Analysis of Type I Collagen Gene Mutations In Osteogenesis Imperfecta

by

Robert Sztrolovics

Department of Medicine, Division of Experimental Medicine

McGill University

and

Shriners Hospital for Crippled Children

Montreal, Quebec

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

June 1992

ISBN 0-315-80397-5

Abstract

This study has optimized techniques for the detection and identification of mutations in the genes encoding type I collagen. Two novel polymorphisms were identified which may serve as useful markers in the linkage analysis of inherited disorders resulting from mutations in these genes. The methods were also used to characterize the gene mutations in four patients with osteogenesis imperfecta, a disease associated with bone fragility of heterogeneous severity. The point mutations demonstrated in these patients resulted in four novel glycine substitutions within the triple helical domains of the type I collagen polypeptide chains, and allowed an evaluation of the relationship between disease severity and the nature of the biochemical defects. The description of these mutations has permitted the identification of three regions in the collagen polypeptide chains in which the disease phenotype exhibits a transition between lethal and non-lethal forms, thereby allowing a more accurate prenatal prognosis of phenotypic severity in osteogenesis imperfecta.

Résumé

Une amélioration des techniques de détection et d'identification des mutations au niveau des gènes codant pour le collagène de type I est décrite. Deux polymorphismes nouveaux ont été identifiés qui pourront être utilisés comme marqueurs dans l'analyse de liaison génétique des maladies héréditaires dues à des mutations de ces gènes. Ces techniques ont aussi permis d'identifier de nouvelles mutations chez quatre patients souffrant d'ostéogénèse imparfaite, une maladie caractérisée par une fragilité osseuse de gravité variable. Ces mutations ponctuelles, qui correspondent à quatre substitutions au niveau de résidus glycine dans les domaines hélicoidaux de la molécule de collagène de type I, ont permis d'évaluer la relation entre la gravité de la maladie et la nature du défaut biochimique. Trois régions dans la molécule de collagène ont pu être ainsi définies associées au passage d'une forme non-létale à une forme létale de la maladie. Cette étude permet ainsi de poser un pronostic prénatal plus précis de la gravité phénotypique du sujet atteint d'ostéogénèse imparfaite.

Dedication

I wish to dedicate this thesis to my parents, Mike and Elianne, and to Petra, without whose support and encouragement, this work would not have been possible.

Acknowledgements

I would sincerely like to thank my co-supervisors, Dr. Michel van der Rest and Dr. Peter J. Roughley, in appreciation of the invaluable support, direction and advice they provided during the course of my research project. I also wish to heartily thank Dr. Francis Glorieux for his continuous support, and for his essential role in providing access to tissue biopsies and clinical evaluations of the osteogenesis imperfecta patients in this study. I would also like to thank all the individuals in our laboratory who helped to create an enjoyable and stimulating working environment.

In addition, the following people must be acknowledged for their various contributions in assisting me to carry out my research:

Drs. William G. Cole and John F. Bateman (Royal Children's Hospital, Melbourne, Australia) for providing the facilities and technical advice to allow one of our laboratory to learn the chemical cleavage technique for mismatch detection;

Dr. Peter H. Byers (University of Washington, Seattle, WA), for providing osteogenesis imperfecta cell lines with characterized mutations to assist in establishing the chemical cleavage technique in our laboratory;

Dr. Moy-Fong Chen (Royal Victoria Hospital, Montreal, QC), for providing skin and bone biopsies and clinical data from a type II osteogenesis imperfecta patient (SS333) included in these studies;

Dr. Leon Kruger (Shriners Hospital, Springfield, MA) for providing skin biopsies and clinical data from three of the osteogenesis imperfecta patients included in these studies (PP238, PM244 and JN245);

Miss Judy Grover, for her invaluable technical expertise and for helping to establish the chemical cleavage method in our laboratory;

Mr. Frank Piccolo, for culturing and screening two of the osteogenesis imperfecta cell lines included in these studies (NC323 and MK345), and for performing some of the chemical cleavage analyses on cell line MK345;

Mrs. Luce Dufour, for the initial screening of a large group of osteogenesis imperfecta cell lines for further study;

Mrs. Janet Moir-Brazeau, for her helpful technical advice;

Mr. Edwin Wan, for synthesizing the oligonucleotides;

Ms. Rose Travers, for performing the bone histology and providing the technical details;

Mr. Guy Charette for preparing the photographs of the bone histological sections;

Ms. Jane Wishart and Mr. Mark Lepik, for performing all the other photography and preparing most of the figures; and

Dr. Brigitte Ecarot, for translating the abstract into french.

This work was funded by the Medical Research Council of Canada and the Shriners of North America.

Abbreviations Used

Amino acids:	Ala	Alanine	Ile	Isoleucine
	Arg	Arginine	Leu	Leucine
	Asn	Asparagine	Lys	Lysine
	Asp	Aspartic acid	Met	Methionine
	Cys	Cysteine	Phe	Phenylalanine
	Gln	Glutamine	Pro	Proline
	Glu	Glutamic acid	Ser	Serine
	Gly	Glycine	Thr	Threonine
	His	Histidine	Trp	Tryptophan
	Hyl	Hydroxylysine	Tyr	Tyrosine
	Hyp	Hydroxyproline	Val	Valine

ATP	Adenosine-5'-triphosphate
bp	Base pairs
CB	Cyanogen bromide
cDNA	Complementary DNA
COL_A_	Collagen gene; eg. COL3A1 gene encodes the prepro-α1(III) chain
dATP	2'-Deoxyadenosine-5'-triphosphate
dCTP	2'-Deoxycytidine-5'-triphosphate
DEAE-	Diethylaminoethyl-
dGTP	2'-Deoxyguanosine-5'-triphosphate
DMEM	Dulbecco's Modified Eagle Medium, buffered at pH 7.2 with 44 mM sodium bicarbonate and 15 mM HEPES
DNA	Deoxyribonucleic acid
dNTP	2'-Deoxyribonucleoside-5'-triphosphate
DPM	Disintegrations/min
DTT	Dithiothreitol
dTTP	2'-Deoxythymidine-5'-triphosphate

ECM	Extracellular matrix
EDS	Ehlers-Danlos Syndrome
EDTA	Ethylenediaminetetraacetic acid, disodium salt
ER	Endoplasmic reticulum
FACIT	Fibril-Associated Collagen with an Interrupted Triple helix
FBS	Fetal bovine serum
HA	Hydroxylamine
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
IL-1	Interleukin-1
MEM	Minimal Essential Medium, supplemented with 1 mM sodium pyruvate, and buffered at pH 7.3 with 25 mM sodium bicarbonate
MOPS	4-Morpholinepropanesulfonic acid, sodium salt
mRNA	Messenger RNA
OI	Osteogenesis imperfecta
Oligo-dT	(2'-Deoxythymidine-5'-monophosphate)$_{12\text{-}18}$, sodium salt
OPC	Oligonucleotide purification cartridge
OT	Osmium tetroxide
PBS	Phosphate-buffered saline (pH 7.4)
pC-collagen	Collagen chain still containing the carboxy-terminal propeptide
PCR	Polymerase chain reaction (as patented by Cetus Corporation)
PEG	Polyethylene glycol
PIPES	1,4-Piperazinebis(ethanesulfonic acid), sodium salt
pN-collagen	Collagen chain still containing the amino-terminal propeptide
Prepro-α	Prepro-α collagen
Pro-α	Pro-α collagen
RFI	Replicative form I

RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNase	Ribonuclease
RNasin	Ribonuclease inhibitor
RPM	Revolutions/min
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SED	Spondyloepiphyseal dysplasia
ssDNA	Single-stranded DNA
TAE	40 mM Tris/20 mM acetic acid/1 mM EDTA
TBE	89 mM Tris/89 mM boric acid/2.5 mM EDTA
TE	10 mM Tris-HCl (pH 8.0)/1mM EDTA
TGF-β	Transforming growth factor-β
TNF-α	Tumor necrosis factor-α
tRNA	Transfer RNA
UDP	Uridine-5'-diphosphate

List of Figures

List of Tables

Table of Contents

Page

Page

Chapter 1

Introduction

1.1 Composition and Function of the ECM

The ECM is a complex mixture of macromolecules surrounding a variety of cell types located in the connective tissues (1). The most extensively studied of these macromolecules have been the collagens, which constitute the major structural elements of connective tissues, including skin and bone (2). Additional types of macromolecules found in the matrix include proteoglycans, adhesive glycoproteins and other specialized proteins, such as elastin (1). The distinct mechanical properties of various connective tissues are determined by the specific composition and organization of their diverse matrix macromolecules.

Since the initial description of the major components of various connective tissues, it has generally been believed that the primary role of the ECM is to provide tissues with the required mechanical properties, such as tensile strength, resistance to compressive loads, extensibility and cohesiveness (3). The matrix also constitutes a structural molecular scaffold for the cellular organization of the connective tissues. However, recent studies have begun to demonstrate the importance of components of the ECM in a wide variety of cellular processes, such as differentiation, migration and the determination of cell shape (4, 5). The studies have revealed an intimate relationship or continuity between the ECM and the membrane receptors, cytoskeleton and biosynthetic machinery of connective tissue cells (5). These findings indicate a critical role for certain matrix components in the embryonic developmental program (6).

The diverse roles of some of these ECM constituents have been examined by studying genetic disorders characterized by abnormalities of the connective tissues. In particular, the significant advances in our knowledge of the structures of the collagens and their genes have allowed great progress in the understanding of a number of such disorders, such as osteogenesis imperfecta (OI) and certain forms of Ehlers-Danlos Syndrome (EDS). Since

the study of one of these genetic disorders, OI, constitutes the central focus of this thesis, the introduction will hereafter concentrate on the collagenous components of the ECM.

1.2 Structures and Functions of the Collagens

The collagens are a family of ECM structural proteins which are characterized by the presence of one or more triple helical domains (7). These proteins are found in all connective tissues, and are critical in conferring distinct mechanical properties to a variety of tissues such as bone, skin and cartilage (3). Fourteen different collagen types have been characterized to date, and others have been described in preliminary reports (7; Table 1). Although they contain a common structural feature, the collagens otherwise exhibit remarkable structural and functional diversity. The collagen family excludes proteins such as acetylcholinesterase, the C1q complement component and the mannose-binding protein, which possess collagen triple helices (8-10) but are not structural elements of the ECM.

The collagen triple helix is a coiled coil structure consisting of three polypeptide chains, each of which forms a left-handed poly-proline-like helix with three residues per turn (3, 11, 12; Fig. 1). The chains are in turn wound around each other to form a tight right-handed superhelix whose interior can only tolerate the presence of the hydrogen atom side chain of glycine residues. Each of the three polypeptides has the side chains of every third amino acid residue facing the helix interior. Therefore, in order for a continuously stable triple helix to form, each chain must possess a $(Gly\text{-}X\text{-}Y)_n$ repeating sequence. The presence of any non-glycine residue at the first position of the tripeptide profoundly affects the formation of a stable triple helical structure. A continuous collagen triple helix appears to exhibit a structure which can be likened to a molecular rod.

Table 1. Summary of the collagen types.

Type	Chains	Trimers Formed	Assemblies Formed	Tissue Source(s)
I	α1(I), α2(I)	$[\alpha1(I)]_2\alpha2(I)$	Fibrils	Skin, bone, dentin, tendon, ligament
		$[\alpha1(I)]_3$		Dentin
II	α1(II)	$[\alpha1(II)]_3$	Fibrils	Hyaline cartilage, vitreous, nucleus pulposus
III	α1(III)	$[\alpha1(III)]_3$	Fibrils	Skin, blood vessels
IV	α1(IV), α2(IV)	$[\alpha1(IV)]_2\alpha2(IV)$	Sheets	Basement membranes
	α3(IV), α4(IV), α5(IV)	?		Glomerular basement membranes
V	α1(V), α2(V), α3(V)	$[\alpha1(V)]_2\alpha2(V)$	Fibrils	Skin, bone
		$[\alpha1(V)]_3$		Hamster lung cell line, liver
		α1(V)α2(V)α3(V)		Placenta, uterus
VI	α1(VI), α2(VI), α3(VI)	α1(VI)α2(VI)α3(VI)	Beaded filaments	Skin, blood vessels, cartilage, etc.
VII	α1(VII)	$[\alpha1(VII)]_3$	Anchoring fibrils	Skin, mucosa
VIII	α1(VIII), α2(VIII)	?	Sheets	Descemet's membrane, endothelial cells
IX	α1(IX), α2(IX), α3(IX)	α1(IX)α2(IX)α3(IX)	Fibril-associated	Hyaline cartilage, vitreous
X	α1(X)	$[\alpha1(X)]_3$	Sheets	Growth plate
XI	α1(XI), α2(XI), α3(XI) [1]	α1(XI)α2(XI)α3(XI)	Fibrils	Hyaline cartilage
XII	α1(XII)	$[\alpha1(XII)]_3$	Fibril-associated	Embryonic skin and tendon
XIII	α1(XIII)	?	?	Endothelial cells
XIV	α1(XIV)	$[\alpha1(XIV)]_3$	Fibril-associated	Fetal skin and tendon

[1] The α3(XI) chain appears to be identical to the α1(II) chain except for post-translational modifications.

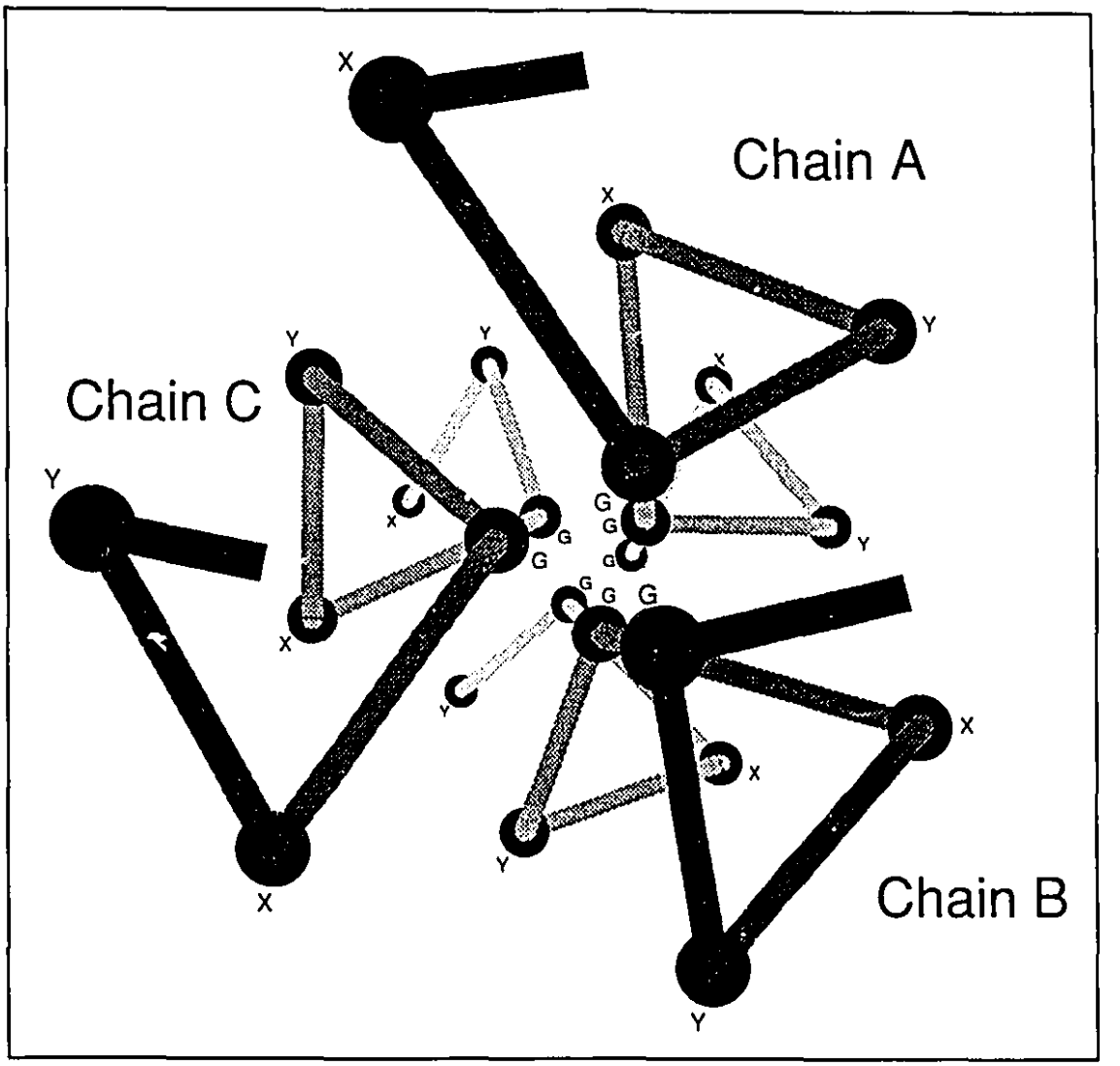


Fig. 1. Structure of the collagen triple helix. Three polypeptide chains, A, B and C, each forming a left-handed poly-proline-like helix, associate into a right-handed superhelix, whose central axis is perpendicular to the plane of the figure. The chains project out of the plane of the figure in an amino- to carboxy-terminal direction. The side group on every third residue of each chain faces the interior of the triple helix, which can only tolerate the presence of the hydrogen atom of glycine residues. Therefore, in order for the chains to form a stable triple helix, they must possess sequences with a $(Gly\text{-}X\text{-}Y)_n$ repeating structure. Reproduced from van der Rest and Garrone (7), with permission.

The X and Y positions of the Gly-X-Y sequence are not restricted to particular amino acids, but are often occupied by proline and hydroxyproline, respectively (3). These imino acid residues, whose peptide bonds limit free rotation, are indispensable for the formation of the poly-proline-like helices. In addition, Y-position hydroxyproline residues contribute further to helix stability, by participating in interchain hydrogen bonding (3). Since, the X- and Y-position residues have their side chains directed outward from the surface of the helix, they allow collagens to participate in many intermolecular interactions.

Different collagen types contain varying numbers of triple helical domains, which also vary in length and are separated by interruptions in the Gly-X-Y repeating sequences of the polypeptide chains (7, 13, 14). These helix interruptions appear to create flexible hinge-like regions between adjacent triple helical domains, and are likely important for their specific functions in the ECM. In addition, some collagen types possess triple helical domains containing short (less than five residues) imperfections in the Gly-X-Y sequences of the polypeptide chains (7, 15, 16), such as Gly-X-Gly-X-Y, which may function as relatively rigid kinks in the triple helical domains.

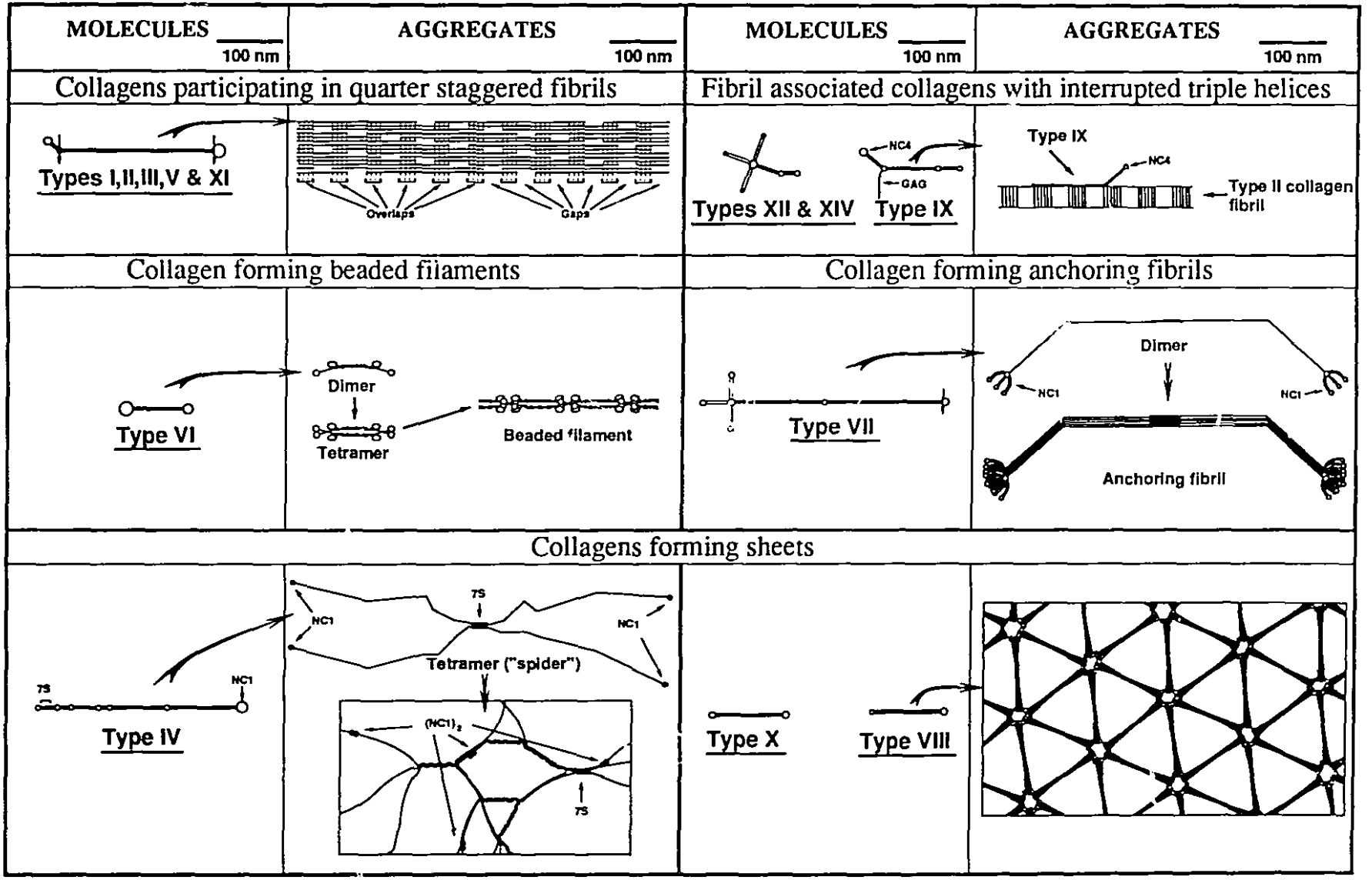
Collagen polypeptide chains are designated by the symbol "α", followed by a chain number, and then by a roman numeral identifying the collagen type. For instance, the designation α3(IV) refers to the α3 chain of type IV collagen. The numbering of the collagen types and their constituent chains is arbitrary. The collagen proteins can consist either of homotrimers involving only one type of chain, such as $[\alpha 1(V)]_3$, or of heterotrimers involving two or three different chain types, such as $[\alpha 1(V)]_2\alpha 2(V)$ or $\alpha 1(V)\alpha 2(V)\alpha 3(V)$ (7, 17). Although the chains of one collagen type may be found in different combinations in different tissues, chains of one collagen type do not normally associate with those of another type, with few exceptions. Each different type of chain is encoded by a distinct gene, such that the constituent chains of the 14 collagen types described to date are encoded by over 25 different genes (7, 18; Table 1).

In addition to containing structural similarities, all of the collagens are involved in the formation of higher-order molecular assemblies by self-interaction, and/or by interactions with other collagen types and other ECM components. On the basis of the types of assemblies formed, the collagens can be divided into distinct groups (7; Fig. 2). The most extensively characterized and most abundant collagens belong to the group which forms quarter-staggered fibrillar structures (I, II, III, V and XI), while other types form sheets (IV, VIII and X), beaded filaments (VI) or anchoring fibrils (VII). Collagen types IX, XII and XIV are referred to as fibril-associated collagens with interrupted triple helices (FACITs), due to their interactions with fibrils of type I or type II collagen.

Fibrillar Collagens

The feature common to these collagens is their ability to form large fibrillar aggregates by lateral association of their long triple helical domains. The most widely accepted model suggests that, within the collagen fibril, individual molecules are assembled into an array in which they are staggered by a distance of 67 nm, which corresponds to 234 amino acid residues (3, 19). This distance is also referred to as one D unit, and accounts for approximately one-quarter the length of the collagen monomer. The collagen monomers consist of a 300-nm (4.4 D units) long triple helical domain, which is flanked by short non-helical ends called telopeptides (20). Some fibrillar collagen types (such as types III, V and XI) may also possess a globular domain at the extreme amino-terminal end (21). The molecules are not precisely end-to-end but are separated by gaps of approximately 40 nm. The association between adjacent collagen molecules is mediated by interactions between homologous domains within the triple helix, and involves charged and hydrophobic side chains of X- and Y-position residues (22, 23). The precise alignment of alternating clusters of charged and hydrophobic residues results in the characteristic striated patterns observed for positively-stained collagen fibrils under the electron microscope (24).

Fig. 2. Structures of the collagen types and of their supramolecular aggregates. The fourteen collagen types which have been characterized to date are divided into five groups based on the types of aggregates formed in the extracellular matrix. The structures of the collagen monomeric molecules are shown, as well as the manner in which they associate to form the various types of molecular aggregates. The collagen monomers are oriented with their amino termini towards the left side of the figure. The small vertical arrows across the fibrillar collagen molecules indicate the regions which undergo proteolytic processing prior to molecular aggregation. Adapted from van der Rest and Garrone (7), with permission.


MOLECULES
100 nm
AGGREGATES
100 nm
MOLECULES
100 nm
AGGREGATES
100 nm
Collagens participating in quarter staggered fibrils
Fibril associated collagens with interrupted triple helices
Types I,II,III,V & XI
Overlaps
Gaps
Types XII & XIV
NC4
GAG
Type IX
Type IX
NC4
Type II collagen fibril
Collagen forming beaded filaments
Collagen forming anchoring fibrils
Type VI
Dimer
Tetramer
Beaded filament
Type VII
Dimer
NC1
NC1
Anchoring fibril
Collagens forming sheets
7S
NC1
Type IV
NC1
7S
NC1
Tetramer ("spider")
(NC1)2
7S
Type X
Type VIII

The striation pattern of a particular collagen type is identical in different tissues, indicating that it is strictly determined by the protein sequence. In addition, the patterns observed for the different collagen types are also similar (25). In contrast, the three-dimensional packing of collagen molecules within fibrils differs between collagen types (2), which may reflect differences in the degree of glycosylation of the collagen monomer (2, 26).

An alternative model for the structure of collagen fibrils proposes that they exhibit an intermediate substructure consisting of microfibrils containing five or more strands (27, 28). Thus, rather than forming by the polymerization of individual molecules, fibrils would be assembled by the incorporation of discrete microfibillar aggregates.

The most extensively characterized member of this group, type I collagen, is a heterotrimer consisting of two $\alpha 1(I)$ chains and one $\alpha 2(I)$ chain, and is the most abundant protein in vertebrates (2, 29). Low quantitites of type I collagen consisting of $[\alpha 1(I)]_3$ have also been detected in certain tissues. Type I collagen is the major collagenous component of most connective tissues, including skin, bone, tendon, dentin and ligament. Type II collagen is a homotrimer, $[\alpha 1(II)]_3$, and is the predominant collagen of hyaline cartilage, the vitreous of the eye, and the nucleus pulposus of intervertebral discs (2, 29). The remainder of the fibrillar collagens are minor components of most connective tissues, and can be divided into two groups according to whether their tissue distributions parallel that of type I or type II collagen (7). Type III collagen is a disulfide-bonded homotrimer (3, 30) which is distributed in most type I collagen-containing tissues with the major exception of bone (2). Although type III collagen is a minor constituent in most of these tissues, it may be a major component in others, such as that surrounding large arteries. Type V collagen is a molecule which can consist of homotrimers or heterotrimers, depending on the tissue or cell line studied, and is usually found in tissues expressing type I collagen (17). Type XI collagen is a heterotrimer consisting of three different chains (31), and which is found in type II collagen-containing tissues (32). The delineation

between collagen types V and XI is unclear since it now appears that they may represent isoforms of the same functional molecule. For example, heterotrimers consisting of α1(V), α2(V) and α1(XI) chains have been described in bovine bone (33).

Although the fibrillar collagens are able to form homotypic fibrils (ie. containing only one type of collagen) *in vitro*, it appears that the fibrils found in tissues are heterotypic in nature (ie. composed of two or more collagen types) (34, 35). For example, in a tissue such as skin, collagen types I, III and V are incorporated into the same heterotypic fibrils, which can then be regarded as copolymers. It is believed that the minor fibrillar collagens are involved in regulating fibrillogenesis, the process by which fibrils are generated (34). Type III collagen may be important in controlling the diameter of type I collagen fibrils (34). The presence of its bulky amino-terminal globular domain may limit the lateral growth of fibrils by interfering with the association of additional type I collagen molecules with the growing fibril. Collagen types V and XI seem to be predominantly located within the cores of type I and type II collagen fibrils (35, 36), respectively, and may therefore be involved in priming fibrillogenesis.

Collagens Forming Sheet-Like Structures

The presence of type IV collagen is restricted to basement membranes, sheet-like molecular networks which serve as barriers between different tissues (37). The type IV collagen monomer consists of a long triple helical region containing a number of imperfections in the Gly-X-Y sequence, and a large carboxy-terminal globular domain (7, 38, 39). The monomers aggregate in a complex process to form a mesh-like network which prevents the passage of high molecular weight molecules (40). The network also involves interactions with other ECM components, such as nidogen (37, 40), laminin (37, 40) and perlecan (41). Thus far, five different chains have been described for type IV collagen (7), although the form present in most basement membranes appears to be

$[\alpha 1(IV)]_2\alpha 2(IV)$ (42).

Type VIII collagen is the major collagenous constituent of Descemet's membrane (43, 44), a structure separating corneal endothelial cells from the corneal stroma (7, 15). Although two chain types have been found in Descemet's membrane type VIII collagen (16, 45), the nature of the homotrimeric or heterotrimeric molecules formed has not been established. Type VIII forms a mesh-like network with a hexagonal arrangement (46), as a result of end-to-end and lateral interactions. The monomer consists of a relatively short interrupted triple helical region flanked by amino- and carboxy-terminal non-helical domains (7, 15, 47).

Type X collagen is a cartilage matrix component with very restricted tissue distribution, being expressed solely by hypertrophic chondrocytes in the growth plate during endochondral ossification (7, 48). It is structurally homologous to type VIII collagen, consisting of a short interrupted helical region flanked by terminal non-helical domains (49, 50). However, type X collagen is a homotrimeric molecule. Type X collagen appears to form sheet-like structures which are similar to those formed by type VIII collagen (51).

Collagen Forming Beaded Filaments

Type VI collagen is a heterotrimer comprised of three different chains (52), and consisting of a short interrupted triple helical region flanked by large amino- and carboxy-terminal globular domains (7, 53-55). Intermolecular interactions involving the triple helical region and globular domains result in the formation of linear structures called beaded filaments (56, 57). The beads result from the large terminal globular domains which are located at regular intervals along the length of the filament. The filaments exhibit a ubiquitous tissue distribution, being observed in all unmineralized connective tissues (55). The function of these structures has not yet been clearly established.

Collagen Forming Anchoring Fibrils

Type VII collagen is a homotrimer consisting of a long interrupted triple helical region and bulky amino- and carboxy-terminal non-helical domains (7, 58). The amino-terminal non-helical domain is particularly large, appearing to consist of three extensions terminating in small globular domains. The monomers associate in an end-to-end fashion to form dimers, which subsequently undergo non-staggered lateral association into structures called anchoring fibrils (59). The ends of the fibrils interact with basement membranes below epithelial cell layers, and with anchoring plaques found in the underlying connective tissue (60). These interactions are mediated at least in part by type IV collagen, which is present in these structures. Therefore, the anchoring fibrils, in tissues such as skin and oral mucosa, are important in anchoring the epithelial cell layer to the underlying connective tissue.

FACITs

The collagens comprising this group are non-fibrillar collagens which are believed to be involved in mediating interactions between fibrils and other ECM components (7). The most extensively-characterized member of this group is type IX collagen (7, 13), a heterotrimer of three different chains (61, 62), which is distributed in type II collagen-containing tissues such as hyaline cartilage and the vitreous of the eye. Type IX collagen contains three helical domains separated by interruptions in the Gly-X-Y sequence (14), and a large amino-terminal globular domain is sometimes present depending on the tissue source (63, 64). A glycosaminoglycan chain can be attached to the α2(IX) chain at one of the helix interruptions (62, 64). Type IX collagen associates with type II collagen fibrils (65, 66), and is thought to interact with other matrix components, such as proteoglycans, perhaps involving the amino-terminal globule (67). Type XII collagen is a homotrimer which exhibits some structural similarities to type IX collagen (7, 68, 69). Consequently, type XII collagen may be the analogous fibril-associated component of tissues containing

type I collagen. This minor collagen has been found in tissues such as embryonic tendon and skin, and periodontal ligament. The most recent addition to the collagen family, type XIV collagen, is a homotrimer displaying strong structural homology to type XII collagen, and has been described in fetal bovine skin and tendon (70). The possible different roles of related collagen types XII and XIV in type I collagen-containing tissues has yet to be established.

1.3 Biosynthesis of Fibrillar Collagens

The biosynthesis of the fibrillar collagens is a complex pathway, consisting of a large number of events occurring within the cell's nuclear and secretory compartments, and in the extracellular space. A knowledge of these events (reviewed in 71, 72) is critical to the understanding of the pathological effects of mutations in disorders such as OI.

Transcription

Like most eukaryotic genes, the fibrillar collagen genes are transcribed by the enzyme RNA polymerase II, and the RNA transcripts are capped and polyadenylated in the same manner as other eukaryotic RNAs (73). As a result of the very complex exon-intron structures of these genes (described in section 1.4), extensive splicing of the primary RNA transcripts is required to generate the mRNAs which are subsequently transported into the cytoplasm. The prepro-α1(I) and prepro-α2(I) mRNAs are present at steady-state levels which are in a 2:1 ratio (74). Each of the genes for type I collagen has been demonstrated to give rise to mRNAs exhibiting size heterogeneity, due to the use of different polyadenylation signals (75, 76). These mRNA size variations are therefore reflected in the lengths of the 3' untranslated region. The COL1A1 gene gives rise to two mRNAs, while

the COL1A2 gene gives rise to three.

