp fragment was cloned and sequenced and found to contain the 3bp deletion. This result clearly shows that the 3bp deletion found in the cell line LW did not cause mis-splicing of exon 19.

4.11 DISCUSSION

LW was diagnosed as suffering from type III osteogenesis imperfecta, the progressively deforming form of the disease, as classified by Sillence *et al.* (1979). She is the daughter of phenotypically normal, non-consanguineous parents.

Cyanogen bromide (CB) mapping of the α1(I) polypeptides synthesized by fibroblasts from LW, indicated overmodification of the type I collagen polypeptides, which started within α1(I)CB8, (α2(I)CB4).

RNase A protection analysis was used to detect a mutation within the mRNA coding for the α2(I) polypeptide, isolated from the cell line LW. The method confirmed the results of the cyanogen bromide mapping, placing a mutation within α2(I)CB4. Sequencing this region revealed a heterozygous 3bp deletion at the 3' end of exon 19. The deletion was inframe and resulted in the loss of amino acid 255, which is a valine. It is suggested that the deletion may have been brought about by slippage during replication.

After confirming parentage using 3 single locus minisatellite probes, it was determined that the mutation was not to be found in DNA isolated from lymphoblasts from either parent. The deletion may be a *de*

98

Figure 4.9

PCR amplification of the two α2(I) mRNA species from the fibroblast cell line LW.

cDNA was synthesized from mRNA isolated from the cell line LW. This cDNA was amplified using the polymerase chain reaction. The method used for the cDNA synthesis and the PCR amplification is described in section 2.15.2.

The two α2(I) mRNA species are represented on the gel as bands which are 132bp and 135bp in length. The two bands running further up the gel are the result of heteroduplex formation amongst the PCR products.

Lane 1 - PCR product

Lane 2 - φX174/*HaeIII*

Figure 4.9
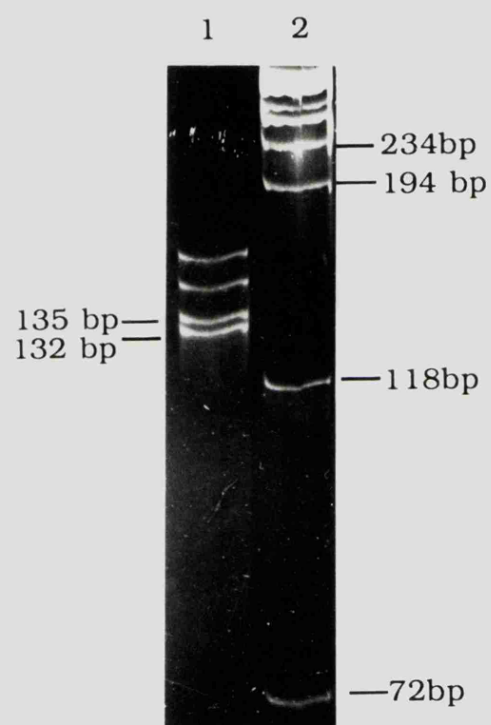
novo mutation or inherited from a germline mutation present in either parent. Germline mutations occur relatively frequently among cases of type II osteogenesis imperfecta (Byers *et al.*, 1988). In at least one case, this is accompanied by somatic cell mutation in the phenotypically normal father. In this reported case, the mutation is detected at a 1:4 ratio with the normal allele in DNA extracted from the father's white blood cells (Cohn, *et al.*, 1990). Such somatic cell mosaicism was not observed in the white blood cell DNA from LW's parents. Either way, germline or *de novo* mutation, the deletion is a good candidate for the cause of the type III OI phenotype because the protein and nucleic acid data agree.

This is the first single amino acid deletion to be characterized in a collagen gene, therefore little is known about the effect of such a mutation on phenotype. There have been relatively few silent mutations characterized in collagen genes, (Weil *et al.*, 1990; Zafarullah *et al.*, 1990). This type of mutation would be predicted to be less well tolerated than most mutations because it disrupts the Gly-X-Y repeat sequence, the sequence which determines the collagen helix formation.

Normal type I collagen helix formation involves the three α-chains lining up in a staggered array. Each α-chain is staggered by one amino acid. This staggered arrangement allows the glycines in each α-chain of Gly-X-Y repeats to be in the centre of the helix, an essential part of normal helix formation. The staggered arrangement also allows optimal hydrogen bond formation between the 3 α-chains. Hydrogen bonds form between the NH-group of the glycines and the CO-group of the residues in the X position of an adjacent chain. The charge distribution along the triple helical domains of normal α1(I) and α2(I) chains is virtually identical. When in this 1 amino acid staggered arrangement, the α-chains form intramolecular charge interactions. These were once thought to be of minor importance to the helix formation, however Willing *et al.* (1988) suggested that a disruption of these interactions, in the absence of any force disrupting the Gly-X-Y sequence, slows down helix formation and thus allows overmodification of the α-chains 5' to the mutation. The presence of a 3bp in frame deletion in the helix disrupts all these arrangements. It is suggested that, in a effort to return the helix to its normal structure as much as possible, a helix is formed with a kink within in it, similar to that described by Vogel *et al.* (1988). This prediction is based on the knowledge that the α2(I) polypeptide chain will be 1 amino acid shorter than the two α1(I) polypeptide chains. If the two α1(I) chains looped out of helix formation at the position corresponding to the

deletion in the α2(I) chain, then the normal helical staggered arrangement could be restored, along with normal hydrogen bonding and side chain charge interactions. The mutation which resulted in the kinked molecule found by Vogel was present in a proband diagnosed as suffering from type II OI. The kinked molecules were formed as a result of a heterozygous glycine to cysteine substitution in the proα1(I) polypeptide. It was suggested that the incorporation of these kinked molecules into the collagen fibrils, instead of the normal rod-like helical structures, resulted in the lethal phenotype. The formation of the kink in these procollagen molecules caused the disruption of the amino procollagen cleavage site which may have contributed to the phenotype, however it is unlikely to have caused it alone. In general, retention of the amino terminal propeptide results in an EDS VII phenotype (Weil et al., 1990; Eyre et al., 1985).

The phenotype of LW is less severe than that of the proband observed by Vogel et al. (1988). However it is more severe than expected from a heterozygous mutation at the amino end of the α2(I) polypeptide. The mutations found so far in this part of COL1A2 have resulted in an EDS or mild OI phenotype (Weil et al., 1990; Kuivaniemi et al., 1988b) The kinked molecule found by Vogel et al. (1988) was the result of a substitution at position 748 of the α1(I) chain, which was approximately where the kink is located. Therefore the kink was almost in the centre of the helix. A kink at position 255 may be better tolerated and lead to a less severe phenotype. About 25% of the type I molecules from the proband observed by Vogel et al. were kinked. It would be expected that proportionally fewer molecules would have a kink in them if the kink was formed as a result of a mutation in the α2(I) polypeptide. Each type I molecule being made up of two α1(I) polypeptides and one α2(I) polypeptide.

Experiments to determine the rate of secretion of procollagen molecules containing this mutated α2(I) chain have not been performed. However, even if most of the abnormal procollagen molecules are degraded by the cells then, because of the crystal-like manner in which the fibrils are formed, the few that are incorporated into the collagen fibrils will be sufficient to disrupt fibril formation and thus cause an abnormal phenotype.

Kinks are found in normal collagen molecules types IV, VI, and IX but none of these molecules aggregate to form fibrils as type I collagen does. No doubt these kinks in the collagen molecules contribute to their function.

The deletion found during this work results in the replacement of a Gly-X for the usual Gly-X-Y in some of the α2(I) polypeptides in the type I

molecules. A Gly-X is found in the type VI polypeptides, but this motif is found in all three type VI polypeptides in the same place (Furthmayr *et al.*, 1983). It is suggested that this Gly-X and a longer helical interruption found in all three type VI polypeptide allows the supertwisting between dimers in the formation of tetramers (Chu *et al.*, 1990), even though Furthmayr *et al.* (1983) did not find any kinks or obvious sites of flexibility in the α-chain helix.

It remains to be seen if this mutation in the α2(I) collagen chains of LW causes a kinked type I molecule to form. Analysis of type I collagen molecules using electron microscopy and rotary shadowing, such as that carried out by Vogel *et al.* (1988) would show the presence of such a molecule, if it exists. A second procedure which may prove useful in assessing the presence or absence of kinked type I collagen molecules is described by Kobayashi *et al.* (1990). The method involves the extraction of collagen from fibroblasts. The extracted collagen is made to form lateral aggregates of collagen molecules. Normal collagen molecules are see by electron microscopy to be tightly packed along the whole of the α-chain. Molecules containing a mutation are seen to be tightly packed at the carboxyl terminus but are loose and spreading at the amino terminus. This procedure has the disadvantage that it does not define the exact site of the mutation. The authors did not known the nature of the mutation with which they were working so it was not determined whether this spreading of α-chains was due to overmodification or the presence of a kink in the α-chains.

These results do not rule out the possibility of LW being a compound heterozygote, with a second mutation in her other COL1A2 allele, as has been seen to occur in another case of OI (de Wet *et al.*, 1983). Nor do they rule out the possibility of LW being a combined heterozygote, with a second mutation in a second gene ie. COL1A1.

Until recently, type III osteogenesis imperfecta was considered to be an autosomal recessive condition but, this result along with other biochemical data (Wenstrup *et al.*, 1990a), shows that this is not always the case. Compared with type II OI, very few of the mutations causing the type III OI phenotype have been characterized (Pack *et al.*, 1989; Pihlajaniemi *et al.*, 1984).

The results presented here suggest that the presence of this deletion does not cause alternative splicing. It would be predicted that, because there was no change in the 5' splice site consensus sequence, no change in splicing would occur.

If this work had been started a year ago, it is probable that the chemical cleavage method (Cotton *et al.*, 1988) would have been employed in the detection of the mutations. At that time chemical cleavage was expected to detect 100% of mutations in collagen genes (Cotton, 1989). This overestimation of the sensitivity of this method has since been revised (Bhattacharyya and Lilley, 1989). It seems that the chemical cleavage method is prone to similar problems as RNase A mapping. The detection of any mutation is dependent on the environment in which the mutation is found. The sequence of the RNA or DNA around the mutation may influence whether the radioactive probe is available for cleavage by RNase A or modified and cleaved by the relevant chemicals.

If this work had been started today, the single stranded conformation polymorphism method (Orita *et al.*, 1989) or Hydrolink gels (Keen *et al.*, 1991) would probably be employed. These methods overcome the problems inherent in RNase A mapping and chemical cleavage, by not requiring the mismatched base to be available for cleavage.

It remains to be seen if these newer methods live up to expectations and detect all possible mutations in collagen genes.

Time limitations and an awareness that RNase A mapping detects only 60-70% of mutations led to the cessation of this line of work before more mutations were detected. It is hoped that DNA and RNA preparations produced during this study will be of use for other people ready to look for mutations causing non-lethal OI using a more sensitive method.

CHAPTER FIVE

CONSTRUCTION OF A cDNA LIBRARY POTENTIALLY ENRICHED WITH COLLAGEN SEQUENCES USING THE POLYMERASE CHAIN REACTION

5.1   INTRODUCTION

To date, 15 different collagens have been characterized. They are coded for by over 20 genes. Many of the collagen proteins were isolated before the genes coding for them (Bentz *et al.*, 1983). However more recently collagen cDNA clones have been isolated before the corresponding proteins (Pihlajaniemi *et al.*, 1987; Gordon *et al.*, 1987; Pihlajaniemi *et al.*, 1990a). The proteins coded for by the most recently isolated clones have not yet been characterized.

The aim of this project was to use the polymerase chain reaction to create a cDNA library enriched with collagen coding sequences, with a view to isolating from it novel collagen cDNA sequences. It was hoped that the library would be enriched enough with collagen sequence containing clones to avoid screening by hybridization, and allow screening by random sequencing of clones. The library was to be produced using the polymerase chain reaction, exploiting the repetitive nature of the cDNA coding for collagens in designing the primer.

Many uses have been found for the polymerase chain reaction since it was described by Saiki *et al.* (1988), some with the isolation of specific cDNA clones or families of cDNA clones in mind. Lee *et al.* (1988) used the PCR to generate cDNA probes which were used to isolate a clone containing a full-length porcine urate oxidase cDNA. The primers used by Lee and coworkers were designed using the amino acid sequence of the porcine urate oxidase. A mixed population of primers was used to allow for codon degeneracy. Other methods of generating specific cDNA clones using PCR rely on at least one primer sequence being specific. Frohman *et al.* (1988) and Ohara *et al.* (1989) used a specific primer and a non-specific primer (oligo dT) to synthesize the 3' end or the 5' end of a known cDNA sequence.

The intention here was to generate cDNA clones for which no cDNA sequence had yet been determined, therefore no sequence-specific primer

103

was available from which the cDNA synthesis could be extended. The clones would code for proteins for which the amino acid sequence has not been determined, not even from another species. It was hoped that more than one novel collagen could be isolated from the library so the primer would have to be designed to amplify many collagen or collagen-like sequences. Therefore, the PCR primer sequence could not be determined from a single protein sequence.

A primer was designed which, it was hoped, would selectively amplify collagen or collagen-like sequences. The primer sequence was based on the DNA sequence coding for the collagen Gly-X-Y repeat (where X is often a proline and Y is often a hydroxyproline) therefore, it was predicted that collagen sequences within the library would contain mostly helical sequences. The clones would not be regarded as true representatives of collagen cDNA sequences because of the inherent mutation rate of the polymerase chain reaction (Tindall and Kunkel, 1988; Dunning *et al.*, 1988). The library would be regarded as a library of probes. It would be easier to isolate a new collagen by picking out novel collagen probes from a collagen sequence enriched library, then to use these probes to screen a conventional library under stringent conditions, than to screen a conventional library under low stringency conditions using a degenerate oligonucleotide probe. The clones were expected to be short. When given a choice of annealing sites PCR primers tend to amplify the smallest fragment (Harbarth and Vosberg, 1988).

5.2   DESIGNING THE PRIMER

Degenerate PCR primers were designed which were expected to select for collagen repetitive sequence. Too much sequence homology with known collagen sequences would lead to a restriction in the types of collagen clones in the library. Conversely, too little sequence homology would lead to only a small proportion of the clones in the library containing collagen sequence. Two populations of degenerate primers were synthesized (RANcDNA-1 and RANcDNA-2). RANcDNA-1 was designed first. This primer was not ideal as explained in sections 5.2.1 and 5.2.2. The lessons learned from attempts at producing a library using RANcDNA-1 were put to good use when the second primer (RANcDNA-2) was designed.

5.2.1   THE 3' END OF THE PRIMER

The sequence at this end of a PCR primer determines its specificity (Erlich, 1989). Intially this end of the primer was designed in the knowledge that the DNA sequence coding for a Gly-X-hydroxyproline repeat would be GGN NNN CCN, where N represents any base. It should be pointed out that hydroxyproline is coded for in DNA sequence as a proline which is hydroxylated as part of post-translational modification. This sequence (GGN NNN CC) was incorporated into the 3' end of primer RANcDNA-1 (Figure 5.1). This primer was able to amplify cDNA sequence, using the method described in section 5.3.1. The amplified cDNA sequence population collagen coding sequence. Southern blotting and hybridization showed that the collagen sequence within the PCR product was limited (results not shown). With this in mind a new, more collagen specific primer was designed. The cDNA sequences coding for the helical regions of the $\alpha 1(I)$, $\alpha 2(I)$ and $\alpha 1(III)$ collagens were lined up and any sequence motifs common to all three were noted. The sequences of these three helical regions were chosen because they are long and uninterrupted and so were simple to line up. The Lineup programme from the Genetics Computer Group program suite (Kanehisa *et al.*, 1984) aligns the required sequences then displays the bases common to two or three of the sequences at each position, so producing a consensus sequence. The most common sequence found within this consensus sequence was :-

5'   3'

G G N N N C C N G G

(The primers were synthesized so that the positions denoted as N above were an equal ratio of G, A, T and C bases)

The sequence of RANcDNA-2 is that of RANcDNA-1 with an extra 3 bases at the 3' end (NGG). It was found 29 times in the consensus helical sequence, therefore it occurred at least 29 times in each of the three collagens, at least 87 times in all three sequences. It occurred in either strand, therefore it could be used as both the forward and reverse primer (Figure 5.2). The sequence occurred throughout the helical consensus sequence. A sequence was picked at random from the EMBL database (accession number = J02769, sequence for human 4F2 antigen heavy chain mRNA complete cds), this sequence was scanned for the occurrence of the putative primer

Figure 5.1

The sequence of RANcDNA-1

The sequence of the first primer used to construct the library (RANcDNA-1). The primer sequence included a *NotI* restriction endonuclease site and an *SstII* site. The positions of the sequences coding for glycine and proline (hydroxyproline) are shown.
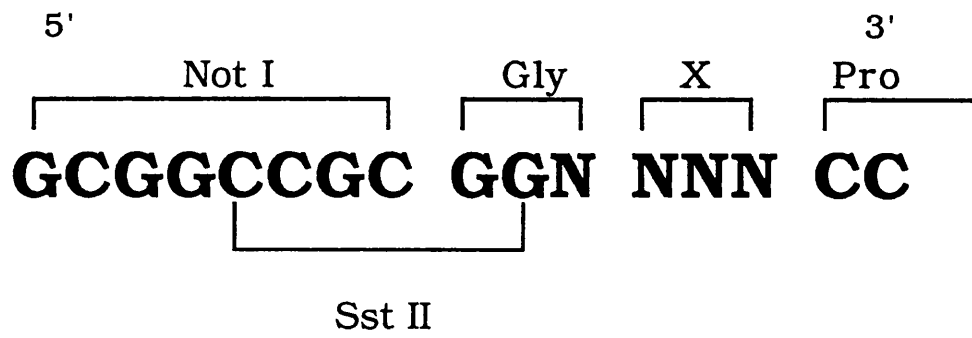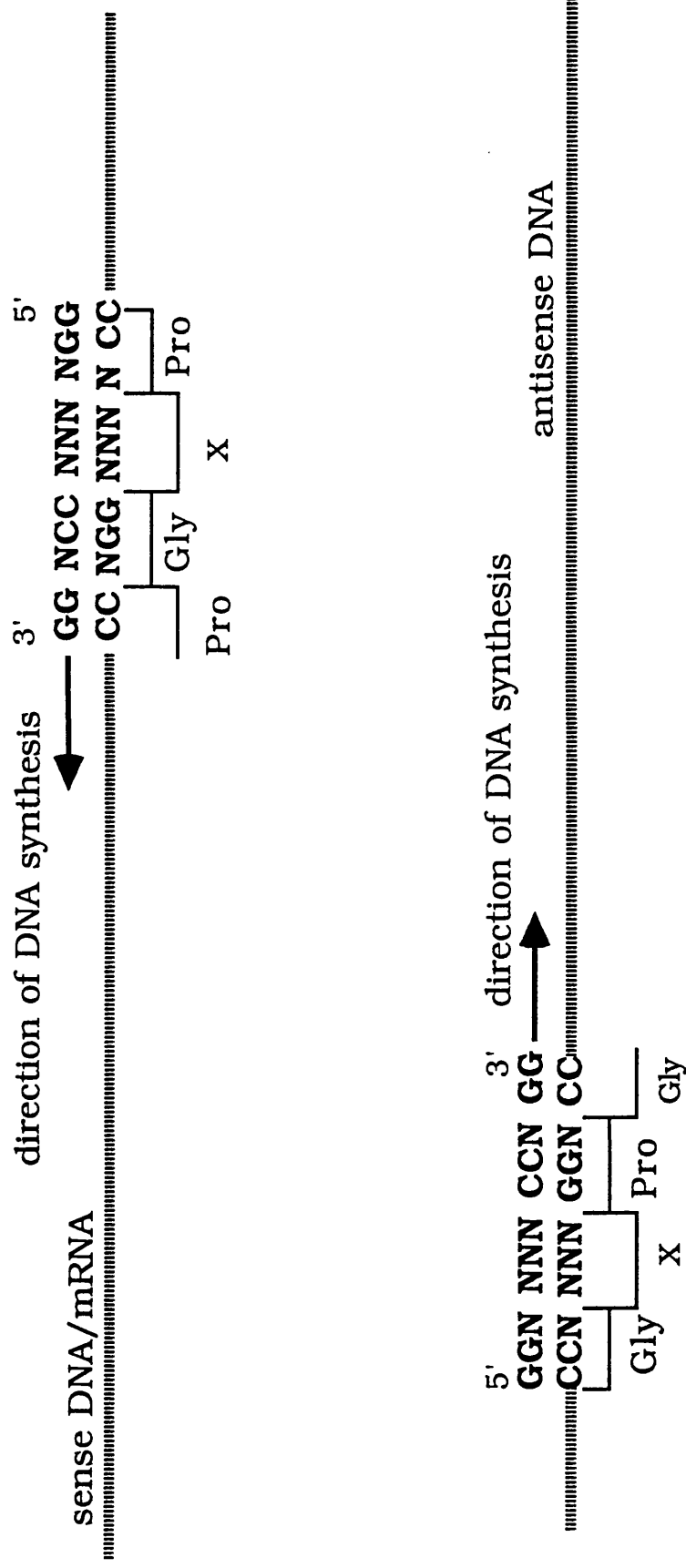
Figure 5.1

Figure 5.2

Primer annealing

A map to show how the 3' end of the primer RAN-cDNA-2 should be able to anneal and prime DNA synthesis from either strand of DNA template, in either direction.