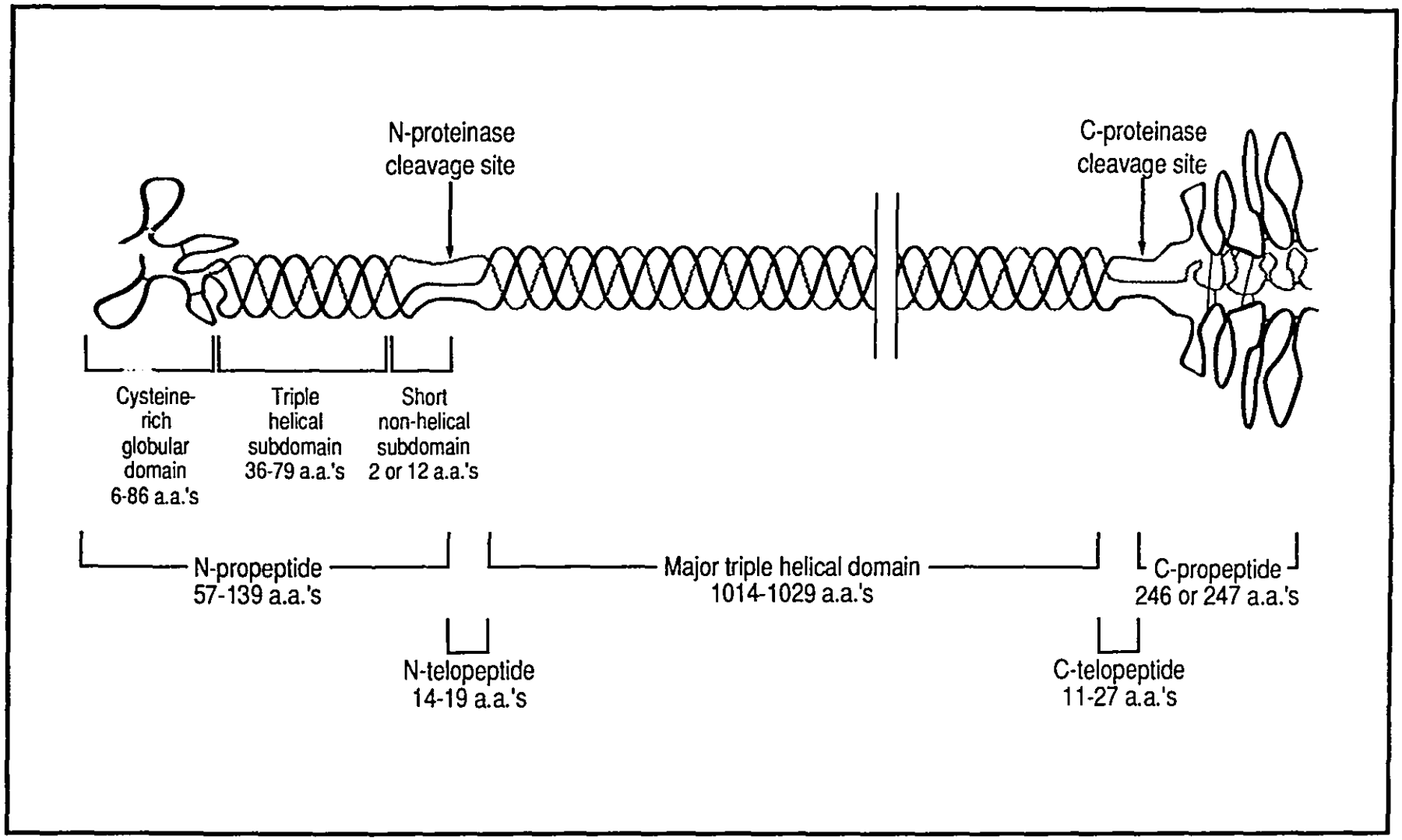
Translation of mRNAs

The mRNAs transcribed from the COL1A1, COL1A2, COL2A1 and COL3A1 genes are translated into polypeptides ranging from 1365 to 1467 residues in length (30, 77-80, 81, 82). In addition to containing the triple helical and telopeptide regions of the collagen chains found in tissues, they contain additional structural domains which are required for cellular biosynthesis and assembly. These primary translation products are called prepro-α chains, and consist of the following domains, in an amino- to carboxy-terminal direction: signal peptide, amino-terminal propeptide and telopeptide, main triple helical domain, and carboxy-terminal telopeptide and propeptide (Fig. 3). The domain descriptions which follow are based on comparisons of the prepro-α chains for collagen types I, II and III, whose structures have been characterized in the most detail (71).

The signal peptides are from 21 to 25 amino acid residues in length and are very hydrophobic in nature. In these regards, the signal sequences are similar to those found in other secreted proteins, which are typically 15 to 35 residues in length (83).

The amino-terminal propeptide is heterogeneous in size and structure between the four different chains. It ranges from 57 to 139 residues in length, and generally consists of three subdomains: an amino-terminal globular subdomain, a short triple helix, and a short non-helical subdomain. The globular subdomain contains a large number of cysteine residues which participate in intrachain disulfide bonds. This domain is absent in all pro-α2(I) chains (80) and some pro-α1(II) chains (81). The triple helical subdomain ranges from 36 to 79 residues in length, and contains a 4-amino acid imperfection in the pro-α1(II) chain (81). The non-helical subdomain is either 2 or 12 residues in length, and is bordered on its carboxy-terminal side by the site for the extracellular proteolytic removal of the amino-terminal propeptide.

Fig. 3. Domain structure of the type I procollagen molecule. Type I procollagen consists of two pro-α1(I) chains (shown in black) and one pro-α2(I) chain (shown in grey) which associate following translation. Each chain type consists of an amino-terminal propeptide and telopeptide, a main triple helical domain, and a carboxy-terminal telopeptide and propeptide. The amino-terminal signal peptide sequences present in the prepro-α chains are removed in the ER. The boundaries between the telopeptide and propeptide domains are defined by the sites of extracellular proteolytic cleavage, which result in the formation of the type I collagen monomer. The amino-terminal propeptide domain of each chain contains a short non-helical region and a minor triple helical subdomain. The pro-α1(I) chain also contains a terminal cysteine-rich globular subdomain which is absent in pro-α2(I) chains. The vertical double-line indicates that the length of the helix is not drawn to scale. The lines connecting different chains or different regions of individual chains represent interchain and intrachain disulfide bonds.


N-proteinase cleavage site
C-proteinase cleavage site
Cysteine-rich globular domain 6-86 a.a.'s
Triple helical subdomain 36-79 a.a.'s
Short non-helical subdomain 2 or 12 a.a.'s
N-propeptide 57-139 a.a.'s
Major triple helical domain 1014-1029 a.a.'s
C-propeptide 246 or 247 a.a.'s
N-telopeptide 14-19 a.a.'s
C-telopeptide 11-27 a.a.'s

The amino-terminal telopeptide is from 14 to 19 residues in length and is bordered on its amino-terminal side by the proteolytic cleavage site. The telopeptide contains a lysine residue which participates in extracellular covalent cross-linking (84).

The main triple helical domain of these polypeptides ranges from 1014 to 1029 residues in length, and consists of an uninterrupted Gly-X-Y sequence. This domain contains a site for cleavage by the mammalian collagenase, an enzyme with a critical role in collagen degradation. The cleaved peptide bond is between residue Gly_{775} and residue Ile_{776} or Leu_{776} of the triple helical domain of the α1(I) or α2(I) chain, respectively (77, 79).

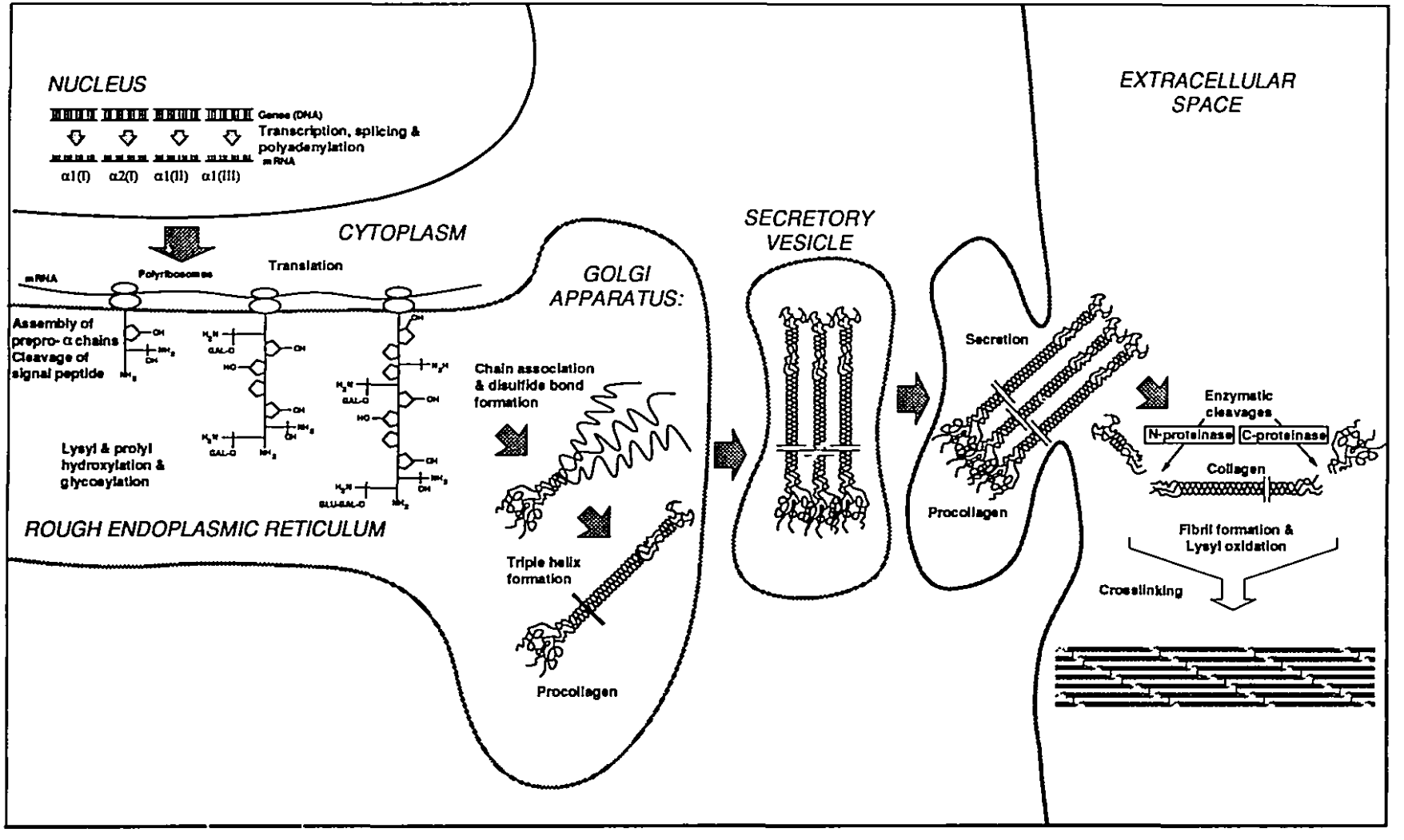
The carboxy-terminal telopeptide is from 11 to 27 residues in length. It is bordered by the site for the extracellular proteolytic removal of the carboxy-terminal propeptide. In addition, all but the pro-α2(I) chain contain a lysine residue which is involved in cross-linking (84).

The carboxy-terminal propeptide is a domain which is very similar in size and structure between the different collagen pro-α chains. It is 246 or 247 residues in length, and contains 7 or 8 cysteine residues at conserved positions. These residues participate in intrachain and interchain disulfide bonds. This domain also contains one or two consensus sites for asparagine N-glycosylation.

Cleavage of Signal Sequences

The primary translation products are synthesized on membrane-bound ribosomes (Fig. 4). The hydrophobicity of the amino-terminal sequence of the nascent polypeptides is thought to direct the polysomes to the ER by interacting with the membrane surface and subsequently penetrating the lipid bilayer. The signal peptide then allows the co-translational translocation of the nascent polypeptide chains into the lumen of the ER (85, 86). The signal peptides appear to be removed co-translationally by an endopeptidase which is likely associated with the ER membrane (87). The sequence requirements of the

Fig. 4. Biosynthesis of the fibrillar collagens. The chains comprising collagen types I, II and III are synthesized as prepro-α precursor polypeptides, and are co-translationally inserted through the membrane of the rough ER. During and after translation, the polypeptide chains undergo extensive processing in the ER lumen, including the removal of their signal sequences, the hydroxylation of proline and lysine residues, and the glycosylation of hydroxylysine residues. When translation is complete and the chains have been sufficiently hydroxylated, they associate into trimeric procollagen molecules, which are packaged into vesicles and secreted into the extracellular space. The terminal propeptide sequences are removed extracellularly by two endoproteinases, resulting in the formation of the collagen monomer. The monomer self-associates to form fibrillar structures which are stabilized by the formation of covalent cross-links.


NUCLEUS
Genes (DNA)
Transcription, splicing & polyadenylation
mRNA
α1(I) α2(I) α1(II) α1(III)
CYTOPLASM
mRNA
Polyribosomes
Translation
Assembly of prepro-α chains
Cleavage of signal peptide
Lysyl & prolyl hydroxylation & glycosylation
ROUGH ENDOPLASMIC RETICULUM
GOLGI APPARATUS:
Chain association & disulfide bond formation
Triple helix formation
Procollagen
SECRETORY VESICLE
Secretion
Procollagen
EXTRACELLULAR SPACE
Enzymatic cleavages
N-proteinase
C-proteinase
Collagen
Fibril formation & Lysyl oxidation
Crosslinking

proteases involved are unclear, since it appears that a single enzyme may act on substrates containing different sequences around their cleavage sites.

Hydroxylation of Proline and Lysine Residues

Hydroxyproline and hydroxylysine constitute quantitatively and functionally significant features of fibrillar collagen chains (2, 88). These residues are not incorporated during translation, but are formed by the hydroxylation of incorporated proline and lysine residues. The hydroxylations occur on the nascent polypeptide chains during and after translation, and are catalyzed by three separate enzymes (72). Peptidyl proline-4-hydroxylase, peptidyl proline-3-hydroxylase and peptidyl lysine hydroxylase are all found within the ER or associated with its membrane, and require Fe^{2+}, α-ketoglutarate, molecular oxygen and ascorbic acid for activity (72, 89). The enzymes catalyze the oxidative decarboxylation of α-ketoglutarate, with the incorporation of an oxygen atom into the hydroxyl group of the proline or lysine residue. Ascorbic acid is required for hydroxylation, but is not consumed stoichiometrically. The ascorbate is believed to be necessary to maintain the 2+ oxidation state of the enzyme-bound iron atom. Thus iron chelators such as α,α'-dipyridyl interfere with normal collagen biosynthesis.

The sequence requirements for proline 4-hydroxylation and lysine hydroxylation appear to be the tripeptide sequences -X-Pro-Gly- and -X-Lys-Gly-, respectively, where the identities of the X-position residues and of neighboring residues affect the rate of hydroxylation (90). These sequence preferences result in the selective proline 4-hydroxylation and lysine hydroxylation at the Y position of the collagen -Gly-X-Y- tripeptide repeat. However, there are exceptions, such that lysine residues within the amino- and carboxy-terminal telopeptides are hydroxylated despite being present in -X-Lys-Ser- or -X-Lys-Ala- tripeptide sequences. The sequence preferences for proline 3-hydroxylation appear to be the tripeptide -Pro-4-Hyp-Gly, with the rate of hydroxylation

being affected by neighboring amino acid residues. The chains of type I collagen each contain only one such residue (2). The hydroxylation of specific proline or lysine residues appears to be determined by the different affinities of the enzymes for the surrounding sequences. The activities of the three hydroxylating enzymes are dependent on the conformation of the substrate, since they can only act on the nascent polypeptide chains prior to the formation of the triple helix. As soon as a sufficient number of Y-position proline residues have been 4-hydroxlated (approximately 100 residues per chain), the triple helix spontaneously forms, thereby preventing further hydroxylation. Unhydroxylated chains can only form stable triple helices at temperatures below approximately 25°C (91). The formation of a triple helix at physiological temperature, however, requires the hydroxylation of over 90% of all Y-position proline residues. The level of hydroxylation of Y-position lysine residues in the α1(I) and α2(I) chains is about 20% and 30%, respectively (2).

O-Glycosylation of Hydroxylysine Residues

Some of the Y-position hydroxylysine residues are glycosylated with α-D-galactose and sometimes subsequently with α-D-glucose. The disaccharides are covalently linked to the hydroxyl oxygen atom of hydroxylysine residues by the sequential action of two glycosyltransferases found within the ER (72, 92). Peptidyl galactosyl transferase first transfers a galactose moiety to a hydroxylysine residue. In the second step, peptidyl glucosyl transferase tranfers a glucose sugar to the peptide-bound galactose. Both of these enzymes utilize UDP-sugars as donors for the glycosylation reactions. The enzymes require a divalent cation such as Mn^{2+}. The sequence requirements for glycosylation appear to be the tripeptide -X-Hyl-Gly-, such that the oligosaccharides are selectively conjugated to Y-position residues. The neighboring amino acid residues may affect the rates of the glycosylation reactions.

The role of the carbohydrates is unclear at present. However, several possibilities have been postulated. Since these bulky groups are located on the surface of the collagen molecule and are likely hydrated, they may have a role in the packing of the molecules into fibrils, such as regulating fibril diameter (2,). Although glycosylated hydroxylysine residues may participate in initial cross-link formation, the carbohydrate may have an effect on cross-link maturation, and therefore on fibril stability (84). However, the degree to which the carbohydrate affects these processes is unclear. Glycosylation may also affect collagen degradation and turnover, by regulating the sensitivity of fibrils to collagenase cleavage (72). In addition, it is possible that these carbohydrate structures provide a scaffold for cellular attachment and migration.

N-Glycosylation of Asparagine Residues

The carboxy-terminal propeptide domain, and in some chains the amino-terminal propeptide domain, contains high-mannose oligosaccharide structures (2, 93), which are first assembled on the lipid carrier dolichol phosphate in the membrane of the ER, in common with other N-linked glycosylations (94, 95). The lipid-linked structure, consisting of N-acetylglucosamine, mannose and glucose residues is synthesized by the action of a number of different glycosyltransferases. The oligosaccharide is then transferred to specific asparagine residues by the action of an enzyme called oligosaccharide transferase (94, 95). The consensus acceptor sequence for N-glycosylation is -Asn-X-Ser/Thr-. The acceptor sequences for the pro-α1(I) and pro-α2(I) chains are Asn-Val-Thr and Asn-Ile-Thr, respectively. Once transferred, the oligosaccharide may then be further processed in the rough ER and Golgi (72).

The functions of the N-linked oligosaccharides have not been clearly established, although a role in the rate of intracellular transport and of secretion has been postulated (96). In addition, the carbohydrate moieties could be important in the process of chain

recognition and association, which leads to the assembly of the procollagen molecule, although this has not yet been demonstrated.

Assembly of Chains into Procollagen

Following translation, the completed pro-α chains are able to associate into trimeric procollagen molecules. It is not clear whether the chains associate while still membrane-bound, or following release into the ER lumen. The demonstration of a lag period between the end of translation and the formation of triple helical procollagen molecules, indicates that the chains do not associate until they have been completely synthesized. Although the exact sequences involved in the initial chain association have not yet been identified, the chains are known to initially become associated through regions within the carboxy terminal propeptides (91). These sequences are therefore critical for ensuring the correct chain stoichiometry and registration for normal procollagen formation. The initial chain association is followed by a stabilization which is mediated by the formation of interchain disulfide bonds (97). When a sufficient number of Y-position proline residues have been 4-hydroxylated, the triple helical domain is formed by a spontaneous entropy-driven process. Helix formation appears to initiate in the region immediately adjacent to the carboxy telopeptide, and is believed to propagate in a zipper-like fashion towards the amino terminus (98, 99). The helical region adjacent to the carboxy-terminal telopeptide seems to be ideally suited for inititating helix formation, since it contains a number of consecutive Gly-Pro-Hyp tripeptides which are most favorable for helix formation and stability.

Two enzymes may be involved in the folding of prepro-α chains to form stable triple helical molecules. Protein disulfide isomerase may participate to ensure that the "correct" disulfide bonds are formed within the carboxy-terminal propeptide region (72). Since the carboxy-terminal propeptide regions contain a number of cysteine residues, it is possible that the chains can form "incorrect" disulfide bonds during chain association. The isomerase activity, which may reside in the β subunit of peptidyl proline 4-hydroxylase

(89), may then assist in the reformation of the "correct" bonds. In addition, the enzyme prolyl *cis-trans* isomerase may be important in the formation of the triple helix by accelerating the isomerization of prolyl peptide bonds in pro-α chains (100). Only the *trans* isomer is compatible with the conformation of the triple helix.

The folding of the pro-α polypeptide chains into procollagen molecules is believed to occur within the lumen of the ER. However, it is still possible that this event may continue during intracellular transport to the Golgi. The formation of a stable triple helix blocks the occurrence of any further post-translational hydroxylation or glycosylation of helical proline and lysine residues (72).

Intracellular Transport and Secretion

In order to be transported and secreted at a normal rate, procollagen chains must be assembled into a stable triple helical molecule (72). Chains which remain unassociated are not secreted efficiently, and are degraded intracellularly. The manner in which procollagen molecules are packaged for secretion is at present unclear. Although, intracellular procollagen molecules are unable to form fibrils, it has been suggested that procollagens may be secreted in the form of aggregates which are distinct from the fibrillar structures found in the ECM (101).

Proteolytic Removal of Propeptide Sequences

Following secretion, the terminal propeptide sequences are proteolytically removed by the action of two enzymes, procollagen N-proteinase and C-proteinase (72, 102, 103). These endopeptidases act at neutral pH and require divalent cations. The N-proteinase appears to be sensitive to changes in the amino-terminal propeptide conformation, which in turn requires the presence of a stable triple helix. Thus, normal cleavage of the amino-terminal propeptide requires that the three chains form a stable triple helix. Numerous *in*

vitro studies have presented conflicting results concerning the sequence in which the propeptides are removed, indicating processing via either a pN-collagen or pC-collagen intermediate (104, 105). It is therefore possible that the proteinases act independently of each other, with the order in which the propeptides are cleaved depending on the relative activities of the proteinases in the cell culture system or tissue examined. There appear to be collagen type-specific differences in the rates at which the amino-terminal propeptides are removed. For example, type III procollagen, which requires a different N-proteinase from procollagen types I and II, undergoes very slow processing at the amino terminus (106). This is believed to be related to the ability of different collagen types to form fibrils of different diameters, and is supported by the observation that pN-collagen forms abnormally thin fibrils (107).

In addition to their role in preventing intracellular fibrillogenesis, the propeptides may have other important functions, such as in the regulation collagen synthesis. For instance, the amino-terminal propeptide has been demonstrated to participate in the feedback inhibition of type I procollagen synthesis *in vitro* (108). This function has been localized to the cysteine-rich globular subdomain of the propeptide.

Fibrillogenesis

The removal of the terminal propeptides from the secreted procollagen molecules results in the production of the collagen monomer, which can then form fibrils by spontaneous lateral association (109, 110). No enzyme appears to be required for fibrillogenesis. The collagen molecules are held together by electrostatic interactions between charged groups and by interactions of hydrophobic surfaces of adjacent molecules (23, 111). Since the side chains of all X- and Y-position residues extend from the surface of the helix (3), the collagen molecule is capable of a very large number of such interactions. The molecules align themselves within the fibril (or microfibril) with a quarter staggered array, which is

precisely determined by the distribution of charged and hydrophobic amino acids along the triple helix. The highest alignment of polar and hydrophobic residues from adjacent molecules occurs when the collagen monomers are shifted by 234 amino acids relative to each other, for collagen types I and III (3, 111).

Removal of the carboxy propeptide is absolutely required for fibrillogenesis to take place. However, pN-collagen is able to form fibrils, although the fibrils formed are thinner and less tightly packed (107). This suggests that the amino-terminal propeptide may be involved in regulating fibril diameter. Furthermore, the order in which the propeptides are removed may be equally important, since this would determine whether pN-collagen or pC-collagen is formed as an intermediate. Only the former is able to form fibrils.

Crosslinking of Collagen Molecules

Following fibrillogenesis, the collagen fibrils are further stabilized to mechanical and thermal disruption by the formation of a number of different types of intramolecular and intermolecular crosslinks (84, 112). The crosslinks contribute to the mechanical tensile strength required by various connective tissues. The process of fibril crosslinking consists of three phases (72, 84): 1) oxidative deamination of a telopeptide lysine or hydroxylysine residue by peptidyl lysine oxidase to create an allysine or hydroxyallysine residue possessing a reactive aldehyde group; 2) reaction of the aldehyde group with another reactive group (aldehyde or amine) within the same collagen molecule or within an adjacent 4-D staggered collagen molecule; and 3) maturation of the crosslinks from reducible to non-reducible forms. The sites involved in the initial formation of immature reducible crosslinks include the lysine (or hydroxylysine) residue within each of the amino- and carboxy-terminal telopeptides, and the hydroxylysine residues at positions 87 and 930 of the triple helical domain (84). The only exception is the $\alpha 2(I)$ chain which lacks the carboxy-terminal telopeptide site. Additional residues are involved in the formation of the

mature cross-links, including histidines surrounding the newly formed difunctional crosslinks, and the hydroxylysine residues of adjacent collagen molecules.

The process of collagen cross-linking begins with the action of peptidyl lysine oxidase, a copper-dependent enzyme which converts the ε-amino group of a lysine or hydroxylysine residue to a reactive aldehyde group (113, 114). The enzyme acts preferentially on collagen in fibrils and cannot efficiently oxidize residues in monomeric collagen (115). The sequences adjacent to the helix hydroxylysine residues involved in cross-links are conserved, suggesting that they may constitute the enzyme-binding site (116). After binding, the enzyme then oxidizes a lysine or hydroxylysine residue in the telopeptides of an adjacent 4-D staggered molecule.

Two major pathways have been described for cross-link formation, depending on whether the reactive telopeptide residue participating in the initial reaction is allysine or hydroxyallysine (84). In the allysine pathway, a reactive allysine residue may undergo an aldol condensation with another such residue of an adjacent α chain in the same molecule, resulting in the formation of an intramolecular allysine aldol crosslink. An alternative reaction of allysine residues, or the major reaction of hydroxyallysine residues in the hydroxyallysine pathway, is the condensation with a helix lysine or hydroxylysine residue of an adjacent 4-D staggered molecule, resulting in intermolecular aldimine (or ketoamine) crosslinks. In weight-bearing tissues, such as bone and cartilage, the hydroxyallysine pathway predominates (84, 117), whereas the allysine pathway predominates in skin. These tissue-specific differences in crosslink formation appear to be determined at the level of hydroxylation of the telopeptide lysine residue (118, 119), since this is the major difference between the two pathways.

The reducible aldol and aldimine (or ketoamine) crosslinks initially formed mature into non-reducible trifunctional and tetrafunctional crosslinks. This occurs because aldol and aldimine (or ketoamine) cross-links still contain groups which can further react with the amine groups of nearby lysine, hydroxylysine, or histidine residues. The major mature

crosslinks resulting from the hydroxyallysine pathway are the trifunctional residues hydroxylysyl pyridinoline (HP) and lysyl pyridinoline (LP) (120-122), which are postulated to arise from the condensation of two ketoamine cross-links and the elimination of a hydroxylysine residue. An additional cross-link, the Ehrlich chromogen (EC), has been demonstrated in tendon, and is thought to result from the condensation of a ketoamine cross-link with an allysine residue (123). One of the mature cross-links resulting from the allysine pathway is the trifunctional residue histidino-hydroxylysinonorleucine (HHL) (124), from the reaction of an aldimine cross-link with a histidine residue.

The manner in which molecules are packed within the collagen fibril may determine the type of cross-links formed, by affecting the distances between reactive groups. Thus, factors such as the degree of glycosylation, which varies between collagen types and between tissues, may have a role in determining the types of cross-links formed.

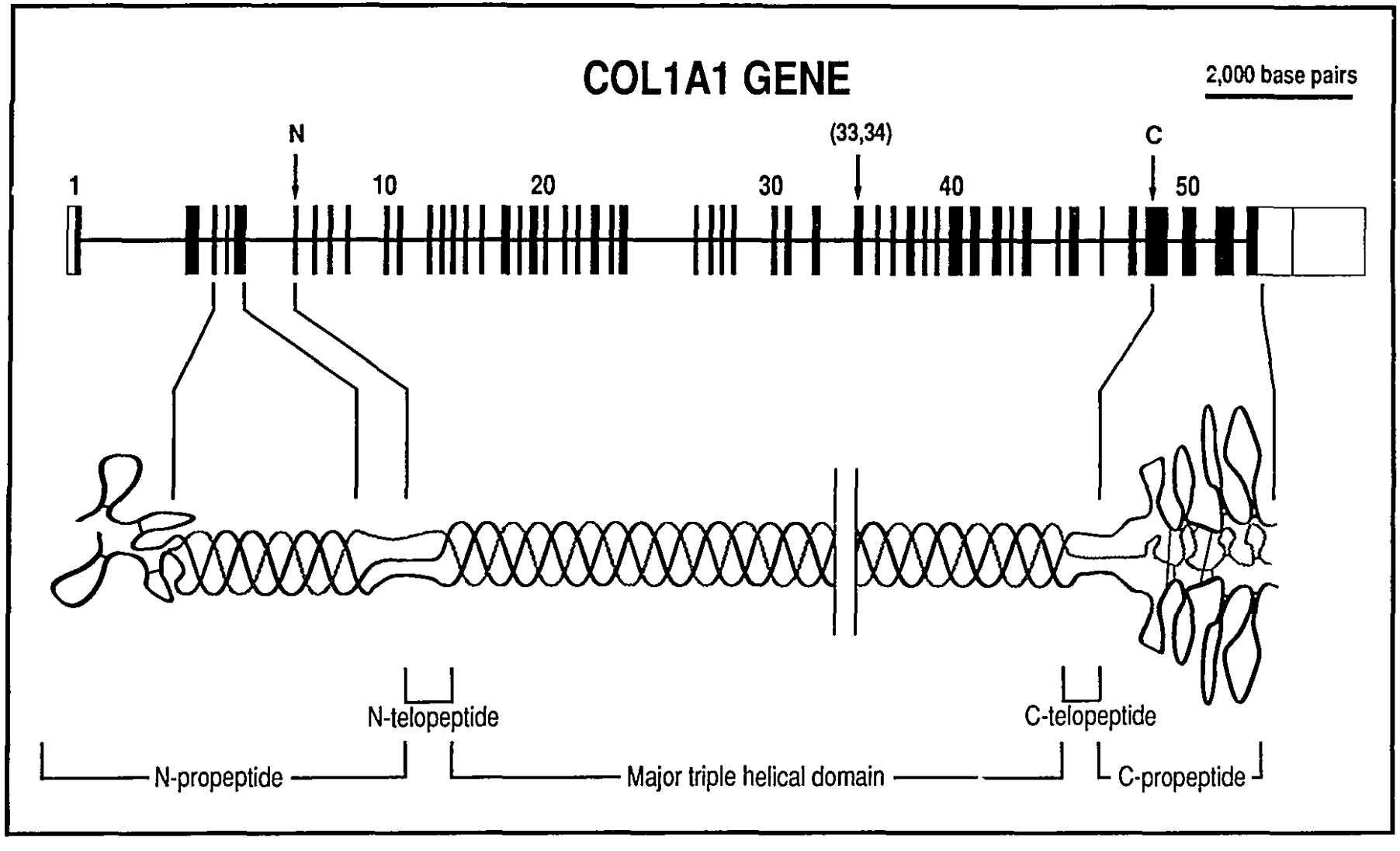
The importance of crosslinks in stabilizing collagen fibrils is illustrated by the connective tissue abnormalities induced in animals which are copper-deficient or which are treated with inhibitors of lysyl oxidase (125, 126). In addition, the skeletal manifestations of two genetic disorders in man have been attributed to abnormalities in collagen cross-linking. Menkes' Kinky Hair Syndrome is an X-linked disorder associated with defective intestinal absorption of copper, and its accumulation in mucosal cells (127). The disorder results in characteristically abnormal hair, neuronal degeneration and osteoporosis. The copper which is present within cells appears to be in a location or form which renders it unusable by copper-dependent enzymes. Homocystinuria is a disorder of methionine metabolism characterized by osteoporosis, kyphosis and scoliosis. In this disorder, the accumulation of homocysteine may interfere with the oxidation of lysine residues, or with the maturation of immature cross-links (128).

1.4 Structures of Fibrillar Collagen Genes

Collagen genes are designated by the letters "COL" followed by the collagen type number, the letter "A", and the number corresponding to the specific prepro-α chain. This introduction will focus on what is known about the genes encoding the chains for the fibrillar collagen types I, II and III, whose structures have to date been described in the most detail. In fact, the sequence of the human COL1A1 gene, whose structure is shown in Fig. 5, has been reported in its entirety (129, 130). In addition, cDNA sequences for the prepro-α1(I), prepro-α2(I), prepro-α1(II) and prepro-α1(III) chains have been published, allowing a detailed comparison of their structural domains (30, 77-80, 81). Characteristic features of the COL1A1, COL1A2, COL2A1 and COL3A1 genes are their large size, ranging from 16 to 44 kb, and their large number of exons, ranging from 51 to 54 (18, 71). Since the differences in the sizes of the coding sequences of these collagen genes are relatively small, the large discrepancies in gene sizes are accounted for by the variability in intron sizes.

The major triple helical domains of collagen types I, II and III are encoded by 43 or 44 exons. A large proportion of these exons are 54, 108 or 162 bp in size (coding for 18, 36 or 54 amino acids), suggesting that they were derived by the amplification of a common ancestral 54-bp DNA sequence (131), followed by sequence divergence as a result of point mutations. Exons with sizes of 45 or 99 bp are believed to have arisen by a process of homologous recombination between 54- or 108-bp exons. An important feature of exons encoding the triple helical domain is that they all begin with a complete glycine codon and end with the complete codon of a Y-position residue. The differences in the number of residues in the triple helical domains of the various collagen chains are due mainly to the presence of different numbers of Gly-X-Y triplets encoded by the exons covering the helix-telopeptide junction regions. With only two minor exceptions, the patterns in the sizes of

Fig. 5. Structure of the COL1A1 gene. The exon-intron structure of the human COL1A1 gene is illustrated, and the various regions of the pro-α1(I) chain encoded by the exons are indicated. The gene is represented by the horizontal line, and the positions and sizes of the 51 exons are denoted by vertical bars. Exon 33/34 (108 bp) has a sequence homologous to exons 33 and 34, each 54 bp in length, of the COL1A2 gene. The solid portions of the bars indicate translated sequences, whereas the open portions represent the 5' and 3' untranslated regions. The presence of two open regions within the last exon indicates the use of two different polyadenylation signals in the prepro-α1(I) mRNA. Most of the exons encoding the triple helical domain are 54 bp in length, or multiples of 54 bp.


COL1A1 GENE
2,000 base pairs
N
(33,34)
C
1
10
20
30
40
50
N-telopeptide
C-telopeptide
N-propeptide
Major triple helical domain
C-propeptide

the exons encoding the triple helical domains are identical between different genes. The COL1A1 gene contains 43 exons coding for the triple helix, due to the replacement of two 54-bp exons, found in the COL1A2, COL2A1 and COL3A1 genes, with a single 108-bp exon (130). In addition, one COL3A1 helix-encoding exon is lengthened by 9 bp, encoding for an additional complete Gly-X-Y triplet (30).