The 5' sequence of the primer including the *EcoRI* restriction site is not shown in this figure because it is not expected to anneal to the template sequence.

Figure 5.2

sequence. The putative primer sequence occurred 5 times. It was predicted that the primer would selectively anneal to collagen helical coding DNA.

5.2.2   THE 5' END OF THE PRIMER

This end of the primer was designed with ease of cloning in mind. Previous workers had found that blunt end ligation of PCR products into vectors is very inefficient. Clark (1988) suggested that this is due to an extra adenine being attached to each end of the PCR product DNA. In general, ligation using DNA with 'sticky ends' is more efficient than if the DNA has blunt ends (Revie *et al.*, 1988).

The primer RANcDNA-1 was synthesized with a *NotI* restriction endonuclease recognition site incorporated into its 5' end (Figure 5.1). A *NotI* site was chosen because it has an 8bp recognition sequence, and as such is infrequently found in DNA sequence. Digestion of any PCR product, produced using a primer with a *NotI* site at its 5' end, with *NotI*, should yield sticky ended products that are rarely cut internally. The *NotI* recognition sequence had the added advantage that it was composed of G and C bases and therefore bore some similarity to collagen sequences. The *NotI* recognition sequence is G C G G C C G C.

The RANcDNA-1 primer was used in a PCR to amplify single stranded cDNA. The cDNA was synthesized as described in section 2.19.1. The PCR was carried out as described in section 5.4.1. The PCR product was electrophoresed through a 0.8% agarose gel containing ethidium bromide. The DNA in the gel was seen to be a smear. The molecular weight of the DNA was up to 6kb (Results not shown). An attempt was made to clone this PCR product. It was digested with the restriction endonuclease *NotI* and ligated with pBluescript SK⁻ which had also been digested with *NotI*. Unfortunately the PCR product was very inefficiently cut by *NotI*. This was probably due to the position of the *NotI* recognition site within the primer, it occupied the last 8 bases of the primer. Other workers have shown subsequently that some restriction endonucleases, including *NotI*, will not cut PCR products if their recognition site is at the very end of the DNA fragment (Jung *et al.*, 1990). There was a *SstII* site included in RANcDNA-1 (Figure 5.1). An attempt was made to digest the PCR product with *SstII*, also without success. The *SstII* site was 4 bases from the 5' end of the primer. *SstII* efficiently cut other PCR products when the *SstII* site was 10 bases from

the 5' end of the PCR primer (see Chapter four).

It was suggested that the restriction endonuclease, *SstII*, was not digesting the PCR product because residual *Taq* DNA polymerase was remaining attached to the ends of the PCR product and preventing the endonuclease from interacting with the DNA. To overcome this problem the PCR solution containing the amplified cDNA was treated with proteinase K. It was extracted with phenol then chloroform as described in section 2.5.1. The DNA was ethanol precipitated as described in section 2.5.2. The pellet was dissolved in 272µl of distilled water to which 20µl of 2.5M NaCl and 20µl of 10% SDS were added with 10µl of proteinase K (10mg/ml). The digest mix was incubated at 37°C for 30 minutes. After phenol extraction and ethanol precipitation the DNA was resuspended in digest buffer and incubated with *SstII* in the presence of spermidine. This proteinase K treatment did not enable the *SstII* to digest the PCR product. Proteinase K digestion of PCR product as a way of enhancing restriction endonuclease digestion has since been described in Crowe *et al.* 1991.

Jeffreys *et al.* (1990) have shown that digestion of PCR products by *EcoRI* was efficient, if the *EcoRI* site was incorporated into the primer with an 8 base 5' extension:-

```
5'              EcoRI   3'
T C A C C G G T G A A T T C
```

It was decided that cloning of the PCR product might be easier if this sequence was incorporated into the 5' end of the degenerate primer. Internal cleavage of PCR products with *EcoRI* should not be a problem since analysis of the cDNA sequences coding for the helical domain of α1(I), α2(I) and α1(III) collagens revealed very few *EcoRI* sites.

The final sequence of primer RAN-cDNA-2 was :-

```
5'                              3'
T C A C C G G T G A A T T C G G N N N N C C N G G
          EcoRI      Gly   X   Pro  Gly
```

Theoretically, this primer will anneal to itself and form dimers which will amplify (Figure 5.3). If self-priming did occur it may inhibit PCR amplification of the cDNA template, by occurring in preference to the required amplification. A control was set up to test for self-priming. The control PCR was set up as described in section 2.15.1, in the absence of template. 30 cycles of 1 minute at 95°C, 1 minute at 55°C and 10 minutes at 70°C were performed in a Techne Programmable Dri-block. One tenth

Figure 5.3

Primer self-annealing

A diagram to show how the primer RAN-cDNA-2 may be able to anneal to and prime DNA synthesis from itself. Such DNA synthesis would result in many DNA fragments which would be 37bp long. It may also lead to inhibition of the amplification of larger DNA fragments which use cDNA as a template.

Figure 5.3

5'

TCACCGGGTGAATTCGGNNNNCCNGG⟶ 3'

3' ⟵GGNCCNNNNGGCTTAAGTGGCCACT 5'

⟶ = DIRECTION OF DNA SYNTHESIS
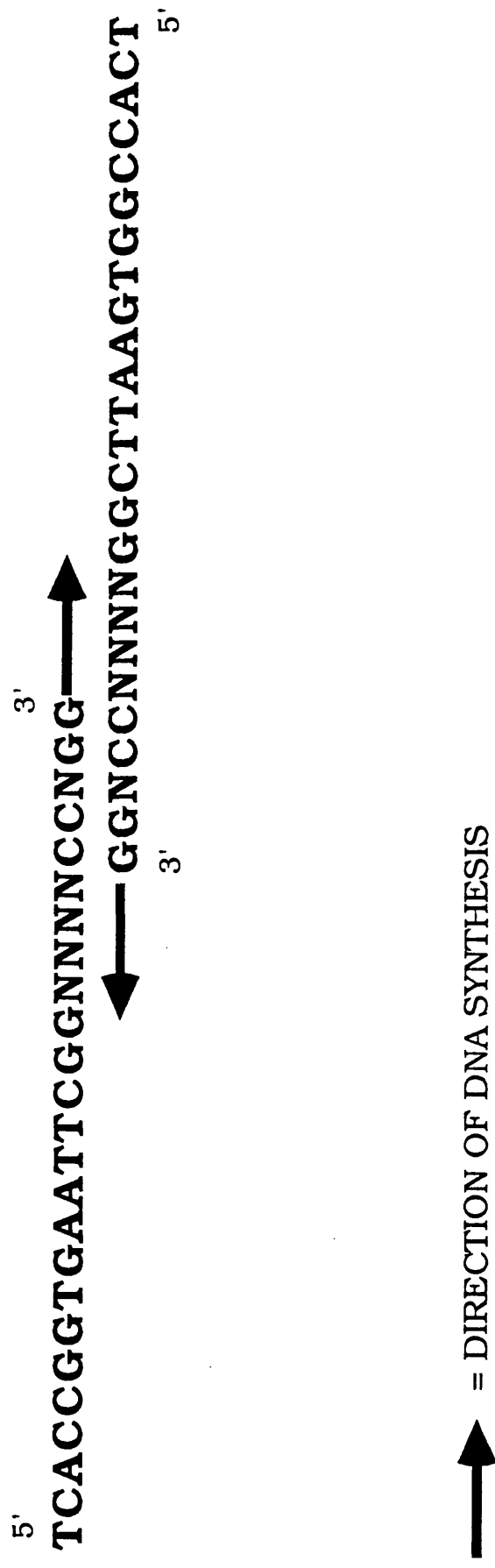
volume of the PCR was electrophoresed through a 0.8% agarose gel. No amplified DNA was seen in the gel (results not shown). This result does not mean that the primer was not annealing to itself and priming. However, if the primer was priming off itself it was not producing fragments large enough to be seen on an agarose gel. From this it was concluded that when template is included in a PCR using RAN-cDNA-2 as the primer, any DNA synthesized which was large enough to be seen on an agarose gel was not the result of the primer priming off itself, but of the primer priming off template DNA.

5.3 AMPLIFICATION OF cDNA USING THE PRIMER RANcDNA-2

5.3.1 AMPLIFICATION OF cDNA USING THE PRIMER RAN-cDNA-2, SINGLE STRANDED cDNA AND THE METHOD DESCRIBED IN SECTION 2.15.1

The primer RANcDNA-2 was used in a PCR with single stranded cDNA (sscDNA) as the template. The sscDNA was synthesized from poly(A$^+$) RNA isolated from human chorion membrane. It was synthesized as described in section 2.19.1.

1μl of this sscDNA (1ng) was added to a PCR with 100 pmoles of RANcDNA-2 primer. The PCR conditions are described in section 2.15.1. 30 temperature cycles of 95°C for 1 minute, 55°C for 1 minute, and 70°C for 10 minutes were run. It was thought that as a general rule 1kb of DNA sequence is amplified per minute of elongation time (Innis *et al.*, 1988). However recently, Rychlik *et al.* (1990) have disputed this. They suggested a figure in the region of 300-500 base pairs per minute. In either case a 10 minute elongation step was considered excess to requirement, since the longest fragment that was expected to be amplified was approximately 3.5kb (the length of the cDNA sequence that codes for the fibrillar collagen helical domain). The very long elongation time was to make sure that all the DNA fragments had time to be synthesized to completion. Saiki *et al.* (1988) showed that under the reaction conditions they used, increased elongation time resulted in a decrease in primer specificity. However, the annealing temperature they used was 40°C. When they later increased the annealing temperature to 55°C primer specificity was resumed.

One tenth of the PCR product was electrophoresed through a 0.8%

agarose gel. The PCR product was seen on the gel as a smear (result not shown). This was the result which was expected, the products of the reaction were not size selected.

5.3.2 AMPLIFICATION OF cDNA USING THE PRIMER RNA-cDNA-2 USING THE METHOD DESCRIBED IN SECTION 2.15.2

Synthesizing sscDNA by the method described in section 2.19.1 is laborious, even more so when the steps to isolate poly(A$^+$) RNA are included. The method also requires a relatively large number of cells from which the poly(A$^+$) RNA can be isolated. It was for these reasons that the method described in section 2.15.2 was tried. The method was carried out using the primer RAN-cDNA-2 and total cellular RNA extracted from fibroblasts. RNA extracted from fibroblasts was used because it was easily obtained. The PCR method described in section 2.15.2 enables PCR amplification of mRNA via a cDNA intermediate to take place in one tube. The method is much simpler than the previous method, does not involve poly(A$^+$) isolation and only needs a small quantity of RNA to be successful. The 100μl reaction mix was put through 30 cycles of 95°C for 1 minute, 55°C for 1 minute, and 70°C for 10 minutes. One tenth volume of the PCR mix was electrophoresed through a 0.8% agarose gel (Figure 5.4A). The gel was Southern blotted and the filter was hybridized with the radiolabelled insert from the α2(I) clone, Hp2010. The PCR product was up to 2kb in length and hybridized to the α2(I) probe very strongly (Figure 5.4B). The autoradiograph in Figure 5.4B was the result after 5 hours exposure. Similar filters were hybridized with the type V collagen probe, Hf511 (Weil *et al.*, 1987) and the type VI collagen probe, p24 (Chu *et al.*, 1987). Figure 5.5 shows the resulting autoradiographs, a strong signal was the result of type VI hybridization (Figure 5.5B), a faint signal was the result of type V hybridization (Figure 5.5A). The primer was able to amplify mRNA via a cDNA intermediate using this quicker and easier method. The hybridization experiments clearly show that cDNA coding for α2(I), α3(VI) and to a lesser extent α2(V) collagen, were contained within the PCR product.

Figure 5.4

Amplification of mRNA via a cDNA intermediate using RANcDNA-2 as the primer

A

  The product of a PCR amplification using RAN-cDNA-2 as the primer and fibroblast mRNA as the template.

From left to right the lanes on the gel were loaded with:-

One tenth volume of PCR product.
The PCR was carried out using the primer RAN-cDNA-2. The template was total cellular RNA isolated from from human skin fibroblasts and the method used was that described in section 2.15.2
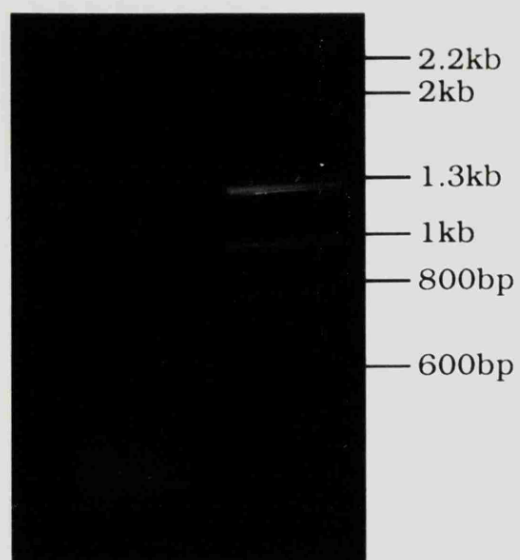
λ/*HindIII* and φX174/ *HaeIII* markers

B

  The results of Southern blot and hybridization analysis of the gel described above. The filter was hybridized with radiolabelled α2(I) collagen cDNA sequence (Hp2010 insert).

Figure 5.4

A

— 2.2kb
— 2kb

— 1.3kb

— 1kb
— 800bp

— 600bp

B

— 4.4kb

— 2.2kb
— 2kb

— 1.3kb
— 1kb
— 800 bp
— 600bp
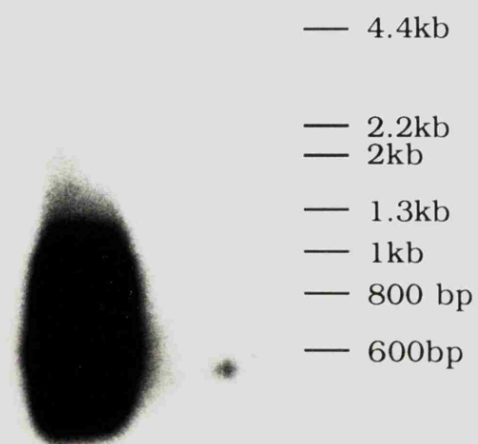
Figure 5.5

<u>The polymerase chain reaction product contains sequences which hybridize
to type V and type VI collagen probes.</u>

Two one tenth volumes of the PCR product produced using
RANcDNA-2 as the primer and fibroblast mRNA as the template were
electrophoresed through a 0.8% agarose gel. The gel was Southern blotted
and the resulting filters were hybridized with the probes described below

<u>A</u>   the α2(V) clone Hf511 (Weil *et al.*, 1987)

<u>B</u>   the α3(VI) clone p24 (Chu *et al.*, 1987)

Figure 5.5

A

2kb —

1.3kb —
1kb —
800 bp —

600 bp —

B

2.2kb —

1.3kb —

1kb —
800 bp —

600 bp —

5.4 *Sau96I* DIGESTION OF THE PCR PRODUCT

To determine how much of this PCR product was collagen or collagen-like sequence, it was digested with the restriction endonuclease *Sau96I*. The recognition site for *Sau96I* is GGNCC. This sequence is frequently found in cDNA coding for the collagen helix. When such DNA is digested with *Sau96I*, short DNA fragments which are multiples of 9bp in length are produced (Gordon *et al.*, 1987).

An aliquot of the PCR product was cleaned by phenol extraction and ethanol precipitated. Half of the cleaned PCR product was digested with *Sau96I*. This digested DNA was electrophoresed through a 0.8% agarose gel alongside the cleaned undigested PCR product.

A slight decrease in DNA concentration was observed after *Sau96I* digestion (Figure 5.6A), this result was disappointing. It was predicted that, a high collagen cDNA content in the PCR product would be accompanied by a large decrease in the DNA concentration after *Sau96I* digestion. That the *Sau96I* had digested most of the α2(I) collagen coding sequence present in the PCR product was clearly seen when the gel was Southern blotted and hybridized with an α2(I) cDNA probe. (Figure 5.6B). Very little of the collagen sequence which hybridized to the probe in the undigested lane was present in the digested DNA lane. The band at the bottom of the digested DNA lane on the autoradiograph probably represents the probe hybridizing to small collagen fragments.

5.5 CLONING THE PCR PRODUCT

The bacteriophage lambda is often used as the vector for libraries because the efficiency of packaging λ particles and *E. coli* infection is greater than bacterial transformation. This increased efficiency results in a larger number of clones in the library and so rare sequences are included. Lambda clones are more difficult to work with than plasmid clones. There should be no need to clone this PCR DNA into λ in order that rare collagen sequences should be included in the library. If the primer selects for collagen sequences, the copy number of rare collagen sequences will be increased during the PCR step.

A PCR was carried out as described in 5.3.2, the product was phenol extracted and ethanol precipitated, the pellet (3.5μg of DNA) was

Figure 5.6

Sau96I digestion of the PCR product

A

A photograph showing one tenth volume of PCR product before and after Sau96I digestion. The PCR product was amplified fibroblast cDNA sequence, using the primer RANcDNA-2.

Lane 1   one tenth volume of PCR product digested with Sau96I

Lane 2   one tenth volume of PCR product

Lane 3   λ/HindIII and φX174/HaeIII markers

B

Southern blot and hybridization analysis of the gel photographed above, showing loss of collagen sequence after digestion with Sau96I.
The filter was hybridized with an α2(I) cDNA probe (Hp2010 insert).

Lane 1 - PCR product digested with Sau96I

Lane 2 - undigested PCR product
(The marker lane was not included in the photograph.)

Figure 5.6

A
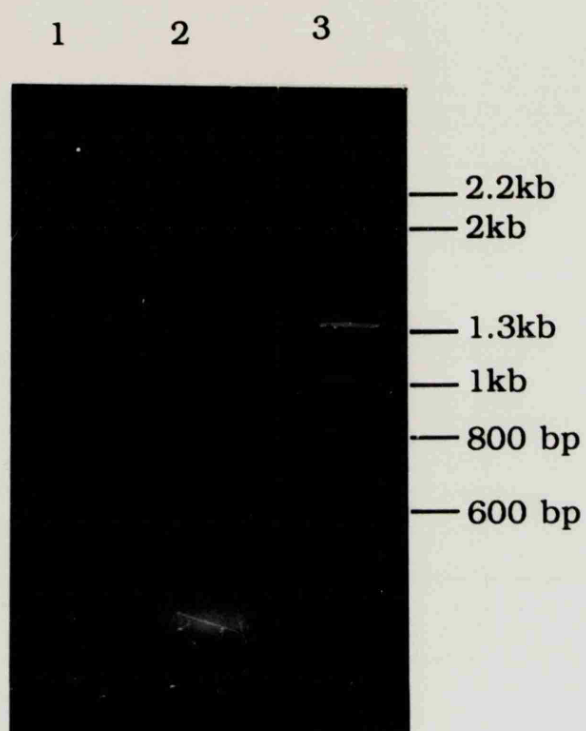
B
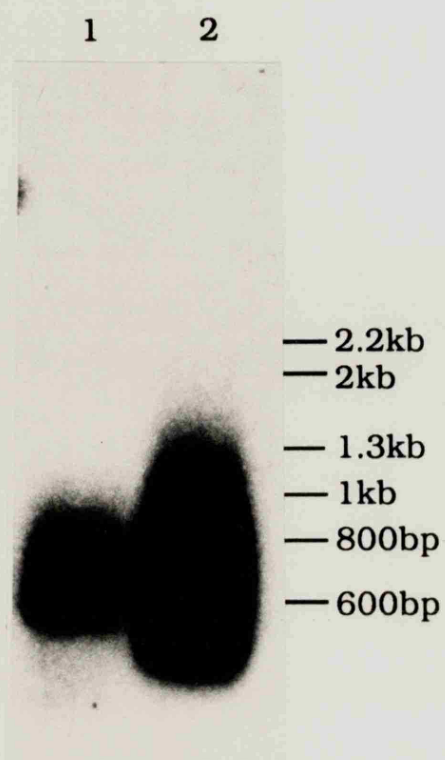
dissolved in 35μl of distilled water. The cleaned PCR product was to be digested with *EcoRI* and ligated to pBluescript SK⁻, which had also been digested with *EcoRI*. The vector DNA was treated as described in section 2.9 and 2.10 to prepare it for ligation. The consequence of digestion of the PCR product with *EcoRI* would be 'sticky ended' PCR product and small DNA fragments cleaved from each end of the PCR product. It was predicted that these small DNA fragments may inhibit the ligation reaction. Other workers have found that small molecular weight molecules (presumably dNTPs from the PCR mix) can inhibit ligation reactions (Scharf *et al.*, 1986). Inhibition of the ligation reaction was predictable in this case because the small DNA fragments will have *EcoRI* sticky ends. They may religate to the digested PCR product or ligate to the digested vector, thus rendering the sticky ends of the DNA fragments blunt. To remove any small molecular weight molecules an aliquot of digested PCR product was passed through a spin-dialysis column as described in section 2.18.3. Dubose and Hartl (1990) found that passing PCR product through a spin-dialysis column removed primer DNA. Indeed, before the PCR product was passed through the column an aliquot of φX174/*HaeII* markers was passed through such a column. DNA fragments up to 118bp in length were removed from the DNA solution by this treatment.

Of the 35μl (3.5μg) of cleaned PCR product:-
10μl were digested with *EcoRI* then heated at 65°C for 10 minutes to denature the enzyme.
10μl were digested with *EcoRI* then passed through a spin dialysis column (section 2.18.3), after being made up to 25μl with sterile water.

      The digested PCR products were each ligated to 50ng of pBluescript SK⁻ which had been digested with *EcoRI*, treated with alkaline phosphatase and prepared as described in section 2.10.

The ligation mixes consisted of:-

Ligation S
50ng pBluescript SK⁻/*EcoRI* (phosphatased)
1μg *EcoRI* digested and spin-dialysed PCR product
20μl of 5 x ligase buffer (section 2.9)
1μl T4 DNA ligase (1 unit)
sterile water to 100μl

111

Ligation U

50ng pBluescript SK⁻/*EcoRI* (phosphatased)

1μg *EcoRI* digested PCR product

20μl 5 x ligase buffer

1μl T4 DNA ligase

sterile water to 100μl

Each ligation mix was left overnight at room temperature then cleaned by phenol extraction and ethanol precipitation. Each DNA pellet was dissolved in 10μl of sterile water.

1μl of each ligation mix was transformed by electroporation into *E. coli* XL-1 blue (section 2.11.4). Aliquots of the transformed bacteria were spread onto Luria agar plates supplemented with tetracycline, amplicillin, X-gal and IPTG and left to grow overnight at 37°C.

Blue and white colonies were the result of both transformations.

The efficiency of transformation of the two ligations were as follows:-

Ligation S = 340 colonies per μg insert DNA

Ligation U = 2.4 x 10^3 colonies per μg insert DNA

A vast excess of insert to vector DNA was used as stipulated in Revie *et al.* (1988). These were disappointingly low transformation rates compared to the theoretical rates published in Dower *et al.* (1988). The transformation rate of U was increased later by halving the amount of DNA in the electroporation. The new transformation rate was 3.9 x 10^5 colonies per μg of insert DNA, but this was still low. The lower transformation rate obtained when the insert was cleaned by spin-dialysis was unexpected, since the digested PCR product used in ligation U was not treated in any way to remove the primers or any other small molecular weight DNA. It was expected that these small DNAs would interfere with the ligation reaction therefore reducing the ligation and thus the transformation efficiency of U.

The transformation rate using the spin-dialysed PCR product may have been much less efficient than the second transformation, U, but the white:blue colony ratio was better with ligation S. However, the better blue:white colony ratio was not good enough to make up for the low transformation rate.

BLUE : WHITE COLONY RATIO

Ligation U	1 : 1.8
Ligation S	1.2 : 1

5.6   <u>ANALYSIS OF THE INSERTS OF SOME OF THE CLONES</u>

12 white colonies were picked from each transformation experiment. The colonies were grown overnight and mini preparations of the plasmids were isolated as described in section 2.13.2. 2.5µl of each plasmid preparation were digested with *EcoRI* to release the insert. The digested DNA was electrophoresed through a 1.5% agarose gel dissolved in TBE buffer. The gel photograph (Figure 5.7) shows that all the plasmids isolated contained inserts. On the whole the inserts from ligation U were larger than those from ligation S. The gel was Southern blotted. The filter was hybridized with a radiolabelled α2(I) cDNA probe (HP2010 insert) under the conditions described in section 2.12.4. Autoradiography of the hybridized filter revealed a clone insert which hybridized to the probe (clone U10, Figure 5.8). This clone was one of the few containing a double insert. It was grown overnight in 100ml of Luria broth containing antibiotic. The resulting culture was used as the substrate for a DNA preparation as described in section 2.18.1. 15µg of isolated plasmid DNA were alkali-denatured and spin-dialysed as described in section 2.18.2 and 2.18.3. 8.5µl (5.1µg) aliquots of the denatured DNA were sequenced (section 2.18.4), using the M13 reverse or forward primers. The clone had to be sequenced in both directions because there was no way of telling which direction would reveal the collagen sequence. Only one insert gave a positive signal upon hybridization with the α2(I) clone, as shown in (Figure 5.8).

One of the inserts in clone U10 was found to have a sequence homologous to bases 2568 to approximately 2768 of the cDNA coding for α1(I) collagen (Figure 5.9). The α1(I) collagen cDNA map shown in Figure 5.9 was compiled from sequence information published in Tromp *et al.* (1988) and Bernard *et al.* (1983a).

This result illustrates that it is possible to amplify collagen sequence cDNA using total cellular RNA isolated from fibroblast cells, the primer RAN-cDNA-2 and the method described in section 2.15.3. It also shows that this amplified material is amenable to cloning.

Figure 5.7

The size distribution of the inserts from some of the clones produced when ligations U or S were transformed into *E.coli.*

A

A photograph of an agarose gel which had been stained with ethidium bromide. It shows the presence and size distribution of inserts from clones produced using ligation U.

Lane 1 - λ/*HindIII* and φX174/*HaeIII* markers

Lanes 2-13 - DNA from 12 white colonies from transformation U, each digested with *EcoRI* to release its insert(s).

Lane 14 - DNA from one white colony from transformation S which had been digested with *EcoRI.*

B

A photograph of an agarose gel which had been stained with ethidium bromide it shows the presence and size distribution of inserts from clones produced using the ligation S (spin-dialysed).

Lane 1 - λ/HindIII and φX174/HaeIII markers.

Lane 2-12 - DNA from 11 white colonies digested with *EcoRI* to release their insert(s)

Figure 5.7
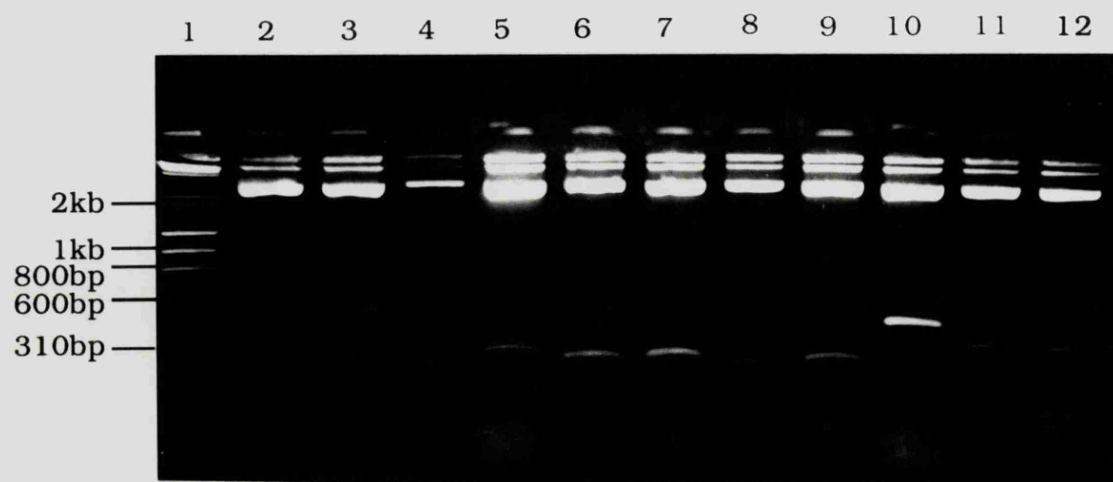
A
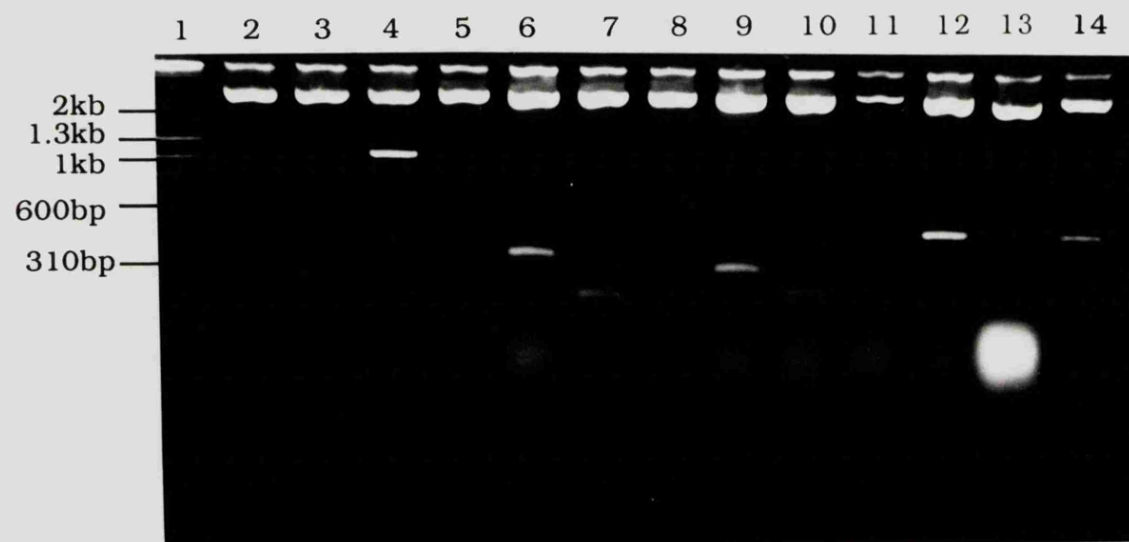
B

Figure 5.8

<u>One of the clones screened, contained sequence which hybridized to a
collagen probe.</u>

Southern blot and hybridization analysis of the gels shown in
Figure 5.7A and 5.7B. The numbering of the signals on this autoradiograph
is slightly different from that of the gel lanes in Figure 5.7. The marker lanes
have been removed from this autoradiograph and the first plasmid lane has
been designated number 1.

The autoradiograph shows only one insert (from clone U10, in lane 10 of the
autoradiograph and lane 11 of the photograph in Figure 5.7A) which
hybridized to the α2(I) cDNA probe (Hp2010).

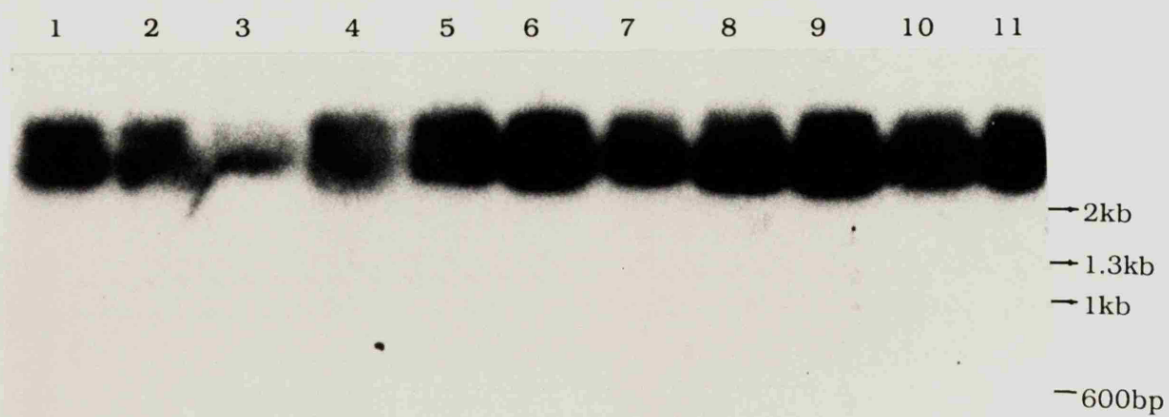V = vector DNA which hybridized to the probe.

Figure 5.8

A
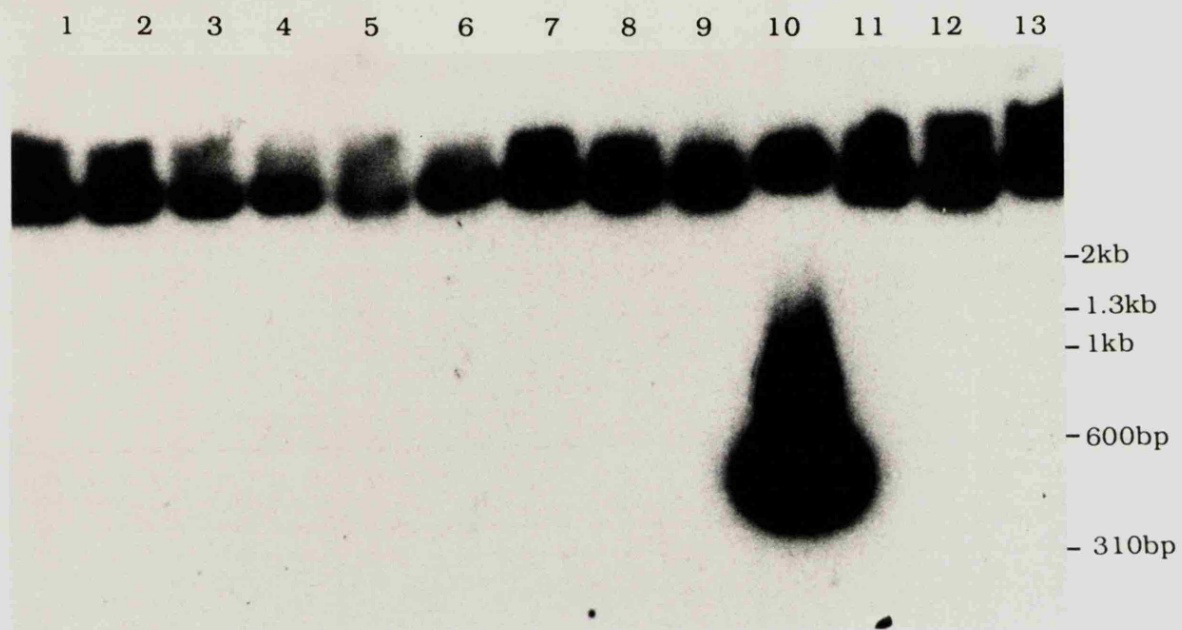
B

Figure 5.9

The positions of some of the collagen clones isolated from the library with respect to the full-length cDNAs coding for α1(I) and α2(I) collagens.

S = signal peptide sequence

T = telopeptide sequences

P = propeptide sequences

Following usual protocol the sequence coding for the amino terminus of the protein is on the left of the map and the sequence coding for the carboxyl portion of the protein is represented by the right of the map.

The nucleotides are numbered from the start of transcription (Chu *et al.*, 1985; Dickson *et al.*, 1985)

Clone 6 contained bases 86-386 of the sequence of the α3(VI) clone p24 published in Chu *et al.* (1988).

Note:- the sizes of the inserts are estimates based on the sizes determined by gel electrophoresis. Approximately 180 bases of each insert were sequenced to identify the sequence.