The domain with the highest degree of conservation of exon organization between collagen chains is the carboxy-terminal propeptide (18). The differences in size of this domain between collagen chains are due to variations in the lengths of the two exons coding for the region adjacent to the triple helical domain. The junction exon codes for the end of the triple helical domain, the carboxy-terminal telopeptide and the start of the carboxy-terminal propeptide. The sizes of the two exons coding for the more carboxy-terminal propeptide regions are more highly conserved, reflecting the stringent conservation of protein sequence in this functionally-important domain.

The patterns in the organization of exons encoding the amino-terminal propeptide domains are much more heterogeneous than of those encoding the other domains. This indicates a less stringent requirement for conservation of the sequences and sizes of exons encoding this domain. The amino-terminal propeptide is encoded by 5 to 8 exons, some of whose sizes vary considerably between collagen genes. The exon encoding the cysteine-rich globular domain present in the pro-α1(I) and pro-α1(III) chains is replaced by two exons of much smaller size in the COL1A2 gene (18, 80), explaining the decreased size and cysteine content of the pro-α2(I) chain. Furthermore, this exon appears to be differentially spliced in pre-mRNA transcribed from the COL2A1 gene (18, 81). The triple helical subdomain of the amino-terminal propeptide varies in length between collagen chains, due to variations in the number (2 to 5) and sizes of the exons encoding this region. Further variations in the sizes of the signal peptides and amino-terminal telopeptides, between different collagen chains, are likewise determined by differences in the sizes of

exon 1 and the amino-terminal junction exon, respectively.

Despite the structural differences described above, the structures of the COL1A1, COL1A2, COL2A1 and COL3A1 genes are nonetheless remarkably similar, indicating the likelihood of a common ancestor. It appears that this ancestral gene existed before the divergence of vertebrates and invertebrates. It is also apparent that the strong conservation in structure exhibited at the protein level for fibrillar collagen types I, II and III, is likewise displayed at the level of gene organization.

1.5 Regulation of Fibrillar Collagen Synthesis

Regulation of Gene Transcription

The control of fibrillar collagen gene expression appears to be exerted predominantly at the level of transcription (18). The promoter DNA sequences upstream of the COL1A2 and COL2A1 gene transcription start sites have been shown to be sufficient to confer tissue-specific expression in transgenic mice (132, 133). A number of specific transcription factor-binding sites have been identified in the COL1A1 and COL1A2 gene promoters (134, 135). In some cases, the corresponding DNA-binding factors have also been characterized (136). These promoter elements and their corresponding *trans*-acting factors are likely involved in the co-ordinate tissue-specific expression of the different fibrillar collagen genes (137). For example, collagen types I and III are usually co-expressed in the same tissues. However, in mature bone, type I collagen is by far the major collagen expressed. These tissue-specific differences in co-ordinate gene regulation may be explained by the presence of distinct promoter elements in the genes for type I and type III collagen (18), or perhaps by the differential expression of the corresponding DNA-binding factors. The responses of the fibrillar collagen genes to cytokines and hormones is likely

also mediated by such various promoter elements. For example, mediators such as TGF-β (138), IL-1 (139), TNF-α (140) and glucocorticoids (141), which affect the levels of the type I collagen mRNAs, probably result in alterations in the pattern of binding of various *trans*-acting factors to positive and negative promoter elements. There is significant evidence for the role of a specific DNA-binding protein and promoter element in the induction of COL1A2 gene expression by TGF-β in fibroblastic cells (142).

Other regulatory elements, located within the first intron of the COL1A1, COL1A2 and COL2A1 genes, have also been demonstrated to control expression in a tissue-specific fashion (143-145). The detailed dissection of this enhancer sequence in the COL1A1 gene has identified the presence of discrete positive and negative *cis*-acting regulatory elements which likely interact with different *trans*-acting DNA-binding factors and with the upstream promoter (146, 147). Some of the enhancer elements in these genes are probably involved in tissue-specific expression, since inactivation of the COL1A1 gene by retroviral insertion within the first intron eliminates expression in some tissues of the Mov-13 mouse (148), yet has no effect on odontoblast type I collagen synthesis (149).

Although ascorbic acid has long been known to result in elevated prepro-α1(I) and prepro-α2(I) mRNA levels (150), the mechanism by which this is achieved has not been determined. It has been suggested to result from the peroxidation of lipids, and the subsequent formation of adducts with various cellular transcription factors (151). The modified factors may then induce collagen gene expression.

Regulation of mRNA Stability

Another manner in which steady-state mRNA levels may be regulated is the control of mRNA stability (152). By this mechanism, two mRNA species which are synthesized at identical rates, could reach very different steady-state levels depending the ratios in the rates at which they are degraded. Although the mRNA sequences conferring stability or lability

have not been well characterized, they are likely to be found within the untranslated regions. There is some evidence that under certain conditions such a mechanism may also be involved in the response of type I collagen genes to TGF-β (153) and ascorbic acid (154).

Regulation of Translation

The pro-α1(I) and pro-α2(I) chains are synthesized in a 2:1 ratio, indicating that the corresponding mRNA species, which are also present in a 2:1 ratio, are translated with equal efficiency (74). Thus, the appropriate stoichiometry of these two chains appears to be determined primarily by the steady-state levels of their mRNAs. However, the COL1A1, COL1A2 and COL3A1 genes exhibit the conservation of a sequence of approximately 50 bp within the first exon (155). The proximity of this sequence to the translation start sites suggests that it may have a role in regulation at the translational level. However, this has not yet been established.

Effects of Post-translational Modifications

Post-translational modifications are critical for the normal biosynthesis of the fibrillar collagens. Specific modifications which may affect the assembly, transport and secretion of these molecules include 4-hydroxylation of triple helical proline residues and N-glycosylation of the carboxy-terminal propeptide. It is unlikely that differential N-glysosylation is involved in the regulation of collagen synthesis. However, the rate of proline 4-hydroxylation, an important determinant in the rate of procollagen assembly and secretion, is critically dependent on the presence of ascorbate as a cofactor. Thus, ascorbic acid may be involved in the regulation of type I collagen synthesis at several levels, including gene transcription, mRNA stability, and post-translational hydroxylation.

1.6 Osteogenesis Imperfecta and Type I Collagen Gene Mutations

Introduction

Osteogenesis imperfecta (OI) is a group of related skeletal dysplasias, in which the hallmark clinical feature is fragility of the skeleton, and which are associated with mutations in the COL1A1 and COL1A2 genes (156-158). These disorders exhibit an extensive heterogeneity in phenotypic severity and clinical presentation. The disorders range in severity from mild, with the infrequent occurrence of fractures, to very severe, resulting in perinatal death. This heterogeneity is reflected by the great variability with which a number of other connective tissues are affected in different individuals.

Classification

The extensive heterogeneity in the clinical presentation of OI patients described in this century's medical literature has led to the use of a wide variety of synonymous medical terms, such as van der Hoeve's, Eddowe's and Adaire-Deighton's syndromes (159). This diversity in medical nomenclature hindered the study and comparison of patients afflicted with these disorders. Therefore, in 1979, in order to simplify the study of these patients, Sillence proposed a classification system (160-162) which has since become widely used by clinicians and researchers. In 1986, the system was formally adopted by the International Congress of Human Genetics (163). The system classifies patients according to the frequency of fractures, the degree of skeletal deformity, radiological appearance and the presence of other clinical features such as dentinogenesis imperfecta. The use of these criteria allows patients to be classified into one of four groups numbered I to IV (Table 2). The significant clinical heterogeneity of patients within any one of these groups has required their subclassification into distinct subgroups (156).

Table 2. Sillence system for the classification of OI patients.

OI Type	Clinical Features	Inheritance
I	Mild to moderate bone fragility	Dominant
	Mild deformity (some cases)	
	Short stature	
	Hearing loss	
	Blue sclera	
	Dentinogenesis imperfecta rare	
II	Perinatal death usual	Dominant/
	Extreme bone and tissue fragility	Recessive
	Severe deformity of ribs and long bones	
	Poor skeletal growth	
	Blue sclera	
III	Severe bone fragility	Dominant/
	Progressive skeletal deformity	Recessive
	Short stature	
	Normal or pale blue sclera	
	Dentinogenesis imperfecta common	
IV	Mild to moderate bone fragility and deformity	Dominant
	Short stature	
	Hearing loss uncommon	
	Normal or blue sclera	
	Dentinogenesis imperfecta common	

Type I OI

Type I OI is the mildest and most common form of the disease, with a reported frequency of approximately 1 per 15,000 to 20,000 births (158). This form was demonstrated to account for approximately 40% of classified OI families in Australia and North America (156). All patients with this form of OI exhibit a triad of clinical features, consisting of blue sclera, adult onset hearing loss, and osteopenia giving rise to increased bone fragility. Joint hyperextensibility and easy bruising of the skin may also be present. In addition, there may be mild skeletal deformity in the form of bowing of the long bones due to the occurrence of multiple fractures. The frequency of fracturing decreases dramatically at puberty, but rises again after the fourth decade of life. Other clinical manifestations, which are exhibited to varying extents by different patients, include short stature, frontal bossing, mild kyphosis and scoliosis, and cardiovascular defects. The variability in the presence of dentinogenesis imperfecta has resulted in the subclassification of patients into groups IA (with normal dentition) and IB (with opalescent teeth). Patients with OI type IB appear to be less common, and exhibit fractures earlier in life and more frequently than those with type IA. This subgroup of patients also displays shorter stature and a higher likelihood of adult onset hearing loss.

This form of OI is inherited in a dominant fashion.

Type II OI

Type II OI is the most severe form of the disease. It occurs at a frequency of 1 per 20,000 to 60,000 births (158), and was shown to account for approximately 25% of classified OI families in Australia and North America (156). This form is characterized by the presence of extremely fragile bones, usually resulting in death *in utero*, or shortly after birth from respiratory failure. Consequently, this form is commonly known as the perinatal lethal type of OI. The fragility of the connective tissues is illustrated by the

frequency with which the avulsion of limbs occurs during delivery. Probands with type II OI are heterogeneous in appearance, and are classified into three subgroups based on detailed radiological observations (156, 164).

Type IIA babies are small for their gestational age, and are often born prematurely. Approximately 20% of type IIA babies are stillborn, while the remainder perish shortly after birth. These babies exhibit large crania, blue sclera, small chest cavities, and short limbs. There is also extreme bowing and angulation of the long bones, particularly of the lower limbs. Radiological examination demonstrates the presence of a small thoracic cage, with thick short ribs exhibiting a continuously beaded or wavy appearance. The long bones are broad and rectangular, or crumpled due to the occurrence of multiple fractures within individual bones. There is extensive evidence of previous fractures at various stages of healing.

Type IIB babies are phenotypically comparable to type IIA babies, but significant differences in their survival and radiological appearances exist. The survival of type IIB babies appears to be more variable than that of type IIA babies, with some dying in the perinatal period, while others may survive into early childhood (156). The increased lifespan of some of these babies appears to be due to their diminished respiratory difficulties, in comparison to type IIA babies. This is supported by the radiological features of type IIB babies, whose ribs are shortened but exhibit few or no fractures. However, the radiological appearance of the long bones is similar to that of type IIA babies, being broad and crumpled in nature. The extended lifespan of these patients complicate the differential diagnosis of this form of OI from the severe deforming type described below (type III).

Type IIC babies are rarely encountered, and are extremely small for their gestational age. They usually perish in the perinatal period, similar to type IIA babies. Upon radiological examination, type IIC babies exhibit hypomineralized skulls and thin,

discontinuously beaded ribs. The long bones appear slender and display extensive bowing and fracturing.

Type II OI was originally believed to be recessively inherited (161, 164). However, it is now accepted that most cases of type II OI result from sporadic dominant mutations, and that only few cases are due to recessive mutations (156, 165).

Type III OI

Type III OI is another severe type of OI and is the rarest form, accounting for only approximately 3% of classified OI families in Australia and North America (156). The expected lifespan of patients with this form of OI is extremely variable, with some dying within the perinatal period, while others survive well into adulthood. The more severe cases of type III OI may be difficult to distinguish from patients with type IIB who survive past the perinatal period (156). Type III OI babies are usually born with multiple fractures, bowing of the limbs and decreased length, but with weights within normal range. A hallmark feature of this OI type is the severe progressive deformity, which results from the frequent occurrence of postnatal fractures and an apparent increase in the malleability of the skeleton. The skeletal deformity includes severe bowing of the limbs, abnormality of the chest wall, frontal bossing, and kyphoscoliosis. By the end of the first decade of life, these patients are typically bed-ridden or confined to a wheelchair. Another striking feature of these patients is the very poor growth rate, giving rise to significantly short stature. Other clinical manifestations, such as dentinogenesis imperfecta and blue sclera, which usually pale throughout life, may be present in some patients. Radiological examination reveals severe osteopenia with evidence of multiple fractures, some of which may be in the process of healing. The long bones are typically short and severely deformed.

Type III OI appears to consist of dominant and recessive forms (158, 166). In addition, some cases may arise from mutations in genes other than those encoding type I

collagen (158), although none of these have yet been identified.

Type IV OI

Type IV OI accounts for approximately 30% of a group of classified OI families in Australia and North America (156), and is associated with osteopenia resulting in variable bone fragility. A striking feature of this form of OI is the great clinical heterogeneity between affected individuals, even within the same family. Some of these patients exhibit fractures at birth. The frequency of fractures after birth is heterogeneous between patients, and decreases dramatically after puberty. The rate of fracture healing appears to be normal (as in type I OI patients). Some type IV OI patients develop mild to moderate deformity of the long bones, even though the frequency of fractures does not appear to be significantly higher than that observed in type I OI patients. Patients with type IV OI also exhibit mild to moderately short stature. Other clinical features which are exhibited by some patients include light blue sclera which pale by adulthood, and joint hypermobility. Adult onset hearing loss is less common than in type I OI. Type IV OI patients are subclassified into groups A and B, depending on whether dentinogenesis imperfecta is present. Type IVA patients have normal dentition, while type IVB patients have opalescent teeth. Dentinogenesis imperfecta is much more common in type IV OI patients than in type I OI patients. Radiological features of these patients include osteopenia, and the presence of fractures and bowing of the long bones.

This form of OI is dominantly inherited (156, 158, 161).

Other Rare Forms of OI

Maroteaux et al have described a rare group of perinatally symptomatic OI patients, which appear to be distinct from Sillence types II and III (167, 168). In contrast to the severe Sillence types, these patients present with short bowed thighs, but with minimally affected upper limbs, skull and trunk. Radiological examination reveals mild osteopenia

and curvature of the femora and tibia, which otherwise are normal in length. The long bones of the upper limbs usually appear normal. The interesting features of these patients are the significant improvements in the bowing of the long bones and in bone fragility which occur with ageing. As a consequence, these patients have been referred to as exhibiting a regressive form of OI. The growth rate of these patients is almost normal, with some patients attaining the lower limits of the range of normal stature. Blue sclera are usually present at birth, and dentinogenesis imperfecta is sometimes present.

Type I Collagen Gene Defects in Osteogenesis Imperfecta

The association of OI with defects in type I collagen was based on two types of evidence. Since most OI mutations are sporadic (165, 169), the availability of large families for linkage analysis has been limited. The analysis of a number of families with OI demonstrated that the disease was linked to the genes encoding type I collagen (170, 171). However, a small proportion of the families studied failed to exhibit linkage of the disorder with these genes (158, 172). It therefore appears that a small percentage (<10%) of OI families possess defects not affecting either of the type I collagen genes.

The development of very sensitive techniques in the area of molecular biology has allowed the identification of more than 70 specific type I collagen gene mutations in OI patients (173). The mutations which have thus far been described can be divided into two distinct groups according to their specific effects on type I collagen biosynthesis (157, 158, 169). One group of mutations result in an abnormality in the quantity of type I collagen which is synthesized, while the second group results in the biosynthesis and secretion of structurally abnormal collagen which may subsequently be incorporated into the ECM. The distinct differences in the effects of these two types of mutations on the biosynthesis of type I collagen are fundamental in determining their phenotypic outcomes and the possible management of the disease.

Quantitative Defects

The common feature of this type of defect is the reduced synthesis and secretion of type I collagen, which otherwise is structurally normal (174-176). These defects are usually associated with abnormally low steady-state levels of the prepro-α1(I) mRNA (175, 176). The types of mutations which appear able to produce low mRNA levels include those giving rise to non-functional COL1A1 genes (174), such as large genomic deletions or mutations affecting the promoter regions. Low prepro-α1(I) mRNA levels may also result from mutations affecting message splicing, transport, stability or translation (176, 177), which have been suggested to account for a large proportion of quantitative defects in type I collagen synthesis (177). Despite indications that the above types of mutations are responsible for at least some quantitative OI defects, no specific mutations have yet been demonstrated.

Quantitative defects can also arise from mutations which do not yield low prepro-α1(I) mRNA levels, but which result in the synthesis of structurally abnormal pro-α1(I) chains which cannot associate with their normal counterparts. A frameshift mutation has been shown to result in pro-α1(I) chains which cannot be incorporated into procollagen (178), due to the structural abnormality of the carboxy-terminal propeptide structural elements required for chain association.

All of these types of mutations result in an approximate 50% decrease in the biosynthesis of type I procollagen by fibroblasts (174, 175). These defects are often detected as alterations in the ratio of type I to type III procollagen synthesized (169, 175), since the defect in type I collagen metabolism does not usually affect the rate of type III procollagen synthesis. The procollagen that is secreted does not exhibit any abnormalities in the levels of post-translational hydroxylation and glycosylation, suggesting that its chains are able to form a triple helix at a normal rate, and with a normal thermal stability. Any chains which are not incorporated into triple helical procollagen are degraded

intracellularly, irrespective of whether they are structurally abnormal (178) or excess normal chains (174). Since the procollagen secreted into the extracellular space is structurally normal, the collagen fibrils formed upon propeptide processing are likely also structurally normal. However, the decreased quantity of type I procollagen that is assembled and secreted results in a decreased amount of fibrillar collagen in skin and bone (174, 179).

Almost all quantitative defects appear to be linked to the COL1A1 gene (157, 158, 176, 180). A possible explanation for this is that similar mutations in the COL1A2 gene may sometimes give rise to milder phenotypes, such as osteopenia with little effect on bone fragility (181). This may in turn be explained by the fact that excess pro-α1(I) chains, resulting from COL1A2 mutations, are able to form homotrimers which are then secreted into the ECM. In contrast, the excess pro-α2(I) chains, resulting from COL1A1 gene mutations, do not appear to be able to form homotrimers, and are likely degraded intracellularly.

Defects resulting in the biosynthesis of decreased amounts of structurally normal type I collagen are associated with type I OI, the mildest form of the disease (157, 158). This indicates that a 50% decrease in the quantity of collagen in bone results in mild bone fragility. Although quantitative defects appear to be limited to type I OI patients (169), not all type I OI patients possess these types of defects. Thus, some cases of mild OI are due to defects affecting collagen structure (158, 157, 169, 173, 180).

Structural Defects

The vast majority of OI mutations identified to date result in the biosynthesis and secretion of structurally abnormal type I procollagen (173). Although the types of defects observed are diverse in nature, they exhibit similarities in their effects on the biosynthesis and structure of type I collagen.

Glycine Substitutions

The most common mutations observed are point mutations resulting in the substitution of glycine residues within the triple helical domain of either the pro-α1(I) or pro-α2(I) chain (173). The point mutation of glycine codons, with the sequence GGN, can give rise to eight different amino acid substitutions. The possible substituting amino acids are cysteine, aspartate, glutamate, arginine, serine, tryptophan, valine and alanine. Thus far, all of these possible substitutions have been observed with the exception of glycine → tryptophan and glycine → glutamate substitutions (173). The substitutions have significant effects on the biosynthesis, structure and stability of the type I procollagen molecule. Since glycine residues are required at every third position within the triple helical domain of each chain in order for triple helix formation to occur normally (11, 12), their substitution results in the delaying of this process, starting from the position of the defect. Therefore, in procollagen molecules incorporating one or two mutant pro-α1(I) chains or one mutant pro-α2(I) chain, helix formation initiates adjacent to the carboxy propeptide and appears to occur normally until the location of the helical defect (182). As a consequence of the delay in normal helix formation, the regions of the pro-α chains on the amino terminal side of the glycine substitution appear to be exposed to the post-translational modification enzymes for a lengthened period of time (182-185). This results in the excessive hydroxylation of Y-position proline and lysine residues, and the glycosylation of hydroxylysine residues in the helical regions amino-terminal to the defect (184). Since virtually all Y-position proline residues are 4-hydroxylated in normal type I procollagen, the effects of delayed helix formation on post-translation modification predominantly involve lysine residues (183, 185). A defect in either the pro-α1(I) or pro-α2(I) chain results in delayed helix formation, and therefore gives rise to the post-translational overmodification of **both** chain types (182). The post-translational overmodification may not only serve as a marker of OI, but may also contribute to the pathogenesis of the disease, although this has not been shown.

In addition to interfering with procollagen assembly, the presence of an amino acid side

chain, other than the hydrogen atom of glycine, within the interior of the triple helix usually results in a heterotrimer with decreased thermal stability (182, 185). The degree to which thermal stability is affected appears to depend on the type of collagen chain involved. Glycine substitutions within the pro-α1(I) triple helical domain result in the assembly of two populations of abnormal procollagen molecules differing in thermal stability, since procollagen molecules containing two mutant chains appear to be more destabilized than molecules incorporating a single mutant chain (186). However, similar defects in the pro-α2(I) triple helical domain can only give rise to abnormal procollagen molecules containing a single mutant chain. Thermal stability probably also depends on the identity of the substituting amino acid, and on the location of the substitution within the triple helix (173). However, insufficient data is currently available to accurately assess the importance of these parameters on the thermal stability of the procollagen molecule.

The decreased thermal stability of abnormal procollagen molecules appears to confer increased sensitivity to intracellular proteolytic degradation (183, 185; Fig. 6). This results in the degradation of the abnormal pro-α chains and of the normal chains associated with them. The term "protein suicide" has been coined to describe this process, by which a mutation in one COL1A1 allele, resulting in a glycine substitution in 50% of pro-α1(I) chains synthesized, can give rise to the degradation of 75% of all pro-α1(I) and pro-α2(I) chains synthesized, as a consequence of their random association into trimeric structures (187). In contrast to COL1A1, a similar mutation in one COL1A2 allele can give rise to the degradation of 50% of all pro-α1(I) and pro-α2(I) chains synthesized. These estimates of the extent of chain degradation may represent minimum values, since there appears to be some cooperativity in this process (188). Thus, the degradation of more than these amounts of collagen occurs due to the destruction of normal procollagen molecules, in addition to those containing mutant chains. This may be the result of the manner in which procollagen molecules are associated during intracellular packaging.

The increased rate of intracellular degradation results in a decrease in the amount of type I procollagen secreted by OI skin fibroblasts (182-185). This is supported by experiments demonstrating that skin and bone from OI patients contain decreased amounts of collagen (184). The process by which a cell sorts normal and abnormal collagens for secretion does not appear to be perfect, since some normal procollagen is likely degraded intracellularly (188), while some abnormal procollagen is secreted (182-185). However, in general, normal procollagen molecules assembled in OI fibroblasts are secreted relatively efficiently, while those containing abnormal chains are preferentially retained intracellularly and degraded (185). The efficiency with which abnormal molecules are secreted appears to correlate with thermal stability, since molecules containing two abnormal pro-α1(I) chains are intracellularly retained to a greater extent than those containing only one (186).

Once secreted into the extracellular space, normal procollagen molecules undergo proteolytic removal of both terminal propeptides. However, the proteolytic processing of procollagen molecules containing one or two mutant chains may not occur normally due to the conformational changes induced by the helical glycine substitution. The carboxy propeptide is likely removed normally due to its folding prior to triple helix formation. However, the conformation of the amino-terminal propeptide, which is required for normal N-proteinase activity, may to some extent be disturbed by helical glycine substitutions. This is probably due to an altered chain registration in the triple helix on the amino-terminal side of the substitution. Thus, in at least some cases helical glycine substitutions interfere with the normal proteolytic removal of the amino-terminal propeptide (189).

The presence of abnormal collagen or pN-collagen molecules in the extracellular space seems to interfere with the process of fibrillogenesis, and is likely mediated by three factors. Firstly, the inefficient removal of the amino-terminal propeptides can result in the production of abnormally thin fibrils (107). Secondly, since the precise three-dimensional orientation of charged and hydrophobic side groups on the outer surface of the collagen

molecule, required to ensure a stable and structurally correct lateral association, is likely altered by the conformation imposed by the glycine substitution, the fibrils formed may have an abnormal structure and stability. Finally, the configuration of the collagen molecule also appears to be important in determining the structure of the fibrils formed. Fibrils formed *in vitro* from collagen molecules containing chains with Gly → Cys substitutions have been shown to exhibit a dendritic appearance (190), in addition to alterations in diameter and rate of assembly (191). The presence of a glycine substitution introduces a kink in the collagen triple helix (189), which is believed to result in the branching structure of the dendritic fibrils (Fig. 6). Although it is unclear what proportion of helical glycine substitutions result in collagen molecules with such a kinked structure, it is probable that the disruption of the triple helical conformation near the site of the substitution is significant enough to affect the manner in which the molecules assemble into fibrils.

The histological examination of bone from a small number of type II OI patients has demonstrated clear abnormalities in the quantity and organization of mineralized matrix (192, 193). These changes are likely the result of the decreased secretion of type I procollagen and/or the presence of structurally altered molecules in the ECM. However, since no patients with the non-lethal forms of OI, and whose biochemical defects have been established, have been examined histologically, it is at present unclear whether these changes in bone architecture can be correlated with phenotypic severity and the nature of the biochemical defect.

In-Frame Deletions and Insertions

Other types of mutations which have been described result in the deletion of varying numbers of residues in the triple helical domain of either chain (173). These deletions can be produced either by the genomic deletion of one or more exons (194, 195), or by

mutations at consensus splice sites which result in exon skipping due to the use of the analogous site of a neighboring intron (196, 197). Since each exon coding for helix amino acid residues begins with a glycine codon and ends with a Y-position residue codon (18), the Gly-X-Y tripeptide repeat structure of the helical domain is not perturbed by these types of mutations. Deletions can also occur within individual exons, such that one or more codons are omitted, without the production of a frameshift or termination mutation (196). These types of deletions only maintain the Gly-X-Y structure if the number of codons deleted from the mRNA is a multiple of three. All of the mutations resulting in the deletion of codons from the mRNA result in the shortening of either the pro-α1(I) or pro-α2(I) chain.

Mutations resulting in the insertion of amino acids into the triple helical domain of either the pro-α1(I) or pro-α2(I) chain have also been described (173). These can be produced by the duplication of segments of genomic DNA containing one or more exons (198, 199), or by mutations affecting consensus splice sites (200). The latter type of mutation results in the use of an alternative splice site within the intron, and causes the retention of intronic sequences within the mRNA. When the retained intron sequences do not result in premature termination or a shift of the reading frame, the chains can be incorporated into procollagen molecules. Exon duplications maintain the Gly-X-Y repeat chain structure, while mutations resulting in the retention of intron sequences do not.

The effects of deletions and insertions on the biosynthesis, structure and stability of type I collagen appear to possess similarities to those produced by glycine substitutions. They delay triple helix formation (201), resulting in post-translational overmodification (201, 202). The presence of an abnormal chain thermally destabilizes the procollagen molecule, giving rise to increased intracellular degradation (201), a decreased rate of secretion (202), and an elevated ratio of type III/type I procollagen secreted by fibroblasts (203). In addition, procollagen molecules containing a shortened or lengthened chain are resistant to amino-terminal processing due to the abnormal conformation of the propeptide

(201, 204). The persistence of the propeptide and the altered distribution of charged and hydrophobic amino acid residues then probably interferes with fibrillogenesis.

Deletions and insertions may be divided into two groups depending on whether or not they result in interruptions in the Gly-X-Y repeating structure of the procollagen chains. These defects may exhibit distinct biochemical manifestations. However, this is unclear since virtually all described deletions and insertions maintain the Gly-X-Y chain structure (173). For those deletions and insertions maintaining the Gly-X-Y chain structure, it is possible that the majority of the triple helix exhibits a normal rate of folding, until the region adjacent to the amino-terminal propeptide. Thus, any thermal destabilization and post-translational overmodification may be limited to the amino terminus of the helix, where the polypeptide chains are out of register. In addition, in the case of deletions, the degree of helix destabilization may correlate with the size of the region deleted, since thermal stability is proportional to the length of the triple helix.

The effects of deletions and insertions which fail to maintain the Gly-X-Y chain structure are likely more similar to those of glycine substitutions than to those of deletions and insertions which maintain the Gly-X-Y chain structure. The former types of deletions and insertions likely result in kinked molecules, either by interrupting the Gly-X-Y repeating sequence, or by introducing sequences which are not compatible with triple helix formation, respectively. Thus, these deletions and insertions may result in greater helix destabilization, which is likely contributed by two regions. These consist of the helical regions immediately adjacent to the defect, as well as at the amino terminus.

Translational Frameshift Mutations

A small number of mutations, resulting in translational frameshifts in the pro-α1(I) and pro-α2(I) chains, have also been observed (173). Those affecting most regions of the pro-α chains result in quantitative defects characteristic of type I OI, by introducing premature

termination codons, or by altering the polypeptide sequence and preventing chain association (178). However, those occurring in certain regions of the carboxy-terminal propeptide allow abnormal chains to associate with their normal counterparts, but interfere with triple helix formation (205, 206). This is perhaps due to an altered carboxy-terminal propeptide conformation resulting from the formation of incorrect disulfide bonds. Similarly to structural defects within the triple helix, this type of translational frameshift within the carboxy-terminal propeptide gives rise to thermal destabilization, intracellular degradation and poor secretion (205). However, since frameshift mutations interfere with helix formation at an early stage, they may result in the virtually complete proteolytic degradation of molecules containing abnormal chains. Thus, in contrast to helix structural defects, very low quantities of the abnormal molecule may be found in the ECM.

Frameshift mutations interfering with helix formation appear to give rise to bone architectural changes which are similar to those produced by glycine substitutions. Bone from a type II OI patient harboring a carboxy-terminal propeptide frameshift mutation was shown to exhibit quantitative and organizational abnormalities in mineralization (207).

Phenotypic Severity Rules

Structural defects result in OI phenotypes with a great range of severity, from mild osteopenia (208) to perinatal death (209), and in a genetic predisposition to post-menopausal osteoporosis (210-213). However, the great majority of patients harbouring structural defects exhibit phenotypes which are more severe than those associated with quantitative defects. The more severe expression of structural mutations is mediated by two factors (Fig. 6). Firstly, the decrease in procollagen secretion resulting from structural defects is more significant than for quantitative defects, due to the degradation of molecules containing abnormal chains ("protein suicide"). Secondly, only structural defects result in the incorporation of abnormal collagen molecules into fibrils, thereby interfering with the

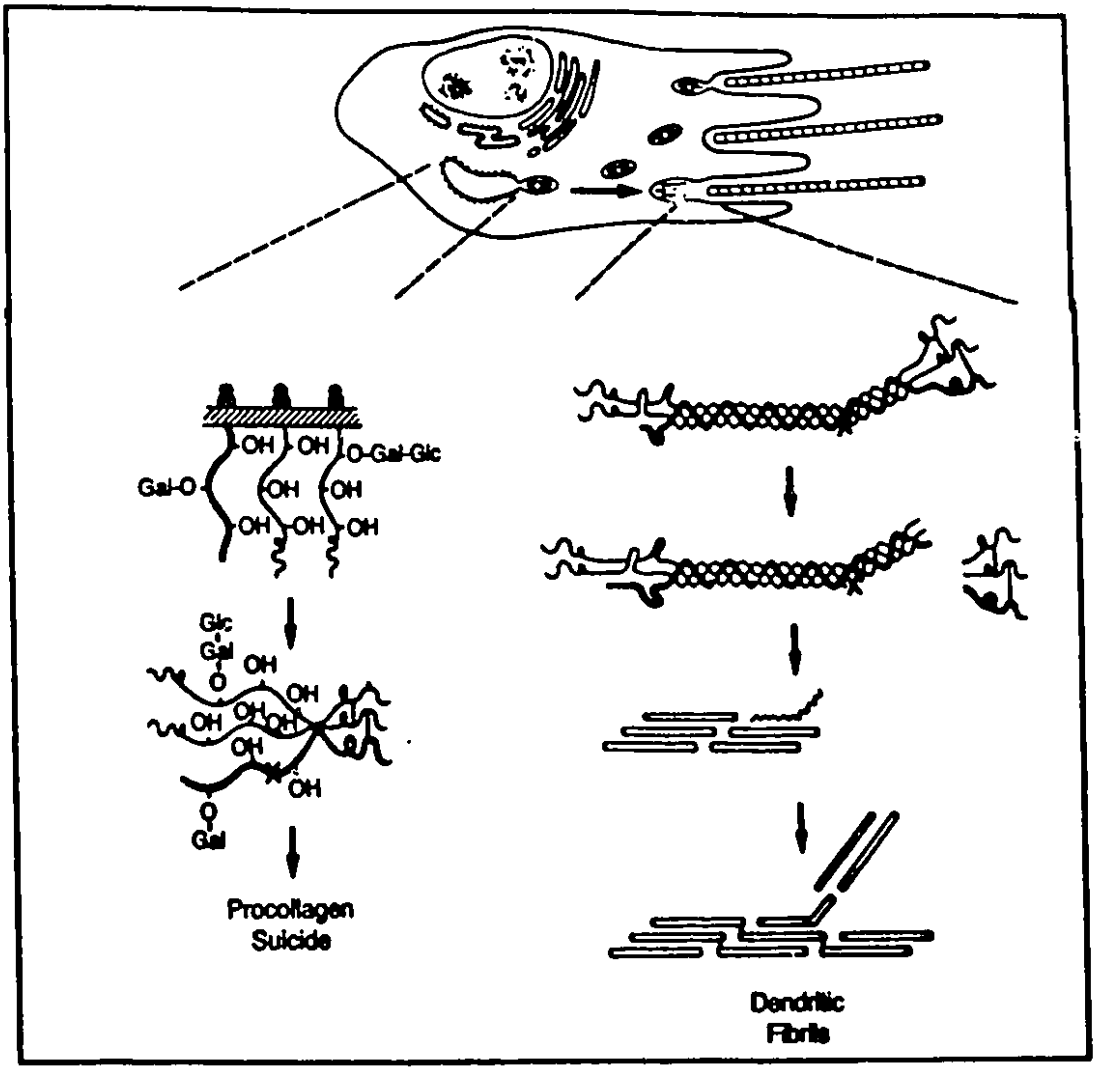


Fig. 6. Effects of structural collagen mutations on biosynthesis. The diagram illustrates the effects of a structural defect within the procollagen triple helix. The defect may retard helix formation and thermally destabilize the procollagen molecule, giving rise to post-translational overmodification and susceptibility to the intracellular degradation of its constituent normal and abnormal chains ("protein suicide"). In addition, for those procollagen molecules that are secreted, the incorporation of an abnormal chain may disrupt the conformation of the triple helix, resulting in the formation of kinked molecules. The conformation of the amino-terminal propeptide may also be disrupted, thereby interfering with its extracellular proteolytic removal. The abnormal shape of the collagen molecule and the persistence of the propeptide may then give rise to structurally altered fibrils in the matrix. The relative extents to which these abnormalities occur is dependent on the specific nature and location of the defect (adapted from Prockop (324)).

structural integrity of the ECM. The relative extent to which these two factors contribute to the OI phenotype is unclear, but likely depends on the exact nature and location of the defect involved.