The position of the clone Hf404 (Chu *et al.*, 1982) is shown on the map, this clone was used as part of the probe cocktail.
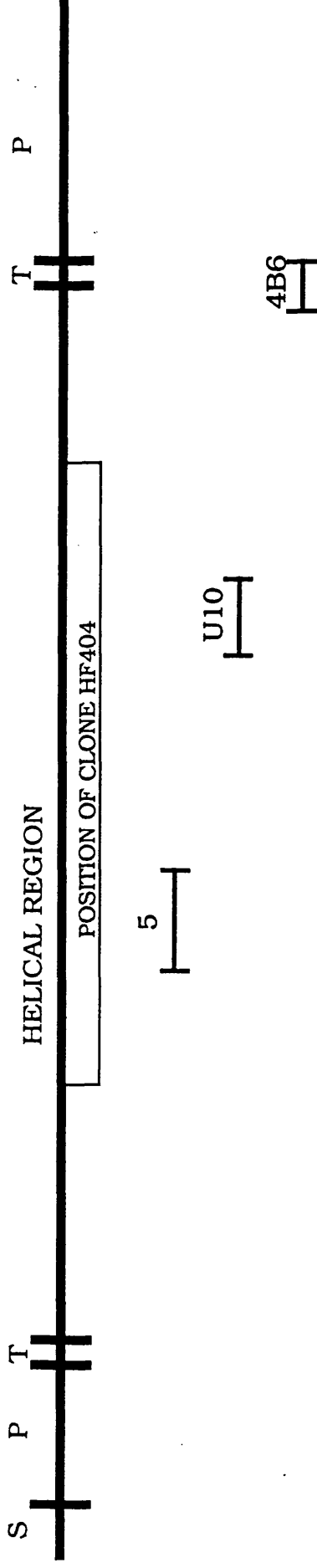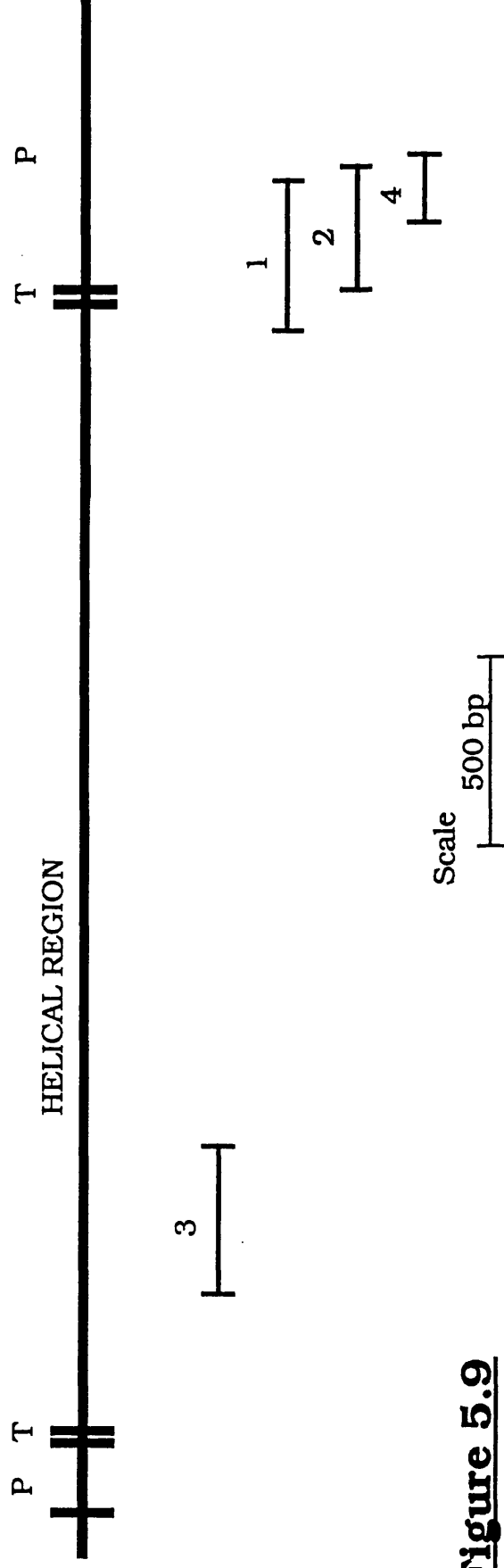
Figure 5.9

5.7   ANALYSIS OF CLONES FROM LIGATION U

Due to the fact that the inserts from the clones from ligation U were on the whole larger than those from the clones from ligation S and the transformation rate of ligation U was better than that of ligation S, clones from ligation U were subject to further analysis.

The remaining ligation mix U was diluted to half its former DNA concentration with water. 2 x 1μl of this ligation mix were transformed by electroporation into two 40μl aliquots of competent *E. coli* XL-1 blue. Samples of the transformed cells were spread onto Luria agar plates containing antibiotic, X-gal and IPTG. The cells were incubated for 24 hours at 37°C. 1056 white colonies were picked individually and used to inoculate separate 100μl aliquots of freezing medium contained in the wells of microtitre plates (described in section 2.20.1). After inoculation of the medium the plates were incubated for 2 hours at 37°C, before freezing at -80°C until required.

288 colonies were transferred onto duplicate pieces of nylon filter as described in section 2.20.1. After overnight incubation the filters were treated to fix the colony DNA to them (section 2.20.1).
The filters were hybridized with a cocktail of collagen clones consisting of 1ng of insert from each of the following cDNA clones:-

HF404	α1(I)(Tromp *et al.*, 1988)	
HP2010	α2(I)(Kuivaniemi *et al.*, 1988a)	
HF511	α2(V)	(Weil *et al.*, 1987)
p(III)-55	α1(III)	(Miskulin *et al.*, 1986)
p18	α1(VI)	(Chu *et al.*, 1987)
p1	α2(VI)	(Chu *et al.*, 1987)
p24	α3(VI)	(Chu *et al.*, 1987)

After hybridization the filters were washed in 3 x SSC, 0.1% SDS at 65°C. From duplicate plates of 288 colonies, 6 positives were seen above the background signal. The background signal was quite high, even when the filters were later washed with 0.1 x SSC, 0.1% SDS.

DNA preparations were made of the plasmids from the six positive colonies. When the plasmid DNA was digested with *EcoRI* to release the inserts all 6 clones were found to contain inserts, two contained double inserts (Figure 5.10). The DNA preparations were used to partially sequence the clones. The clones were sequenced using the M13 reverse sequencing primer. The two clones with double inserts were partially sequenced with
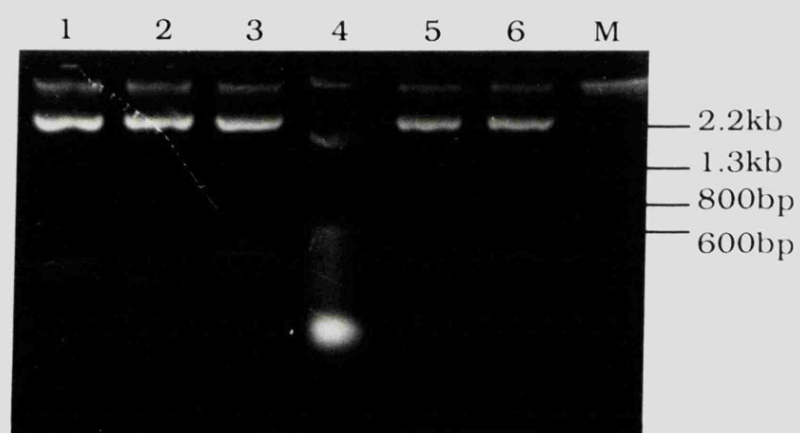
Figure 5.10

Six clones from the library which contain collagen coding sequence.

      EcoRI digested plasmid DNA isolated from 6 clones from the library. Hybridization analysis revealed that these clones all contained sequences complementary to the probe cocktail.
The photograph indicates the approximate size of the inserts.

Figure 5.10

both the M13 forward and reverse primers.

The sequence obtained for each insert was scanned through the EMBL database. All 6 clones were found to contain an insert with sequence which was homologous to one of the probes in the cocktail. The sequence obtained from each clone was cDNA, none of the clone contained intron material. Obviously there was no genomic DNA contamination in the RNA preparation. This supports the findings of Chomczynski and Saachi (1987), who found no contaminating DNA in RNA preparations isolated using their method.

4 clones contained α2(I) sequence, of which three overlapped.

one clone contained α1(I) sequence

one clone contained α3(VI) sequence

The positions of these clones (clones 1-5) in relation to their full-length cDNA sequences are shown in Figure 5.9.

From the 288 clones screened, 6 contained sequences that hybridized to the cocktail of probes (2% of clones). The next step was to look for clones that were not completely homologous to the DNA in the probe cocktail, but which cross-hybridize to it. It was thought that this would determine if the collagen clone content of the library was greater than 2% and may reveal the presence of a new collagen sequence. To detect the clones 1-5, 288 clones were screened as described in section 2.12.4. The hybridization took place at 65°C overnight. The filters were washed with 3 x SSC, 0.5 x SSC then with 0.1 x SSC all at 65°C. The filters were autoradiographed between each set of washings.

A further 384 clones were screened under the conditions described above. From the total of 672 clones screened, 18 clones (2.7%) were found to give a positive signal when hybridized to the probe cocktail under stringent hybridization conditions. Increasing the stringency of the washes from 3 x SSC to 0.5 x SSC resulted in the loss of a positive clone in the second batch of 384 clones (4B6, Figure 5.11). This clone satisfied the criteria for cross-hybridization, it hybridized to the probe until the conditions were made more stringent (increasing the stringency of the washes), then it did not hybridize to the probe. A DNA preparation was made of this clone. The clone was partially sequenced. The clone sequence was homologous to bases 3588-3737 of the α1(I) cDNA sequence (4B6, Figure 5.9). This clone although coding for α1(I) collagen was actually cross-hybridizing to the DNA in the probe cocktail. The α1(I) clone insert in the probe cocktail was from the clone HF404 which contains bases 1355-3197 of proα1(I) (Figure 5.9). The

Figure 5.11

A clone which hybridizes to a cocktail of collagen probes under low stringency conditions but not under high stringency conditions.

Autoradiographs of a library filter after hybridization with a cocktail of collagen probes and low stringency washing (3 X SSC (A)) or after high stringency washing (0.5 X SSC (B)).

The autoradiographs show a signal at position 4B6 (marked with an arrow) at low stringency which disappears after high stringency washing.
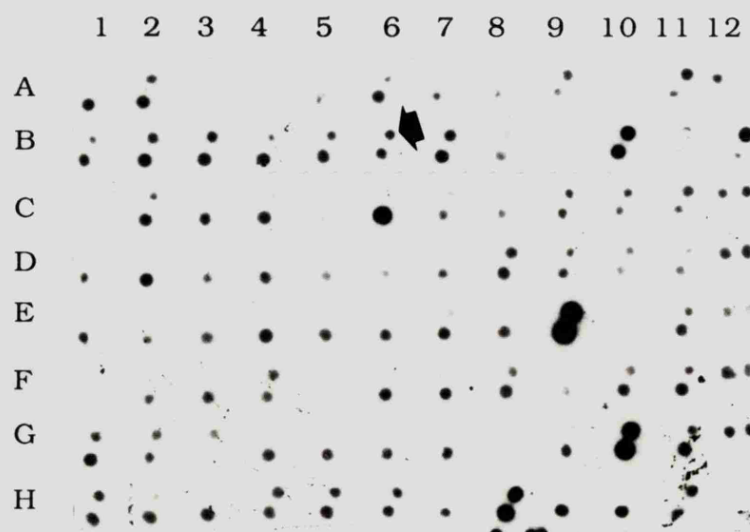
Duplicate transfers of colonies from the library were placed onto the same filter. One batch of colonies were transferred to the filter then the duplicate was placed on the filter, the second set of colonies being placed slightly up and to the right of the first set of colonies.

Figure 5.11

A
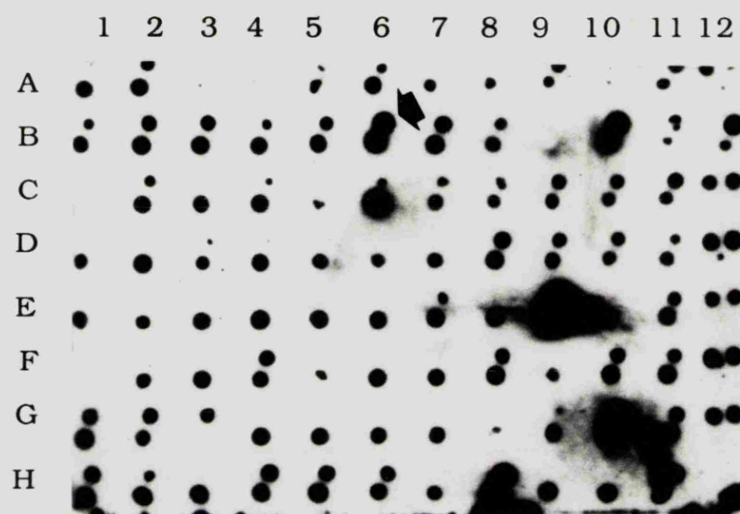
B

sequence of HF404 and that of clone 4B6 do not overlap therefore, cross-hybridization must have occurred between one of the probes in the cocktail and this clone, 4B6.

To encourage more cross-hybridization, in an effort to pull out of the library a novel collagen, the hybridization temperature was decreased to 60°C. Unfortunately this resulted in a increase in the background signal so that it was not possible to differentiate between a positive signal and the background signal. The autoradiograph in Figure 5.8 shows that the α2(I) probe hybridized to the pBluescript SK⁻ vector DNA. The same probe template was used in the probe cocktail which was hybridized to the library. Contamination of the probe template with vector DNA was probably causing the high background signal. 'Cleaner' template DNA was required.

A fragment of DNA synthesized during a PCR would not be contaminated with vector DNA, provided that vector DNA did not contaminate the PCR. A fragment of α2(I) mRNA was amplified via a cDNA intermediate as described in section 2.15.2. The primers used were COL1A2B and COL1A2C (Figure 5.12) and the template was total cellular RNA from human fibroblasts. A 100μl PCR was set up and run through 30 cycles of 95°C for 1 minute, 55°C for 1 minute and 70°C for 1 minute. 10μl of the PCR product were run through a 4% Nusieve agarose gel (result not shown). The expected 212bp fragment was the only product of the reaction. 25μl of the reaction mix were passed through a spin-dialysis column to remove any excess primers. 5μl of this cleaned DNA were used as the template for oligo-labelling (section 2.12.2).

The radiolabelled, PCR-synthesized, α2(I) cDNA probe was used to screen the second batch of 384 colonies at 60°C. These colonies had previously been screened with the probe cocktail at 65°C and 12 positive signals were obtained. After hybridization at 60°C with the PCR—synthesized probe the filters were washed in 3 x SSC, 0.1% SDS at 60°C (low stringency wash). 8 positive signals were detected by autoradiography. 7 of these positives had previously been detected using the probe cocktail, one hybridized to the PCR synthesized probe and not to the probe cocktail. These results probably indicate that this PCR-synthesized probe is not an ideal probe to be used to screen a library such as this. An ideal probe would be one which cross-hybridized to most, if not all, the collagen sequences present in the library. This probe, under the conditions described above, does not achieve this as shown by the fact that the PCR-probe does not hybridize to all the clones detected using the probe cocktail.

Figure 5.12

The sequences of the primers A2A and A2C.

The positions of the primer sequences within the α2(I) collagen cDNA sequence are also shown. The DNA fragment amplified using these primers and cDNA as a template was predicted to be 212bp long.

Figure 5.12
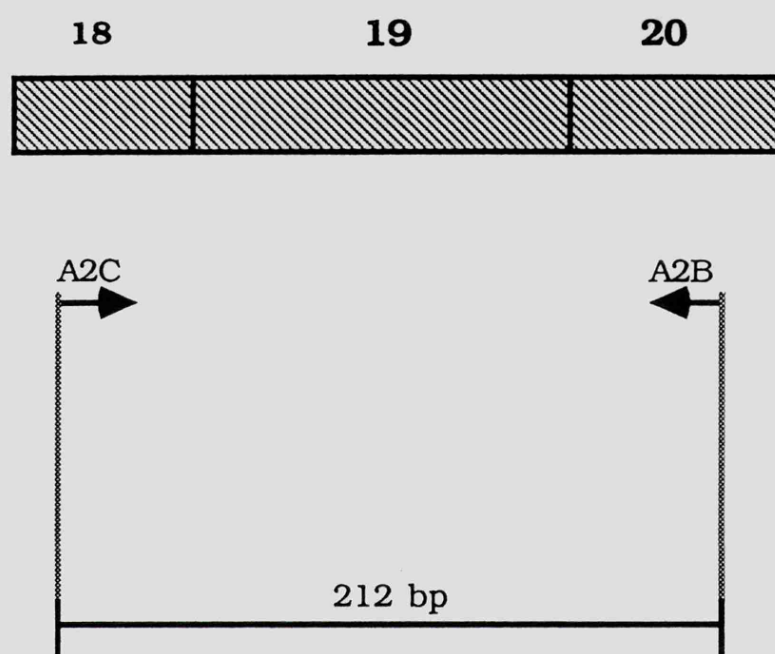
Primer A2B = TCACCGGTGAATTCGGGCTCACCCTTGTTACCGC

Primer A2C = GGTAATCCTGGAGCAACGG

The one clone that hybridized to the PCR-synthesized probe but not to the probe cocktail (clone 8A4) was partially sequenced. The sequence was scanned through the EMBL database and was found not to be homologous to any sequence in the database. It is unlikely that this clone contains cDNA coding for a new collagen because the list of the 20 sequences in the database with sequences which were closest to that of the clone did not include a collagen.

Hybridization of the library filters to the PCR synthesized probe at 57°C did not result in the isolation of any new positives from the library.

5.8 AMPLIFICATION OF COLLAGEN cDNA FROM mRNA ISOLATED FROM A LYMPHOBLAST CELL LINE.

Instead of trying to isolate collagen cDNA sequences using mRNA isolated from tissues such as placenta or fibroblasts it was considered a better idea to produce a library from mRNA isolated from other tissues. Many collagens have been isolated from placental or fibroblast libraries. Using RNA isolated from tissues other than the more conventional placenta or fibroblast as the starting material for a library would result in a library with a different collagen sequence population, and may increase the chances of isolating a novel collagen. This method of producing a cDNA library has the advantage over the more conventional methods in that it does not involve poly (A+) isolation nor any cleaning steps after cDNA synthesis. Both of these steps are time consuming and wasteful of starting material. Using this method to produce a library means that a much smaller amount of starting material is needed. This is useful when more unusual tissues are to be used as the starting material for the library, they may not be as freely available as placenta or fibroblast cells.

Another route to synthesize a cDNA library with a different population of collagen clones than that obtained when using fibroblast RNA as the starting material relies on the results published in Sarkar and Sommer, (1990). The authors published results to suggest that all cells produce at least a basal level of transcripts from all of their genes. It was predicted that RNA isolated from cells which do not synthesize collagen proteins would include transcripts coding for all the different types of collagens at an equal basal level. Amplification of such mRNA would result in an equivalent level of DNA in the PCR product coding for all the collagens.

If the primers selectively amplified collagen sequences then this equivalent level of DNA would be high. With this in mind total cellular RNA was isolated from a transformed lymphoblastoid cell line. The RAN-cDNA-2 primer was used with 1μg of the lymphoblast RNA in a PCR reaction as described in section 2.15.2. One tenth volume of the PCR reaction was electrophoresed through a 0.8% agarose gel.

A similar PCR product was obtained to that from the fibroblast RNA (Figure 5.13A). This gel was Southern blotted and the filter was hybridized with an α2(I) cDNA probe (Hp2010 insert). Some of the DNA in the lymphoblast PCR product hybridized to the α2(I) (Figure 5.13B). This indicated the presence of α2(I) collagen coding sequences in the PCR product even though lymphoblasts are known not to express type I collagen genes (Rossi and de Crombrugghe, 1987; Boast *et al.*, 1990). However, it should be considered that the process of transformation may have led to a change in the lymphoblasts that resulted in collagen gene expression.

5.9   <u>DISCUSSION</u>

The degenerate primer, when used as part of the polymerase chain reaction, was able to amplify mRNA via a cDNA intermediate. The high concentration of non-collagenous sequence contained within the library was disappointing. The high non-collagenous sequence content was first indicated by *Sau96I* digestion of the PCR product. The DNA concentration of the PCR product did not decrease substantially after *Sau96I* digestion. Of course digestion of a DNA sequence to small fragments by *Sau96I* is not an absolute prerequisite for the sequence to code for collgen. Clones coding for the carboxyl propeptide domain (clones 1, 2 and 3, Figure 5.9) will not be digested by *Sau96I*. However, the Southern blot and hybridization analysis of the *Sau96I* digested PCR product showed that most of the collagen sequence within the PCR product, which hybridized to the α2(I) probe, was digested to small molecular weight fragments by the enzyme. The lower than expected collagen sequence content of the PCR product was also indicated by the fact that only 2.7% of the clones in the library were found to be complementary to the cocktail of collagen probes. For the library to be used as it was intended, as library rich in collagen sequences which could be screened for the presence of new collagen sequences by random sequencing, the collagen sequence content would have to be improved.

Figure 5.13

Amplification of collagen sequence DNA using cDNA synthesized from lymphoblast mRNA as the template.