The characterization of a large number of defects, particularly glycine substitutions and deletions, within the triple helical domains of the α1(I) and α2(I) chains has allowed the identification of general patterns in OI phenotypic severity relative to the location and type of defect (173, 214). These general patterns have been termed the "phenotypic severity rules".

In general, defects within the triple helical domain of the α1(I) chain are more deleterious than identical defects occurring at equivalent positions within the α2(I) chain (173, 214). This is referred to as the "chain rule". This observation can be explained by the 2:1 stoichiometry of pro-α1(I) and pro-α2(I) chains in the type I procollagen heterotrimer, and the ability of pro-α1(I) chains to form homotrimers.

The analysis of Gly → Cys or Gly → Arg mutations described within the α1(I) triple helical domain (173, 214) demonstrates that substitutions near the carboxy terminus are more deleterious than identical substitutions near the amino terminus. In fact, there appears to be a fairly continuous gradient of decreasing phenotypic severity in a carboxy-to-amino terminal direction. This is known as the "phenotypic gradient rule". This pattern can be explained by considering the carboxy to amino terminal directionality of triple helix formation. At present it is unclear whether deletions and insertions also exhibit this pattern.

Substitutions of helix glycine residues with bulky or charged amino acids, such as arginine and aspartate, are more deleterious than substitutions with smaller amino acids, such as serine and cysteine (173, 214). Bulky amino acids appear to retard triple helix formation to a greater extent than those with smaller side chains, probably due to their greater steric hindrance. Deletions and insertions of varying sizes may perhaps be considered the analog of substitutions involving different replacing amino acids. In

general, it appears that large deletions and insertions (multi-exon) are more deleterious than those involving a single exon or a number of codons within an individual exon (173, 214).

Although most of the glycine substitutions described to date are consistent with these phenotypic severity rules, there are a number of substitutions which produce phenotypes which are unexpectedly mild (215, 216) or severe (217). This indicates that the α1(I) and α2(I) triple helical domains consist of regions which exhibit different levels of tolerance to certain types of substitutions. Defects involving different regions of the helix may consequently have distinct effects on the stability of the entire procollagen molecule, independently of the phenotypic gradient rule. In addition, since relatively few deletions and insertions have been identified to date, it is unclear to what extent they adhere to the phenotypic severity rules. The identification of additional OI mutations will be required to clearly establish the relationship between type I collagen defects and phenotypic severity.

Transgenic Mouse Models

The most direct evidence for the role of type I collagen defects in the pathogenesis in OI has come from the study of mouse models for the disease. The Mov-13 mouse strain, heterozygous for a COL1A1 gene inactivated by retroviral insertion in the first intron, is a model for type I OI (218). The presence of the nonfunctional COL1A1 allele results in a quantitative defect in type I collagen synthesis, and is associated with biochemical and phenotypic characteristics of type I OI patients. These include bone fragility, decreased collagen in skin and hearing loss.

Transgenic mouse strains, harboring mutated type I collagen genes, have also been examined, and represent models for the forms of OI associated with structural defects. COL1A1 transgenes possessing point mutations resulting in a Gly → Cys or Gly → Arg substitution, at position 859 of the triple helical domain, gave rise to offspring exhibiting phenotypic features similar to human OI type II patients (188). More recently, a COL1A1

minigene containing a 41-exon deletion was also shown to result in a lethal OI phenotype (219). These studies provided conclusive evidence for the role of such mutations in producing bone fragility, the hallmark feature of OI. In addition, the model demonstrated that only very low levels of expression of the mutant transgene were required to produce skeletal abnormalities and significant decreases in tissue collagen content (188).

Inheritance of OI Mutations, Genetic Counselling, and Prenatal Diagnosis

Although it was initially believed that a significant proportion of OI cases were recessively inherited (161, 164, 166), it is now generally accepted that most of the mutations are dominantly expressed (156-158). The uncertainty in the mode of inheritance resulted from the observation of recurrence of affected offspring to apparently normal parents, a pattern suggestive of recessivity.

The recurrence of type II OI in a number of families has been explained by parental germ-line mosaicism (165, 220-222). In these families, the number of affected offspring is determined by the fraction of mature parental gametes harbouring the mutation. In addition, the mutation carried in the germ-line may also be found in variable proportions in cells of the parental somatic tissues (220-222), depending on whether and when it occurred during development or gametogenesis. However, the percentage of bone osteoblast cells affected in these parents appears to be insufficient to give rise to the OI phenotype observed in their offspring. The risk of recurrence of lethal OI appears to be approximately 6% (165). In other words, in about 1 out of every 16 type II OI families, there is an increased risk of having other affected offspring. This significant recurrence risk has rendered the process of genetic counselling more difficult, since the demonstration of heterozygosity for a dominant mutation in an affected child is not sufficient to indicate the occurrence of a sporadic mutation. The possibility of parental mosaicism must be considered in order to provide responsible genetic counselling to the parents.

In some cases, the recurrence of mild to severe OI, in offspring born to apparently normal parents, may be due to variable penetrance of the OI phenotype in different individuals within a family (158). Thus, some mutations may result in intrafamilial phenotypic variability, in which the "carrier" parent is so mildly affected that he/she is considered normal. For these families, it would appear that the affected offspring inherited recessive mutations from the parents.

Recessive mutations have been described in only a small number of OI cases. These include a type II OI individual exhibiting compound heterozygosity for a non-functional COL1A2 allele and a COL1A2 allele bearing a structural mutation (223, 224). Skin fibroblasts from this individual secreted only $[\alpha 1(I)]_3$ homotrimer (225). The other described cases of OI type II and type III involved homozygosity for the same recessive mutation (206, 226). Although few examples of recessive mutations have been described to date, they appear to be associated most commonly with a type II or type III OI phenotype.

The prenatal diagnosis of OI is currently performed by the serial ultrasonographic detection of skeletal abnormalities (227). However, this type of examination may only reliably detect the forms of OI which are associated with moderate to severe deformity (228, 229). OI types IV and I, in which the deformity may be relatively mild, may be undetectable or detected late in the pregnancy. An alternative method for prenatal diagnosis is CVS, chorionic villus sampling, which can be performed at approximately 10 weeks of gestation (228). The type I collagen from these cells is then analyzed for signs of post-translational overmodifications, and a response can be obtained within approximately 5 weeks of sampling. In addition, this analysis may be more sensitive than ultrasonography for the detection of the milder forms of OI, since the relatively small changes in post-translational modification often associated with these cases may be detected. This approach may be useful for families in which there is a previously affected child, or in which one parent is clinically affected.

1.7 Other Disorders Due to Mutations in Fibrillar Collagen Genes

In recent years, other connective tissue disorders have been attributed to mutations within fibrillar collagen genes. Although none of these have been as extensively examined as OI, it is likely that in the following years, many further mutations will be described. The disorders for which specific defects have been identified are summarized in Table 3.

Ehlers-Danlos Syndrome Type VII

EDS is a heterogeneous group of disorders characterized by joint laxity, skin fragility and hyperextensibility, and a lack of obvious bone fragility (158, 230). At least eleven distinct forms of EDS have been described (163), which differ in the degree of involvement of the various connective tissues. The common feature of these disorders appears to be a weakening of the tensile strength of the tissues, or an abnormality in collagen biosynthesis. EDS type VII, also referred to as arthrochalasis multiplex congenita, is a disorder associated with short stature and significant joint laxity, commonly resulting in bilateral hip dislocations and scoliosis (158, 230). Other clinical features include a mild propensity to bruising and moderate hyperextensibility of the skin.

The biochemical analysis of skin from such patients demonstrated the persistence of type I pN-collagen chains, indicative of an abnormality in the extracellular processing of procollagen into collagen (104, 231, 232). In addition, examination of the skin and tendons from these patients revealed the presence of thin disorganized collagen fibrils (233). A similar biochemical abnormality has been observed in animals with dermatosparaxis, a recessively-inherited disorder resulting from a deficiency of procollagen N-proteinase activity in the tissues (158). In contrast to EDS type VII, this disease gives rise predominantly to significant skin fragility, with little effect on joint mobility. The skin of dermatosparactic animals contains type I collagen with a normal banding pattern, but

Table 3. Disorders resulting from mutations in fibrillar collagen genes.

Collagen Type	Disorder	Defect Type(s) [1]
I	OI	Nonfunctional gene
		Abnormal mRNA processing
		Glycine substitutions, deletions, insertions in triple helical domain
		Translational frameshift mutations
		Carboxy-terminal propeptide mutations
	EDS type VII	Deletion of exon 6 sequences
II	Achondrogenesis type II	Glycine substitution in triple helical domain
	SED	Deletion, insertion in triple helical domain
	Stickler syndrome	Arg → Stop substitution in triple helical domain
	Predisposition to 1° OA [2]	Arg → Cys substitution in triple helical domain
III	EDS type IV	Glycine substitutions, deletions in triple helical domain

[1] The defect types listed are those which have been identified in the various disorders. The table does not imply that the disorders are restricted to the types of defects given.

[2] Abbreviation used: 1° OA, primary osteoarthritis.

which forms abnormal ribbon-like fibrils (234) which are inadequately cross-linked (235). Despite the significant differences in the patterns of tissue involvement between these disorders, it was initially believed that EDS type VII was recessively-inherited and resulted from a deficiency of procollagen N-proteinase (231). However, recent studies have demonstrated that this form of EDS usually results from dominantly-inherited mutations.

Type I procollagen secreted by fibroblasts from EDS type VII patients was demonstrated to be resistant to cleavage by partially-purified chicken procollagen N-proteinase *in vitro* (236). In addition, fibroblast cultures contained normal levels of procollagen N-proteinase (236). Experiments with a large number of EDS type VII fibroblast cell lines demonstrated that approximately 50% of the pro-α1(I) or pro-α2(I) chains synthesized were shorter than chains synthesized by normal fibroblasts (237-239). This indicated heterozygosity for mutations resulting in the deletion of approximately 20 amino acid residues near the procollagen N-proteinase cleavage site. Protein sequence analysis of abnormal chains synthesized by EDS type VII fibroblasts identified deletions of the 24 or 18 amino acids comprising the amino-terminal propeptide/helix junction region of the pro-α1(I) or pro-α2(I) chain, respectively. For each chain, this region is encoded by exon 6 of the corresponding gene. The availability of cDNA probes and sensitive molecular biology techniques has permitted the identification of the exact molecular defects in a number of these patients. All of the defects described to date have been point mutations at the exon 6/intron 6 splice junction of the COL1A1 or COL1A2 gene (240-244). The point mutations result in the complete or partial skipping of exon 6, with the degree of consequent missplicing being determined by the exact base position affected. The identification of the biochemical defects in these patients has resulted in a new system for subclassifying EDS type VII patients (163). Patients harbouring mutations in the COL1A1 and COL1A2 genes are now classified as type VIIA and VIIB, respectively. Patients with a deficiency of the enzyme procollagen N-proteinase are classified as type

VIIC, although no such cases have yet been demonstrated.

The phenotypic effect of deleting the amino-terminal amino acid residues encoded by exon 6 indicates the functional importance of this region of the type I procollagen chains. This region contains the procollagen N-proteinase cleavage site. Deletion of this site in mutant pro-α1(I) or pro-α2(I) chains prevents their cleavage by procollagen N-proteinase, but allows cleavage of the normal chains with which they are associated. The cleaved amino-terminal propeptides of these normal pro-α chains remain non-covalently associated with the uncleaved amino-terminal propeptide of the abnormal chain (245). As a consequence, the bulky globular amino-terminal propeptide persists and interferes with normal fibrillogenesis (233). In addition, since the amino-terminal propeptide is believed to be involved in the feedback inhibition of type I collagen biosynthesis (108), fibroblasts from EDS type VII patients appear to have an elevated rate of collagen synthesis (238).

Another functionally important site located in the amino-terminal propeptide is a hydroxylysine residue which participates in intermolecular cross-linking (84). At the present time, it is unclear whether the extreme joint laxity observed in these patients is due to the effects of these mutations on collagen fibril structure, collagen cross-linking, or a combination of the two. It is interesting to note that a patient with OI, bearing a deletion of an α2(I) region containing cross-linking lysine residue 87 of the triple helix, exhibits some of the clinical symptoms of EDS type VII, in addition to the bone fragility (204, 246, 247). This suggests that the deletion of either of these cross-linking lysine (or hydroxylysine) residues, possibly in combination with the changes in amino-terminal propeptide processing, results in a decrease in the tensile strength of type I collagen fibrils, which is then reflected by joint laxity. The absence of significant bone fragility in EDS type VII patients suggests that the formation of the intermolecular cross-link involving the telopeptide hydroxylysine residue is not critical for normal bone strength. In addition, the identification of the defects in EDS type VII demonstrates that not all type I collagen gene

mutations result in OI.

Chondrodysplasias

The chondrodysplasias are a heterogeneous group of over 100 inherited disorders characterized by the deficient growth and development of the skeleton (248). They are associated with short stature, and with the involvement of abnormalities of the ears and eyes which vary in severity between the different disorders.

Achondrogenesis type II (Langer-Saldino) and hypochondrogenesis are forms of lethal neonatal dwarfism in which there is a severe defect in endochondral bone formation. The SEDs are heterogeneous dwarfing disorders characterized by abnormalities of the vertebral bodies of the spine, abnormal epiphyses, and ocular problems such as myopia, vitreous degeneration, cataract formation and retinal detachment (248). These disorders are now believed to constitute a spectrum of dwarfing disorders, in which achondrogenesis type II and the SEDs represent the severe and mild ends, respectively. These chondrodysplasias were anticipated to result from defects in type II collagen, since a common feature of the tissues affected (cartilage, vitreous of the eye and nucleus pulposus of the spine) was the expression of this matrix component (2, 29).

Morphological examination of growth plate cartilage from patients with these dwarfing disorders demonstrated clear differences from the normal cellular architecture (249). The cartilage was hypercellular, and contained sparse quantities of matrix. Increased numbers of vascular vessels, surrounded by fibrous tissue, were also evident. The biochemical analysis of the cartilage matrix demonstrated abnormally low amounts of type II collagen, and the predominant expression of type I collagen (250). Ultrastructural analysis of chondrocytes revealed the presence of dilated rough ER compartments, presumably due to the intracellular accumulation of a structurally altered type II collagen (251, 252). The type II collagen extracted from hyaline cartilage exhibited post-translational overmodification, of

which the degree appeared to correlate with clinical severity (253, 254). These observations suggested that some of these patients harboured mutations resulting in the biosynthesis of structurally abnormal pro-α1(II) chains, which interfered with the normal assembly and secretion of type II procollagen. Furthermore, they indicated that type II collagen defects exhibit a phenotypic gradient similar to that observed in OI, such that defects nearer to the carboxy terminus of the type II procollagen molecule result in more severe phenotypes.

To date, COL2A1 mutations have been identified in few chondrodysplasia patients. One achondrogenesis type II patient was demonstrated to possess a point mutation giving rise to a glycine → serine substitution at position 943 of the α1(II) triple helix (255). Other mutations have been described in patients with SED. One of these was a 45-bp intraexon duplication which lengthened half of the pro-α1(II) chains by 15 amino acids (256). The other was a 390-bp genomic deletion resulting in the deletion of 36 amino acid residues, encoded by exon 48, from the triple helical domain (257). The similarity between these mutations and those which have been identified in OI suggests that their effects on procollagen biosynthesis are probably also comparable. However, since type II collagen is a homotrimeric molecule, and since abnormal chains are able to randomly associate with their normal counterparts, only approximately one-eighth of assembled type II procollagen molecules are structurally normal. Thus, heterozygosity for a structural defect in type II collagen results in more extensive "protein suicide" and accounts for the very low tissue levels of type II collagen exhibited by these chondrodysplasia patients.

Type II collagen defects have also been implicated in other chondrodysplasias. Stickler syndrome (arthro-ophthalmopathy) is an autosomal dominant disorder associated with premature osteoarthritis, severe ocular problems, sensorineural hearing loss and mild SED (248). Linkage analysis of a number of families has demonstrated cosegregation of specific COL2A1 alleles with the Stickler syndrome (258-261). In addition, one study has identified a point mutation resulting in the substitution of an arginine codon with a stop

codon at position 732 of the α1(II) triple helical domain (262). By analogy to OI mutations, the point mutation in this individual likely gives rise to a quantitative defect in type II collagen synthesis.

The role of type II collagen gene mutations in the chondrodysplasias is supported by transgenic mouse models. Mice harboring a mutant transgene, containing a 12-exon deletion (263), or a point mutation resulting in a Gly → Cys substitution (264), were shown to exhibit the phenotypic and pathological features of a severe chondrodysplasia.

Kneist dysplasia is a disorder in which patients exhibit short stature, joint swelling, flat face, depressed nasal bridge, cleft palate, hearing loss and ocular problems (248). These patients exhibit a degeneration of the growth plate and epiphyseal cartilage ECM (265), which contains abnormal type II collagen fibrils and decreased amounts of the carboxy-terminal propeptide (266). The propeptide has been demonstrated to be involved in the process of cartilage matrix calcification during endochondral ossification (267). Although no COL2A1 mutations have yet been described, the data suggest that Kneist patients possess defects in the intracellular processing of type II procollagen.

Mutations in the gene encoding type II collagen have also been implicated in the genetic predisposition to primary osteoarthritis, a disorder characterized by a progressive degeneration of the articular cartilage of joints, in three multi-generation families (268, 269). These studies demonstrated linkage between the COL2A1 gene and primary osteoarthritis, which appeared to be inherited in an autosomal dominant fashion in these families. Subsequent studies identified the mutation in one of these families to be a point mutation resulting in an arginine → cysteine substitution at position 519 of the α1(II) triple helical domain (270). However, the effect of this substitution is probably not limited to the genetic predisposition to primary osteoarthritis, since members of this family also exhibit the clinical characteristics of a mild chondrodysplasia.

Ehlers-Danlos Syndrome Type IV and Familial Arterial Aneurysms

EDS type IV , or the ecchymotic form, is the most severe form of EDS (158, 230). Unlike the other forms of EDS, it is associated with the presence of only minimal skin hyperextensibility and joint hypermobility, which mainly affects the digits. The skin of these patients is velvety and translucent, allowing the visualization of underlying blood vessels. In addition, the skin is fragile and easily bruised, resulting in the ecchymoses which are characteristic of EDS type IV. However, the primary concern of these patients is their predisposition to aneurysms leading to the rupture of large arteries and hollow organs, particularly the colon, which commonly results in life-threatening hemorrhages. The clinical manifestations of EDS type IV exhibit some overlap with those observed in familial arterial aneurysm cases. In this disorder, patients are predominantly affected with fragility of the vascular organs with little or no effects on skin or joints.

It was initially believed that EDS type IV consisted of autosomal dominant and recessive forms (271-273). However, it is now generally accepted that most patients possess dominantly-inherited mutations (272, 274-276).

The biochemical analysis of tissues from EDS type IV and familial arterial aneurysm patients revealed abnormally low quantitites of type III collagen, levels that were 10-20% of those observed in normal individuals (271, 277, 278). In addition, skin fibroblasts from most patients secreted decreased amounts of type III procollagen into the culture medium, indicating a defect in its biosynthesis or secretion (275, 277, 279-281). Some EDS type IV fibroblast cell lines exhibit dilated ER compartments, indicative of an intracellular accumulation of structurally altered type III procollagen (277, 281). In contrast, fibroblasts from other patients appear to secrete normal amounts, despite a demonstrated deficiency of type III collagen in their tissues (279, 282). However, fibroblasts from one of these patients appeared to secrete type III procollagen with an increased sensitivity to proteolysis. This anomaly would explain the low levels of type III

collagen in the tissues of this patient. The collagen fibrils in the tissues were also abnormal in structure, exhibiting decreased diameters as compared to those in normal individuals (277).

The association between defects in type III procollagen and EDS type IV has also been demonstrated by the linkage analysis of a number of families (283, 284). These studies showed linkage between specific COL3A1 alleles and the segregation of EDS type IV.

The availability of prepro-α1(III) cDNA probes has permitted the identification of specific mutations in a number of EDS type IV patients, as reviewed recently by Kuivaniemi et al (173). The same types of mutations observed in OI have also been demonstrated in the COL3A1 genes of EDS type IV and arterial aneurysm patients. The defects which have been described to date include point mutations resulting in glycine substitutions within the α1(III) triple helical domain (285-288). Large genomic multi-exon deletions (274, 289) and mutations affecting mRNA splicing (290-293), resulting in the deletion of varying numbers of amino acid residues in the triple helical domain, have also been observed. The similarity of these defects to those observed in OI patients suggests that they may have similar effects on procollagen assembly, stability, degradation and secretion. However, similarly to type II collagen, the homotrimeric structure of type III collagen results in more extensive "protein suicide", in comparison with type I collagen defects, and accounts for the very low tissue levels of type III collagen exhibited by EDS type IV patients.

An EDS type IV patient whose fibroblasts secreted normal amounts of type III collagen was shown to harbour a mutation resulting in a glycine → serine substitution at position 790 of the α1(III) triple helical domain. This defect did not appear to impair helix stability enough to affect the rate of secretion, yet exposed a nearby arginine residue which constitutes a protease cleavage site (285).

Individuals from a family exhibiting a predisposition to arterial aneurysms were found

to possess a point mutation which resulted in the outsplicing of exon 20 from the prepro-α1(III) mRNA (290, 291). This indicates that some familial arterial aneurysm cases result from defects similar to those observed in EDS type IV patients. The reasons for the differences in the involvement of non-vascular tissues between these two disorders is at present unclear.

Based on these observations, it appears that the defects present in EDS type IV patients can be separated into two groups. The first group consists of defects resulting in impaired secretion of type III procollagen molecules containing structurally altered chains. The second group consists of defects resulting in a normal rate of secretion but in decreased resistance to extracellular degradation. For both types of defects, it is likely that a small proportion of the abnormal type III collagen remains in the ECM and is incorporated during fibrillogenesis, thereby altering the structure of the collagen fibrils.

The demonstration of type III collagen mutations in EDS type IV patients is consistent with the pattern of tissue involvement in this disorder. Tissues expressing type III collagen at significant levels, such as skin, gut and the vascular system, are adversely affected, while tissues with little or no type III collagen are unaffected (2). The lack of evident bone abnormalities in EDS type IV patients is due to the absence of significant amounts of type III collagen in this tissue. The phenotypic effects of defects in type III collagen demonstrate the vital role of this ECM component in providing various soft tissues with the ability to withstand mechanical stresses.

1.8 Role of Type I Collagen in the Structure of Bone Mineral

A feature which distinguishes bone from other connective tissues such as skin, ligaments and tendons, is the presence of a mineral phase which is intimately associated with the organic phase (294). This feature is responsible for conferring the rigidity to mineralized bone in comparison to unmineralized tissues. The mineral phase accounts for

approximately 75% of dry bone mass; the remaining 25% comprises the organic phase, of which about 90% consists of collagen (295). Other matrix components such as phosphoproteins, adhesive glycoproteins and proteoglycans constitute the remaining 10% of the organic phase (295, 296). Type I collagen is the major collagenous constituent of bone, with minor quantities of type V collagen also being present (2). Type I collagen appears to play a unique role in bone. Rather than being the major contributor of tensile strength, as in tissues such as skin and tendon, type I collagen serves as a molecular scaffold for mineralization (294). It is estimated that at least 50% of the mineral present in bone is localized within the hole zones of type I collagen fibrils (295), the locations in which the crystals appear to form and grow (294). The balance of the bone mineral may be located between collagen molecules within fibrils, or surrounding individual fibrils (297).

In contrast to the extensive advances in the understanding of the structure of type I collagen fibrils in bone, many questions concerning the mineral crystal structure have remained unanswered, in part due to the extremely small dimensions involved. The electron microscopic examination of mineralized collagen fibrils has revealed that the carbonate apatite (dahllite) crystals, which constitute the basic building blocks of the mineral phase (294), are in the shape of plates (298), rather than needles as initially believed. The plates are approximately 50 x 25 nm in size and 2-3 nm thick (298, 299). One theoretical model for the three-dimensional packing of collagen molecules predicts a hexagonal arrangement, and the alignment of hole zones for an undetermined distance transverse to the longitudinal axis of the fibril (300). This would create parallel layers of transverse grooves with approximate cross-sectional dimensions of 37 x 2 nm, and spaced at 4 nm intervals. This model is supported by the close agreement between the theoretical estimates of crystal size and the values obtained from the analysis of the mineralized regions of turkey tendons. Within an individual collagen fibril, crystals are arranged in parallel layers separated by 4-5 nm (301). Since the fibrils within an individual bone

lamella have their long axes in a roughly parallel direction, the mineral crystals within an entire lamella are oriented in the same direction along one axis, taking into account the rotation of fibrils relative to each other (294). Since the collagen fibrils in adjacent bone lamellae are oriented in different directions, the mineral crystals are also. This feature therefore provides bone with the ability to withstand mechanical forces in different directions.

The close association between the mineral and organic phases in bone indicates that the organization of collagen molecules within the collagen fibrils is an extremely important determinant of the structure of the mineral crystals. It is therefore conceivable that any changes in collagen structure which affect fibril architecture, such as those observed in OI, would also have profound effects on mineral structure and bone strength. In addition, the decreased amount of collagen in the matrix likely gives rise to a diminished quantity of mineral. However, to date the correlation between abnormalities in collagen structure and bone mineral structure has not been significantly addressed.

1.9 Approaches to the Identification of Disease Mutations

Recent advances in the area of molecular biology techniques, most notably the development of PCR (302, 303), have made possible an acceleration in the characterization of molecular defects in numerous genetic disorders, and of sequence variation present in the normal population. In addition, these techniques have allowed the detection of many such biochemical defects pre-natally, as well as the possibility of genetic counselling for parents of affected individuals (304, 305).

The task of identifying mutations and polymorphisms, however, still remains considerable, since this may involve the detection of a sequence change as subtle as a single

base substitution in genes that are many kilobase pairs in length. Genetic defects such as genomic insertions and deletions or splicing alterations are more readily detected due to their more obvious effects on gene and mRNA structure Since the sequencing of large regions of DNA is a long and tedious process, a more practical and rapid technique is needed to scan long genetic sequences for all mutations and polymorphisms. Many such techniques have been described and have recently been reviewed (306, 307). However, each technique has its own limitations and advantages.

One of the most commonly employed techniques is the detection of RFLPs (308, 309). The disadvantages of this method are that only mutations affecting restriction sites are detected and that many enzymes may have to be tested before a change is observed. Another method which has been described involves changes in the mobility of DNA:DNA homoduplexes or heteroduplexes upon electrophoresis on a denaturing gradient gel, which detects sequence changes resulting in altered helix stability in the low-melting region of a DNA molecule (310-312). However, this procedure may not detect sequence changes occurring in the more thermally stable regions having high GC contents. This limitation can be circumvented by the addition of GC clamps, very stable helical sequences, to the end of the molecule of interest (313-315). A variety of enzymatic protection methods have also been described which are able to detect mismatched bases in DNA:DNA, RNA:DNA and RNA:RNA heteroduplexes. S1 nuclease (316, 317), RNases A, T1 and T2 (318-320), and ABC nuclease following carbodiimide treatment (321) have all been used with only partial success, due to their inability to detect all possible combinations of base-pair mismatches.

One of the most recently published methods is the chemical cleavage analysis of DNA:DNA heteroduplexes, as described by Cotton et al (322). The technique has also been modified for the analysis of RNA:DNA heteroduplexes (205, 323). Both techniques take advantage of the increased reactivity of specific mismatched bases with reagents such

as HA and OT, which preferentially modify mismatched C and T bases, respectively. By using both DNA strands as probes, all possible base-pair mismatches can be detected. While the use of RNA:DNA heteroduplexes only allows the detection of 2 out of 4 possible base changes, it avoids the need for additional steps to generate cDNA, which can be performed if information on the other base changes is required.

1.10 Aims of the Thesis

The aims of this thesis are to:

1) develop a strategy for the detection and identification of type I collagen gene mutations in OI.

A more thorough understanding of the manner in which type I collagen defects result in the bone fragility characteristic of OI will require the identification of the mutations in a large number of patients. Therefore, there is a need for a methodological approach which can rapidly detect and identify such mutations. In addition, since a large proportion of the defects in OI patients are point mutations, the approach must be sensitive enough to detect single base substitutions.

2) investigate the frequency and distribution of sequence polymorphisms within the coding regions of the COL1A1 and COL1A2 genes.

In the course of searching for disease mutations, it is important to consider that not all mutations give rise to disease. Thus, it is critical to be able to distinguish between those sequence changes resulting in disease, and those present within the normal population (ie. polymorphisms). A knowledge of the polymorphisms occurring within the COL1A1 and COL1A2 gene coding regions will therefore assist in identifying those mutations which are

disease related. In addition, polymorphisms will provide important information about the regions of the collagen polypeptide chains which are tolerant of mutations, and which are therefore less critical for the normal function.

3) identify the mutations in a group of OI patients exhibiting a diverse range of clinical severity, in order to examine the relationship between specific biochemical defects in type I collagen and their phenotypic expression.

In recent years, the description of the mutations in over 70 OI patients have allowed the identification of patterns in phenotypic severity relative to the nature and location of defects, particularly for glycine substitutions. However, when considering the diversity in the types (ie. glycine substitutions, deletions, etc.) and locations (ie. different chains, different residues, etc.) of the defects possible in type I collagen, it is clear that the OI mutations which have been described to date represent only a tiny fraction of these. Thus, much remains to be answered concerning the phenotypic manifestations of defects which have not yet been observed (eg. Gly $\rightarrow$ Trp substitutions, etc.). In addition, the growing number of exceptions, which are incompatible with the "phenotypic severity rules", underlines the incomplete understanding of the manner in which these defects determine clinical outcome. The characterization of additional mutations, particularly in patients exhibiting a wide range of OI phenotypes, will contribute to this understanding, and will allow more accurate prognosis of the disease severity in OI patients.

4) examine the effects of different type I collagen gene mutations at the level of bone architecture.

To date, relatively few OI patients have been examined at the bone histological level. As a consequence, the relationship between phenotypic severity and the underlying changes in bone architecture is poorly understood. In addition, since few of those patients which

have been examined have also had their collagen defects identified, the nature of the correlation between the biochemical defects and the abnormalities in bone architecture is unknown. This relationship will be examined by identifying the collagen defects in a group of patients exhibiting different forms of OI, and wherever possible, a comparison of bone architecture will be made.

Chapter 2

Materials and Methods

2.1 Materials

T4 polynucleotide kinase, Klenow fragment of E. coli DNA polymerase I and Moloney Murine Leukemia Virus reverse transcriptase were from New England Biolabs. Calf intestinal alkaline phosphatase, dNTPs, ATP, oligo-dT, M13mp18 RFI DNA and NICK columns were from Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer Cetus. Taq DNA polymerase, L-[5-^{3}H]proline, [α-^{32}P]dCTP, [α-^{32}P]dGTP, [α-^{32}P]dTTP, [α-^{35}S]thio-dATP and Amplify were from Amersham. The T7 DNA polymerase sequencing kit was from Bio/Can, and RNasin was from Promega. T4 DNA ligase and T4 DNA polymerase were from Boehringer-Mannheim. HaeIII-digested ϕX174 DNA standards and NACS•52 PREPAC columns were from Gibco-BRL. OPC columns were from Applied Biosystems, and the NA45 DEAE-cellulose membrane was from Schleicher & Schuell. HA hydrochloride, diethylamine and OT were from Aldrich Chemicals. Piperidine, pyridine and light mineral oil were from Fisher Scientific. Pepsin, cyanogen bromide and lysozyme were from Sigma Chemicals. EcoLite (+) scintillant cocktail was from ICN-Flow.

2.2 Sources of Human Fibroblast Cell Lines

The fibroblast cell lines utilized in these studies, summarized in Table 4, were obtained from two sources. Control cell lines C10, C19 and C40 were obtained from the National Institute of General Medical Sciences Human Genetic Mutant Cell Repository (accession numbers GM3349A, GM3377A and GM0321, respectively; Camden, NJ, U.S.A.), and were derived from three clinically normal individuals with normal karyotypes. All other fibroblast cell lines were derived by outgrowth from the explant culture of skin biopsies

Table 4. Individuals from whom fibroblast cell lines were derived.

A. Control individuals

Cell Line	Diagnosis	Sex	Race	Age at Biopsy	Current Age
C10	Normal	M[1]	Black	10 y	U
C19	Normal	M	Caucasian	19 y	U
C40	Normal	F	Caucasian	40 y	U
BD268	Normal	M	Caucasian	31 y	36 y
YL278	Normal	M	Caucasian	23 y	27 y

B. OI patients and parents

Cell Line	Diagnosis	Sex	Race	Age at Biopsy	Current Age
GD199	OI type IA	M	Caucasian	2 y	10 y
SL211	OI type IVB	F	Caucasian	1 y	7 y
ML226 [2]	Normal	M	Caucasian	25 y	30 y
JL227 [2]	Normal	F	Caucasian	25 y	30 y
PP238	OI type IV	M	Caucasian	1.5 y	6 y
PM244	OI type I	F	Caucasian	5 y	10 y
JN245	OI type IV	M	Caucasian	10 y	14 y
MB252	OI type IA	M	Caucasian	10 y	15 y
DP266	OI type III	F	Caucasian	5 y	10 y
NC323	OI type III	F	Caucasian	7 y	9 y
SS333	OI type IIA	F	Caucasian	20 w gest.	D
MK345	OI type III	F	Native Indian	4 w	D

[1] Abbreviations used: M, male; F, female; y, years; w, weeks; w gest., weeks of gestation; U, unknown; D, deceased.

[2] Cell lines ML226 and JL227 were derived from the asymptomatic parents of patient SL211.

collected from additional control individuals (BD268 and YL278), or from the OI probands and parents described below. The skin biopsies were aseptically collected by qualified medical personnel, with appropriate consent having been obtained.

2.3 Patients' Clinical Descriptions

The OI patients included in this study were examined and followed up by physicians at the Shriners Hospitals in Montreal, QC and Springfield, MA, and at the Royal Victoria Hospital, Montreal, QC. The classification of the patients into the various OI types and subtypes was performed by qualified physicians, and was based on physical examinations, radiological findings, and the presence or absence of other clinical features (such as scleral coloration and dentinogenesis imperfecta). The OI classifications and phenotypic features of the patients included in this study are summarized in Table 5. Proband SS333 was a fetus aborted at a gestational age of 20 weeks, due to skeletal abnormalities detected by ultrasonography. The fetus was diagnosed with short-limbed dwarfism, and exhibited the characteristic physical and radiological appearances of type IIA OI. Cytogenetic analysis of skin fibroblasts from the fetus found no evidence of chromosomal abnormalities. MK345 was a patient with severe type III OI, who died from respiratory failure at the age of four weeks, as a consequence of fractures suffered during natural delivery. Although this proband perished in the perinatal period, the radiological findings were not consistent with the classical appearance of type II OI, but with that of severe type III OI. MK345 was the second affected offspring of clinically unaffected parents, whose first child also died in the perinatal period.