A

A photograph to show the DNA amplified when primer RANcDNA-2 is used in a PCR with cDNA synthesized from total cellular RNA isolated from a lymphoblast cell line or skin fibroblasts.

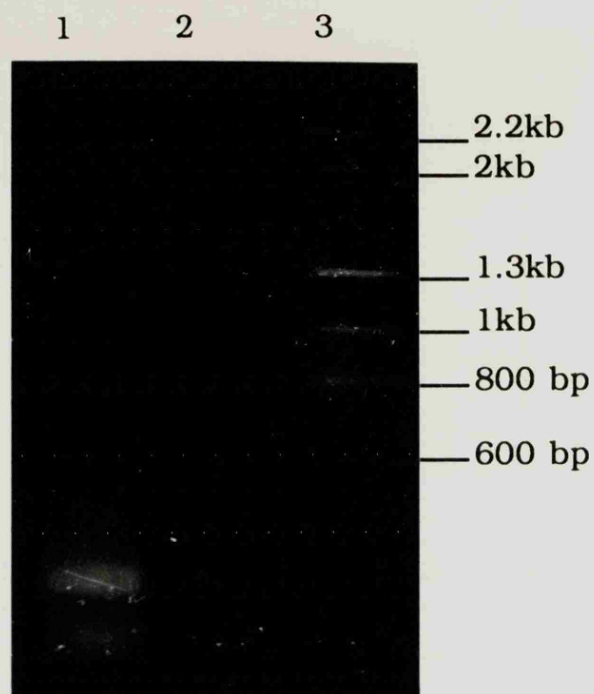Lane 1   One tenth volume of PCR product using skin fibroblast RNA as the template

Lane 2   One tenth volume of PCR product using transformed lymphoblastoid RNA as the template.
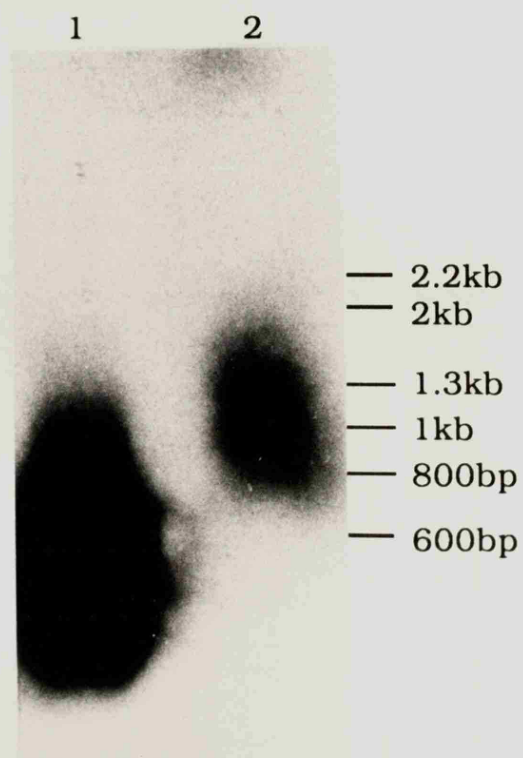
Lane 3   λ/*HindIII* and φX174/*HaeIII* markers

B

An autoradiograph showing the presence of DNA complementary to an α2(I) cDNA probe (Hp2010 insert) in PCR products described above.

Figure 5.13

A

B

There are a number of potential ways to increase the collagen enrichment of the library. One such way could be to increase the stringency of the PCR conditions ie. to increase the annealing temperature. Although an annealing temperature of 55°C is considered quite stringent, it is predicted that an increase in the stringency of the PCR conditions would still allow collagen sequence amplification. Analysis of some of the sequences to which the primers annealed, to synthesize the clones which were sequenced, revealed mismatched bases at the 3' end of the primer (Table 5.1). Lee *et al.* (1988) determined that, under their experimental conditions, up to 20% of the PCR primer may be mismatched and amplification will still take place. From the results in Table 5.1 it is clear that none of the clones were synthesized from completely homologous priming. This leads to the question, how much primer mismatching led to the production of the background sequences within the library population? From the sequence data collected so far this question cannot be answered. Of the seven non-collagenous sequence containing inserts which were sequenced, none contained sequence with enough homology to any sequence in the EMBL database to identify the insert, therefore the sequence which the primer annealed to could not be determined. If the stringency of the annealing step is increased, the collagen sequences amplified by way of annealing via mismatched primers may be lost along with the non—specific sequences. Only sequences which anneal to the whole of the collagen-like 3' end of the primer would be amplified. The loss of the collagen sequences amplified via mismatched primers should not be a problem. It has already been established that there are many sites within at least the type I and III collagen helical sequences which are capable of annealing to the whole of the 3' end of the primer sequence (excluding the EcoRI site and the 5' extension). Obviously more experimental work will have to be carried out into the ideal conditions for amplification of collagen cDNA sequence using this primer.

Another way of increasing the stringency of the annealing step is by the use of so called PCR enhancers such as TMAC (Hung *et al.*, 1990) or formamide (Sarkar *et al.*, 1990). Both these substances are reported to enhance PCR primer specificity.

A novel strategy which has been shown to increase the specificity of degenerate primers is described by Dietz *et al.* (1990). The authors used double stranded instead of single stranded primers. The complementary primers are said to keep the PCR primers from annealing to non-specific sites in the template DNA.

Table 5.1

A comparison of the sequences of the primers used in the PCR to synthesize 8 collagen sequence DNA fragments, and the sequences of the template cDNA they annealed to.

The sequence of each of the primers used to synthesize the clones is shown (lower case) with the sequence that the template sequence that the primers must have annealed to to synthesize the clones (upper case). The comparison of the two sets of sequences reveals that upto 36% of the primer bases can be mismatched and amplification still occurs.

Table 5.1

CLONE NUMBER	Sequence of primer used to create the clone (lower case) / Sequence which the primer annealed to to create the clone (upper case)	% mismatched bases
1	g t t c g c c t g g C U G U A G G A C C	36%
2	g g t g t g c c g g g C A A C C C G G G C C	18%
3	g g g c g c c c t g g C C C G U G G A C C C	27%
4	g a g t g c c g g g A A C C C G G C C	36%
5	g g c g a g c c t g g C C A G A C G G A C C	27%
6	g g g a a t c c c g g G A G A G A G G A C C	54%
U10	g g g a c c g g g g g C C C U G G U C C C C	9%
4B6	g g c c c g c c g g g C U A G U C G G G C C	36%

However any route which makes the PCR more collagen specific may reduce the ability of the primer to amplify novel collagens sequences, which is after all the aim of the project. A careful balance between increased collagen enrichment and decreased non-specific annealing will have to be reached.

Successful cloning of the PCR product very much depended on the position of the endonuclease site which was incorporated into the primer. The lack of digestion with the endonuclease *NotI* was probably due to the positioning of the *NotI* site at the 5' end of the primer. Digestion of the PCR product was achieved by incorporating an *EcoRI* site into the primer. Using *NotI* as the enzyme to create sticky ended PCR fragments had the advantage that *NotI* has an 8bp recognition site, it would therefore be unlikely to cut the PCR product internally. *EcoRI* has a 6bp recognition site however, it rarely cuts within the collagen helical sequences. No *EcoRI* site is found in cDNA coding for the helical region of α1(I) or α2(I), and only one is found in the cDNA coding for the helical region of α1(III). These sequences each contain one long stretch of helical coding sequence. It would be expected that sequences coding for collagens containing interrupted helical sequences may contain *EcoRI* sites more frequently. From the type I and type II sequence data it could be predicted that the *EcoRI* sites, if there are any, would be within the non-helical sequences. This problem is greatly outweighed by the knowledge that the recognition site is easily cut.

The rate of transformation of the ligated PCR product was disappointingly low, as was the proportion of clones containing inserts, 50%. The transformation efficiency may be the result of low ligation efficiency.

The low proportion of clones with inserts may be due to incomplete phosphatase activity. If the digested vector was not phosphatased effectively, it would be able to recircularize without including an insert. Even given that the alkaline phosphatase treatment was not 100% effective, the high insert concentration in the ligation reaction would be expected to increase the recombinant clone content above 50%.

Another reasons for the high recircularized vector content of the library could be, that circulating primer ends produced during *EcoRI* digestion of the PCR product could be ligating to the *EcoRI* digested vector. If blunt ends were produced this would allow the vector to ligate to itself in preference to the insert material. This is unlikely because the addition of primer ends to the vector would result in an insert 16bp in length. This would cause a frameshift in the β-galactosidase gene and therefore a white

colony. All of the white colonies analysed had inserts larger than 18bp.

The problem with the phosphatase and the possible problem with the digested primer ends may be overcome by employing the method described in Holton and Graham, (1991). This method is based on the fact that *Taq* polymerase can add a single non-template directed deoxyadenosine residue to the 3' end of a duplex PCR product (Clark *et al.*, 1988). Holton and Graham digested the vector, to which the PCR product was to be ligated, with *EcoRV* to create a linear molecule with blunt ends. They then used dideoxythymidine triphosphate (ddTTP) and terminal transferase to generate a single 3'-overhang T. A vector treated in this way will not ligate to itself but should ligate to insert material which possesses a 3' overhang adenosine. This method relies on the fact that all PCR molecules have a terminal extra adenosine. The authors do not give any indication as to the transformation rate they obtained using this method. Perhaps the method could be improved still further if half the vector population had this extra terminal adenosine and half was digested to create a blunt end and dephosphorylated. This would allow for a mixture of PCR product ends.

The low transformation rate may be caused by the presence of single stranded DNA in the PCR product which may be interfering with the ligation reaction. 30 PCR cycles may be too many. During the 30 cycles, each including 10 minutes elongation time, the *Taq* polymerase activity may decrease leading to the production of incomplete DNA fragments. Incomplete PCR products will not include both primer sequences so would not be available as template for further rounds of DNA replication. The addition of extra Taq polymerase might prevent the production of ssDNA in the PCR product by encouraging the complete synthesis of DNA fragments. A decrease in the elongation time of the PCR reaction might help to preserve the *Taq* polymerase activity until the end of the reaction. Even give that all these precautions are taken to discourage single stranded DNA synthesis, it may still persist. Olsen and Eckstein (1989) found that even when PCR product is seen on an agarose gel to be a single band, on an acrylamide gel it may be seen to be many bands due to premature termination of the PCR.

All 24 of the clones picked in the first instance for DNA isolation and *EcoRI* digestion contained inserts. The insert sizes ranged from 1.2kb to 130bp. One insert sequence hybridized to the α2(I) probe. Further screening of the library resulted in the isolation of 6 more clones containing sequence coding for α1(I), α2(I) or α3(VI) collagen cDNA. The isolation of the α3(VI) sequence is encouraging because this sequence was not included when the

helical sequences were aligned to design the primer.

The results of analysis by Southern blot and hybridization of the PCR product indicated the presence of much more type I and type VI collagen sequence than type V. These results confirm the notion that the library displays a collagen clone population bias which reflects the skin fibroblast mRNA population. This is also the case with libraries produced by more conventional methods.

Because of the bias in the library towards clones containing cDNA sequences which code for collagens synthesized by fibroblasts, it is considered a better idea that this work should progress with a starting material other than RNA isolated from skin fibroblasts. A library produced using fibroblast RNA as the starting material will result in the isolation of clones containing collagen sequences which have already been characterized ie. types I and IV. RNA isolated from cells which are known to produce the collagen of interest or in the case of the search for a new collagen, cells which do not synthesize collagen at all may be preferable. Using cells which do not produce collagen would rely on the knowledge that Chelly *et al.* (1988) and Sarkar and Sommer (1989) published results which indicate that all cells express a basal level of all their genes. The clinical applications of PCR amplification of basal level mRNA have been published in Roberts *et al.*, 1990 and Berg *et al.*, 1990. If a cell line was chosen that was known not to produce collagens, except at this background level, then PCR amplification of the mRNA from the cells in the manner described in this chapter would presumably result in an equal quantity of cDNA clones coding for each collagen. This would make the isolation of a novel collagen easier. That is if the PCR conditions are improved so that primers select for collagen sequences. Care has to taken in the choice of cells from which the RNA is to isolated. For many years it was thought that macrophages did not produce type I collagen however, Vaage and Lindbald (1990) recently published results that proved that macrophages do indeed synthesize the type I collagen protein.

PCR amplification using this primer and RNA isolated from a transformed lymphoblastoid cell resulted in the production of sequences which hybridize to an α2(I) cDNA collagen probe. Lymphoblasts are cells which are known not to produce type I collagen and yet there is sequence included in the PCR product which hybridizes to an α2(I) collagen cDNA probe. The concentration of collagen sequence DNA in the PCR product is lower than that in the PCR product using fibroblast RNA as the template.

This is not surprising, especially since the concentration of DNA produced by the PCR was lower when lymphoblast RNA was used as the template (Figure 5.13). Was this due to the lower collagen mRNA available as starting material? Chelly *et al.* (1988) showed that PCR amplification of a mRNA sequence from cells that were known not to produce the relevant protein was much less productive than amplification of the same mRNA sequence from cells that were known to express it. The use of this basal level mRNA as template material for the library could theoretically pose a problem; an increased number of amplification cycles would be required to get the collagen sequence DNA concentration up to a level were there is enough DNA to clone. This may cause an increase in the occurrence of PCR artefacts. This would be a major problem if the starting material consisted of a single molecule of template mRNA. Efforts have been made to quantify the basal mRNA of cells (Chelly *et al.*, 1988; Chelly *et al.*, 1990; Berg *et al.*, 1990). All the workers found a similar basal RNA concentration; approximately one transcript molecule per 10^3 cells. Any problems that arise because of high level PCR artefacts may be solved by increasing the concentration of starting RNA.

Contamination of the probe cocktail with vector DNA caused problems when the hybridization conditions were made less stringent. The signal produced when the contaminating vector DNA hybridized to the clones masked the signal produced by the collagen probes hybridizing to the clones. This problem was overcome by using cDNA sequence that had not been in contact with vector DNA as probe template. The probe was 212bp of α2(I) cDNA. It was synthesized using the primers, A2B and A2C (Figure 5.12). This DNA fragment was chosen because the primers had already been synthesized, not because its sequence was considered particularly ideal for the job of screening the library. The inadequacy of this probe to distinguish between collagen sequences and other DNA sequences is displayed in its inability to hybridize, at 60°C, to all the clones which hybridized to the probe cocktail at 65°C. Yet this probe hybridized to a clone which was found to contain a sequence which was not homologous to any sequence in the EMBL database and was not recognizably collagen-like.

Hybridization of some of the filters with the α2(I) PCR synthesized probe at 57°C resulted in high background signal. This may be due to the probe hybridizing to the primer sequence which would be present in all the clones.

Described in this chapter is a very simple method for the production of a library of cDNA probes. The aim of the project was to produce a library enriched with enough collagen sequence containing cDNA clones to make it feasible to sequence clones at random as a way of isolating a clone containing the sequence of a novel collagen. This was not achieved. However, the foundation has been laid for the production of such a library in the future. If the specificity of the primer is improved and the transformation efficiency is increased, the way will be clear to use this quick and easy method to produce a library enriched with cDNA collagen clones

6.0 FINAL DISCUSSION

The study of the genes coding for members of the collagen gene family is of fundamental importance because these proteins are found ubiquitously in most multicellular organisms from sponges to man. Their biological functions are many and diverse and so the study of the collagens may help in our understanding of the aetiology of connective tissue diseases and cellular processes such as chemotaxis and oncogenesis.

We do not seem to have exhausted the number of collagen proteins to characterize. However, the methods employed to isolate and characterize the collagens have changed. The first collagens characterized were those found most abundantly in the human body ie. types I and III. These proteins were isolated before the genes or the cDNA coding for them, since they were so abundant that they were easily extracted from tissues. More recently collagens have been characterized as cDNA clones without isolation of the protein. These collagens are difficult to isolate from tissue because they are present in low concentrations.

It is perhaps more productive to screen cDNA libraries for novel collagen sequences than to try to isolate the proteins. However, if these new collagens are synthesized by cells at very low concentrations then the concentration of the mRNAs coding for the them may also be very low. Therefore the library must be enriched with collagen coding sequences in order that cDNA clones coding for these rare collagens may be isolated. The method described in this thesis is potentially useful in this respect. Although it was not perfected during the time allowed for this project, work is continuing to improve it.

The isolation of new collagen cDNA sequences may help in the search for the causes of the more common connective tissue disorders. However, before we can understand the causes of these diseases, we must first understand the causes of the less common types. Genetically, the rarer connective tissue disorders are easier to investigate because they are inherited in a Mendelian manner and are, in most cases, the result of single mutations. Information gained about the causes of diseases such as osteogenesis imperfecta is most powerful when the results of many isolated cases are brought together. The effect of collagen gene mutations on phenotype may then be defined. In this search for a trend it is important to know which amino acid variants are phenotypically silent. Until recently, due to lack of published data, it was thought that no collagen sequence

125

amino acid changes were associated with the normal phenotype. Recent data has corrected this, silent mutations have been found in COL1A3 in a number of phenotypically normal individuals. The question should now be asked, are these mutations really silent or do they contribute to the many factors which are thought to cause conditions such as osteoarthritis or aortic aneurysms?

So what are the implications of this research for those individuals who suffer from connective tissue disorders? The development of simpler mutation detection methods makes the process of prenatal diagnosis more amenable and predictions about how mutations affect phenotype will aid genetic counselling. Preventative surgery may be used to prolong the life expectancy of patients known to have a predisposition to aortic aneurysm or relieve the symptoms of those with severe osteoarthritis. However the future holds little joy for the patients suffering from osteogenesis imperfecta. Hopes of using gene therapy to restore the health of patients suffering from OI are hampered by the fact that a return to a normal phenotype involves not only the insertion of a new, functionally normal gene but also the inactivation of the abnormal allele. Only a small amount of abnormal collagen molecules is needed to cause an abnormal phenotype. Collagen genes are expressed early in fetal life thus it would be difficult to pre-empt the onset of disease with gene therapy. A better course of action would be to employ *in vitro* fertilization techniques. Fertilized ovum could be screened for the presence of the disease causing mutation and only those not carrying the mutation implanted in the womb.

REFERENCES

Ainsworth, P.J., Surh, L.C. and Coulter-Mackie, M.B. (1991) Diagnostic single strand conformational polymorphism (SSCP): A simplified non-radioisotopic method as applied to a Tay-Sachs B1 variant. Nucleic. Acids. Res. 19, 405-406.

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

Ala-Kokko, L., Baldwin, C.T., Moskowitz, R.W. and Prockop, D.J. (1990) Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with mild chondrodysplasia. Proc. Natl. Acad. Sci. USA 87, 6565-6568.

Ala-Kokko, L., Kontusaari, S., Baldwin, C.T., Kuivaniemi, H. and Prockop, D.J. (1989) Structure of cDNA clones coding for the entire prepro α1(III) chain of human type III procollagen. Biochem. J. 260, 509-516.

Anderson, I.J., Goldberg, R.B., Marion, R.W., Upholt, W.B. and Tsipouras, P. (1990) Spondyloepiphyseal dysplasia congenita: Genetic linkage to type II collagen (COL2A1). Am. J. Hum. Genet. 46, 896-901.

Anon. (1984) Methods to fill in recessed 3'-ends of DNA. BRL Focus 6, 6.

Armour, J.A.L., Povey, S., Jeremia, H.S. and Jeffreys, A.J. (1990) Systemic cloning of human minisatellites from ordered array chromatid libraries. Genomics 8, 501-512.

Aslanidis, C. and de Jong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic. Acids. Res. 18, 6069-6074.