Table 5. Clinical features of the OI patients.

Patient	OI type	Sclera	Teeth	Stature	Deformity
GD199	IA	Blue	Normal	Short[1]	Mild
SL211	IVB	Blue	DI[2]	Normal	Mild
PP238	IV	U	U	U	U
PM244	I	U	U	U	U
JN245	IV	U	U	U	U
MB252	IA	Blue	Normal	Short	Mild
DP266	III	Normal	DI	Short	Moderate
NC323	III	Normal	DI	Short	Severe
SS333	IIA	U	N/A	Small	Severe
MK345	III	Blue	N/A	Small	Severe

[1] The term "short stature" refers to heights below the third percentile.

[2] Abbreviations used: DI, dentinogenesis imperfecta; U, unknown; N/A, not applicable.

2.4 Skin Explant and Fibroblast Cell Culture

Skin explants were cultured in 60-mm plastic petri dishes with MEM/10% FBS containing 50 units/ml penicillin-G and 25 µg/ml gentamycin sulfate. Outgrowing cells and subsequent confluent subcultures were passaged (1:2 to 1:4) by incubating the cell monolayers with 0.25% trypsin in Puck's medium, recovering the cells by centrifugation at 2000xg for 5 min at 4°C, and thereafter culturing in plastic flasks with MEM/10% FBS lacking antibiotics. Fibroblast monolayers intended for the isolation of total cellular RNA were grown to early confluence, and in order to induce the expression of type I collagen mRNAs, they were then incubated in medium supplemented with 50 µg/ml L-ascorbic acid for 72 h, with daily changes of medium (325). In some experiments, fibroblast monolayers were grown to late confluence in the absence of ascorbic acid stimulation. However, unless specifically indicated, fibroblast monlayers were **routinely** stimulated with ascorbic acid prior to isolation of total RNA.

2.5 Metabolic Labelling of Type I Procollagen

The method used is based on that described by Barsh and Byers (202). Fibroblasts were seeded into six 35-mm wells at a density of 2.5 x 10^5 cells/well, and allowed to attach for 24 h in MEM/10% FBS. The cell monolayers were rinsed and incubated for 4 h with 1 ml/well of serum-free low-glucose DMEM containing 50 µg/ml L-ascorbic acid, which was then replaced with 900 µl/well of the same medium containing 55.5 µCi/ml L-[5-^{3}H]proline (15-40 Ci/mmol). Following a 16-h incubation, the culture media were harvested and adjusted to 25 mM EDTA, 1 mM p-aminobenzamidine, 1 mM phenylmethylsulfonylfluoride and 10 mM N-ethylmaleimide. Chicken skin type I collagen

(dissolved in 50 mM Tris-HCl, pH 7.5) was then added as a carrier to a final concentration of 90 µg/ml. Proteins were precipitated overnight at 4°C with ammonium sulfate at 30% saturation, and recovered by centrifugation at 16,000xg for 30 min at 4°C. Following a wash with 30%-saturated ammonium sulfate, the pellet was vacuum-dried and resuspended in 1.0 ml of 0.5 M acetic acid for subsequent pepsin digestion.

2.6 Pepsin Digestion of Labelled Procollagen

L-[5-^{3}H]proline-labelled procollagen was digested with 50 µg/ml porcine gastric mucosa pepsin in 0.5 M acetic acid for 4 h at 4°C. Digestions were terminated by adding NaOH to a final concentration of 0.5 M, and incubating for 5 min on ice. Collagen was then precipitated for 16 h at 4°C following the addition of an equal volume of 4 M NaCl/0.5 M acetic acid, and recovered by centrifugation at 16,000xg for 30 min at 4°C. Pellets were resuspended in 200 µl of SDS-PAGE sample buffer (125 mM Tris-HCl, pH 6.8/2% SDS/10% glycerol/0.002% bromophenol blue).

2.7 Electrophoretic Analysis of Collagen Polypeptide Chains

Samples of pepsinized media were dialyzed against 100 volumes of SDS-PAGE sample buffer, and assayed for ^{3}H-radioactivity by liquid scintillation counting. Samples of pepsinized media, containing 45,000 DPM of radioactivity, were analyzed under non-reducing conditions to permit the separation of α1(I) and α1(III) chains. Under these conditions type III collagen chains migrate as disulfide-bonded homotrimers, $[\alpha 1(III)]_3$, due to the presence of cysteine residues within the triple helical domain (30). Samples

were uenatured at 100°C for 3 min, and applied to a discontinuous SDS-polyacrylamide gel consisting of a 5% resolving gel (11 cm X 14 cm X 1 mm) and 3.5% stacking gel, each containing 2 M urea. Gels were fixed and stained by soaking for 30 min in 0.1% Coomassie Brilliant Blue R-250/40% methanol/10% acetic acid, destained overnight in 15% methanol/7.5% acetic acid, and subsequently incubated for 30 min in Amplify, before being dried and exposed to Kodak X-OMAT film for fluorographic visualization of proteins.

2.8 Cyanogen Bromide Peptide Mapping

Samples of pepsinized media, containing 650,000 DPM, were electrophoresed as described above, and the unfixed gel strips were stored at -80°C until used for peptide mapping. Resolved collagen chains were subjected to cyanogen bromide cleavage within the polyacrylamide gel matrix using a modification of the method described by Barsh et al (326). Briefly, one-dimensional gel strips were thawed at room temperature, equilibrated with 70% formic acid, and then incubated with 100 mg/ml cyanogen bromide in 70% formic acid under nitrogen for 3 h at room temperature. The strips were washed with 5 changes of 500 mM Tris-HCl (pH 6.8)/30% glycerol/0.002% bromophenol blue over a period of 90 min, or until the indicator dye remained blue, and were then equilibrated with SDS-PAGE sample buffer for 15 min. Each gel strip was placed on top of a discontinuous SDS-polyacrylamide gel consisting of an 8-16% gradient resolving gel (11 cm X 14 cm X 1 mm) and 5% stacking gel. After electrophoresis, the gels were fixed in destaining solution, and analyzed by fluorography.

2.9 Isolation of Total Fibroblast RNA

Total RNA was isolated by an acid guanidine isothiocyanate-phenol extraction method based on that described by Chomczynski and Sacchi (327). Fibroblast cell monolayers (25-30 x 10^6 cells in 5 or 6 T150 flasks) were washed twice with ice-cold PBS, extracted directly in 4 ml/T150 flask of a denaturing solution containing 4 M guanidine isothiocyanate, 25 mM sodium citrate, 0.5% sodium lauroyl sarcosine and 0.1 M 2-mercaptoethanol, and transferred to a 50-ml conical tube. Subsequently, 2 M sodium acetate, pH 4.0 (400 µl/flask), phenol (4 ml/flask) and 49:1 chloroform:isoamyl alcohol (800 µl/flask) were sequentially added with thorough vortexing after each reagent addition. The emulsion was placed on ice for 15 min, and centrifuged at 10,000xg for 20 min at 4°C. The aqueous phase was precipitated with one volume of isopropanol at -20°C for 1 h, and centrifuged at 10,000xg for 20 min at 4°C. The pellet was dissolved in 1.2 ml of the denaturing solution, and reprecipitated in 1.5-ml eppendorf tubes with one volume of isopropanol at -20°C for 15 min. The RNA was collected by centrifugation at 16,000xg for 7 min at 4°C, rinsed with 1 ml of absolute ethanol, vacuum-dried and resuspended in 400 µl of TE. In some experiments, the RNA was then further purified by re-extracting with phenol:chloroform at either pH 8.0 or 4.0, and recovered by ethanol precipitation.

2.10 Transformation of E. coli Strain DH1

Transformations were always performed using competent bacterial cells which were freshly prepared by a procedure based on that described by Hanahan (328). A dense bacterial culture was prepared by innoculating 5 ml of SOB medium (329) with an aliquot of a stock culture of E. coli DH1 (frozen at -80°C in medium containing 15% glycerol; strain genotype described in reference 330), and incubating overnight at 37°C in a rotary

shaker (275 RPM). The overnight culture was utilized to innoculate (1:50) two fresh 25-ml cultures, one of which served to monitor bacterial cell growth until an OD_{550} of between 0.45 and 0.55. The other culture was then chilled on ice, and centrifuged at 1000xg for 12 min at 4°C. The bacterial pellet was gently resuspended in 8 ml of an ice-cold buffer containing 10 mM Tris-HCl (pH 8.0) and 50 mM $CaCl_2$, and incubated on ice for 20 min. The bacteria were recentrifuged and resuspended in 2 ml of the same buffer. The transformation samples, consisting of 5 µl duplicates of ligation reaction mixtures, 1 ng of supercoiled plasmid DNA, or no DNA, were combined with 200 µl of the bacterial suspension, incubated on ice for 1 h, and heat-shocked at 42°C for 90 s. Pre-warmed SOB medium (3 ml) was then added to each tube, and the cultures were incubated at 37°C for 1 h in a rotary shaker (175 RPM). Aliquots of the cultures (200 µl) were then applied to plates containing SOB agar supplemented with 50 µg/ml ampicillin. The plates were incubated overnight at 37°C, and thereafter stored at 4°C. Transformation efficiencies were monitored by counting the number of colonies obtained with pBR322 plasmid DNA.

2.11 Large-Scale Isolation of Plasmid DNA

The method used is an adaptation of procedures described by Sambrook et al (331-333). A dense bacterial culture was prepared by innoculating 5 ml of TB medium (329, 334) containing 50 µg/ml ampicillin with an aliquot of a stock culture of the bacterial clone (frozen at -80°C in broth containing 15% glycerol), and incubating overnight at 37°C in a rotary shaker (275 RPM). A portion of the overnight culture (600 µl) was utilized to innoculate 200 ml of fresh TB medium, which was then incubated at 37°C for 16-18 h. The bacterial culture was cooled on ice, and centrifuged at 1800xg for 20 min at 4°C. The cells were resuspended in 60 ml of a solution containing 50 mM D-glucose, 25 mM Tris-

HCl (pH 8.0), 10 mM EDTA and 3 mg/ml chicken egg white lysozyme, and incubated at room temperature for 10 min. The bacteria were lysed by combining the suspension with 120 ml of freshly-prepared 0.2 M NaOH/1.0% SDS, and incubating on ice for 10 min. Chromosomal DNA and protein were precipitated by adding 60 ml of ice-cold 3 M potassium acetate/2 M acetic acid, and incubating on ice for 10 min. Following centrifugation at 4000xg for 15 min at 4°C, the supernatant was filtered through cheesecloth, and precipitated with an equal volume of isopropanol at -20°C for 1 h. Plasmid DNA and RNA were recovered by centrifugation at 16,000xg for 20 min at 4°C, and were resuspended in 6 ml of water. The solution was combined with 7 ml of 5 M LiCl/5 mM MOPS, and reprecipitated with an equal volume of isopropanol. After recentrifugation, the pellet was rinsed with 70% ethanol, resuspended in 8 ml of TE, and combined with 8.8 g of cesium chloride. The cesium chloride was dissolved by heating at 65°C for 2 min, and 800 µl of 10 mg/ml ethidium bromide was added. The mixture was centrifuged in two sealed polyallomer tubes at 80,000 RPM for 18 h at 22°C in a Beckman VTi80 rotor. The plasmid DNA fractions were collected into a fresh polyallomer tube, which was then filled with a solution of equal density (prepared by combining 21 ml of TE, 2 ml of 10 mg/ml ethidium bromide and 22 g of cesium chloride) and recentrifuged for 6 h under identical conditions. Plasmid DNA was again collected, transferred to a 15-ml conical tube, and extracted with equal volumes of water-saturated isopropanol until the aqueous phase lost its pink color. The DNA was diluted with two volumes of water, precipitated with ethanol, and resuspended in 2 ml of 10 mM Tris-HCl, pH 9.0/1 mM EDTA/0.1 M NaCl. Following phenol:chloroform extraction, the DNA was reprecipitated with ethanol, resuspended in TE, and quantitated by absorbance at 260 nm.

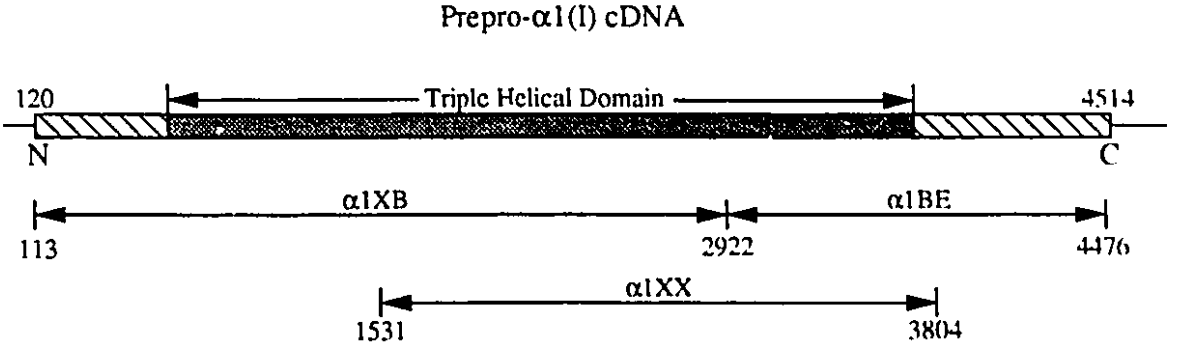
2.12 Electroelution of DNA from Agarose Gels

The method was based on procedures which have been described by Sambrook et al (335, 336). A restriction enzyme digest of plasmid DNA, containing the fragment of interest, was electrophoresed on a preparative 0.8% agarose gel in TBE/0.25 µg/ml ethidium bromide, and resolution was monitored using a long-wave ultraviolet light source. When satisfactory resolution was obtained, the band of interest was excised and placed in a dialysis bag with approximately 1 ml of 0.5X TBE/0.25 µg/ml ethidium bromide. The DNA was eluted from the gel slice by electrophoresis at 200V, and the eluate inside the dialysis bag was transferred to a 15-ml conical tube. The dialysis bag was rinsed with two 1-ml aliquots of 0.5X TBE and the rinses were pooled with the eluate. The sample of eluted DNA was extracted with phenol:chloroform, and then concentrated to a volume of 0.5-1.0 ml by repeated extraction with two volumes of isobutanol. Following ethanol precipitation, the DNA was resuspended in 500 µl of 10 mM Tris-HCl, pH 7.5/1 mM EDTA/0.1 M NaCl. The DNA was then heated at 65°C for 5 min, and purified using a NACS•52 PREPAC DEAE-resin column as recommended by the supplier. The DNA was reprecipitated with ethanol, resuspended in TE, and quantitated by absorbance at 260 nm.

2.13 Preparation and Labelling of the Probes Used in Chemical Cleavage Analysis

The probes used in chemical cleavage mismatch detection were prepared using full-length human prepro-α1(I) and prepro-α2(I) cDNA plasmid constructs generously provided by Dr. F. Ramirez (Mount Sinai School of Medicine, New York, NY), and are summarized in Fig. 7 and Table 6. E. coli strain DH1 bacteria were transformed with the constructs, and large-scale plasmid preparations were obtained. The isolated plasmids

A



B

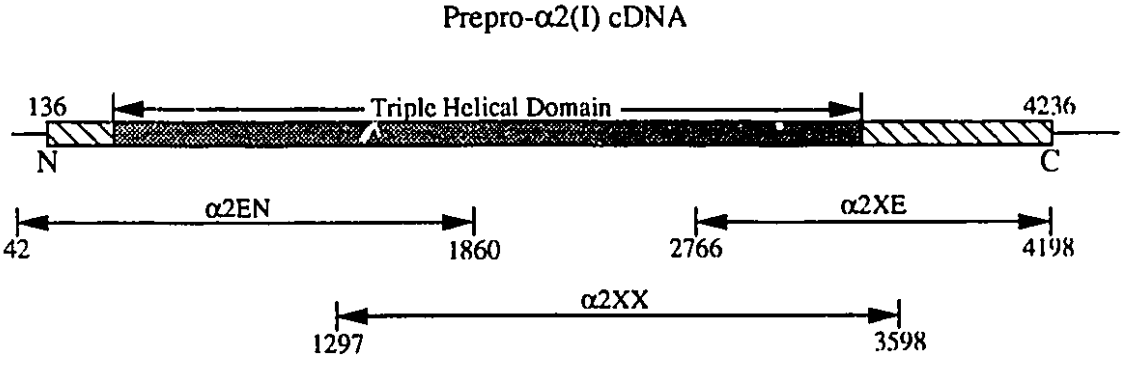


Fig. 7. Schematic diagram of the cDNA probes used in chemical cleavage mismatch analysis. The prepro-α1(I) (A) and prepro-α2(I) (B) probes are represented by double-headed arrows below the bar diagrams of the corresponding full-length cDNAs. cDNA coding regions are indicated by bars, and untranslated regions are indicated by plain lines. Regions encoding the main triple helical domains (grey shading) and terminal domains (diagonal shading) are also indicated. cDNA sequence positions are relative to the transcription start sites (78, 80). The probe abbreviations (ie. α1XB, α2XX, etc.) refer to the restriction fragments used, and are summarized in Table 6.

Table 6. cDNA probes used in chemical cleavage mismatch analysis.

Probe [1]	cDNA Region Covered [2]	Approx. Size (kbp)	Annealing Temperature (°C) [3]	Labelled With
α1XB	113-2922	2.8	61	^{32}P-dCTP
α1BE	2922-4476	1.6	60	^{32}P-dGTP
α1XX	1531-3804	2.3	63	^{32}P-dTTP
α2EN	42-1860	1.8	60	^{32}P-dGTP
α2XE	2766-4198	1.4	55	^{32}P-dGTP
α2XX	1297-3598	2.3	60	^{32}P-dTTP

[1] The restriction enzyme fragments used as probes are identified by the following abbreviations: α1XB, XbaI-BamHI; α1BE, BamHI-EcoRI; α1XX, XhoI-XhoI; α2EN, EaeI-NcoI; α2XX, XhoII-XbaI; α2XE, XbaI-EcoRI.

[2] The numbers given are the positions of the first bases of the corresponding restriction endonuclease recognition sites within the prepro-α1(I) and prepro-α2(I) cDNA sequences relative to the transcription start sites (78, 80).

[3] The annealing temperatures used in chemical cleavage mismatch analysis with these probes were determined empirically. However, approximations were obtained using the mathematical relationship between %GC content and the melting temperatures of DNA:DNA and RNA:DNA heteroduplexes, which has been described by Dean (337).

were digested with the appropriate restriction endonucleases, and the probe fragments were purified by electroelution of the DNA from preparative agarose gels.

The probes were end-labelled by incubating 100 ng of probe DNA with 30 μCi [α-^{32}P]dCTP, dGTP or dTTP (>3000 Ci/mmol), and 10 units of Klenow fragment in the presence of 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT and with or without 20 μM each of the three unlabelled dNTPs. The α-^{32}P-labelled dNTP used for each probe is listed in Table 6, and corresponded to the first base expected to be incorporated at the recessed 3' end of the non-coding strand. Incubations were performed in a final volume of 20 μl at 22°C for 30 min. The samples were adjusted to a volume of 50 μl and 25 mM EDTA, and the labelled probes were separated from unincorporated label by gel filtration on NICK columns pre-equilibrated with TE. Probe specific activities were 5-10 x 10^6 or 2-5 x 10^7 DPM/μg DNA in the absence or presence of the unlabelled dNTPs, respectively.

2.14 Chemical Cleavage Mismatch Detection

The method was based on a modification (323) of the HOT chemical cleavage technique described by Cotton et al (322). Total fibroblast RNA (10 or 25 μg) was combined with an end-labelled cDNA probe (5-20 ng) from Table 6, and the nucleic acids were ethanol-precipitated in an ethanol-dry ice bath for 5 min (as for all subsequent precipitations in this procedure). The RNA/DNA was recovered by centrifugation for 15 min at 4°C, vacuum-dried and resuspended in 50 μl of a buffer containing 80% (v/v) formamide, 40 mM PIPES (pH 6.4), 1 mM EDTA and 0.4 M NaCl. The sample was denatured at 90°C for 10 min, and incubated for 2 h at the annealing temperature listed in Table 6. The RNA:DNA heteroduplex was ethanol-precipitated, resuspended in 26 μl of water, and divided into 4 aliquots of 6.5 μl. For the detection of C-base mismatches, the RNA:DNA samples were

combined with 20 µl of 6.7 M HA hydrochloride/3.2 M diethylamine, and incubated at 37°C for 0, 15, 30 and 60 min. For the detection of T-base mismatches, the RNA:DNA samples were combined with 2.5 µl of 100 mM Tris-HCl, pH 7.5/10 mM EDTA/15% (v/v) pyridine and 15 µl of 0.4% OT, with incubations then carried out at 37°C for 0, 5, 10 and 15 min. The chemical reactions were terminated by the addition of 200 µl of a solution containing 0.3 M sodium acetate (pH 5.5), 0.1 mM EDTA and 25 µg/ml yeast tRNA. The RNA:DNA was ethanol-precipitated, resuspended in 50 µl of 1.0 M piperidine, and incubated at 90°C for 30 min. The RNA:DNA was again ethanol-precipitated, and the pellets were resuspended in 5 µl of water. Equal volumes of loading buffer, consisting of 95% (v/v) formamide, 20 mM EDTA (pH 8.0), 0.05% bromophenol blue and 0.05% xylene cyanole FF, were added. After overnight storage at -20°C, two aliquots of the samples were denatured at 90°C for 3 min, and applied to a 5% polyacrylamide/7 M urea gel in 0.5X TBE. Samples were electrophoresed alongside end-labelled (338) HaeIII-digested ϕX174 DNA molecular weight standards at 75 W for 1 to 4 h. Gels were covered with plastic wrap and packed with Kodak X-OMAT film overnight at -80°C.

2.15 First Strand cDNA Synthesis

First strand cDNA was prepared by incubating 25 µg of total fibroblast RNA in a 50-µl reaction mixture containing 2.5 µg of oligo-dT or random hexamer primers, 50 units of RNasin, 100 units of Moloney Murine Leukemia Virus reverse transcriptase, 1 mM dNTPs, 50 mM Tris-HCl (pH 8.5), 8 mM $MgCl_2$, 30 mM KCl, 10 mM DTT and 100 µg/ml bovine serum albumin at 42°C for 90 min. In order to melt secondary structure, the RNA and primers were pre-mixed in a volume of 35 µl, incubated at 65°C for 5 min, and cooled on ice prior to addition of the other reagents. Following cDNA synthesis, alkaline

hydrolysis of the RNA was performed by adjusting to 50 mM NaOH and 12.5 mM EDTA, and incubating at 65°C for 60 min. The sample was adjusted to neutral pH by the addition of Tris-HCl (pH 7.5) and HCl to concentrations of 200 mM and 40 mM, respectively. First strand cDNA was then ethanol-precipitated and resuspended in 50 µl of water.

2.16 PCR Amplification

The PCR method used was based on that originally described by Saiki et al (303). A 5-µl sample of a first strand cDNA preparation, or 5 ng of plasmid DNA, was incubated with 2.5 units of Taq or AmpliTaq DNA polymerase in the presence of 50 mM KCl, 1.5 or 2.0 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.5), 200 µM dNTPs and 150 pmol each of the sense and antisense primers in a final volume of 100 µl. When Taq DNA polymerase was used in amplification, the reaction mixtures also contained 0.01% gelatin and 0.1% Triton X-100. Amplifications were performed with a 100-µl light mineral oil overlay, and using a Perkin-Elmer Cetus DNA Thermal Cycler, with a 3-min incubation at 94°C, followed by 35 cycles of 94°C for 45 s, 55°C or 60°C for 45 s, and 72°C for 90 s. The final PCR cycle was followed by an incubation at 72°C for 5 min. The oligonucleotide primers and PCR conditions used are detailed in Tables 7 and 8, respectively. When amplifying cDNA for cloning and subsequent sequence analysis, the PCR reactions were performed in duplicate to control for the possible occurrence of Taq DNA polymerase replication errors. Furthermore, control reactions, to which no template was added, were always included for each primer pair used, in order to detect any possible crossover contamination of the PCR reagents. All primers were synthesized on an ABI 381A DNA Synthesizer, purified on OPC columns as recommended by the supplier, and resuspended in water at a concentration of 100 µM.

Table 7. Oligonucleotide primers used in the PCR.

Primer	Sequence (5' → 3')	Gene	Exon(s)
COLL-1	GGAGACCAAACTCACCAT	COL1A2	30
COLL-2	GGAAAAGAAGGTCCTGTC	COL1A2	24
COLL-7	TAGCACCAGTGTCTCCTTT	COL1A1	20
COLL-8	TGGTGCTCGTGGAAATGAT	COL1A1	15
COLL-9	CTGATGGAAACAAGGGTGAA	COL1A2	31, 32
COLL-10	ACCGACTCGCCACGTTCA	COL1A2	36
COLL-11	TCCAAGTGCGACCATCTTTT	COL1A2	48
COLL-12	CGGTTACCCTGGCAATATTG	COL1A2	44
COLL-15	AAGTCGAGGGCCAAGACGAA	COL1A1	1
COLL-16	CTCGTCACAGATCACGTCATC	COL1A1	2

Table 8. PCR conditions and the sizes of the amplified products for each primer pair.

Primer Pair	Annealing Temperature (°C)	Concentration of $MgCl_2$ (mM)	Size of Amplified cDNA Fragment (bp)
COLL-1/COLL-2	55	2.0	358
COLL-7/COLL-8	60	1.5	386
COLL-9/COLL-10	60	1.5	305
COLL-11/COLL-12	60	1.5	364
COLL-15/COLL-16	60	1.5	152

The enzymes Taq and AmpliTaq DNA polymerase are hereafter collectively referred to in the text as Taq DNA polymerase, with no distinction being made between them.

2.17 Blunt-End Cloning of PCR Products into the M13mp18 vector

Extraction of Taq DNA Polymerase

For some preparations of insert DNA, the PCR products were extracted immediately after amplification in order to remove the Taq DNA polymerase prior to blunt-ending. These PCR mixtures were adjusted to 100 µl with TE, extracted with phenol:chloroform, ethanol-precipitated, and resuspended in 36 µl of water for subsequent blunt-ending.

Blunt-Ending of PCR Products

Some PCR products were incubated with either T4 DNA polymerase or Klenow fragment to fill-in recessed 3' ends and to remove 3' overhangs synthesized by Taq DNA polymerase. When performed directly after the PCR, fresh dNTPs were added to 80 µl of PCR mixture at final concentrations of 100 µM, together with either 8 units of T4 DNA polymerase or 10 units of Klenow fragment. When performed after the extraction of Taq DNA polymerase, or following DEAE-cellulose gel purification of the PCR product, the DNA was incubated with 4 units of T4 DNA polymerase or 10 units of Klenow fragment, in the presence of 100 µM dNTPs, 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$ and 1 mM DTT in a final volume of 50 µl. T4 DNA polymerase reactions were incubated at 12°C for 15 min (339), and Klenow fragment reactions were performed at 37°C for 30 min. Following the enzyme incubations, samples were adjusted to 100 µl with TE, extracted with phenol:chloroform and precipitated with ethanol. The DNA was then resuspended in 10 µl of water for subsequent DEAE-cellulose gel purification, or 36 µl for subsequent

phosphorylation.

DEAE-Cellulose Gel Purification of PCR Products

All PCR products were purified from 1.0% or 2.0% agarose gels in TAE, using NA45 DEAE-cellulose membranes by a method similar to that described by Sambrook et al (340). The samples were electrophoresed at 100-150V until sufficient resolution was achieved, and the bands corresponding to the PCR products of expected sizes were electrophoresed at 150-200V onto freshly prepared NA45 membranes. The membranes were rinsed briefly in TAE, and then in a buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM EDTA and 0.1 M NaCl. The DNA was then eluted by placing the membranes in eppendorf tubes containing 200 µl of 50 mM Tris-HCl, pH 8.0/10 mM EDTA/2 M NaCl, and incubating at 65°C for 1 h. The membranes were subjected to two such serial elutions which were then pooled into single samples. The samples of eluted DNA were then extracted with phenol:chloroform, and combined with equal volumes of TE to dilute the NaCl concentration to 1 M. The DNA was precipitated with 2 volumes of ethanol (without the addition of sodium acetate), and resuspended in 36 µl of water for subsequent blunt-ending or phosphorylation.

Phosphorylation of 5' Ends

All PCR products were 5'-phosphorylated by a modification of the method described by Maniatis et al (341). DNA in 36 µl of water was combined with 4 µl of 200 mM Tris-HCl, pH 9.5/10 mM spermidine hydrochloride/1 mM EDTA, heated at 70°C for 5 min, and cooled on ice. The samples were mixed with 5 µl of 500 mM Tris-HCl, pH 9.5/100 mM $MgCl_2$/50 mM DTT/50% glycerol, 2 µl of 100 mM ATP (pH 7.0) and 20 units of T4 polynucleotide kinase in a final volume of 50 µl, and were incubated at 37°C for 30 min. The samples were adjusted to 100 µl with TE, extracted with phenol:chloroform, ethanol-precipitated, and resuspended in 10 µl of water.

Preparation of M13mp18 RFI Vector and Ligation to PCR Products

SmaI-digested M13mp18 RFI (342, 343) vector DNA (5 µg) was dephosphorylated by incubating at 65°C for 30 min with 20 units of calf intestinal alkaline phosphatase, 10 mM Tris-HCl (pH 8.5) and 10 mM NaCl, in a final volume of 200 µl. Vector DNA was then extracted with phenol:chloroform, ethanol-precipitated, and resuspended in water at a concentration of 50 ng/µl.

Blunt-end ligations were performed by incubating 4 µl of each phosphorylated PCR product with 50 ng of dephosphorylated vector in the presence of 1 mM ATP (pH 7.0), 1 mM DTT, 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5% PEG 8000 and one Weiss unit of T4 DNA ligase, in a final volume of 10 µl at 16°C for 20 h (344). Ligation efficiencies were monitored by including a control consisting 50 ng of SmaI-digested vector and no insert DNA.

Transformation of E. coli Strain 71/18

The transformation of E. coli strain 71/18 (genotype described in reference 330) was always performed using freshly competent bacterial cells. A dense culture of 71/18 bacteria was prepared by innoculating 5 ml of 2X YT medium (329) with an aliquot of a bacterial stock (frozen at -80°C in broth containing 15% glycerol), and incubating overnight at 37°C in a rotary shaker (275 RPM). The overnight culture was then utilized to prepare two fresh cultures. The first culture was used to provide fresh log-phase bacteria for the subsequent plating of transformants (see below), and was prepared by innoculating 20 ml of 2X YT medium with one drop of the overnight culture, and incubating at 37°C until required for plating (approximately 4 h). The second culture was used to provide competent bacteria for transformation, and was prepared by innoculating 100 ml of 2X YT medium with 1 ml of the overnight culture, and incubating at 37°C until an OD_{660} of between 0.3 and 0.4. At an appropriate OD_{660}, 60 ml of the second culture was centrifuged at 2500xg for 10 min at

4°C. The bacteria were resuspended in 30 ml of ice-cold 50 mM $CaCl_2$, and incubated on ice for 40 min. Following recentrifugation, the bacteria were resuspended in 6 ml of ice-cold 50 mM $CaCl_2$. The DNA samples, consisting of 5-µl duplicates of the ligation reaction mixtures or of duplicate transformation positive controls (1 ng of M13mp18 RFI plasmid DNA), were combined with 300 µl of the bacterial suspension, incubated on ice for 40 min, and heat-shocked at 42°C for 2 min. The samples were than combined with 40 µl of 20 mg/ml X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside), 40 µl of 23.8 mg/ml isopropyl-β-D-thiogalacto-pyranoside, 200 µl of the fresh log-phase 71/18 bacterial culture (see above) and 3 ml of molten 0.8% agar in B broth (343). The contents were poured onto plates containing 1.2% agar in B broth, allowed to set at room temperature for 15 min, and incubated overnight at 37°C. The numbers of clear and blue plaques on each plate were recorded, and the plates were stored at 4°C. Transformation efficiencies were monitored by counting the number of blue plaques obtained with M13mp18 RFI DNA.

Screening of M13 Recombinants

M13 recombinants were screened for the presence of DNA inserts by the electrophoretic analysis of their ssDNA genomes, as described by Messing (343). A dense culture of 71/18 bacteria was prepared by innoculating 5 ml of 2X YT medium with an aliquot of a bacterial stock (frozen at -80°C in broth containing 15% glycerol), and incubating overnight at 37°C in a rotary shaker (275 RPM). M13 recombinants were then propagated in liquid culture by picking clear plaques into 1.5-ml aliquots of a 1:100 dilution of the overnight culture of 71/18 bacteria. The 1.5-ml cultures were incubated at 37°C for 6 h in a rotary shaker (275 RPM), and were centrifuged in eppendorf tubes at 16,000xg for 15 min at 4°C. The M13 culture supernatants were transferred to fresh tubes, and 20-µl samples were combined with 1 µl of 2% SDS and 3 µl of 15% Ficoll (type 400)/0.25% bromophenol blue/0.25% xylene cyanole FF/1X TBE. Following incubation at 65°C for 5

min, the samples were electrophoresed on a 0.7% agarose gel in TBE, alongside wild type M13 ssDNA and M13 recombinant ssDNAs containing inserts of known sizes. The presence of an insert can be detected by a shift in electrophoretic mobility.

2.18 Isolation of ssDNA Templates for Sequencing

ssDNA templates were prepared by a method based on that described by Messing (343). M13 culture supernatants (1.2 ml) were combined with 300 µl of freshly prepared 20% PEG 8000/2.5 M NaCl, and were incubated at room temperature for 30 min. Viral particles were recovered by centrifugation at 16,000xg for 15 min at 4°C, and were resuspended in 200 µl of 20 mM Tris-HCl, pH 8.0/0.1 mM EDTA/10 mM NaCl. Samples were combined with 100 µl of phenol:chloroform, heated at 65°C for 5 min and centrifuged for 5 min. The aqueous phases were then extracted twice with chloroform, and precipitated with ethanol. The ssDNA was resuspended in 10 µl of TE, and quantitated by absorbance at 260 nm.

2.19 DNA Sequencing

M13 recombinant clones of interest were sequenced by the Sanger dideoxy chain termination method (345), using a commercially available T7 DNA polymerase kit. Briefly, template ssDNA (1 µg) was combined with 1.6 pmol of M13 universal sequencing primer (5' GTAAAACGACGGCCAGT 3'), heated at 65°C for 2 min, and annealed at 37°C for 30 min. Labelling reactions, using [α-^{35}S]thio-dATP (>1000 Ci/mmol), were incubated and terminated as recommended by the supplier. Reactions were stopped by the

addition of 95% (v/v) formamide/20 mM EDTA/0.05% bromophenol blue/0.05% xylene cyanole FF, divided into two aliquots, and stored at -20°C. Samples were denatured at 90°C for 3 min, and electrophoresed on a 6.0% polyacrylamide/7 M urea wedge gel at 75 W for 2 to 6 h. Gels were fixed for 5 min in 5% acetic acid/5% methanol, and dried for autoradiography with Kodak X-OMAT film.