Bächinger, H.P., Morris, N.P., Lunstrum, G.P., Keene, D.R., Rosenbaum, L.M., Compton, L.A. and Burgeson, R.E. (1990) The relationship of the biophysical and biochemical characteristics of type VII collagen to the function of anchoring fibrils. J. Biol. Chem. 265, 10095-10101.

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.

Bateman, J.F., Lamande, S.R., Dahl, H-H., M, Cohn, D., Mascara, T. and Cole, W.G. (1989) A frameshift mutation results in a truncated nonfunctional carboxyl-terminal proα1(I) peptide of type I collagen in osteogenesis imperfecta. J. Biol. Chem. 264, 10960-10964.

Battacharyya, A. and Lilley, D.M.J. (1989) Single base mismatches in DNA. Long and short range structure probed by analysis of axis trajectory and local chemical reactivity. J. Mol. Biol. 209, 583-597.

Beighton, P. and Curtis, D. (1985) X-linked Ehlers-Danlos syndrome type V: the next generation. Clin. Genet. 27, 472-478.

Beighton, P. (1970) Serious ophthalmological complications in the Ehlers-Danlos syndrome. Brit. J. Ophthal. 54, 263-267.

Beighton, P., de Paepe, A., Danks, D., Finidori, G., Gedde-Dahl, T., Goodman, R., Hall, J.G., Hollister, D.W., Horton, W., McKusickVa, et al. (1988) International nosology of heritable disorders of connective tissue, Berlin. Am. J. Med. Genet. 29, 581-594.

Bennett, V.D. and Adams, S.L. (1987) Characterization of the translational control mechanism preventing synthesis of α2(I) collagen in chicken vertebral chondrocytes. J. Biol. Chem. 262, 14806-14814.

Bennett, V.D. and Adams, S.L. (1990) Identification of a cartilage-specific promoter within intron 2 of the chick alpha2(I) collagen gene. J. Biol. Chem. 265, 2223-2230.

Bentz, H., Morris, N.P., Sakai, L.Y., Hollister, D.W. and Burgeson, R.E. (1983) Isolation and partial characterization of a new human collagen with

an extended triple-helical structural domain. Proc. Natl. Acad. Sci. USA 80, 3168-3172.

Benya, P.D. and Padilla, S.R. (1986) Isolation and characterization of type VIII collagen synthesized by cultured rabbit corneal endothelial cells. J. Biol. Chem. 261, 4160-4169.

Berg, L-P., Weiland, K., Millar, D.S., Schlosser, M., Wagner, M., Kakar, V.V., Reiss, J. and Cooper, D.N. (1990) Detection of a novel point mutation causing haemophilia A by PCR/direct sequencing of ectopically-transcribed factor VIII mRNA. Hum. Genet. 85, 655-658.

Bernard, M.P., Myers, J.C., Chu, M.-l, Ramirez, F., Eikenberry, E.F. and Prockop, D.J. (1983a) Structure of a cDNA for the proα2 chain of human type I procollagen. Biochemistry 22, 1139-1145.

Bernard, M.P., Chu, M.-L., Myers, J.C., Ramirez, F., Eikenberry, E.F. and Prockop, D.J. (1983b) Nucleotide sequences of complementary deoxyribonucleic acids for the pro α1 chain of human type I procollagen. Biochemistry 22, 5213-5223.

Birk, D.E., Fitch, J.M., Babiarz, J.P. and Linsenmajer, T.F. (1988) Collagen type I and type V are present in the same fibril in the avian corneal stroma. J. Cell Biol. 106, 999-1008.

Blanton, S.H., Sarfarazi, M., Eiberg, H., De, G.J., Farndon, P.A., Kilpatrick, M.W., Child, A.H., Pope, F.M., Peltonen, L., Francomano, C.A., Boileau, C., Keston, M. and Tsipouras, P. (1990) An exclusion map of Marfan syndrome. J. Med. Genet. 27, 73-77.

Boast, S., Su, M.-W., Ramirez, F., Sanchez, M. and Avveolimento, E.V. (1990) Functional analysis of cis-acting DNA sequences controlling transcription of the human type I collagen genes. J. Biol. Chem. 265, 13351-13356.

Bonadio, J. and Byers, P.H. (1985) Subtle structural alterations in the chains of type I procollagen produced in osteogenesis imperfecta type II. Nature 316, 363-366.

Bonadio, J., Ramirez, F. and Barr, M. (1990) An intron mutation in the human alpha 1(I) collagen gene alters the efficiency of pre-mRNA splicing and is associated with osteogenesis imperfecta type II. J. Biol. Chem. 265, 2262-2268.

Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. and Colombatti, A. (1990) Structural and functional features of the alpha 3 chain indicate a bridging role for chicken collagen VI in connective tissues. Biochemistry 29, 1245-1254.

Bornstein, D. and McKay, J. (1988) The first intron of the $\alpha 1$(I) collagen gene contains several transcriptional regulatory elements. J. Biol. Chem. 263, 1603-1606.

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

Bornstein, P. and Sage, H. (1989) Regulation of collagen gene expression. Prog. Nucleic Acid res. Mol. Biol. 37, 67-106.

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

Buluwela, L., Forster, A., Boehm, T. and Rabbitts.T.H. (1989) A rapid procedure for colony screening using nylon filters. Nucleic. Acids. Res. 17, 452.

Burbelo, P., Martin, G. and Yamada, Y. (1988) $\alpha 1$(IV) and $\alpha 2$(IV) collagen genes are regulated by a bidirectional promoter and shared enhancer. Proc. Natl. Acad. Sci. USA 85, 9674-9682.

Burbelo, P.D., Horikoshi, S. and Yamada, Y. (1990) DNA methylation and collagen IV gene expression in teratocarcinoma cells. J. Biol. Chem. 265, 4839-4843.

Burgeson, R.E. and Hollister, D.W. (1979) Collagen heterogeneity in human cartilage: Identification of several new collagen chains. Biochem. Biophys.

Res. Commun. 87, 1124-1131.

Burgeson, R.E., Lunstrum, G.P., Rokosova, B., Rimberg, C.S., Rosenbaum, L.M. and Keene, D.R. (1990) The structure and function of type VII collagen. Ann. NY Acad. Sci. 580, 32-43.

Byers, P.H., Siegel, R.C., Holbrook, K.A., Narayanan, A.S., Bornstein, P. and Hall, J.G. (1980) X-linked cutis laxa. Defective cross-linked formation in collagen due to decreased lysyl oxidase activity. N. Engl. J. Med. 303, 61-63.

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. (1989) Inherited disorders of collagen gene structure and expression. Am. J. Med. Genet. 34, 72-80.

Byers, P.H. (1990) Brittle bones-fragile molecules: Disorders of collagen gene structure and expression. Trends Genet. 6, 293-300.

Cawthorn, R.M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. and White, R. (1990) A major segment of neurofibromatosis type I gene: cDNA sequence, genomic structure and point mutations. Cell 62, 193-201.

Chelly, J., Kaplan, J-C., Maire, P., Gautron, S. and Kahn, A. (1988) Transcription of the dystrophin gene in human muscle and non-muscle tissues. Nature 333, 858-860.

Chelly, J., Concordet, J-P., Kaplan, J-C. and Kahn, A. (1989) Illegitimate transcription: Transcription of any gene in any cell. Proc. Natl. Acad. Sci. USA 86, 2617-2621.

Chelly, J., Montarras, D., Pinset, C., Berwald-Netter, Y., Kaplan, J-C. and

Kahn, A. (1990) Quantitative estimation of minor mRNAs by cDNA-polymerase chain reaction. Application to dystrophin mRNA in cultured myogenic and brain cells. FEBS Lett. 187, 691-698.

Chomozynski, 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 characterization of cDNAs specific for the human pro α1(I) collagen chain. Nucleic. Acid. Res. 10, 5925-5934.

Chu, M-L., Willaims, C., Pepe, G., Hirsch, 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., de Wet, W., Bernard, M., Ding, J.-F., Morabito, M., Myers, J., Williams, C. and Ramirez, F. (1984) Human proα1(I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. Nature 310, 337-340.

Chu, M-L., de Wet, W., Bernard, M. and Ramirez, F. (1985a) Fine structural analysis of the human pro-α1(I) collagen gene. J. Biol. Chem. 260, 2315-2320.

Chu, M-L., Weil, D., de Wet, W., Bernard, M., Sippola, M. and Ramirez, F. (1985b) Isolation of cDNA and genomic clones encoding human pro α1(III) collagen. J. Biol. Chem. 260, 4357-4363.

Chu, M-L., Mann, K., Deutzmann, R., Pribula-conway, D., Hsu-chen, Bernard, M.P. and Timpl, R. (1987) Characterization of three constituent chains of collagen type VI peptide sequences and cDNA clones. Eur. J. Biochem. 168, 309-317.

Chu, M-L., Conway, D., Pan, T., Baldwin, C., Mann, K., Deutzmann, R. and Timpl, R. (1988) Amino acid sequence of the triple helical domain of human collagen type VI. J. Biol. Chem. 263, 18601-18606.

Chu, M-L., Pan, T-C., Conway, D., Saitta, B., Stokes, D., Kuo, H-J., Glanville, R.W., Timpl, R., Mann, K. and Deutzmann, R. (1990) The structure of type VI collagen. Ann. NY Acad. Sci. 580, 55-63.

Clarke, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic. Acids. Res. 16, 9677-9686.

Cohn, D.H., Apore, 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.

Cohn, D.H., Starman, B.J., Blumberg, B. and Byers, P.H. (1990) Recurrence of lethal osteogenesis imperfecta due to parental mosaicism for a dominant mutation in a human type I collagen gene (COL1A1). Am. J. Hum. Genet. 46, 591-601.

Cole, W.G., Chan, D., Cambers, 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.

Constantinou, C.D., Pack, M., Young, S.B. and Prockop, D.J. (1990a) Phenotypic heterogeneity in osteogenesis imperfecta: The mildly affected mother of a proband with a lethal variant has the same mutation substituting cysteine for α1-glycine 904 in a type I procollagen gene (COL1A1). Am. J. Hum. Genet. 47, 670-679.

Constantinou, C.D., Pack, M., Young,S.B. and Prockop, D.J. (1990b) A substitution of cysteine for glycine 904 in COL1A1 in a proband with lethal osteogenesis imperfecta and her asymptomatic mother. Ann. NY Acad. Sci. 580, 540-541.

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.

Cotton, R.G.H. (1989) Review article: Detection of single base changes in nucleic acids. Biochem. J. 263, 1-10.

Cotton, R.G.H. and Campbell, R.D. (1989) Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences. Nucleic. Acids. Res. 17, 4223-4233.

Cox, G.N., Fields, C., Kramer, J.M., Rosenzweig, B. and Hirsh, D. (1989) Sequence comparisons of developmentally regulated collagen genes of *Caenorhabditis elegans*. Gene 76, 331-344.

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

Crowe, J.S., Cooper, H.J., Smith, M.A., Simms, M.J., Parker, D. and Gewert, D. (1991) Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic. Acids. Res. 19, 184.

D'Alessio, J.M., Noon, M.C., Ley, H.L. and Gerard, G.I. (1987) One tube double stranded cDNA synthesis using cloned M-MLV reverse transcriptase. Focus 9, 1-4.

D'Alessio, M., Ramirez, F., Suzuki, H.R., Solursh, M. and Gambino, R. (1989) structural and developmental expression of a sea urchin fibrillar collagen gene. Proc. Natl. Acad. Sci. USA 86, 9303-9307.

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.

Daw, S., Nicholls, A.C., Williams, E.M., Sykes, B. and Pope, F.M. (1988) Autosomal recessive osteogenesis imperfecta: Excess postranslational modification of collagen not linked to either COL1A1 or COL1A2. J. Med. Genet. 25, 275.

Dean, M., White, M.B., Amos, J., Gerrard, B., Stewart, C. Khaw, K-T and Leppert, M. (1990) multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. Cell 61, 863-870.
Deitz, H.C., Pyeritz, R.E., Hon, B.D., Cadle, R.G., Hamosh, A., Schwartz, J., Meyers, D.A. and Francomana, C.A. (1991) The Marfan syndrome locus: confirmation of assignment to chromosome 15 and identification of tightly linked markers to 15q15-q21.1. Genomics 9, 335-362.

Denhardt. (1966) A membrane filter technique for the detection of complementary DNA. Biochem. Biophys. acta, 641-646.

Diaz-Ruiz, J.R. and Kaper, J.M. (1979) Cucumber mosaic virus-associated RNA 5. VI. Characterization and denaturation-renaturation behaviour of the double-stranded form. Biochim. Biophys. Acta 564, 275-288.

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

Dietz, U.H., Garies, M. and Bertling, W.M. (1990) Use of primer pools for specific gene amplification. Nucleic. Acids. Res. 18, 5914.

Differante, N., Leachman, R.D. and Angelini, P. (1975) Lysyl oxidase deficiency in Ehlers Danlos syndrome type V. Conn. Tis. Res. 3, 49-53.

Dombrowski, K.E. and Prockop, D.J. (1988) Cleavage of type I and type II procollagens by type I/II procollagen N-proteinase. J. Biol. Chem. 263, 16545-16552.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E.coli* by high voltage electroporation. Nucleic. Acids. Res. 16, 6127-.

Dublet, B., Oh, S., Sugrue, S.D., Gordon, M.K., Gerecke, D.R., Olsen, B.R. and van der Rest, M. (1989) The structure of avian type XII collagen. J. Biol. Chem. 264, 13150-13156.

Dublet, B. and van der Rest, M. (1987) Type XII collagen is expressed in embryonic chick tendons. J. Biol. Chem. 262, 17724-17727.

Dubose, R. and Fondhartl, D.L. (1990) Rapid purification of PCR products for DNA sequencing using sepharose CL-6B spin columns. BioTechniques 8, 271-273.

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.

Emanuel, B.S., Cannizzaro, L.A., Seyer, J.M. and Myers, J.C. (1985a) Human $\alpha1(III)$ and $\alpha2(V)$ 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 location of the human elastin gene. Am. J. Hum. Genet. 37, 873-882.

Erlich, H.A., ed. PCR Technology. Principles and applications for DNA amplification. USA: Stockton Press, 1989.

Exposito, J-Y. and Garrone, R. (1990) Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearances of two collagen gene families. Proc. Natl. Acad. Sci. USA 87, 6669-6673.

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., Apon, S., Wu, J.J., Ericsson, L.H. and Walsh, K.A. (1987) Collagen type IX: Evidence for covalent linkages to type II collagen in cartilage. FEBS Lett. 220, 337-341.

Eyre, D.R., Weis, M.A. and Moskowitz, R.W. (1991) Mutation in an inherited form of osteoarthritis associated with mild chondrodysplasia. J. Clin. Invest. 87, 357-361.

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.

Ferre, F. and Gardino, F. (1989) Preparation of crude cell extract suitable for amplification of RNA by the polymerase chain reaction. Nucleic. Acids. Res. 17, 2141.

Fleischmajer, R., Perlish, J.S., Burgeson, R.E., Shaikh-Bahal, F. and Timpl, R. (1990) Type I and type III collagen interactions during fibrillogenesis. Ann. NY Acad. Sci. 580.

Francomano, C.A., Liberfarb, R.M., Hirose, T., Maumenee, I.H., Streeten, E.A., Meyer, D.A. and Pyeritz, R.E. (1987) The Stickler syndrome: evidence for close linkage to the structural gene for type II collagen. Genomics 1, 293-296.

Francomano, C.A., Streeton, E.A., Meyers, D.A. and Pyeritz, R.E. (1988) Exclusion of fibrillar procollagens as causes of Marfan syndrome. Am. J. Hum. Genet. 29, 457-462.

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.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene

specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85.

Furthmayr, H., Weidemann, H., Timpl, R., Odermatt, E. and Engel, J. (1983) Electron-microscopical approach to a structural model of intima collagen. Biochem. J. 211, 303-311.

Furuto, D.K. and Miller, E.J. (1983) Different levels of glycosylation contribute to the heterogeneity of α1(III) collagen chains derived from transplantable rat chondrosarcoma. Arch. Biochem. Biophys. 226, 604-611.

Ganguly, A. and Prockop, D.J. (1990) Detection of single-base mutations by reaction of DNA heteroduplexes with a water-soluble carbodiimide followed by primer extension: Application to products from the polymerase chain reaction. Nucleic. Acids. Res. 18, 3933-3939.

Genovese, C., Brufsky, A., Shapiro, J. and Rowe, D. (1989) Detection of mutations in human type I collagen mRNA in osteogenesis imperfecta by indirect RNase protection. J. Biol. Chem. 264, 9632-9637.

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

Gibbs, R.A. (1990) DNA amplification by the polymerase chain reaction. Anal. Chem. 62, 1202-1214.

Godfrey, M., Menashe, V., Weleber, R.G., Koler, R.D., Bigley, R.H., Lourien, E., Zonana, J. and Hollister, D.W. (1990a) Cosegregation of elastin-associated microfibrillar abnormalities with the Marfan phenotype in families. Am. J. Hum. Genet. 46, 652-660.

Godfrey, M., Olson, S., Burgio, R.G., Martini, A., Valli, M., Cetta, G., Hori, H. and Hollister, D.W. (1990b) Unilateral microfibrillar abnormalities in a case of assymmetric Marfan syndrome. Am. J. Hum. Genet. 46, 661-671.

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.

Gordon, M.K., Gerecke, D.R., Dublet, B., Van, der R.M., Sugrue, S.P. and Olsen, B.R. (1990) The structure of type XII collagen. Ann. NY Acad. Sci. 580, 8-16.

Grange, D.K., Gottesman, G.S., Lewis, M.B. and Marini, J.C. (1990) Detection of point mutations in type I collagen by RNase digestion of RNA/RNA hybrids. Nucleic. Acids. Res. 18, 4227-4236.

Griffin, C.A., Emanuel, B.S., Hansen, J.R., Caunee, W.K. and Myers, J.C. (1987) Human collagen genes encoding basement membrane $\alpha 1(IV)$ and $\alpha 2(IV)$ chains map to the distal long arm of chromosome 13. Proc. Natl. Acad. Sci. USA 84, 512-516.

Gunwar, S., Saus, J., Noelken, M.E. and Hudson, B.G. (1990) Glomerular basement membrane. Identification of a fourth chain, Alpha 4, of type IV collagen. J. Biol. Chem. 265, 5466-5469.

Gussow, D. and Clackson, T. (1989) Direct clone characterization from plaques and colonies by the polymerase chain reaction. Nucleic. Acids. Res. 17, 4000.

Harbarth, P. and Vosberg, H.-P. (1988) Enzymatic amplification of myosin heavy-chain mRNA sequences in vitro. DNA 7, 297-306.

Hartsfield, J.K. and Kouseff, B.E. (1990) Phenotypic overlap of Ehlers-Danlos syndrome type IV and VIII. Am. J. Med. Genet. 37, 465-470.

Hawkins, J.R. The molecular genetics of inherited connective tissue disorders [Dissertation]. University of Leicester, 1989.

Hawkins, J.R., Superti-Furga, A., Steinmann, B. and Dalgleish, R. (manuscript in preparation).

Hayman, A.R., Koeppel, J., Winterhalter, K.H. and Trueb, B. (1990) The triple-helical domain of alpha 2(VI) collagen is encoded by 19 short exons that are multiples of 9 base pairs. J. Biol. Chem. 265, 9864-9868.

Helmsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. and Galas, D.J. (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic. Acids. Res. 17, 6545-6551.

Helseth, D.L. and Veis, A. (1981) Collagen self assembly in vitro. J. Biol. Chem. 256, 7118-7128.

Henkel, W. and Glanville, R.W. (1982) Covalent cross-linking between molecules of type I and type III collagen. Eur. J. Biochem. 122, 205-213.

Henry, I., Jeanpierre, M., Bernard, M., Weil, D., Grzescik, K.H., Ramirez, F., Chu, M-L. and Junien, C. (1985) The structural gene for fibronectin (FN) maps to 22p323-qter. Cytogenet. Cell. Genet. 40, 650.

Hörlein, D., McPherson, J., Goh, S.H. and Bornstein, P. (1981) Regulation of protein synthesis: Translational control by procollagen-derived fragments. Proc. Natl. Acad. Sci. USA 78, 6163-6167.

Holton, T.A. and Graham, M.W. (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic. Acids. Res. 19, 1156.

Horton, W., Miyashita, T., Kohno, K., Hassell, J.R. and Yamada, Y. (1987) Identification of a phenotype-specific enhancer in the first intron of the rat collagen II gene. Proc. Natl. Acad. Sci. USA 84, 8864-8868.

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

Huerre, J.C., Mattei, M-G., Weil, D., Grzeschik, K.H., Chu, M-L., Sangiorgi, F.O., Sobel, M.E., Ramirez, F. and Junien, C. (1986a) Further evidence for the dispersion of the human fibrillar collagen genes. Am. J. Hum. Genet. 38, 26-37.

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

Hung, T., Mak, K. and Fong, K. (1990) A specificity enhancer for polymerase chain reaction. Nucleic. Acids. Res. 18, 4953.

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

Impraim, C.C., Saiki, R.K., Erlich, H.A. and Teplitz, R.L. (1987) Analysis of DNA extracted from formalin-fixed, paraffin embedded tissues by enzymatic amplification and hybridization with sequence-specific oligonucleotides. Biochem. Biophys. Res. Commun. 142, 710-716.

Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988) DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc. Natl. Acad. Sci. USA 85, 9436-9440.

Irwin, M.I. (1986) Use of monoclonal antibodies to locate the chondoitin sulphate chain(s) to type IX collagen. J. Biol. Chem. 261, 16281-16283.

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

Janeczko, R.A. and Ramirez, F. (1989) Nucleotide and amino acid sequences of the entire human α1(III) collagen. Nucleic. Acids. Res. 17, 6742.

Jeffreys, A.J., Wilson, V., Wood, D., Simons, J.P., Kay, R.M. and Williams, J.G. (1980) Linkage of adult alpha- and beta-globin genes in X. laevis and

gene duplication by tetraploidization. Cell 21, 555-564.

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic. Acids. Res. 16, 10953-10971.

Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60, 473-485.

Jimenez, S.A., Yankowski, R. and Reginato, A.M. (1986) Quantitative analysis of type X collagen biosynthesis by embryonic-chick sternal cartilage. Biochem. J. 233, 357-367.

Jung, V., Pestka, S.B. and Pestka, S. (1990) Efficient cloning of PCR generated DNA containing terminal restriction endonuclease recognition sites. Nucleic. Acids. Res. 18, 6156.

Kadler, K.E., Hojima, Y. and Prockop, D.J. (1987) Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. J. Biol. Chem. 260, 15696-15701.

Kadler, K.E., Hojima, Y. and Prockop, D.J. (1988) Assembly of type I collagen fibrils de novo. J. Biol. Chem. 263, 10517-10523.

Kainulainen, K., Pulkkinen, L., Savalainen, A., Kaitila, I. and Peltonen, L. (1990a) Location on chromosome 15 of the gene defect causing Marfan syndrome. N. Engl. J. Med. 323, 935-939.

Kainulainen, K., Savolainen, A., Palotie, A., Kaitila, I., Rosenbloom, J. and Peltonen, L. (1990) Marfan syndrome: Exclusion of genetic linkage to five genes coding for connective tissue components in the long arm of chromosome 2. Hum. Genet. 84, 233-236.

Kanehisa, M., Klein, P., Greif, P. and Delisa, C. (1984) Computer analysis and structure prediction of nucleic acids and proteins. Nucleic. Acids. Res. 12, 417-428.

Karsenty, G. and De, C.B. (1990) Two different negative and one positive regulatory factors interact with a short promoter segment of the alpha 1(I) collagen gene. J. Biol. Chem. 265, 9934-9942.

Keen, J., Lester, D., Inglehearn, C., Curtis, A. and Bhattacharya, S. (1991) Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. TIG 7, 5.

Keitly, C.M., Boot-Handford, R.P., Ayad, S., Shuttleworth, C.A. and Grant, M.E. (1990) Molecular composition of type VI collagen. Biochem. J. 272, 787-795.

Kemp, D.J., Smith, D.B., Foote, S.J. Somaras, N. and Peterson, M.G. (1989) Colorimetric detection of specific DNA segments amplified by polymerase chain reactions. Proc. Nat. Acad. Sci. 86, 2423-2427

Kivirikko, K.I. and Myllyla. (1982) Post-translational enzymes in the biosynthesis of collagen: Intracellular enzymes. Meth. Enzym. 82, 245-304.

Knowlton, R.G., Weaver, E.J., Struyk, A.F., Knobloon, W.H., King, R.A., Norris, K., Shambon, A., Uitto, J., Jemenez, S.A. and Prockop, D.J. (1989) Genetic linkage analysis of hereditary arthro-ophthalmology (Stickler syndrome) and the type II procollagen gene. Am. J. Hum. Genet. 45, 681-688.

Koller, E., Hayman, A. R. and Trueb, B. (1991) The promoter of the chicken a2(VI) collagen gene has features characteristic of house-keeping genes and proto-oncogenes. Nucleic. Acid. Res. 19, 485-491.

Kobayashi, K., Hata, R., Nagai, S., Niwa, J. and Hoshino, T. (1990) Direct visualization of affected collagen molecules synthesized by cultured fibroblasts from an osteogenesis imperfecta patient. Biochem. Biophys. Res. Commun. 172, 217-222.

Kogan, S.C., Doherty, M. and Gitschier, J. (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N. Engl. J. Med. 317, 985-990.

Kontusaari, S., Tromp, G., Kuivaniemi, H., Ladda, R.L. and Prockop, D.J. (1990) Inheritance of an RNA splicing mutation (G+ 1 IVS20) in the type III procollagen gene (COL3A1) in a family having aortic aneurysms and easy bruisability: phenotypic overlap between familial arterial aneurysms and Ehlers-Danlos syndrome type IV. Am. J. Hum. Genet. 47, 112-120.

Kornher, J.S. and Lavik, K.J. (1989) Mutation detection using nucleotide analogs that alter electrophoretic mobility. Nucleic. Acids. Res. 17, 7779-7784.

Kuivaniemi, H., Pelton, L., Palotie, A., Kaitila, I. and Kivirikko, K. (1982) Abnormal copper metabolism and deficient lysyl oxidase activity in a heritable connective tissue disorder. J. Clin. Invest. 69, 730-733.

Kuivaniemi, H., Tromp, G., Chu, M.-L. and Prockop, D.J. (1988a) Structure of full-length cDNA clone for the prepro α2(I) chain of human type I procollagen. Comparison with the chicken gene confirms unusual patterns of gene conservation. Biochem. J. 252, 633-640.

Kuivaniemi, H., Sabol, C., Tromp, G., Sippola-Thiele, M. and Prockop, D.J. (1988b) A 19 base pair deletion in the pro-a2(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 symptomatic mother. J. Biol. Chem. 263, 11407-11413.

Kuivaniemi, H., Kontusaari, S., Tromp, G., Zhao, M., Sabol, C. and Prockop, D.J. (1990) Identical G+1 to A mutations in three different introns of the type III procollagen gene (COL1A3) produce different patterns of RNA splicing in three variants of Ehlers-Danlos syndrome IV. J. Biol. Chem. 265, 12067-12074.

Labermeier, U. and Kenney, M.C. (1983) The presence of EC collagen and type IV collagen in bovine Descemet's membranes. Biochem. Biophys. Res. Commun. 116, 619-625.

Labhard, M.E. and Hollister, D.W. (1990) Segmental amplification of the entire helical and telopeptide regions of the cDNA for human alpha 1(I) collagen. Matrix. 10, 124-130.

Law, M.L., Chan, S.D.H., Berger, R., Jones, C., Kao, F.T., Solomon, E. and Cheah, K.S.E. (1990) The gene for the alpha 2 chain of the human fibrillar collagen type XI (COL11A2) assigned to the short arm of chromosome 6. Ann. Hum. Genet. 54, 23-29.

Lear, A.L., Eperon, L.P., Wheatley, I.M. and Eperon, I.C. (1990) Hierarchy of 5' splice site preference determined in vivo. J. Mol. Biol. 211, 103-115.

Lee, C.C., Wu, X., Gibbs, R.A., Cook, R.G., Muzny, D.M. and Caskey, T. (1988) Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. Science 239, 1288-1291.

Leigh, I.M., Eady, R.A.J., Heagerty, A.H.M., Purkis, P., Whitehead, P.A. and Burgeson, R.E. (1988) Type VII collagen is a normal component of epidermal basement membrane which shows altered expression an recessive dystrophic epidermolysis bullosa. J.Invest. Dermatol. 90, 639-642.

Li, H., Gyllensten, U.B., Cui, X., Saiki, R.K., Erlich, H.A. and Arnheim, N. (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335, 414-417.

Loidl, H.R., Brinker, J.M., May, M., Pihlajaniemi, T., Morrow, S., Rosenbloom, J. and Myers, J.C. (1984) Molecular cloning and carboxyl-propeptide analysis of human type III procollagen. Nucleic. Acids. Res. 12, 9383-9394.

Lorneli, H., Tyagli, S., Pritchard, C.G., Lizardi, P.M. and Kramer, F.R. Quantitative assays based on the use of replicable hybridization probes. Clin. Chem. 35, 1826-1831.

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. Proc. Natl. Acad. Sci. USA 84, 2803-2807.

Lunstrum, G.P., Sakai, L.Y., Keene, D.R., Morris, N.P. and Burgeson, R.E. (1986) Large complex globular domains of type VII procollagen contributes to the structure of anchoring fibrils. J. Biol. Chem. 261, 9042-9048.

LuValle, P., Ninomiya, Y., Rosenblum, N., D and Olsen, B.R. (1988) The type X collagen gene. J. Biol. Chem. 263, 18378-18385.

Maddox, B.K., Sakai, L.Y., Keene, D.R. and Glanville, R.W. (1989) Connective tissue microfibrils. J. Biol. Chem. 264, 21381-21385.

Mankoo, B.S. and Dalgleish, R. (1988) Human pro α1(III) collagen: cDNA sequence for the 3' end. Nucleic. Acids. Res. 16, 2337.

Mann, K., Jander, R., Korsching, E., Kuhn, K. and Rauterberg, J. (1990) The primary structure of a triple-helical domain of collagen type VIII from bovine Descemet's membrane. FEBS Lett. 273, 168-172.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic. Acid. Res. 19, 1154.

Matfield, M.J. Novel lambda cloning vectors [Dissertation]. University of Leicester, 1983.

Mayne, R., Vail, M.S. and Miller, E.J. (1975) Analysis of changes in collagen biosynthesis that occur when chondrocytes are grown in 5-bromo-2'-deoxy-uridine. Proc. Natl. Acad. Sci. USA 72, 451-455.

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

Mercier, B., Gaucher, C., Feugas, O. and Mazurier, C. (1990) Direct PCR from whole blood, without DNA extraction. Nucleic. Acids. Res. 18, 5908.

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

Miller, E.J. (1985) The structure of fibril-forming collagens. Ann. NY Acad. Sci. 580, 1-13.

Miller, E.J. and Gay, S. (1987) Collagens: An overview and update. Meth.

Enzym. 144, 3-41.

Miskulin, M., Dalgleish, R., Kluve-Beckerman, B., Rennard, S., Tolstoshev, P. and Crystal, R.G. (1986) Human type III collagen gene expression is coordinately modulated with the type I collagen genes during fibroblast growth. Biochemistry 25, 1408-1413.

Moro, L. and Smith, B.D. (1977) Identification of collagen α1(I) trimer and normal type I collagen in a polyoma virus-induced mouse tumor. Arch. Biochem. Biophys. 182, 33-41.

Morris, N.P., Keene, D.R., Glanville, R.W., Bentz, H. and Burgeson, R.E. (1986) The tissue form of type VII collagen is an antiparallel dimer. J. Biol. Chem. 261, 5638-5644.

Morton, H.J. (1970) A survey of commercially available tissue culture media. In Vitro 6, 89-108.

Muragaki, Y., Nishimura, I., Henney, A., Ninomiya, Y. and Olsen, B.R. (1990) The alpha 1(IX) collagen gene gives rise to two different transcripts in both mouse embryonic and human fetal RNA. Proc. Natl. Acad. Sci. USA 87, 2400-2404.

Murphy, G. and Ward, E.S. Nucleic acid sequencing: A practical approach. Oxford: IRL Press, 1989 (Howe, C.J. and Ward, E.S., eds.).

Myers, J.C., Jones, T.A., Pohjolainen, E-R., Kadri, A.S., Goddard, A.D., Sheer, D., Solomon, E. and Pihlajaniemi, T. (1990) Molecular cloning of Alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus. Am. J. Hum. Genet. 46, 1024-1033.

Myers, J.C., Loidl, H.R., Seyer, J.M. and Dion, A.S. (1985a) Complete primary structure of the human α2 type V procollagen COOH-terminal propeptide. J. Biol. Chem. 260, 11216-11224.

Myers, R.M., Larin, Z. and Maniatis, T. (1985b) Detection of single base substitutions by ribonuclease cleavage of mismatches in RNA:DNA duplexes.

Science 230, 1242-1246.

Myers, R.M., Fischer, S.G., Maniatis, T. and Lerman, L.S. (1985c) 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. (1985d) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic. Acids. Res. 13, 3131-3145.

Nestler, F.H., Huidt, S., Ferry, J.D. and Veis, A. (1983) Flexibility of collagen determined from dilute solutions viscoelastic measurements. Biopolymers 22, 1747-1758.

Nicholls, A.C., De Paepe, A., Narcisi, D., Dalgleish, R., De Keyser, F., Matton, M. and Pope, F.M. (1988) Linkage of a polymorphic marker for the type II collagen gene (COL1A3) to atypical autosomal dominant Ehlers-Danlos syndrome type IV in a large Belgian family. Hum. Genet. 78, 276-281.

Nishimura, I., Muragaki, Y. and Olsen, B.R. (1989) Tissue-specific forms of type IX collagen-proteoglycan arise from the use of two widely separated promoters. J. Biol. Chem. 264, 20033-20041.

Niyibizi, C., Fietzeki, P.P. and van der Rest, M. (1984) Human placenta type V collagens: Evidence for the existence of an $\alpha1(V)\alpha2(V)\alpha3(V)$ collagen molecule. J. Biol. Chem. 259, 14170-14174.

Niyibizi, C. and Eyre, D.R. (1989) Identification of the cartilage $\alpha1(XI)$ chain in type V collagen from bovine bone. FEBS Lett. 242, 314-318.

Novack, D.F., Casna, N.J., Fischer, S.G. and Ford, 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., Priestley, L.M., Dalgleish, R., Schmidtke, J.,

Zoll, B. and Sykes, B.C. (1987) Segregation of all four major fibrillar collagen genes in the Marfan syndrome. Am. J. Hum. Genet. 41, 1071-1082.

Ohara, O., Dorit, R.L. and Gilbert, W. (1989) One-sided polymerase chain reaction: the amplification of cDNA. Proc. Natl. Acad. Sci. USA 86, 5673-5677.

Oldberg, Å, Franzén, A. and Heinegård. (1986) Cloning and sequence analysis of rat bone sialoprotein (osteopontin) cDNA reveals an Arg-Gly-Asp cell binding sequence. Proc. Natl. Acad. Sci. USA 83, 8819-8823.

Olsen, D.B. and Eckstein, F. (1989) Incomplete primer extension during in vitro DNA amplification catalyzed by Taq polymerase, exploitation for DNA sequencing. Nucleic. Acids. Res. 17, 9613-9620.

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5, 874-879.

Pack, M., Constantinou, C.D., Kalia, K., Neilson, K.B. and Prockop, D.J. (1989) Substitution of serine for α1(I)-glycine 844 in a severe variant of osteogenesis imperfecta minimally destabilizes the triple helix of type I procollagen. J. Biol. Chem. 264, 19694-19699.

Paglia, L.M., Weistner, M., Duchene, M., Oullette, L.A., Horlein, D., Martin, G.R. and Muller, P.K. (1981) Effects of procollagen peptides on the translation of type II collagen messenger ribonucleic acid and on collagen in chondrocytes. Biochemistry 20, 3523-3527.

Pajunen, L., Tamminen, M., Solomon, E. and Pihlajaniemi, T. (1989) Assignment of the gene coding for the alpha 1 chain of collagen type XIII (COL13A1) to human chromosome region 10q11qter. Cytogenet. Cell. Genet. 52, 190-193.

Parker, M.I., Smith, A.A., Boast, S., Su, M.W. and Ramirez, F. (1990) Regulation of the human alpha 2(I) procollagen gene by a trans-acting factor. Ann. NY Acad. Sci. 580, 451-453.

Paterson, C.R. (1990) Osteogenesis imperfecta and other bone disorders in the differential diagnosis of unexplained fractures. J. Royal Soc. Med. 83, 72-74.

Phillips, C.L., Shrago-Howe, A.W., Pinnell, S.R. and Wenstrup, R.J. (1990) A substitution at a non-glycine position in the triple-helical domain of proa2(I) collagen chains present in an individual with a variant of the Marfan syndrome. J. Clin. Invest. 86, 1723-1728.

Pihlajaniemi, T., Dickson, L.A., Pope, F.M., Korhonen, V.R., Nicholls, A. and Prockop, D.J. (1984) Osteogenesis imperfecta: Cloning of a pro-α2(I) collagen gene with a frameshift mutation. J. Biol. Chem. 259, 12941-12944.

Pihlajaniemi, T., Myllyla, S., Seyer, J., Kurkinen, M. and Prockop, D.J. (1987) Partial characterization of a low molecular weight human collagen that undergoes alternative splicing. Proc. Natl. Acad. Sci. USA 84, 940-944.

Pihlajaniemi, T., Pohjolainen, E.-R. and Myers, J.C. (1990a) Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, Alpha 5(IV). J. Biol. Chem. 265, 13758-13766.

Pihlajaniemi, T., Tamminen, M., Sandberg, M., Hirvonen, H. and Vuorio, E. (1990b) The alpha 1 chain of type XIII collagen: Polypeptide structure, alternative splicing, and tissue distribution. Ann. NY Acad. Sci. 580, 440-443.

Pingoud, A. (1985) Spermidine increases the accuracy to type II restriction endonucleases. Eur. J. Biochem. 147, 105-109.

Pinnell, S.R., Krane, S.M., Kenzora, J.E. and Glimcher, M.J., (1972) Heritable disorder with hydroxylysine-deficient collagen: hydroxylysine deficient collagen disease. New Engl. J. Med. 286, 1013-1020.

Pinnell, S.R., Murcal, S., Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S., eds. Disorders of collagen. 5th ed. NY: McGraw-Hill, 1983 (The metabolic basis of inherited disease).

Pollner, R., Fischer, G., Poeschl, E. and Kuehn, K. (1990) Regulation of divergent transcription of the genes coding for basement membrane type IV collagen. Ann. NY Acad. Sci. 580, 44-54.

Pope, F.M., Martin, G.R., Lichtenstein, J.R., Pentinnen, R., Gerson, B., Rowe, D.W. and McKusick, V.A. (1975) Patients with Ehlers-Danlos syndrome type IV lack type III collagen. Proc. Natl. Acad. Sci. USA 72, 1314-1316.