2.20 Histological Analysis of Bone Biopsies

The procedures used were as described by Schenk et al (346). Bone biopsies were fixed in formaldehyde-phosphate buffer (pH 7.1) for 24 to 48 h, and subsequently dehydrated in solutions containing increasing concentrations of ethanol. The biopsies were then cleared using xylenes and embedded in methylmethacrylate. Sections (6 µm thick) were cut using a Reichert-Jung Polycut E microtome, placed onto gelatin-coated slides, and dryed at 50°C for 20 h. The methylmethacrylate was removed using ethylene glycol monoethyl ether acetate, and the sections were stained with Goldner's stain.

2.21 Miscellaneous Techniques

Phenol:Chloroform Extractions

Unless otherwise indicated, these included one extraction with one volume of phenol:chloroform:isoamyl alcohol (25:24:1) and two extractions with chloroform:isoamyl alcohol (24:1). The extractions consisted of thorough vortexing, followed by centrifugation at 16,000xg for 2 min at 4°C. Prior to use, the extraction reagents were pH-equilibrated by combining them with equal volumes of 1 M Tris-HCl (pH 8.0) or 2 M

acetate (pH 4.0), vortexing thoroughly, and separating the phases by centrifugation at 1500xg for 2 min. Unless otherwise specified, extractions were performed using reagents equilibrated at pH 8.0.

Ethanol Precipitations

These were routinely carried out by adding sodium acetate (pH 5.5) to a concentration of 0.3 M, and two volumes of absolute ethanol. Unless otherwise specified, precipitations were incubated overnight at -20°C, and centrifuged at 16,000xg for 30 min at 4°C.

Liquid Scintillation Counting

Aqueous samples were transferred to 5-ml plastic vials, combined with 4.5 ml of EcoLite (+) scintillant cocktail, and vortexed thoroughly. Radioactivity was then measured using an LKB 1219 RackBeta liquid scintillation counter.

Chapter 3

Results

3.1 Optimization of the Chemical Cleavage Technique for Mismatch Detection

The chemical cleavage analysis of mismatches in mRNA:cDNA heteroduplexes is a powerful new technique for the detection of gene mutations. The heteroduplexes are formed by combining total cellular RNA with a cDNA probe whose strands have been labelled at their 3' ends by fill-in with Klenow fragment, heat-denaturing at 90°C, and subsequently incubating at a temperature which is permissive for the formation of RNA:DNA hybrids, but not for DNA:DNA reannealing (337). This is due to the higher melting temperature of RNA:DNA hybrids in comparison with DNA:DNA. Under these conditions, the noncoding cDNA strand hybridizes to a specific mRNA, thereby protecting itself from the modifying effects of reagents such as HA and OT, whereas the coding strand and excess noncoding strand (ie. unhybridized to the mRNA) are not protected from modification (205, 323, 338). However, mismatched C bases (in the case of HA) and mismatched T bases (in the case of OT) are still relatively unprotected and undergo modification, resulting in strand scission upon subsequent piperidine treatment. The labelled cleavage products are then detected and sized on a denaturing polyacrylamide gel, allowing the precise localization of the mismatches, and therefore of the underlying mutations, relative to the known position of the 3' end of the probe.

The successful use of this technique is dependent on the efficient hybridization of the noncoding strand of the probe to its complementary mRNA, so as to maximize the generation of mismatch cleavage products. Incomplete hybridization results in the decreased intensity of mismatch cleavage products, and instead allows the noncoding strand to undergo random cleavage at all C or T bases, producing a heterogeneous mixture of labelled fragments which upon electrophoresis has the appearance of a chemical sequencing ladder. If this background radioactivity is of significant intensity, the detection

of cleavage products resulting from mismatched bases can be completely obscured.

Other experimental factors, pertaining to the preparation and labelling of the cDNA probe, may also adversely affect mismatch detection. These parameters may contribute to the background radioactivity by resulting in heterogeneity in probe size and in the location of the radioactive label.

In order to optimize the detection of mismatches in RNA:DNA heteroduplexes, the effects of varying certain experimental parameters of the chemical cleavage method were examined. The following experiments were performed to assess the importance of such parameters on the detection of a HA-reactive mismatch, observed with a prepro-α2(I) cDNA probe (α2XX), in RNA from the control fibroblast cell line BD2c8.

Effect of the RNA Isolation Method Used

The original method for the chemical cleavage of RNA:DNA heteroduplexes (323) made use of total RNA prepared by a time-consuming method requiring ultracentrifugation through cesium chloride cushions (347). However, a more convenient method is that described by Chomczynski and Sacchi (327), which utilizes a rapid one-step purification consisting of a single phenol-chloroform-isoamyl alcohol extraction. The advantages of this technique are its speed, simplicity and the absence of an ultracentrifugation step. This allows the simultaneous handling of a much larger number of samples, if necessary. However, it is important to verify that RNA prepared by this method is satisfactory for chemical cleavage analysis, in terms of RNA message integrity and purity. Any RNA degradation resulting from residual RNase activity might adversely affect mismatch detection by producing mRNAs with heterogeneous sizes, and DNA contamination might interfere with hybridization of probe to mRNA.

Probe DNA (10 ng) was annealed with total fibroblast RNA (25 μg) isolated by the basic method of Chomczinski and Sacchi, or by modifications of this method in which the

RNA was further purified by subsequent phenol:chloroform extraction at pH 4.0 or 8.0. In all samples, the probe was cleaved to a major product of 500 bases, which increased in intensity with time of HA treatment (Fig. 8). In addition, the background cleavage observed was of a low intensity, indicating satisfactory probe protection with all three RNA samples. It is therefore unnecessary to further purify the RNA, since the addition of the second phenol-chloroform extraction performed at either pH 4.0, to remove protein and DNA, or pH 8.0, to remove only protein, failed to enhance detection of the mismatch (Fig. 8B,C). Furthermore, since the yields of total RNA were decreased by the addition of these extractions, there was no advantage in including them in the isolation procedure. Thus, in addition to being rapid and simple, the basic method of Chomczynski and Sacchi provides RNA message of suitable quality for the detection of mismatches by chemical cleavage analysis.

The absence of a strong band corresponding to the intact probe for the zero-minute time points is unexpected (Fig. 8), as these samples were stopped immediately after addition of the HA reagent. However, since the aliquots of RNA:DNA heteroduplex used for these time points were stored at -80°C during the incubations of the 15-, 30- and 60-minute time points, it is possible that freezing and thawing is responsible for the non-specific degradation of the probe. Results from more recent chemical cleavage experiments with other DNA probes have suggested that better probe recovery in zero-minute time points can be achieved by processing these at the same time as the other time points. The zero-minute time points can then be ethanol-precipitated at -80°C while the other time points are incubating.

The production of a labelled 500-base fragment from the noncoding strand of the probe indicates the presence of a C-base mismatch near position 1800 of the prepro-α2(I) cDNA sequence, approximately 500 bases from the XhoII restriction site which constitutes the 3' end of the noncoding strand (Fig. 9). Since the cell line being analyzed was derived from a

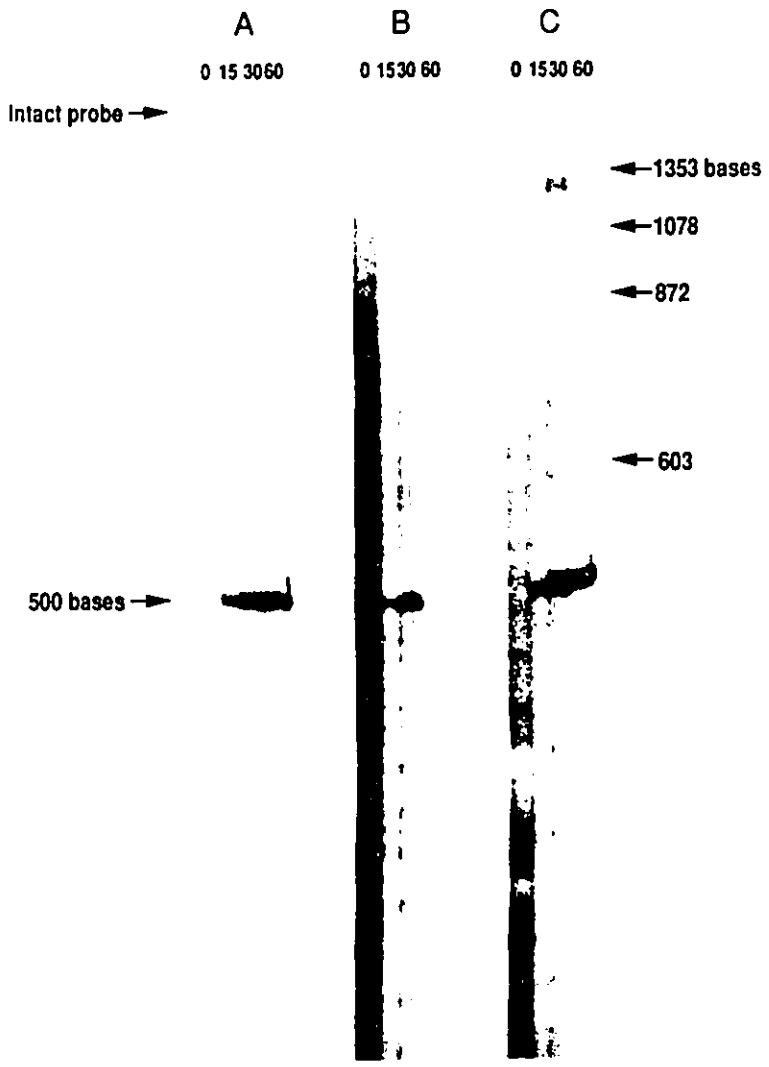


Fig. 8. Effect of the RNA isolation method on mismatch detection. Chemical cleavage analysis was performed using total RNA isolated from ascorbate-stimulated BD268 fibroblasts by the method of Chomczynski and Sacchi (A), or by modifications in which the RNA was further purified by phenol:chloroform extraction at pH 4.0 (B), or pH 8.0 (C). Total RNA (25 μg) was annealed with the α2XX probe (10 ng) labelled in the presence of unlabelled dNTPs. Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 3 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

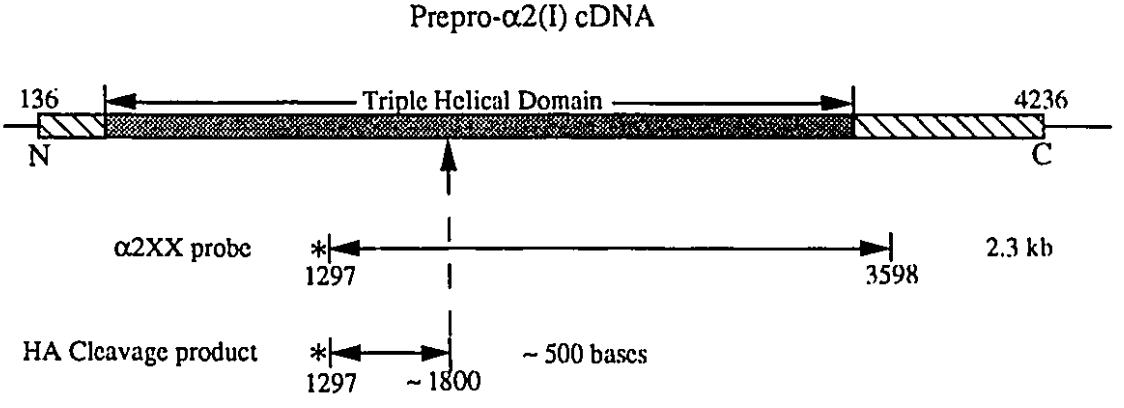


Fig. 9. Schematic diagram of the location of the HA-reactive mismatch detected in cell line BD268. The α2XX probe and the 500-base HA cleavage product are represented by double-headed arrows below the bar diagram of the full-length prepro-α2(I) cDNA. cDNA coding regions are indicated by bars, and untranslated regions are represented by plain lines. The regions encoding the main triple helical domain (grey shading) and the terminal domains (diagonal shading) are also indicated. cDNA sequence positions are relative to the transcription start site (80). The asterisks denote the location of the radioactive label, and the vertical arrow indicates the location of the detected mismatch within the cDNA sequence.

normal individual, the detected mismatch is likely due to a neutral polymorphism. The underlying sequence change resulting in this C-base mismatch is identified in section 3.3.

Effect of the RNA:Probe Molar Ratio Used in Annealing

In theory, increasing the molar ratio of the specific mRNA to cDNA probe during the annealing process, to ensure that the probe is not in great excess, should result in more efficient probe protection, and therefore in a more intense cleavage product relative to the background. One manner in which this may be achieved is to isolate RNA from a tissue or cell line expected to have high levels of the transcript of interest, or to induce its expression in culture. In the case of the mRNAs encoding type I collagen, expression by fibroblast cells can be induced by treatment with L-ascorbic acid (325). However, it is essential to verify that such induction is not associated with increased intracellular turnover, which might result in mRNA size heterogeneity. This would interfere with mismatch detection by giving rise to cleavage products with heterogeneous sizes, and thereby increasing background radioactivity.

In order to determine whether the background radioactivity observed in Fig. 8 could be improved by omitting prepro-α2(I) mRNA induction, probe DNA (10 ng) was annealed with total RNA (25 μg) isolated from ascorbate-stimulated or unstimulated BD268 fibroblasts. The induction of the prepro-α2(I) mRNA resulted in a significantly more intense 500-base cleavage product than did the RNA from unstimulated cells (Fig. 10), indicating that the DNA probe was protected more extensively. Furthermore, induction was not accompanied by more intense background probe cleavage, suggesting that ascorbate treatment does not result in mRNA size heterogeneity which adversely affects mismatch detection. In addition to providing superior probe protection in chemical cleavage analysis, ascorbate treatment resulted in approximately two-fold higher yields of total RNA than when treatment was omitted (data not shown), indicating an increase in

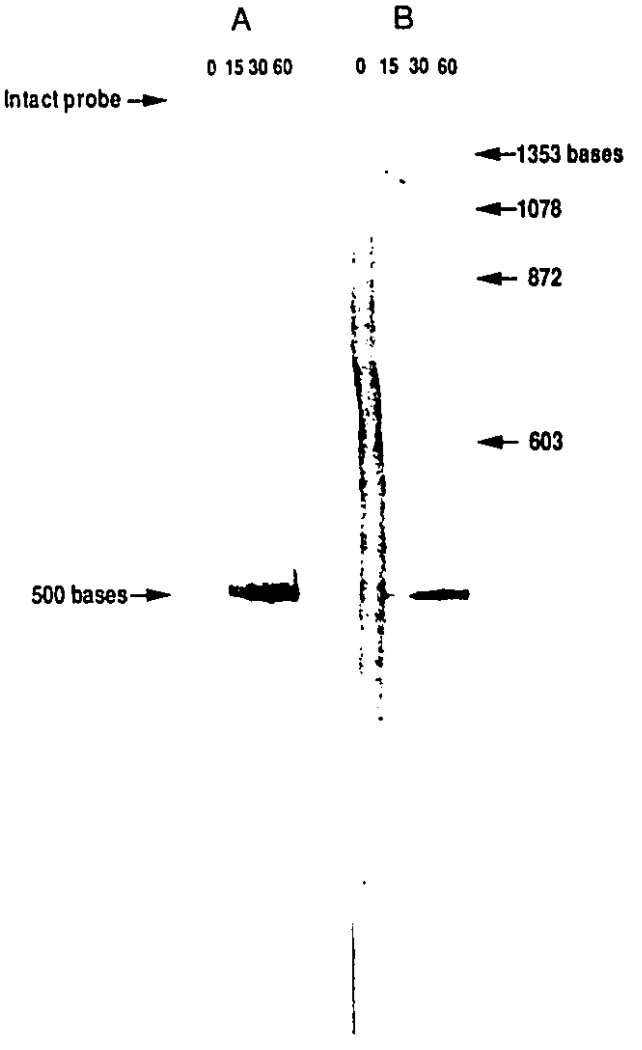


Fig. 10. Effect of ascorbate stimulation of fibroblasts prior to RNA isolation on mismatch detection. Total RNA (25 μg) isolated from ascorbate-stimulated (A) or unstimulated (B) BD268 fibroblasts was annealed with the α2XX probe (10 ng) labelled in the presence of unlabelled dNTPs. Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 3 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

general RNA synthesis. This was also advantageous since it allowed a greater number of experiments to be performed using RNA from the same number of fibroblasts.

The RNA:probe molar ratio used in annealing was also varied by maintaining the quantity of one reagent (total RNA or probe) constant while varying the other. The amount of RNA used was varied by annealing 10 or 25 μg with 19 ng of α2XX probe DNA. When the probe was annealed with 25 μg of RNA, the 500-base product previously observed in Fig. 8, accumulated with HA treatment (Fig. 11). An 870-base fragment appearing at 15 min of treatment at an intensity comparable to that of the 500-base fragment was also detected, but remained fairly constant with prolonged exposure to HA. This cleavage product may not be increasing in intensity with prolonged HA treatment because its 5' end is further from the radioactive label than that of the 500-base fragment, and may therefore be chased into the smaller fragment. Alternatively, the 870-base cleavage product may result from cleavage of the heteroduplex at a region with increased sensitivity to HA modification due to local sequence-specific thermal destabilization. In contrast, when only 10 μg of total RNA was used, neither intact probe nor cleavage products were visible with overnight exposure of samples electrophoresed for 4 h. Samples electrophoresed for a period of 1 h (data not shown), in which radioactivity is less diffusely distributed, displayed a faint smear, suggesting that 10 μg of RNA is not able to adequately protect the probe from random chemical cleavage. Furthermore, the recovery of radioactivity with 10 μg of RNA was much lower than with the use of 25 μg. This was in part due to the loss of ^{32}P label by random cleavage, but was also due to the decreased efficiency of ethanol precipitation, since radioactivity was lost at every precipitation step. The use of larger amounts of RNA counteracts both of these problems by more adequately protecting the probe from chemical cleavage, and by acting as a more efficient carrier during ethanol precipitations.

The effect of varying the quantity of probe used in annealing was also examined, as the

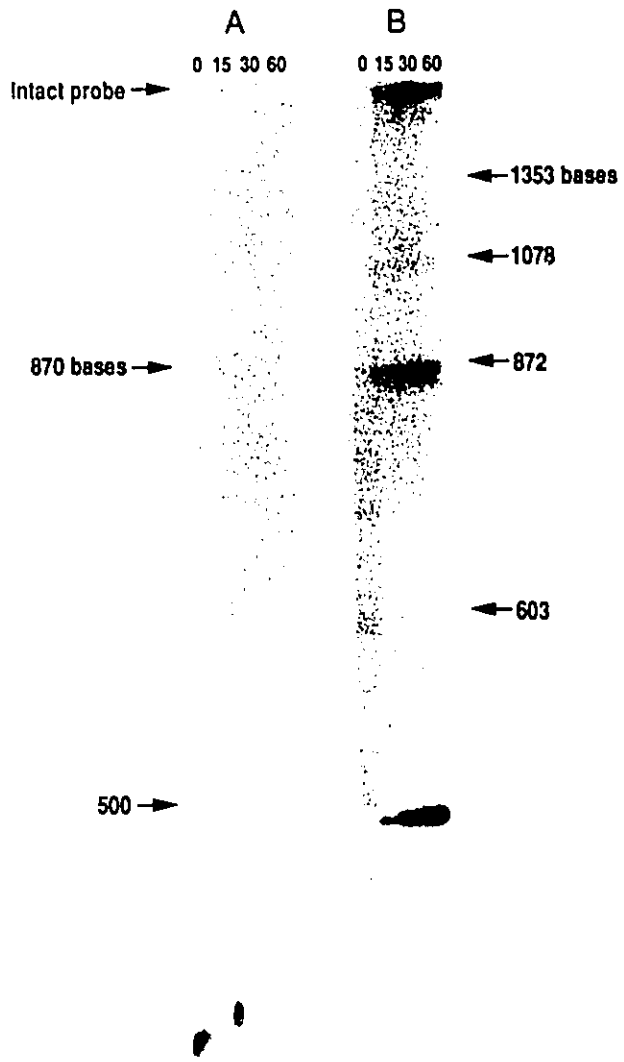


Fig. 11. Effect of the quantity of RNA used in annealing on mismatch detection. α2XX probe (19 ng) labelled in the presence of unlabelled dNTPs was annealed with 10 μg (A), or 25 μg (B) of total RNA from ascorbate-stimulated BD268 fibroblasts. Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

use of excess probe may result in high background levels by increasing the amount of unhybridized probe present. Total RNA (25 μg) was annealed with 5, 10 or 20 ng of α2XX probe DNA. All three samples exhibited the 500-base cleavage fragment increasing in intensity with prolonged exposure to the modifying reagent (Fig. 12). However, the use of decreasing amounts of probe resulted in a much less intense signal, without an apparent effect on the strength of background chemical cleavage. The 870-base fragment detected previously, appeared very faintly when using 20 ng of probe, but was not detected when using lesser amounts, indicating that low probe levels may require long exposure times for the visualization of minor cleavage products. In the case of fibroblast RNA annealed to this particular probe, there was no significant advantage in using less than 20 ng of probe with 25 μg of total RNA. In fact, it is possible that exposure times could be further shortened by the use of even larger amounts of probe, without adversely affecting the intensity background cleavage, as long as probe excess is avoided.

For general use, the optimal quantities of total RNA and probe to be used in annealing must be determined empirically, and will depend on the abundance of the specific message of interest in the preparation of total RNA. When dealing with low levels of expression, it is equally important to consider conditions in which a tissue or cell line may be stimulated to produce increased amounts of the specific mRNA, otherwise excessively large amounts of RNA may be required to protect the probe. However, one must verify that such induction is not associated with increased message turnover that might interfere with mismatch detection.

Effect of the Probe Labelling Method

The α2XX cDNA probes used in the above experiments were labelled by fill-in with the Klenow fragment of E. coli DNA polymerase I and [α-^{32}P]-dGTP, in the presence of a mixture of unlabelled dATP, dCTP and dTTP. This should label the 3' end of each strand

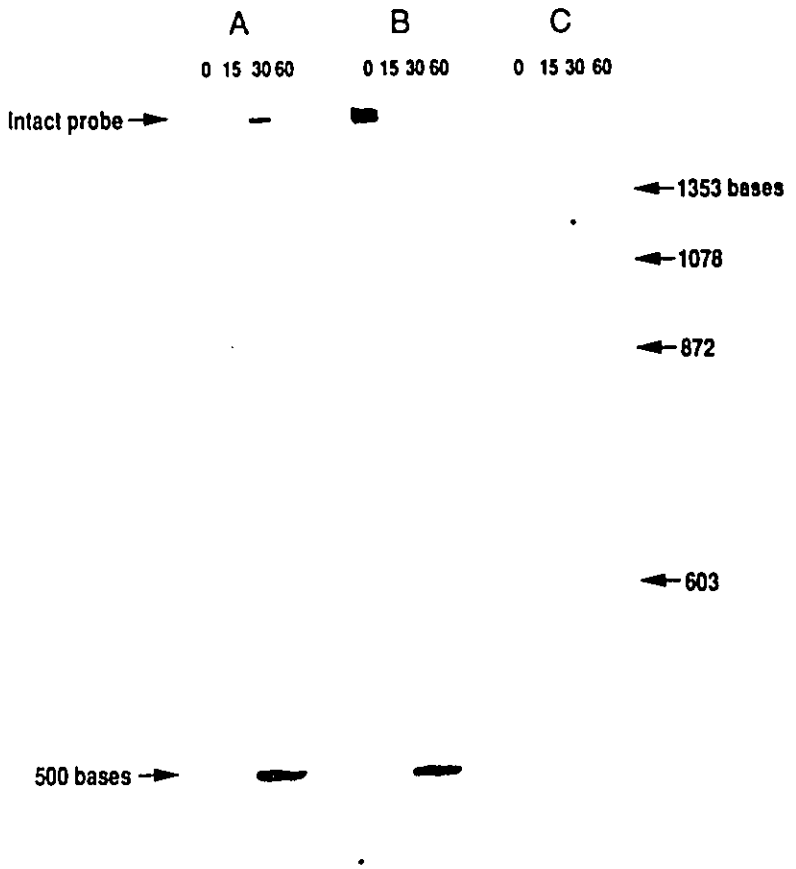


Fig. 12. Effect of the quantity of probe used in annealing on mismatch detection. Total RNA (25 μg) from ascorbate-stimulated BD268 fibroblasts was annealed with 20 ng (A), 10 ng (B), or 5 ng (C) of the α2XX probe labelled in the presence of unlabelled dNTPs. Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

of DNA, since both ends of the probe possess 5' overhangs (Fig. 13A). Since the coding strand of the probe is left unprotected during the chemical modification step, it is extensively degraded upon piperidine treatment, and may significantly contribute to the background radioactivity observed in chemical cleavage analysis. In addition, since nicks can be introduced in the probe during its preparation, Klenow fragment may be permitted to carry out significant labelling of the probe at internal positions, by the process of strand displacement (Fig. 13B). This may in turn result in a significant increase in background radioactivity on chemical cleavage analysis gels due to a heterogeneity in probe size, and in the location of the radioactive label. Omitting the unlabelled dNTPs during the probe labelling procedure should reduce labelling of the 3' end of the coding strand, since the first base expected to be incorporated at the XbaI site is a C, while still permitting the labelling of the 3' end of the noncoding strand. In addition, this omission should reduce the internal labelling of the probe by allowing the incorporation of ^{32}P-dGTP only at nicks occurring opposite C bases in the coding strand. On the other hand, since Klenow fragment possesses a 3' $\rightarrow$ 5' exonucleolytic activity, it may be capable of degrading any available 3' ends (i.e. at the terminal restriction sites and at internal nicks) to positions at which its polymerase activity can then incorporate ^{32}P-labelled dGTP. However, since it is unclear how active Klenow fragment's 3' $\rightarrow$ 5' exonuclease function is under the labelling conditions used, its ability to counteract the potential benefits of omitting the unlabelled dNTPs was unknown. These possibilities have been examined by comparing probes labelled with Klenow fragment in the presence or absence of unlabelled dATP, dCTP and dTTP, in chemical cleavage mismatch detection.

Annealing was performed using 25 μg of total RNA and 20 ng of probe DNA. In practice, the intensity of the 500-base cleavage fragment was comparable for the two probes, but the background chemical cleavage was considerably more intense when using the probe labelled in the presence of the unlabelled dNTPs (Fig. 14). Although the probe

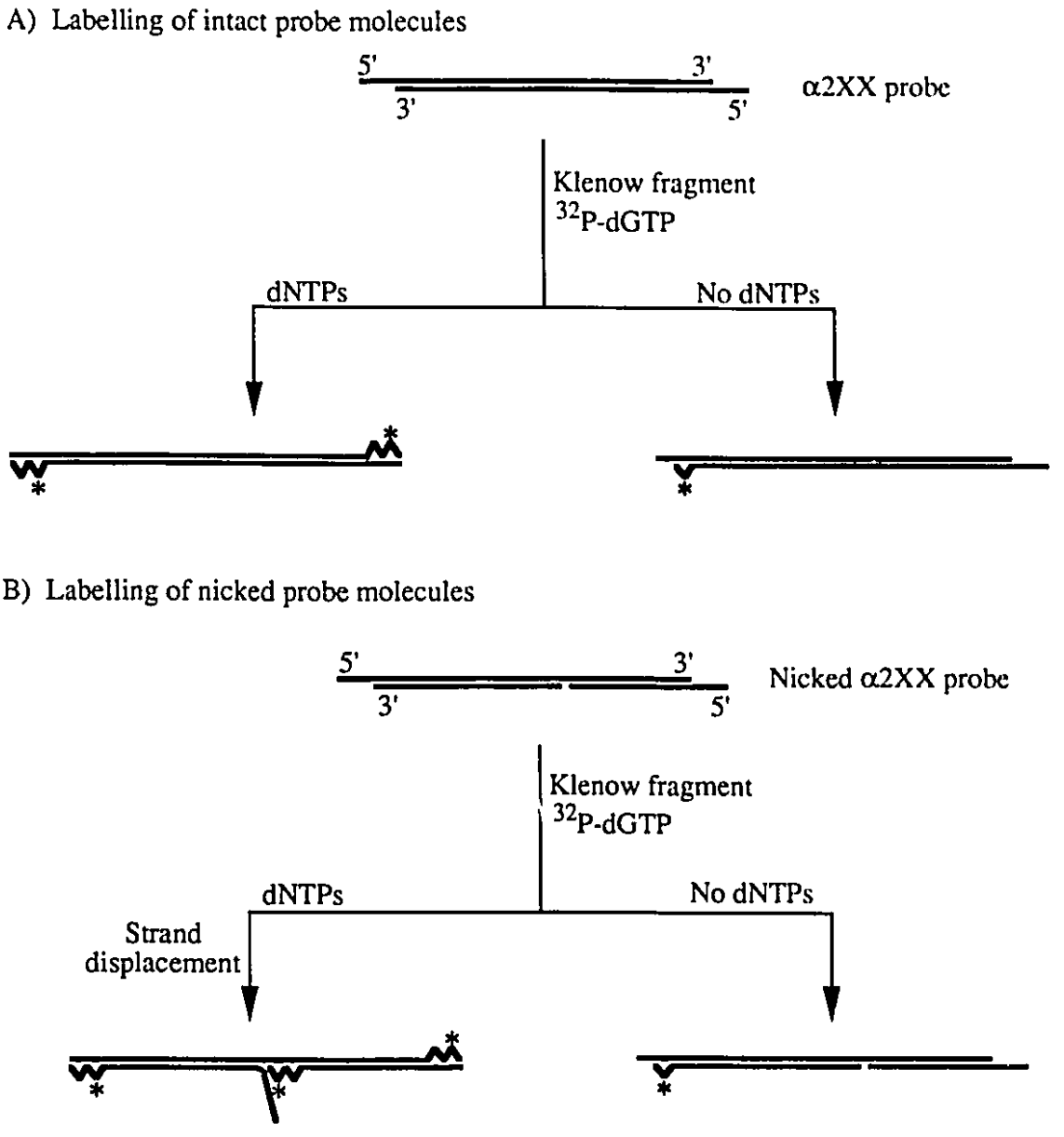


Fig. 13. Schematic diagram of the possible effects of omitting dNTPs on probe labelling. The diagram illustrates the positions at which intact (A) or nicked (B) α2XX probe molecules may be labelled in the presence or absence of unlabelled dNTPs. The double-stranded DNA probe is illustrated as a double line, with the top and bottom lines representing the coding and noncoding strands, respectively. DNA synthesized by Klenow fragment is indicated by zig-zagging lines, and incorporated ^{32}P-dGTP is denoted by asterisks. The omission of dNTPs may result in decreased labelling of the 3' end of the coding strand, and of internal nicks in either strand. Some nicks may be labelled if they occur opposite a C base in the coding strand, or following 3' → 5' exonuclease digestion by Klenow fragment.

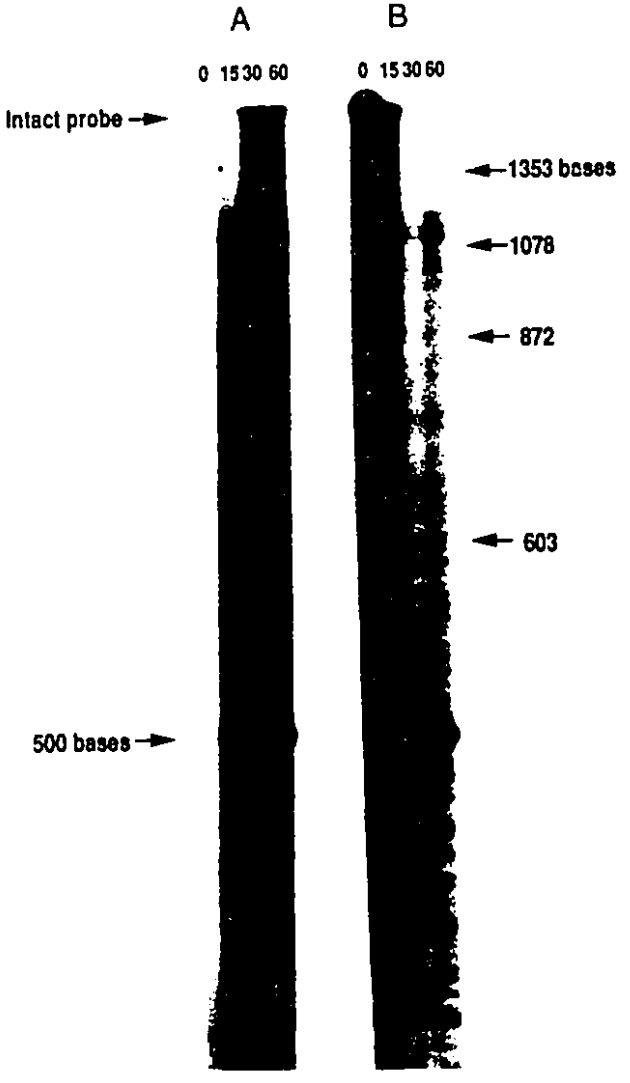


Fig. 14. Effect of the probe labelling method on mismatch detection. Total RNA (25 μg) isolated from ascorbate-stimulated BD268 fibroblasts was annealed with the α2XX probe (20 ng) labelled in the presence (A) or absence (B) of unlabelled dATP, dCTP and dTTP. Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

labelled in the presence of unlabelled dNTPs gave probe specific activities more than three-fold higher, it did not give rise to an enhancement in the intensity of the mismatch cleavage signal, despite the use of a larger amount of radioactivity. This is compatible with the increased probe specific activity arising from labelling at internal nicks and/or at the 3' end of the coding strand, and would account for the increase in background radioactivity. The results demonstrate that the benefits of omitting the three unlabelled dNTPs during probe labelling significantly outweigh the ability of the 3' → 5' exonuclease activity of Klenow fragment to contribute to heterogeneity in the location of the radioactive label. Whether the improvements in background cleavage are mediated by selective labelling of the noncoding strand, by the minimization of internal labelling, or by a combination of the two, is unclear.

In summary, these experiments have allowed the optimization of the conditions for the chemical cleavage mismatch analysis of the prepro-α1(I) and prepro-α2(I) mRNAs synthesized by skin fibroblasts. It was determined that the ascorbate stimulation of fibroblasts and the method of choice for the isolation of total RNA resulted in satisfactory mismatch detection. Furthermore, the use of 25 μg of total RNA and 20 ng of probe DNA in the annealing process, as well as the selective labelling of the 3' end of the noncoding strand of the probe (by the omission of unlabelled dNTPs) were found to significantly improve mismatch detection. Therefore, unless otherwise specified, these conditions were routinely utilized in the subsequent chemical cleavage mismatch experiments described in sections 3.3 and 3.4.