Poschl, E., Pollner, R. and Kuhn, K. (1988) The genes for the α1(IV) and α2(IV) chains of human basement membrane collagen type IV are arranged head to head and separated by a bidirectional promoter of unique sequence. EMBO J. 7, 2687-2695.

Postlethwaite, A.E., Seyer, J.M. and Kang, A.H. (1978) Chemotactic attraction of human fibroblasts to type I,II and III collagens and collagen-derived peptides. Proc. Natl. Acad. Sci. USA 75, 871-875.
Priestley, L., Kumar, D. and Sykes, B. (1990) Amplification of the COL2A1 3' variable region used for segregation analysis in a family with the Stickler syndrome. Hum. Genet. 85, 525-526.

Privalo, P.L. (1982) Stability of proteins: Proteins which do not present in a single cooperative system. Adv. Protein Chem. 35, 1-104.

Prockop, D.J. and Kivirikko, K.I. (1984) Heritable diseases of collagen. N. Engl. J. Med. 311, 376-386.

Prockop, D.J., Constantinou, C.D., Dombrowski, K.E., Hojima, Y., Kadler, K.E., Kuivaniemi, H., Tromp, G. and Vogel, B.E. (1989) Type I procollagen: the gene-protein system that harbors most of the mutations causing osteogenesis imperfecta and probably more common heritable disorders of connective tissue. Am. J. Med. Genet. 34, 60-67.

Prockop, D.J., Baldwin, C.T. and Constantinou, C.D. (1990) Mutations in type I procollagen genes that cause osteogenesis imperfecta. Adv. Hum. Genet. 19, 105-132.

Prockop, D.J. (1990) Mutations that alter the primary structure of type I collagen. The perils of a system for generating large structures by the principle of nucleated growth. J. Biol. Chem. 265, 15349-15352.

Pulkkinen, L., Savolainen, A., Kaitila, I. and Peltonen, L. (1990) Location on chromosome 15 of the gene defect causing Marfan syndrome. N. Engl. J. Med. 323, 935-939.

Pyeritz, R.E. and McKusick, V.A. (1979) The Marfan syndrome: Diagnosis and management. N. Engl. J. Med. 300, 772-777.

Ramirez, F. and di Liberto, M. (1990) Complex and diversified regulation programs control the expression of vertebrate collagen genes. FASEB J. 4, 1616-1623.

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic. Acids. Res. 13.
Reginato, A.M. and Jimenez, S.A. (1991) Biochemical characterization of the native tissue form of type X collagen from embryonic chick sternal cartilage and identification of a chymotrypsin sensitive site within its triple helical domain. Biochem. J. 273, 333-338.

van der Rest, M., Mayne, R., Ninomiyo, Y., Seidal, N.G., Chretian, M. and Olsen, B.R. (1985) The structure of type IX collagen. J. Biol. Chem. 260, 220-225.

Revie, D., Smith, D.W. and Yu, T.W. (1988) Kinetic analysis for optimization of DNA ligation reactions. Nucleic. Acids. Res. 16, 10301-10321.

Roberts, R.G., Bentley, D.R., Barby, T.F.M., Manners, E. and Bobrow, M. (1990) Direct diagnosis of carriers of Duchenne and Becker muscular dystrophy by amplification of lymphocyte RNA. The Lancet 336, 1523-1526.

Rossi, P. and de Crombrugghe, B. (1987a) 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.

Ryan, M.C. and Sandell, L.J. (1990) Differential expression of a cysteine-rich domain in the amino-terminal propeptide of type II (cartilage) procollagen by alternative splicing of mRNA. J. Biol. Chem. 265, 10334-10339.

Rychilik, R.W., Spencer, W.J. and Rhoads, R.E. (1990) Optimization of the annealing temperature for DNA amplification in vitro. Nucleic. Acids. Res. 18, 6409-6412.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.T., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) PCR directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487-491.

Saitta, B., Stokes, D.G., Vissing, H., Timpl, R. and Chu, M.-L. (1990) Alternative splicing of the human alpha 2(VI) collagen gene generates multiple mRNA transcripts which predict three protein variants with distinct carboxyl termini. J. Biol. Chem. 265, 6473-6480.

Sandberg, M., Tamminen, M., Hirvonen, H., Vuorio, E. and Pihlajaniemi, T. (1989) Expression of mRNAs coding for the α1 chain of type XIII collagen in human fetal tissues: Comparison with expression of mRNAs for collagen types I, II and III. J. Cell Biol. 109, 13371-13379.

Sandell, L.J. and Boyd, C.D., eds. Extracellular matrix genes. San Diego: Academic Press Inc., 1990 (Mecham, R.P., ed. Biology of Extracellular Matrix).

Sarkar, G., Kapelner, S. and Sommer, S.S. (1990) Formamide can dramatically improve the specificity of polymerase chain reaction. Nucleic. Acids. Res. 18, 7465.

Sarkar, G. and Sommer, S.S. (1989) Access to a messenger RNA sequence or its protein protein product is not limited by tissue or species specificity. Science 244, 331-334.

Savagner, P., Miyashita, T. and Yamada, Y. (1990) Two silencers regulate the tissue-specific expression of the collagen II gene. J. Biol. Chem. 265, 6669-6674.

Scharf, S.J., Horn, G.T. and Erlich, H.A. (1986) Direct cloning of enzymatically amplified genomic sequences. Science 233, 1076-1078.

Schmid, M., Popp, R.G. and Linsemayer, T.F. (1990) Hypertrophic cartilage matrix. Ann. NY Acad. Sci. 580, 64-73.

Schnieke.A, Harbers, K. and Jaenisch, R. (1983) Embryonic lethal mutation in mice induced by retroviral insertion in the alpha 1(I) collagen gene. Nature 304, 315-320.

Schuldiner, A.R., Scott, L.A. and Roth, J. (1990) PCR-induced (ligase-free) subcloning: a rapid reliable method to subclone polymerase chain reaction (PCR) products. Nucleic. Acids. Res. 18, 1920.

Serghini, M.A., Ritzenthaler, C. and Pinck, L. (1989) A rapid and efficient 'miniprep' for isolation of plasmid DNA. Nucleic. Acids. Res. 17, 3604.

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) λ Zap: a bacteriophage λ expression vector with in vivo excision properties. Nucleic. Acids. Res. 16, 7583-7600.

Shows, T.B., Tikka, L., Byers, M.G., Eddy, R.L., Haley, L.L., Henry, W.M., Prockop, D.J. and Tryggvason, K. (1989) Assignment of the human collagen α1(XIII) chain gene (COL13A1) to the q22 region of chromosome 10. Genomics 5, 128-133.

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-alpha 2 chain of type I procollagen in a mild variant of osteogenesis imperfecta. J. Biol. Chem. 259, 14094-14100.

Smith, L.T. and Sybert, V.P. (1990) Intra-epidermal retention of type VII collagen in a patient with recessive dystrophic epidermolysis bullosa. J.Invest. Dermatol. 94, 261-264.

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 strands and have overlapping promoter regions. J. Biol. Chem. 263, 17217-17220.

Solomon, E., Hiorns, L.R., Spurr, N., Kurkinnen, M., Barlow, D., Hogan, B.L.M. and Dalgleish, R. (1985) Chromosomal assignments of the genes coding for human types II, III and IV collagen: A dispersed gene family. Proc. Natl. Acad. Sci. USA 82, 3330-3334.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Spotila, L.D., Constantinou, C.D., Sereda, L., Riggs, L. and Prockop, D.J. (1990) Substitution of serine for Gly α2-661 in gene for type I procollagen (COL1A2) as cause of post-menopausal osteoporosis. Am. J. Hum. Genet. 47, A237.

Stacey, A., Bateman, J., Choi, T., Mascara, T., Cole, W.G. and Jaenisch, R. (1988) Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered mutant pro-α1(I) collagen gene. Nature 332, 131-136.

Stoker, A.W. (1990) Cloning of PCR products after defined cohesive termini are created with T4 DNA polymerase. Nucleic. Acids. Res. 18, 4290.

Su, M.-W., Lee, B., Ramirez, F., Machodo, M. and Horton, W. (1989) Nucleotide sequence of the full-length cDNA encoding for human type II procollagen. Nucleic. Acids. Res. 17, 9473.

Superti-Furga, A., Gugli, E., Gitzelmann, R. and Steinmann, B. (1988) A

multi-exon deletion in one of the two COL3A1 alleles affacting structure stability and processing of type III procollagen. J. Biol. Chem. 263, 6226-6232.

Superti-Furga, A. and Steinmann, B. (1990) Alterations in intracellular processing of procollagen I in fibroblasts from patients with osteogenesis imperfecta. Proceedings of the 4th International Conference on Osteogenesis Imperfecta.

Svoboda, K.K., Nishimura, I., Sugrue, S.P., Ninomiya, Y. and Olsen, B.R. (1988) Embryonic chicken cornea and cartilage synthesize type IX collagen molecules with different amino terminal domains. Proc. Natl. Acad. Sci. USA 85, 7496-7500.

Sykes, B., Francis, M., Phil, D. and Smith, R. (1977) Altered relation of two collagen types in osteogenesis imperfecta. N. Engl. J. Med. 296, 1200-1203.

Sykes, B., Ogilvie, D., Wordsworth, P., Wallis, G., Mathew, C., Beighton, P., Nicholls, A., Pope, F.M., Thompson, E., Tsipouras, P., Schwartz, R., Jensson, O., Arnason, A., Borresen, A.-L., Heiberg, A., Frey, D. and Steinmann, B. (1990) Consistent linkage of dominantly inherited osteogenesis imperfecta to the type I collagen loci: COL1A1 and COL1A2. Am. J. Hum. Genet. 46, 293-307.

Tate, V.E., Finer, M.H., Boedtker, H. and Doty, P. (1983) Chick pro α2(I) collagen gene: Exon location and coding potential for the prepropeptide. Nucleic. Acids. Res. 11, 91-104.

Theophilus, B.D.M., Latham, T., Grabowski, G.A. and Smith, F.I. (1989) Comparison of RNase A, chemical cleavage and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid b-glucosidase gene. Nucleic. Acids. Res. 17, 7707-7722.

Thomas, J.T., Kwan, A.P.L., Grant, M.E. and Boot-Handford, R.P. (1991) Isolation of cDNAs encoding the complete sequence of bovine type X collagen. Biochem. J. 273, 141-148.

Tikka, L., Pihlajaniemi, T., Henttu, P., Prockop, D.J. and Tryggvason, K.

(1988) Gene structure for the α1 chain of a human short chain collagen (type XIII) with alternatively spliced transcripts and translation termination codons at the 5' end of the last exon. Proc. Natl. Acad. Sci. USA 85, 7491-7495.

Tiller, G.E., Rimoin, D.L., Murray, L.W. and Cohn, D.H. (1990) Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with spondyloepiphyseal dysplasia. Proc. Natl. Acad. Sci. USA 87, 3889-3893.

Tindall, K.R. and Kunkel, T.A. (1988) Fidelity of DNA synthesis by Thermus aquaticus. Biochemistry 27, 6008-6013.

Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic. Acids. Res. 16, 8186.

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.

Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C.T., Jaenisch, R. and Prockop, D.J. (1988) Structure of a full length cDNA clone for the prepro α1(I) chain of human type I procollagen. Biochem. J. 253, 919-922.

Tromp, G. and Prockop, D.J. (1988) Single base mutation in the pro-α2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in frame pro-α2(I) chain. Proc. Natl. Acad. Sci. USA 85, 5254-5258.

Tromp, G., Kleinert, C., Kuivaniemi, H. and Prockop, D.J. (1991) C to T polymorphism in exon 33 of the COL3A1 gene. Nucleic. Acid. Res. 19, 681.

Trüeb, B. and Winterhalter, K.H. (1986) Type VI collagen is composed of a 200kd subunit and two 140 kd subunits. EMBO J. 5, 2815-2818.

Tsipouras, P. (1990) Marfan syndrome: Light at the end of the tunnel. Am. J.

Hum. Genet. 46, 643-645.

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., Euglielmina, P., Cassidy, S.B. and Ramirez, F. (1986b) Ehlers-Danlos syndrome type IV: Cosegregation of the phenotype to a COL3A1 allele of type III procollagen. Hum. Genet. 74, 41-46.

Tsipouras, P., Scher, C., Ramesar, R.S., Martell, R.W. and Beighton, P. (1990) Spondyloepiphyseal dysplasia, mild autosomal dominant type is not due to primary defects of type II collagen. Am. J. Med. Genet. 37, 272-276.

Uitto, J. (1979a) Collagen polymorphism: Isolation and partial characterization of α1(I) trimer molecules in normal human skin. Arch. Biochem. Biophys. 192, 371-379.

Uitto, J. (1979b) Conversion of type II procollagen to collagen. Eur. J. Biochem. 99, 97-101.

Utani, A., Tanaka, T., Nishigori, C., Miyachi, Y., Danno, K., Imamura, S., Hosokawa, M., Takeda, T., Hirayoshi, K. and Nagata, K. (1990) Another mechanism for the defect in type III collagen accumulation in Ehlers-Danlos syndrome type IV: Increased intracellular degradation of the procollagen. Lab. Invest. 63, 181-188.

Vaage, J. and Lindblad, W.J. (1990) Production of collagen type I by mouse peritoneal macrophages. J. Leukocyte. Biol. 48, 274-280.

Valette, F., Mege, E., Reiss, A. and Adesnik, M. (1989) Construction of mutant and chimeric genes using the polymerase chain reaction. Nucleic. Acids. Res. 17, 723-733.

Vaughan, L., Mendler, M., Huber, S., Bruckner, P., Winterhalter, K.H., Irwin, M.I. and Mayne, R. (1988) D-periodic distribution of collagen type IX along cartilage fibrils. J. Cell Biol. 106, 991-997.

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.

Vieira, J. and Messing, J. (1982) The pUC plasmids and M13mp7 derived system for insertion mutagenesis amd sequencing with synthetic universal primers. Gene 19, 259-268.

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. 263, 19249-19255.

Vuorio, E. and De, Crombrugghe, B. (1990) The family of collagen genes. Annu. Rev. Biochem. 59, 837-872.

Wallis, G.A., Starman, B.J., Schwartz, M.F. and Byers, P.H. (1990) Substitution of arginine for glycine at position 847 in the triple-helical domain of the alpha 1(I) chain of type I collagen produces lethal osteogenesis imperfecta. Molecules that contain one or two abnormal chains differ in stability and secretion. J. Biol. Chem. 265, 18628-18633.

Wallis, G.A., Starman, B.J., Zinn, A.B. and Byers, P.H. (1990) Variable expression of osteogenesis imperfecta in a nuclear family is explained by somatic mosaicism for a lethal point mutation in the Alpha 1(I) gene of (COL1A1) of type I collagen in a parent. Am. J. Hum. Genet. 46, 1034-1040.

Weaver, E.J., King, R.A., Norris, K., Knoblock, W.H., Shamban, A., Jimenez, S.A., Prockop, D.J. and Knowlton, R.G. (1989) Linkage analysis of the type II collagen gene (COL2A1) and hereditary artho-ophthalmology (AOM) in three large families. Cytogenet. Cell. Genet. (Abstract) 51, 1103 .

Weber, S., Engel, J., Weidemann, H., Glanville, R.W. and Timpl, R. (1984) Subunit structure and assembly of the globular domain of basement membrane collagen type IV. Eur. J. Biochem. 139, 401-410.

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. (1988) Cloning and chromosomal localization of human genes encoding the three chains of type VI collagen. Am. J. Hum. Genet. 42, 435-445.

Weil, D., D'Alessio, M., Ramirez, F. and Eyre, D.R. (1990) Structural and functional characterization of a splicing mutation in the pro alpha2(I) collagen gene of an Ehlers-Danlos type VII patient. J. Biol. Chem. 265, 16007-16011.

Wenstrup, R.J., Willing, M.C., Starman, B.J. and Byers, P.H. (1990) Distinct biochemical phenotypes predict clinical severity in non-lethal variants of osteogenesis imperfecta. Am.J. Hum. Genet. 46, 975-982.

Westerhausen, A., Kishi, J. and Prockop, D.J. (1990) Mutations that substitute serine for glycine Alpha 1-598 and glycine Alpha 1-631 in type I procollagen. The effects on thermal unfolding of the triple helix are position-specific and demonstrate that the protein unfolds through a series of cooperative blocks. J. Biol. Chem. 265, 13995-14000.

de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M.-L., Dickson, L., Weil, D. and Ramirez, F. (1987) Organisation of the human proα2(I) collagen gene. J. Biol. Chem. 262, 16032-16036.

de Wet, W., Pihlajaniemi, T., Myers, J., Kelly, T.E. and Prockop, D.J. (1983) 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.

Williams, C.J. and Prockop, D.J. (1983) Synthesis and processing of a type I procollagen containing shortened pro-alpha 1(I) chains by fibroblasts from a patient with osteogenesis imperfecta. J. Biol. Chem. 258, 5915-5921.

Willing, M.C., Cohn, D.H. and Byers, P.H. (1990) Franeshift mutation near the 3' end of the COL1A1 gene of type I collagen predicts an elongated pro α1(I) chain and results in osteogenesis imperfecta type I. J. Clin. Invest. 85,

282-290.

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 characterize point mutations in transcribed genes: Amlpification and overexpression of the mutant c-KI-ras allele in human tumour cells. Proc. Natl. Acad. Sci. USA 82, 7575-7579.

Wirtz, M.K., Glanville, R.W., Steinmann, B., Rao, V.H. and Hollister, D.W. (1987) Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids comprising the N-telopeptide region of pro α2(I) chain. J. Biol. Chem. 262, 16376-16385.

Wirtz, M.K., Keene, D.R., Hori, H., Glanville, R.W., Steinmann, B., Rao, V.H. and Hollister, D.W. (1990) In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptides with mutant pNalpha 2(I) collagen chains in native mutant collagen in a case of Ehlers-Danlos syndrome, type VII. J. Biol. Chem. 265, 6312-6317.

Wittner, C.T. and Garling, D.J. (1991) Rapid cycle DNA amplification: time and temperature optimization. Biotechniques 10, 76-83.

Wohllebe, M. and Carmicheal, D.J. (1978) Type I trimer and type I collagen in neutral-salt-soluble lathyritic-rat dentine. Eur. J. Biochem. 92, 183-188.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann. Hum. Genet. 51, 269-288.

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., Avvedimento, V.E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. and de Crombrugghe, B. (1980) The collagen gene: Evidence

for its evolutionary assembly by amplification of a DNA segment containing an exon of 54bp. Cell 22, 887-892.

Yamada, Y., Kühn, K. and de Crombrugghe, B. (1983) A conserved nucleotide sequence coding for a segment of the C-propeptide is found in the same location in different collagen genes. Nucleic. Acids. Res. 11, 2733-2744.

Yang, R.C-A., Lis, J. and Wu, R. (1979) Elution of DNA from agarose gels after electrophoresis. Meth. Enzym. 68, 176-182.

Yanis-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage coning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 102-119.

Yoshioka, H. and Ramirez, F. (1990) Pro-Alpha 1(XI) collagen. Structure of the amino-terminal propeptide and expression of the gene in tumor cell lines. J. Biol. Chem. 265, 6423-6426.

Young, L.S., Bevan, I.S, Johnson, M.A. Blomfield, P.I. Bromidge, T., Maitland, N.J. and Woodman, C.B.J. (1989) The polymerase chain reaction: a new epidemiological tool for investigating human papillomavirus infection. 298, 14-18.

Zafarullah, K., Keinert, C., Tromp, G., Kuivaniemi, H., Kontusaari, S., Wu, Y., Ganguly, A. and Prockop, D.J. (1990) G to A polymorphism in exon 31 of the COL3A1 gene. Nucleic. Acids. Res. 18, 6180.