3.2 Optimization of the Blunt-End Cloning of PCR Products

Chemical cleavage analysis allows the rapid detection of sequence differences in large regions of RNA:DNA heteroduplexes, obviating the need to clone and sequence extensive regions of cDNA. However, since it does not identify the exact nature and location of observed mismatches, it is still necessary to clone and sequence, but the cDNA region of interest has been greatly focused. A convenient way of obtaining segments of cDNA for cloning and subsequent sequencing is the PCR. However, the cloning of PCR-amplified DNA is often problematic (348). The approach of incorporating restriction sites into the PCR primers is complicated by the fact that a number of commonly used endonucleases do not efficiently cleave sites located near DNA termini (349). Furthermore, blunt-end cloning, the most generally applicable approach, has presented problems in the form of very poor cloning efficiencies, which can only partially be explained by the low rate of blunt-end ligation. A more plausible explanation is the observation that Taq DNA polymerase does not leave DNA fragments with blunt ends, due to a non-template dependent terminal transferase-like activity which creates single A-residue 3' overhangs (350). In order to improve the efficiency of PCR product blunt-end cloning into the M13mp18 sequencing vector, several parameters relating to the preparation of insert DNA were examined.

DNA product from six identical 100-µl PCR mixtures, using the COLL-1/COLL-2 primer pair and the full-length prepro-α2(I) cDNA construct as template, was pooled and divided into seven samples of equal volume. Insert DNA was then prepared by seven different methods, and cloned into the M13mp18 vector (342, 343). These methods are summarized in Table 9, and differed in: a) whether blunt-ending was performed; b) the choice of blunt-ending enzyme between Klenow fragment and T4 DNA polymerase; c) the order in which blunt-ending and gel purification were performed; and d) whether the

Table 9. Methods utilized to prepare PCR insert DNA and efficiencies of cloning into the M13mp18 vector.

Method	Extraction of Taq DNA polymerase	Blunt-Ending Enzyme Used	Blunt-ending Before or After Gel Purification	Number of Clear Plaques [1]
1	–[2]	T4	Before	96
2	–	K	Before	305
3	–	–	–	34
4	–	T4	After	73
5	–	K	After	250
6	+	T4	Before	95
7	+	K	Before	550

[1] Cloning efficiencies are reported as mean values from duplicate transformations.

[2] Abbreviations used: –, not performed; +, performed; T4, T4 DNA polymerase; K, Klenow fragment.

extraction of Taq DNA polymerase, immediately after the PCR, was performed.

Both T4 DNA poymerase and the Klenow fragment of E. coli DNA polymerase I possess the 5' → 3' DNA polymerase and 3' → 5' exonuclease activities necessary to fill in recessed 3' ends and to remove 3' overhangs, respectively. Therefore, the treatment of PCR products with either enzyme should enhance the efficiency of ligation to blunt-ended vector DNA. As shown in Table 9, blunt-ending with either enzyme gave higher cloning efficiencies than when blunt-ending was omitted (method 3), but Klenow fragment appeared to be more effective, resulting in a roughly 9-fold higher number of clear plaques (method 2), while T4 DNA polymerase caused an approximate 3-fold increase (method 1). These results support the hypothesis that the ends of DNA fragments synthesized by Taq DNA polymerase are not compatible with efficient blunt-end ligation unless they are first repaired with an enzyme such as Klenow fragment.

Before PCR products can be ligated into the desired vector, they must first be separated from unincorporated PCR oligonucleotide primers. These would interfere with the efficient phosphorylation of the PCR product, since even small quantities of primer DNA (in terms of weight amounts) would constitute a high molar concentration of 5' ends. Therefore, PCR products to be cloned were routinely purified from agarose gels by electrophoresis onto DEAE-cellulose membranes. This results in higher cloning efficiencies than if the gel purification step is omitted (data not shown). The order in which blunt-ending and gel-purification are performed may be important, however, as Klenow and T4 DNA polymerase may be inhibited by agarose and other impurities carried through from electrophoresis. Blunt-ending with T4 DNA polymerase (method 4) or Klenow fragment (method 5) following gel purification gave cloning efficiencies which were 24% and 18% lower, respectively, than when blunt-ending preceded gel purification (methods 1 and 2, respectively). This suggests that blunt-ending of the PCR product is achieved less efficiently when performed after gel purification, possibly due to inhibition of the enzymes

by gel impurities.

Another potential problem which could interfere with efficient blunt-ending, is the presence of residual Taq DNA polymerase following the PCR. To investigate this possibility, the effect of removing the enzyme by phenol-chloroform extraction immediately after amplification was studied. PCR products blunt-ended with T4 DNA polymerase showed virtually identical cloning efficiency with (method 6) or without (method 1) prior extraction of Taq DNA polymerase. However, the extraction of Taq DNA polymerase prior to Klenow fragment blunt-ending (method 7) gave the highest cloning efficiency, which was 80% higher than when the extraction was omitted (method 2). These results suggest that the residual Taq DNA polymerase activity remaining following the PCR may counteract the blunt-ending repairs made by Klenow fragment, mediated by its terminal transferase-like activity.

One possible explanation for the differences in cloning efficiencies observed could be variability in the yields of insert DNA obtained with the different methods. This would result in differences in the vector:insert molar ratio during ligation, and would thereby give rise to differences in the numbers of recombinants observed. However, this possibility was excluded, since the yields of insert DNA obtained with the different methods were compared by agarose gel electrophoresis prior to ligation, and were found to be similar.

Thus, these results indicate that in order to improve the efficiency of blunt-end cloning of PCR products, phenol:chloroform extraction of Taq DNA polymerase should be performed immediately after amplification, and followed by blunt-ending with Klenow fragment. The PCR product of interest should then be gel-purified, prior to phosphorylation of its 5' ends and subsequent ligation to the desired blunt-ended vector. This method of insert DNA preparation (method 7) was found to yield the highest cloning efficiency and was hereafter routinely used for the preparation of PCR products for cloning (in sections 3.3 and 3.4 below).

3.3 Identification of Sequence Polymorphisms in the COL1A1 and COL1A2 Coding Regions

The recent development of powerful molecular biology techniques, such as chemical cleavage mismatch analysis and the PCR, has permitted the rapid detection and characterization of disease-producing mutations in a variety of inherited disorders. However, techniques such as chemical cleavage analysis, are unable to discriminate between neutral polymorphisms and pathological sequence changes. Therefore, in order to demonstrate that an observed sequence change is responsible for producing a disease phenotype (such as OI), it would be necessary to show that such a change is not found in the normal population. This is of course impractical, but a knowledge of the frequency and distribution of polymorphisms in the gene of interest is nevertheless of great value. In addition, the characterization of neutral polymorphisms provides valuable information about the evolutionary conservation of the primary structure of a protein by identifying those amino acid residues which are tolerant of sequence variation. This is in contrast to disease-producing mutations which identify residues that are required for normal function and are intolerant of sequence variation. From a technical point of view, this type of characterization will also help to avoid the time-consuming cloning and sequencing of polymorphisms occurring in the patients of interest. For instance, mismatches detected by chemical cleavage analysis would be pursued only if they were not previously observed in the normal population. This would allow a more rapid identification of disease-related mutations.

In order to assist in the identification of the mutations in OI patients, and to provide information on the conservation of the prepro-α1(I) and prepro-α2(I) polypeptide sequences, the frequency and distribution of polymorphisms in the corresponding cDNA sequences were examined by chemical cleavage analysis using HA and OT. Since these

reagents are only able to detect mismatches resulting in reactive C or T bases, this study was not intended as an exhaustive characterization of all the polymorphisms in these two cDNAs, but rather to get an indication of how frequently they occur, and which polypeptide regions are affected. Assuming random base usage, these reagents should be able to detect at least half of the sequence polymorphisms present. This includes sequence changes resulting in mismatched C or T bases. However, the reagents can in some instances also react with C or T bases which neigbour mismatches. In addition, an understanding of the positions of common HA-reactive polymorphisms will be particularly useful in the identification of OI mutations, since most of these affect glycine codons (GGN) and consequently result in C mismatches. This will make it possible to distinguish between mismatches due to disease-related mutations and those due to polymorphisms.

Detection of Polymorphisms by Chemical Cleavage Mismatch Analysis

The cell lines in the polymorphism study included five derived from control individuals (C10, C19, C40, BD268 and YL278) and two derived from the asymptomatic parents (ML226 and JL227) of an OI patient (SL211). Total RNA (25 μg) isolated from these cell lines was annealed with the cDNA probes summarized in Table 6 (20 ng), and the resulting heteroduplexes were treated with either HA or OT. Chemical cleavage analysis of RNA from these cell lines with HA and OT demonstrated the presence of mismatches at only two cDNA positions. Both mismatches were observed in the prepro-α2(I) cDNA using the α2XX probe.

One mismatch was observed in cell lines C19, C40, ML226, BD268 and YL278 using HA as the modifying reagent (Table 10), and resulted in cleavage products of approximately 500 bases (Fig. 15; data shown for C10, C19, C40 and BD268). This HA-reactive mismatch in BD268 fibroblasts is that which was utilized in the optimization of the chemical cleavage technique (described in section 3.1). Cleavage products of this size indicate the presence of sequence changes resulting in mismatched C bases in the

Table 10. HA- and OT-reactive mismatches detected in normal fibroblast cell lines.

	Detected Mismatches	
Cell Line	α2XX HA	α2XX OT
C10	- [1]	-
C19	+	-
C40	+	-
ML226	+	+
JL227	~	-
BD268	+	+
YL278	+	-

[1] Abbreviations used: -, not observed; +, observed.

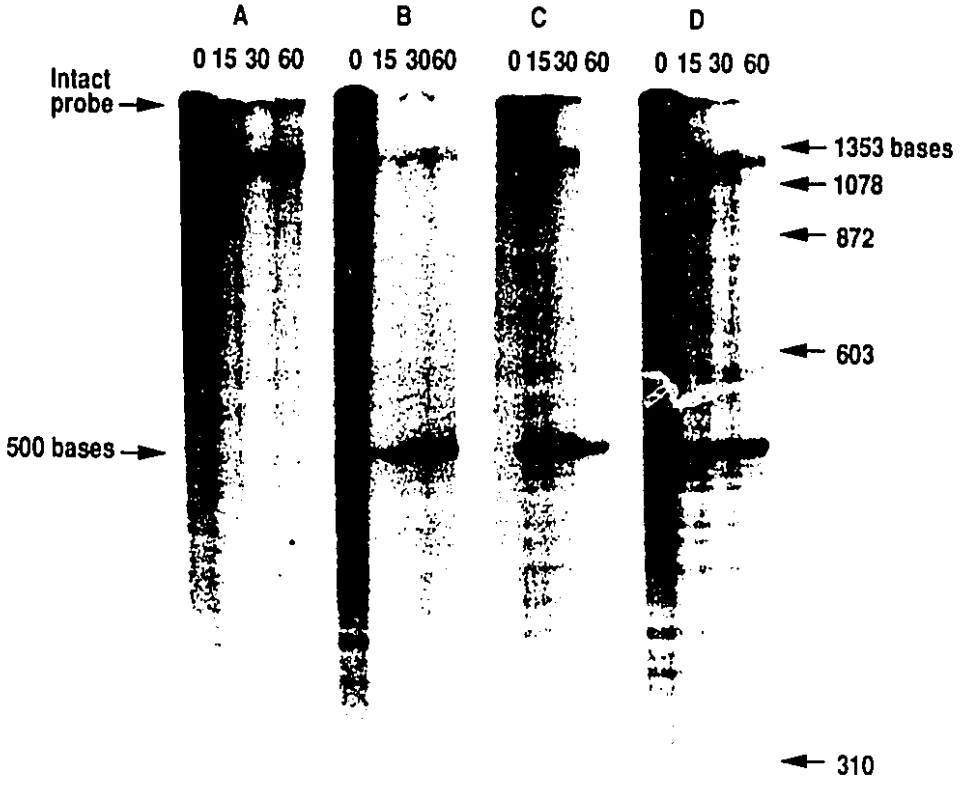


Fig. 15. Chemical cleavage detection of a HA-reactive mismatch in normal fibroblast cell lines. Total RNA (25 μg) from cell lines C10 (A), C19 (B), C40 (C) and BD268 (D) was annealed with the α2XX probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 2 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

mRNA:cDNA probe heteroduplexes, and located near position 1800 of the prepro-α2(I) cDNA sequence. This position corresponds approximately to amino acid residue 555 of the prepro-α2(I) chain, or residue 465 of the α2(I) triple helical domain.

The second mismatch was observed in RNA from cell lines ML226 and BD268 using OT as the modifying reagent (Table 10), and resulted in cleavage products of approximately 300 bases (Fig. 16; data shown for C10, C19, C40 and BD268). Cleavage products of this size indicate sequence changes resulting in mismatched T bases in the mRNA:cDNA probe heteroduplexes, and located near position 1600 of the prepro-α2(I) cDNA sequence . This corresponds approximately to amino acid residue 490 of the prepro-α2(I) chain, or residue 400 of the α2(I) triple helical domain.

The approximate locations of these two mismatches within the prepro-α2(I) cDNA sequence are schematically represented in Fig. 17.

Identification of Polymorphisms by DNA Sequence Analysis

The exact nature of the sequence changes was determined by the sequence analysis of a cDNA region bearing the detected mismatches. Since ML226 and BD268 each exhibited the two mismatches detected with the α2XX probe, these cell lines were selected for the precise identification of the corresponding sequence changes. Furthermore, as the two α2XX mismatches were separated by only approximately 200 bp in the prepro-α2(I) cDNA sequence, these could be identified by the DNA sequence analysis of a single PCR product from each cell line. Total RNA isolated from cell lines ML226 and BD268 was utilized to synthesize oligo-dT-primed first strand cDNA, which then served as a template in the PCR amplification using the primers COLL-1 and COLL-2. The PCR products were cloned into the M13mp18 vector and analyzed by DNA sequencing. Analysis of cDNAs cloned from each of the cell lines identified two sequences differing only at positions 1780 and 1581 of the prepro-α2(I) cDNA.

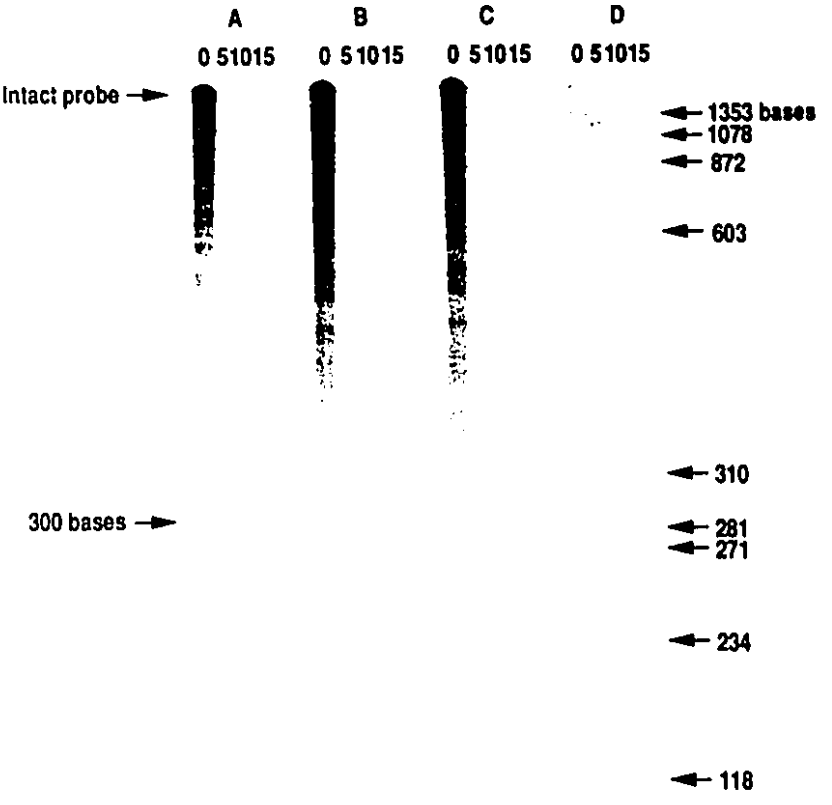


Fig. 16. Chemical cleavage detection of an OT-reactive mismatch in normal fibroblast cell lines. Total RNA (25 μg) from cell lines C10 (A), C19 (B), C40 (C) and BD268 (D) was annealed with the α2XX probe (20 ng). Heteroduplexes were treated with OT for 0, 5, 10 and 15 min, cleaved with piperidine, and electrophoresed for 2 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

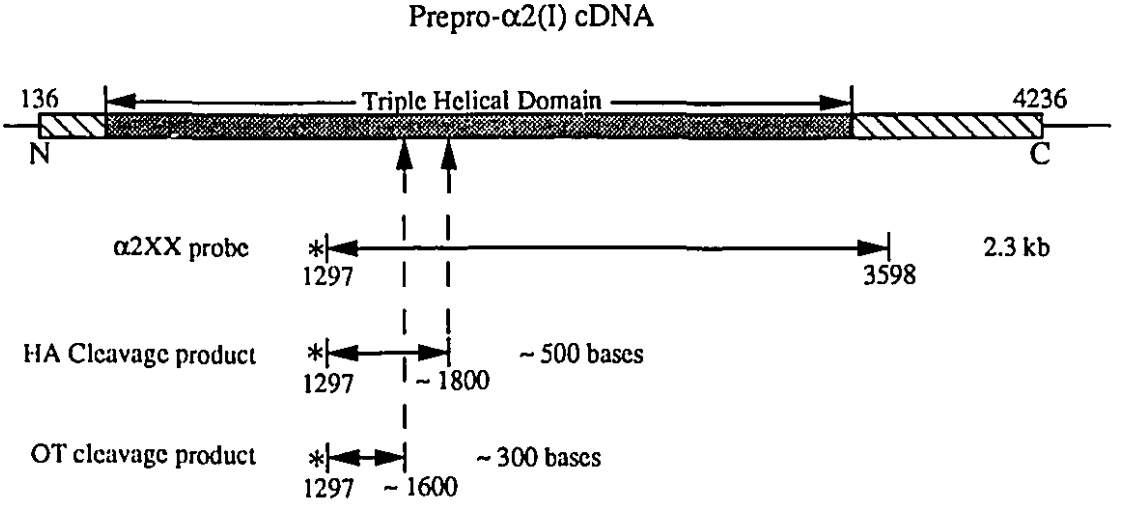


Fig. 17. Schematic diagram of the locations of the HA- and OT-reactive mismatches detected in normal fibroblast cell lines. The α2XX probe and the HA and OT cleavage products are represented by double-headed arrows below the bar diagram of the full-length prepro-α2(I) cDNA. cDNA coding regions are indicated by bars, and untranslated regions are represented by plain lines. The regions encoding the main triple helical domain (grey shading) and the terminal domains (diagonal shading) are also indicated. cDNA sequence positions are relative to the transcription start site (80). The asterisks denote the location of the radioactive label, and the vertical arrows indicate the locations of the detected mismatches within the cDNA sequence.

At position 1780 (within exon 28), one sequence had a G as the first base of a GCT triplet coding for alanine residue 549 of the prepro-α2(I) chain, or residue 459 of the α2(I) triple helical domain (Fig. 18; data shown for BD268). The other identified sequence had a C at this position, thereby converting the GCT alanine codon to a CCT triplet coding for a proline residue. The corresponding region of the full-length cDNA construct, from which the α2XX chemical cleavage probe was prepared, was also cloned and sequenced. The construct was found to have the former sequence, with a G at position 1780 (data not shown). The G → C single base substitution at this position results in a HA-reactive C:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and is located within 20 bp of the site predicted by chemical cleavage analysis. Although this polymorphism affects an EcoRII restriction site, it may not be a useful RFLP due to the frequency of EcoRII cleavage of the repetitive cDNA sequence encoding the triple helical domain. However, the use of alternative methods for detecting the sequence change may nevertheless render this polymorphism a useful marker for the COL1A2 gene.

At position 1581 (within exon 25) of the prepro-α2(I) cDNA, one sequence had an A as the wobble base of a CCA triplet coding for proline residue 482 of the prepro-α2(I) chain, or residue 392 of the α2(I) triple helical domain (Fig. 19; data shown for BD268). However, the other identified sequence had a C at this position, thereby converting the CCA codon to a CCC triplet, though still encoding a proline residue. The corresponding region of the full-length cDNA construct was found to have the former sequence (data not shown). The A → C single base substitution at this position results in an OT-reactive C:T base pair mismatch in the mRNA:cDNA probe heteroduplex, and is located within 20 bp of the location predicted by chemical cleavage analysis. This particular sequence change does not result in an alteration in the primary structure of the prepro-α2(I) polypeptide chain, and is therefore a neutral polymorphism. The sequence variation results in a PvuII RFLP which has recently been described (351), and which may be another useful marker for the

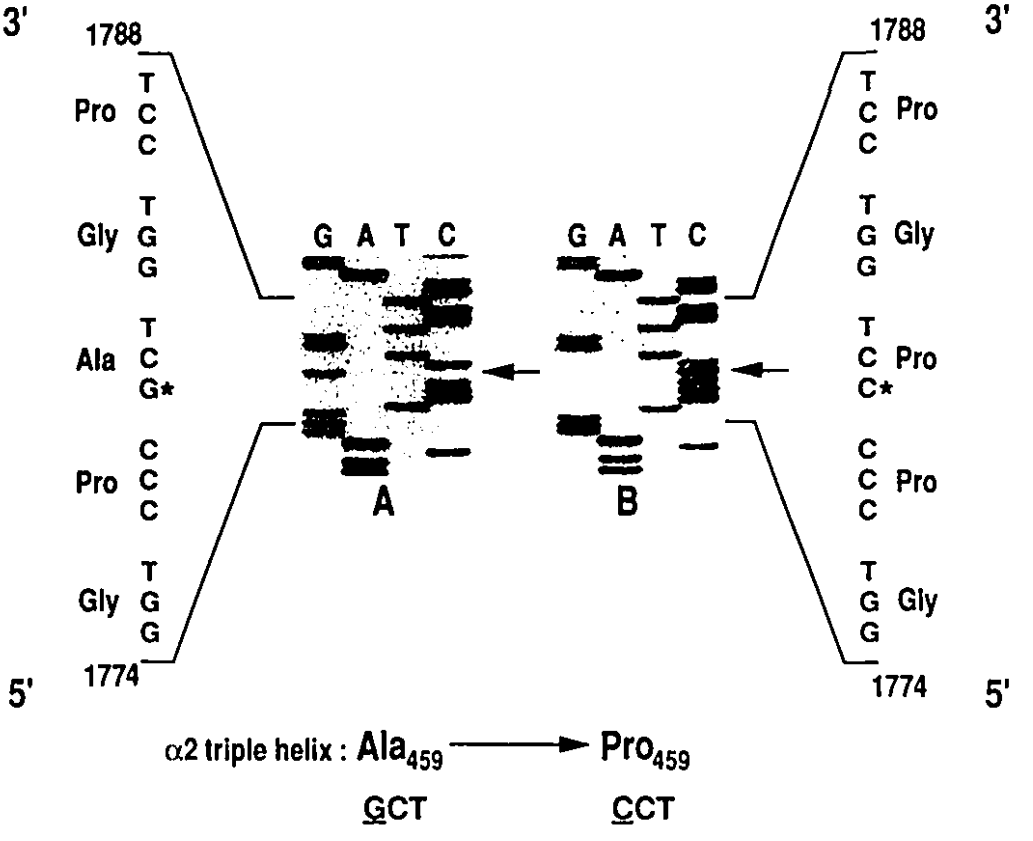


Fig. 18. Sequence analysis of the HA-reactive mismatch detected in cell line BD268. BD268 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-1 and COLL-2 to amplify a region of the prepro-α2(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the two sequenced alleles (A and B) differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (80). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α2(I) triple helical domain. The sequence of allele A corresponds to that of the α2XX cDNA probe.

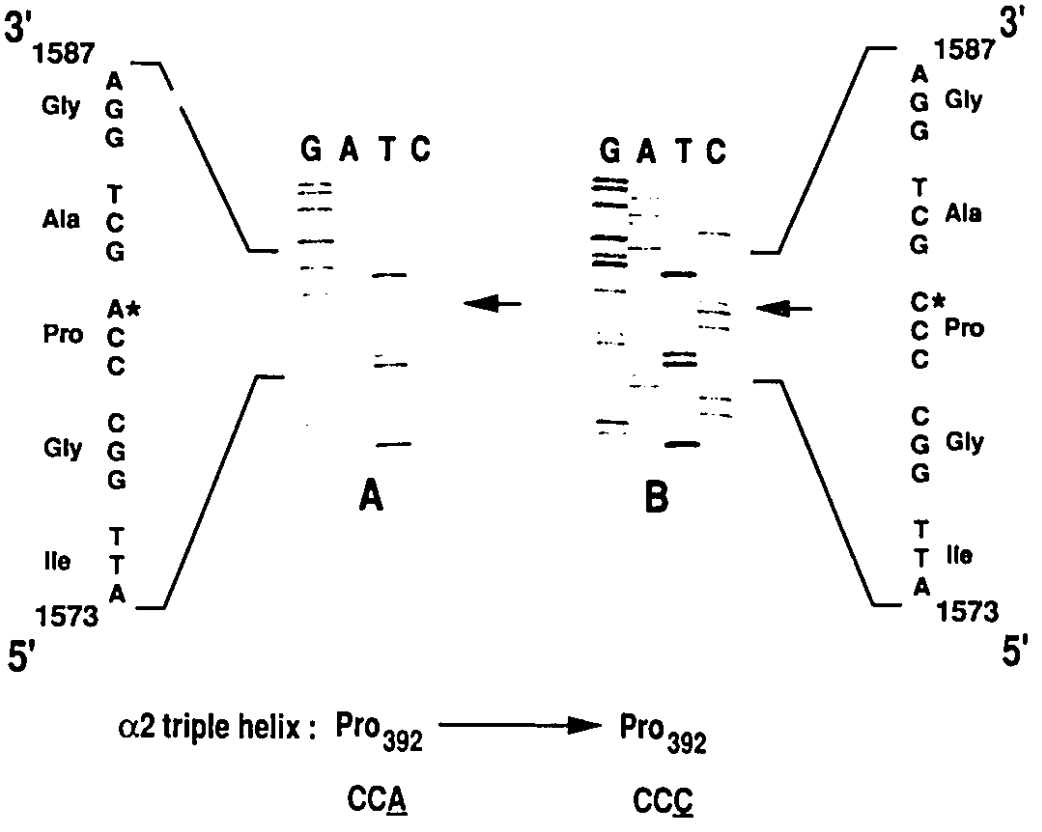


Fig. 19. Sequence analysis of the OT-reactive mismatch detected in cell line BD268. BD268 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-1 and COLL-2 to amplify a region of the prepro-α2(I) cDNA containing the OT-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the two sequenced alleles (A and B) differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (80). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α2(I) triple helical domain. The sequence of allele A corresponds to that of the α2XX cDNA probe.

COL1A2 gene.

The above sequence data, summarized in Table 11, demonstrate that the 500-base cleavage products observed in cell lines ML226 and BD268 resulted from identical sequence polymorphisms in the prepro-α2(I) cDNA. It is therefore probable that the HA cleavage products of approximately the same size, observed in cell lines C19, C40 and YL278, also resulted from the same sequence variation at position 1780. In addition, the representation of the two sequences, among the clones analyzed, indicated that ML226 and BD268 were both heterozygous at each base position.

In summary, although the group studied was quite small, with a sample size of 7 individuals, and sequence changes involving only two of the four bases were examined, it appears from the limited number of mismatches detected, that polymorphisms within the coding regions of the COL1A1 and COL1A2 genes are relatively rare in the caucasian population. As a consequence, the examination of OI cell lines by chemical cleavage analysis is particularly convenient for the identification of the disease-related mutations, since few of the HA-reactive mismatches observed would be expected to be due to neutral sequence polymorphisms. Furthermore, the demonstration of the alanine/proline polymorphism at residue 459 of the α2(I) triple helical domain indicates that the replacement of non-glycine residues with proline, or of proline residues with certain other amino acids, may be tolerated at different positions within the triple helical domain. If an OI patient is found to harbor such a change in amino acid sequence relative to the published sequence, the possibility that it may not be related to the disease phenotype should be considered.

Table 11. DNA sequence analysis of the HA- and OT-reactive mismatches detected in cell lines ML226 and BD268.

Template Sequenced	Detected Mismatch	Affected cDNA Sequence Position[1]	Observed Sequence	Codon/Amino Acid Residue[2]	Representation Among Clones
ML226 cDNA	HA	1780	G	GCT Ala_{549}	2/4
			C	CCT Pro_{549}	2/4
	OT	1581	A	CCA Pro_{482}	2/4
			C	CCC Pro_{482}	2/4
BD268 cDNA	HA	1780	G	GCT Ala_{549}	8/17
			C	CCT Pro_{549}	9/17
	OT	1581	A	CCA Pro_{482}	9/17
			C	CCC Pro_{482}	8/17
α2XX probe	-	1780	G	GCT Ala_{549}	2/2
	-	1581	A	CCA Pro_{482}	2/2

[1] Prepro-α2(I) cDNA sequence positions are relative to the transcription start site (80).

[2] The amino acid residue numbers refer to positions within the prepro-α2(I) chain (80).

3.4 Identification of Type I Collagen Gene Mutations in OI

Although most of the helical glycine substitutions observed in OI patients are compatible with the phenotypic severity rules, there are a number of such defects which result in disease of unexpected mildness or severity. In order to further define the relationship between phenotypic expression and specific type I procollagen defects, a group of OI patients exhibiting a range of clinical severities was examined, with the aim of identifying their mutations. The patients in this study are summarized in Table 4, and included GD199, SL211, PP238, PM244, JN245, MB252, DP266, NC323, SS333 and MK345. In addition, the asymptomatic parents of patient SL211 (ML226 and JL227) were examined.

Electrophoretic Analysis of Collagen Chains

A characteristic feature of most OI fibroblast cell lines is the biosynthesis of type I procollagen whose chains undergo excessive post-translational hydroxylation and glycosylation, due to the presence of structural changes affecting the folding of the triple helical region of the molecule. These overmodifications result in increased molecular weights of the pro-α and α chains which can often be detected by altered mobility in gel electrophoresis. Thus, in order to determine whether fibroblast cell lines derived from the OI patients described above synthesize type I procollagen with alterations in post-translational modification, SDS-PAGE analysis of their type I collagen chains was performed. The fibroblast cell lines were metabolically labelled with ^{3}H-proline, and the proteins secreted into the medium were treated with pepsin, and subsequently analyzed by SDS-PAGE under non-reducing conditions. Pepsin treatment results in the removal of the amino- and carboxy-terminal propeptide sequences from procollagen types I and III which are secreted into the culture medium. However, the triple helical regions of these molecules

are resistant to pepsin cleavage since the digestions are performed at 4°C, which is well below their melting temperatures of approximately 40°C (352). Pepsin treatment is useful in that it also results in the proteolysis of non-collagenous ^{3}H-proline-labelled proteins which are secreted into the culture medium, allowing a selective analysis of pepsin-resistant collagen chains. Under non-reducing electrophoretic conditions, α1(I) and α2(I) chains are easily resolved from α1(III) chains (Fig. 20), since the latter migrate as $[\alpha 1(III)]_3$ homotrimers, due to the presence of intermolecular disulfide bonds near the carboxy terminus of the triple helical domain (30). This allows the excessive post-translational modifications associated with OI to be more readily detected since α1(I) chains with retarded electrophoretic mobility can be readily distinguished from α1(III) chains. Under reducing conditions, α1(III) chains migrate slightly slower than α1(I) chains, making it difficult to distinguish between α1(III) chains and overmodified α1(I) chains.

SDS-PAGE analysis allowed the OI fibroblast cell lines studied to be classified into two groups based on whether their α1(I) and α2(I) chains exhibited any detectable post-translational overmodification compared to normal chains derived from cell lines BD268 and YL278 (Table 12). Group 1 consisted of cell lines GD199, JL227, PM 244, MB252 and MK345, whose α1(I) and α2(I) chains comigrated with their normal counterparts (Fig. 20; data not shown for JL227, MB252 and MK345). The apparently normal electrophoretic mobility of type I collagen chains suggests the absence of extensive post-translational overmodification in these cell lines. It is of interest to note that JL227, the mother of type IV OI patient SL211, exhibits no detectable structural defect in type I collagen.

Group 2 consisted of cell lines SL211, ML226, PP238, JN245, DP266, NC323 and SS333, whose α1(I) and α2(I) chains exhibited retarded electrophoretic mobility compared to their normal counterparts (Fig. 20). For cell lines SL211, ML226, PP238, NC323 and SS333, this was manifested by the presence of significant broadening of the α1(I) and

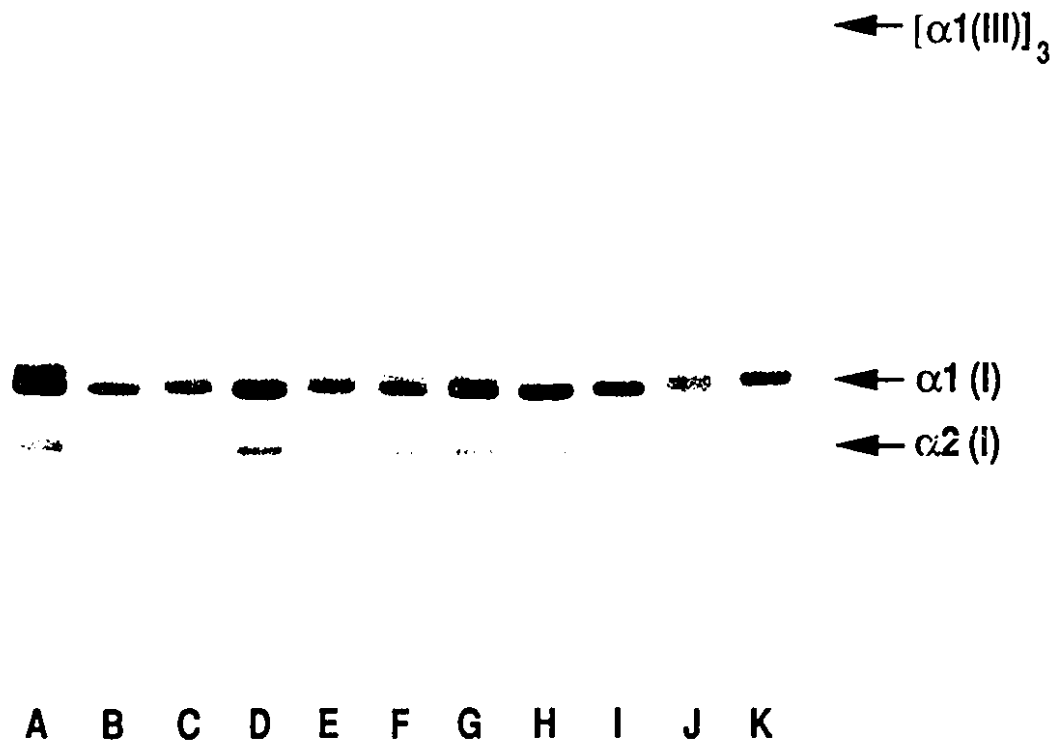


Fig. 20. SDS-PAGE analysis of collagen chains derived from the OI cell lines. Normal and OI fibroblast cell lines were metabolically-labelled with ^{3}H-proline, and the culture media were harvested. Proteins secreted into the media were precipitated with ammonium sulfate, digested with pepsin, and analyzed by SDS-PAGE under non-reducing conditions, followed by fluorography. The cell lines were DP266 (A), PM244 (B), JN245 (C), BD268 (D), NC323 (E), ML226 (F), SL211 (G), YL278 (H), GD199 (I), PP238 (J) and SS333 (K).

Table 12. Classification of OI cell lines according to the electrophoretic properties of their collagen chains.

Group	Electrophoretic Properties	Cell Lines	OI Type
1	Normal mobility of	GD199	IA
	α1(I) and α2(I) chains	JL227	- [1]
		PM244	I
		MB252	IA
		MK345	III
2	Retarded mobility of	SL211	IV
	α1(I) and α2(I) chains	ML226	- [1]
		PP238	IV
		JN245	IV
		DP266	III
		NC323	III
		SS333	II

[1] ML226 and JL227 are the clinically normal parents of patient SL211.

α2(I) bands. For cell lines JN245 and DP266, however, the overmodification appeared to be more extensive, resulting in band splitting, the presence of two distinct bands for each chain. For all of the cell lines, the band broadening or splitting was much more prominent for the α1(I) than for the α2(I) chains. The results indicate that the cell lines in Group 2 secrete two populations of each chain type, one which comigrates with its normal counterparts, and the other which consists of more slowly migrating chains. The degree of retarded electrophoretic mobility of the slowly-migrating chains was variable between the cell lines, being quite subtle for α1(I) chains derived from SS333 (with mild band broadening), but marked for those derived from JN245 and DP266 (with clear band splitting). These differences in the extents of overmodification were likely due to the presence of structural defects at different positions within the triple helical regions of the α1(I) or α2(I) chain. The demonstration of chain overmodification for the type IV OI cell line SL211 was not unexpected. However, the observation of extensive chain overmodification for fibroblasts from her clinically unaffected father, ML226, was somewhat surprising. The similarity in the extent of overmodification between cell lines SL211 and ML226 suggests that the daughter may possess the same defect as her father. However, more sensitive techniques are required to demonstrate that the type I collagen defects in these two individuals are identical.

In summary, the cell lines in this study appeared to fall into two groups based on whether post-translational overmodification of collagen chains was detected. This classification also appeared to divide the cell lines by OI type, since Group 1 contained all of the type I OI patients, while Group 2 contained most of the patients with OI types II, III and IV. Therefore, Group 1 may consist primarily of cell lines which harbor quantitative defects resulting in the decreased biosynthesis of type I procollagen, but which is structurally normal. Alternatively, they may possess structural defects resulting in minimal extents of chain overmodification which cannot be detected by the electrophoretic analysis

of α1(I) and α2(I) chains. This may explain the presence of the type III OI cell line MK345 in this group. In contrast, the cell lines in Group 2 all appeared to possess structural defects in type I procollagen, which resulted in varying degrees of post-translational overmodification and in OI of variable phenotypic severity. There did not appear to be a correlation between the degree of overmodification and phenotypic severity, since SS333 (type II OI) had mild band broadening, while JN245 (type IV OI) exhibited clear band splitting.

Cyanogen Bromide Peptide Mapping

The analysis of collagen chains by one-dimensional SDS-PAGE is a method which is helpful in identifying OI cell lines in which significant post-translational overmodification is present. However, this type of analysis may be incapable of detecting the less extensive overmodification associated with defects near the amino terminus of the α1(I) or α2(I) triple helical domain. In addition, it does not provide accurate information on the position of the defects in the various cell lines studied. In order to determine whether the cell lines from Group 1 possessed amino-terminal helical defects resulting in low levels of chain overmodification, and to obtain more accurate information on the positions of the structural defects in the cell lines from Group 2, collagen chains derived from each of these cell lines were examined by a two-dimensional cyanogen bromide peptide mapping technique. This approach is more sensitive for the detection of overmodification than is the analysis of intact α1(I) and α2(I) chains, because the changes in molecular weight accompanying overmodification are more significant in comparison to the molecular weights of the peptides. In addition, by determining the pattern of peptide overmodification, an approximate location for the structural defect within the triple helical domain of the pro-α1(I) or pro-α2(I) chain can be obtained, since only peptides on the amino-terminal side of

the defect should be affected. ^{3}H-proline-labelled media proteins were digested with pepsin and subjected to first-dimension SDS-PAGE. The resolved α1(I) and α2(I) chains were cleaved with cyanogen bromide within the polyacrylamide gel matrix, and the resulting peptides were separated by second-dimension SDS-PAGE. Cyanogen bromide cleaves polypeptide chains on the carboxy-terminal side of methionyl residues, via a mechanism involving electrophilic attack of the cyanogen moiety on the thioether side chain (353). The α1(I) and α2(I) chains are cleaved into four (CB3, CB6, CB7 and CB8) and two (CB4, CB3.5) major peptides, respectively, which can be visualized by fluorographic analysis of the SDS-PAGE gels. The α2(I) CB3.5 peptide contains an internal methionyl residue whose carboxyl peptide linkage is resistant to cleavage due to the neighboring threonine residue (353).

Analysis of collagen chains from the control cell lines BD268 (Fig. 21) and YL278 (Fig. 22) demonstrated cyanogen bromide peptides migrating as mildly sloping bands with their left (higher molecular weight) edges trailing upwards. This suggests that normal fibroblasts synthesize type I procollagen whose chains contain heterogeneous levels of the post-translational modifications. Furthermore, there appeared to be more heterogeneity in the modification of α2(I) chains since the peptides CB4 and CB3.5 migrated in a significantly more sloped manner than those derived from the α1(I) chain. The α1(I) CB3 peptide is the smallest of the major type I collagen cyanogen bromide peptides, and was usually too faint to be observed.

The analysis of collagen chains derived from some of the cell lines in Group 1 (GD199, JL227, PM244 and MB252) resulted in maps in which the major α1(I) and α2(I) cyanogen bromide peptides migrated similarly to those derived from normal collagen chains (Figs. 23 to 25; data not shown for JL227). This suggests that the type I collagen chains derived from these cell lines were normally modified. The maps for cell lines PM244 and MB252 (Figs. 24 and 25) contained an additional series of peptides which were derived from

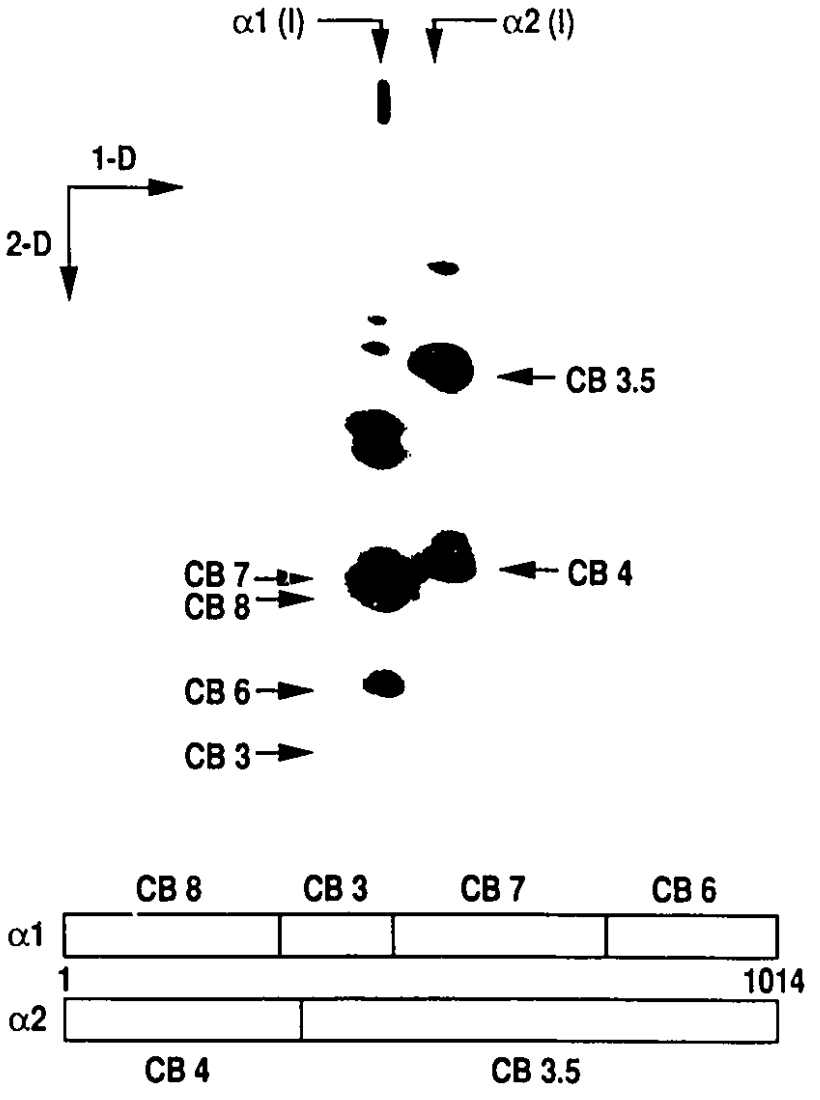


Fig. 21. Cyanogen bromide peptide analysis of collagen derived from cell line BD268. ^{3}H-proline-labelled collagen chains derived from BD268 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the positions of the cyanogen bromide peptides within the α1(I) and α2(I) triple helical domains.

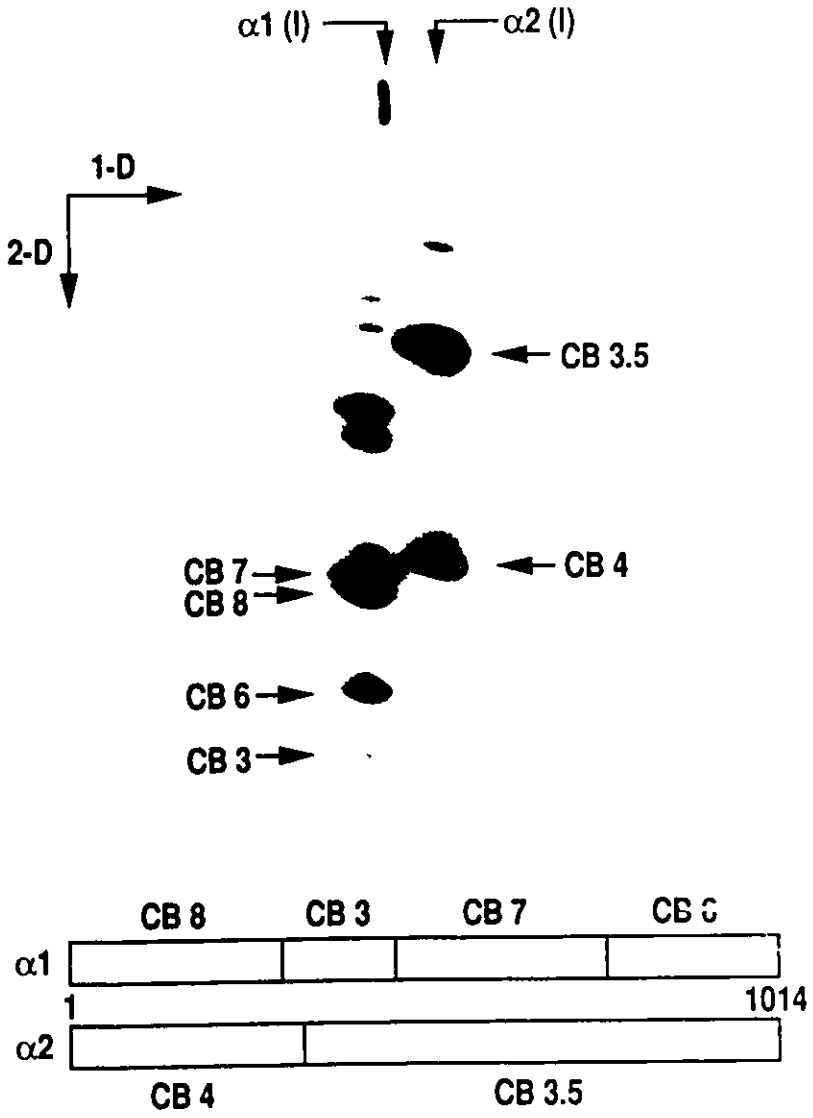


Fig. 22. Cyanogen bromide peptide analysis of collagen derived from cell line YL278. ^{3}H-proline-labelled collagen chains derived from YL278 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the positions of the cyanogen bromide peptides within the α1(I) and α2(I) triple helical domains.

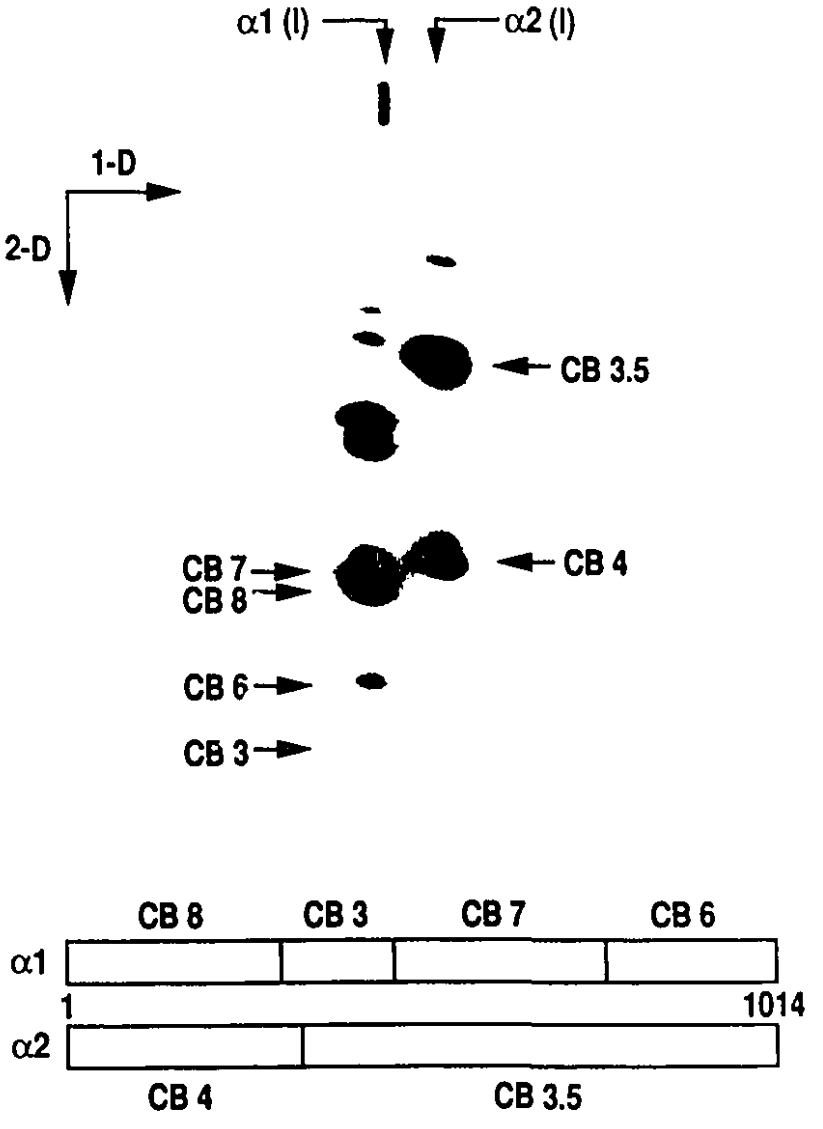


Fig. 23. Cyanogen bromide peptide analysis of collagen derived from OI cell line GD199. ^{3}H-proline-labelled collagen chains derived from GD199 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

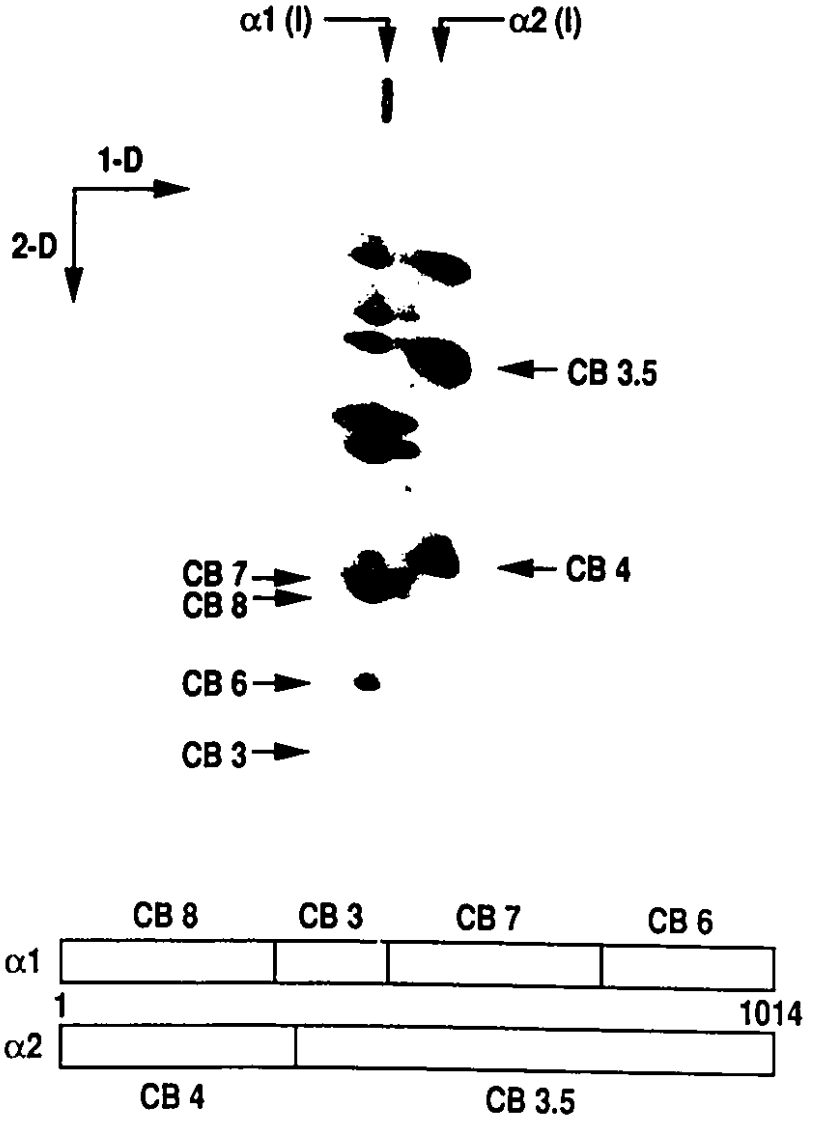


Fig. 24. Cyanogen bromide peptide analysis of collagen derived from OI cell line PM244. ^{3}H-proline-labelled collagen chains derived from PM244 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

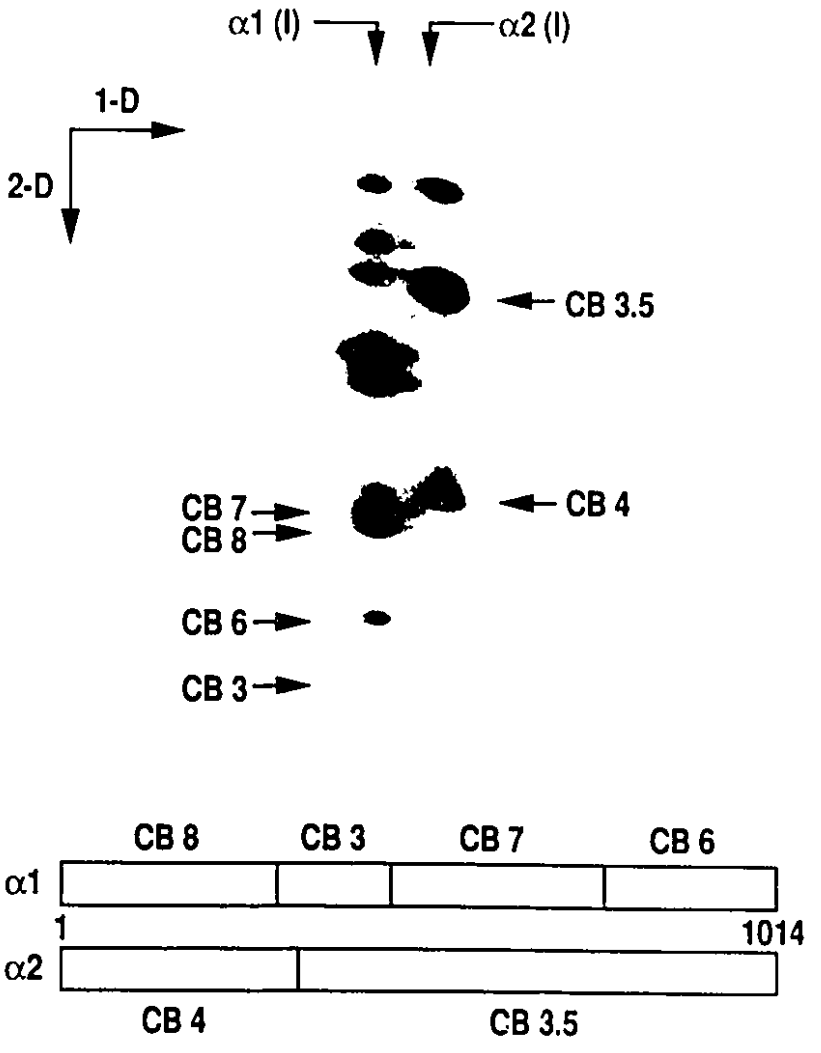


Fig. 25. Cyanogen bromide peptide analysis of collagen derived from OI cell line MB252. ^{3}H-proline-labelled collagen chains derived from MB252 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

chains migrating between the α1(I) and α2(I) chains. The peptides derived from these chains, however, co-migrated with those arising from normally migrating α1(I) chains, with the exception of two additional spots which did not appear to have normal α1(I) peptide counterparts. These features, which were not observed in maps of collagen obtained from control cell lines, likely do not represent disease-related defects, but more likely correspond to partially-digested cyanogen bromide peptides containing the α1(I) CB6 peptide which has been truncated at its carboxy terminus by pepsin digestion. This is indicated by the selective absence of a CB6 peptide spot for these abnormally migrating α1(I) chains.

Analysis of collagen chains derived from the cell lines in Group 2 (SL211, ML226, PP238, JN245, DP266, NC323 and SS333) demonstrated clear differences in the post-translational modification of certain peptides for each cell line, when compared to their normal counterparts (Figs. 26 to 32). Some cyanogen bromide peptides were observed to migrate in a sloped manner similar to their counterparts from normal collagen chains, suggesting that they contained normal levels of post-translational modifications. However, other peptides migrated as V-shaped bands whose right (lower molecular weight) edges sloped in a manner similar to their counterparts from normal collagen chains, but whose left (higher molecular weight) edges trailed upwards at a more severe angle. This is indicative of peptide overmodification, with molecules in the right edge of the band being derived from normally migrating collagen chains, and those in the left edge being derived from more slowly migrating, overmodified collagen chains. In addition, the angle at which the left edge of a specific peptide band trailed upwards was found to be variable between different cell lines, indicating that further distinctions could be made in the degree of overmodification present in that peptide. Helical defects in either the α1(I) or α2(I) chain result in the overmodification of both chains on the amino-terminal side of the defect. However, despite observing clear patterns of overmodification of the α1(I) cyanogen

bromide peptides in some OI cell lines, no clear abnormalities were exhibited by their α2(I) peptides. This is probably due to the greater extent and heterogeneity of the post-translational modification of α2(I) chains in normal collagen. Thus, the overmodification of α2(I) chains results in a less significant change in molecular weight, and is less readily detected than that in α1(I) chains. However, since the α1(I) chain consists of four major cyanogen bromide peptides, in contrast to the two from the α2(I) chain, the analysis of the pattern of α1(I) peptide overmodification is more informative in determining helical defect position.

Collagen chains from PP238, NC323 and SS333 fibroblasts produced maps in which only the amino-terminal peptide of the α1(I) chain, CB8, appeared to be excessively overmodified (Figs. 26 to 28). However, the α1(I) CB8 peptides derived from PP238 and NC323 fibroblasts (Figs. 27 and 28) seemed to be more severely affected than that from SS333 fibroblasts (Fig. 26), suggesting that the defect in the latter cell line is likely within α1(I) CB8, or the corresponding region of the α2(I) chain. For PP238 and SS333 fibroblasts, where the α1(I) CB7 and CB6 peptides appeared normally modified and the effects on α1(I) CB3 were inconclusive, the defects are likely in the α1(I) region extending from the carboxy-terminal portion of CB8 to the amino-terminal portion of CB7, or in the corresponding region of the α2(I) chain. Collagen chains from DP266 fibroblasts exhibited extensive overmodification of α1(I) CB8 and CB3 (Fig. 29). However, α1(I) CB7 was only minimally affected, suggesting that the defect in this cell line is located within this peptide or the corresponding region of the α2(I) chain. Collagen from cell lines SL211 and ML226 produced maps in which the α1(I) peptides CB8 and CB7 were extensively overmodified, whereas α1(I) CB6 appeared only partially affected (Figs. 30 and 31). This indicates that these cell lines possess defects within α1(I) CB6, or the corresponding region of the α2(I) chain. Since the collagen derived from SL211 and ML226 fibroblasts exhibited similar patterns of peptide overmodification, it appears likely

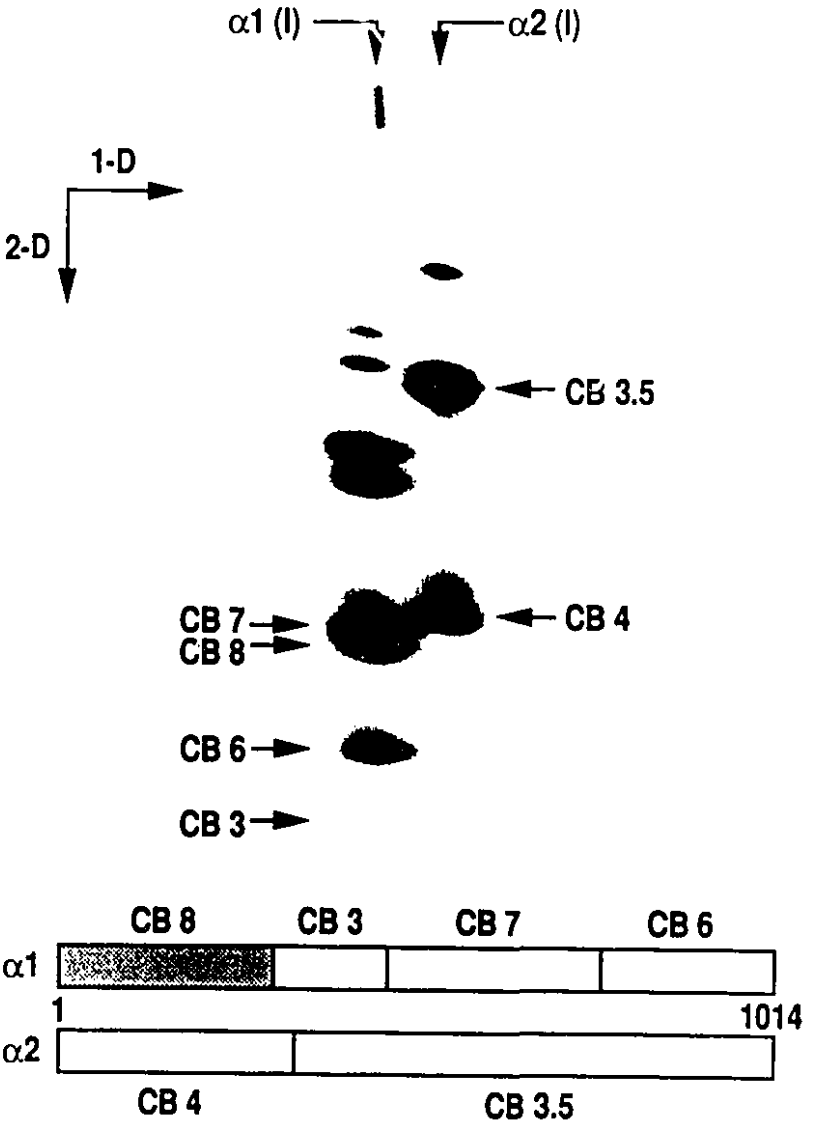


Fig. 26. Cyanogen bromide peptide analysis of collagen derived from OI cell line SS333. ^{3}H-proline-labelled collagen chains derived from SS333 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

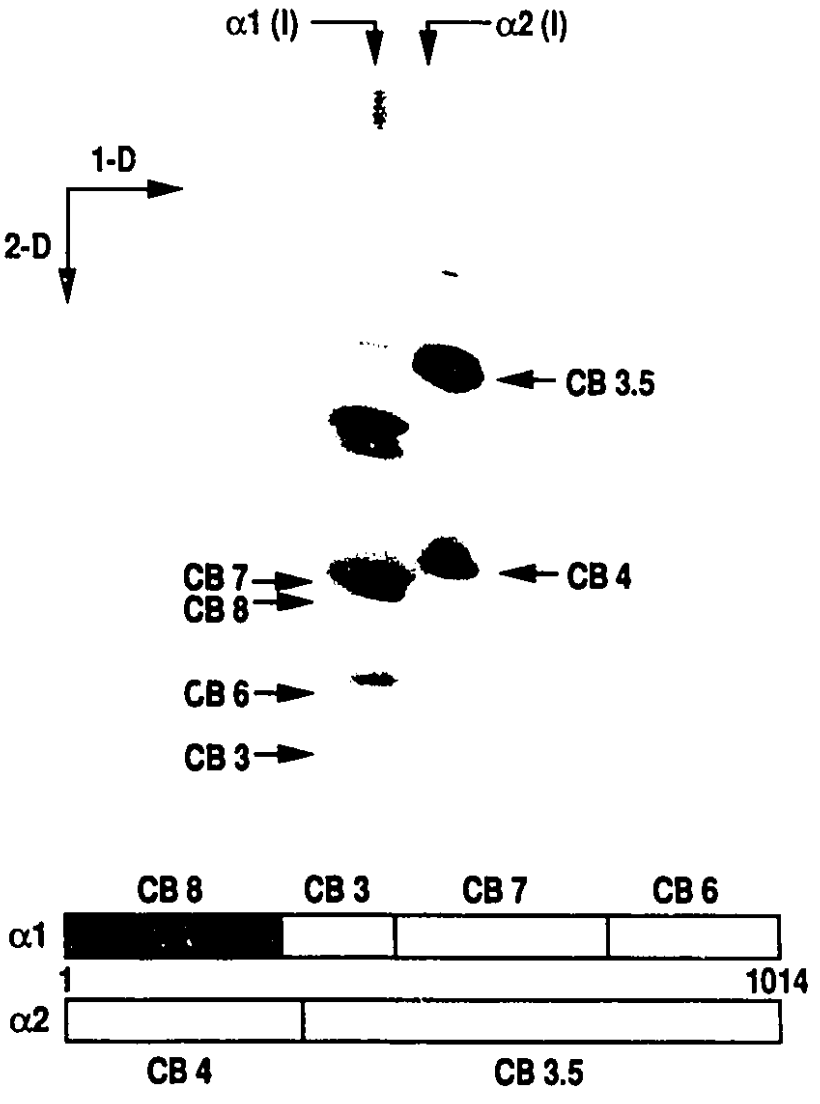


Fig. 27. Cyanogen bromide peptide analysis of collagen derived from OI cell line PP238. ^{3}H-proline-labelled collagen chains derived from PP238 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

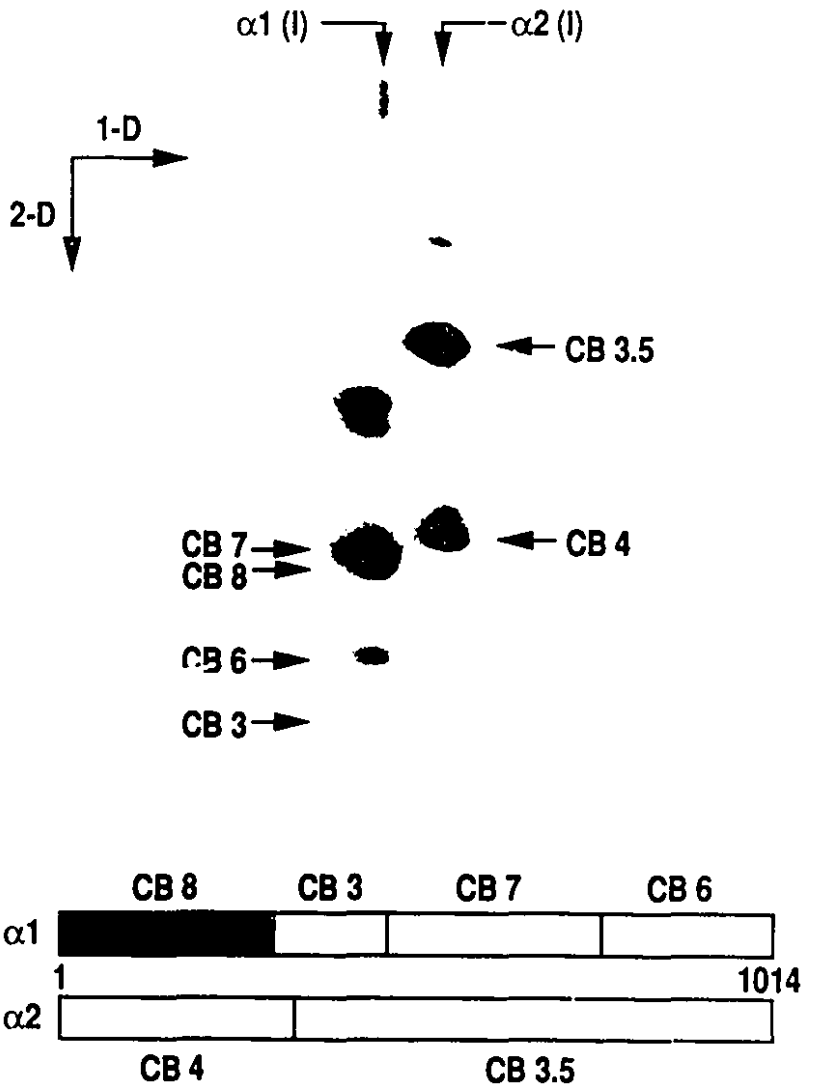


Fig. 28. Cyanogen bromide peptide analysis of collagen derived from OI cell line NC323. ^{3}H-proline-labelled collagen chains derived from NC323 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

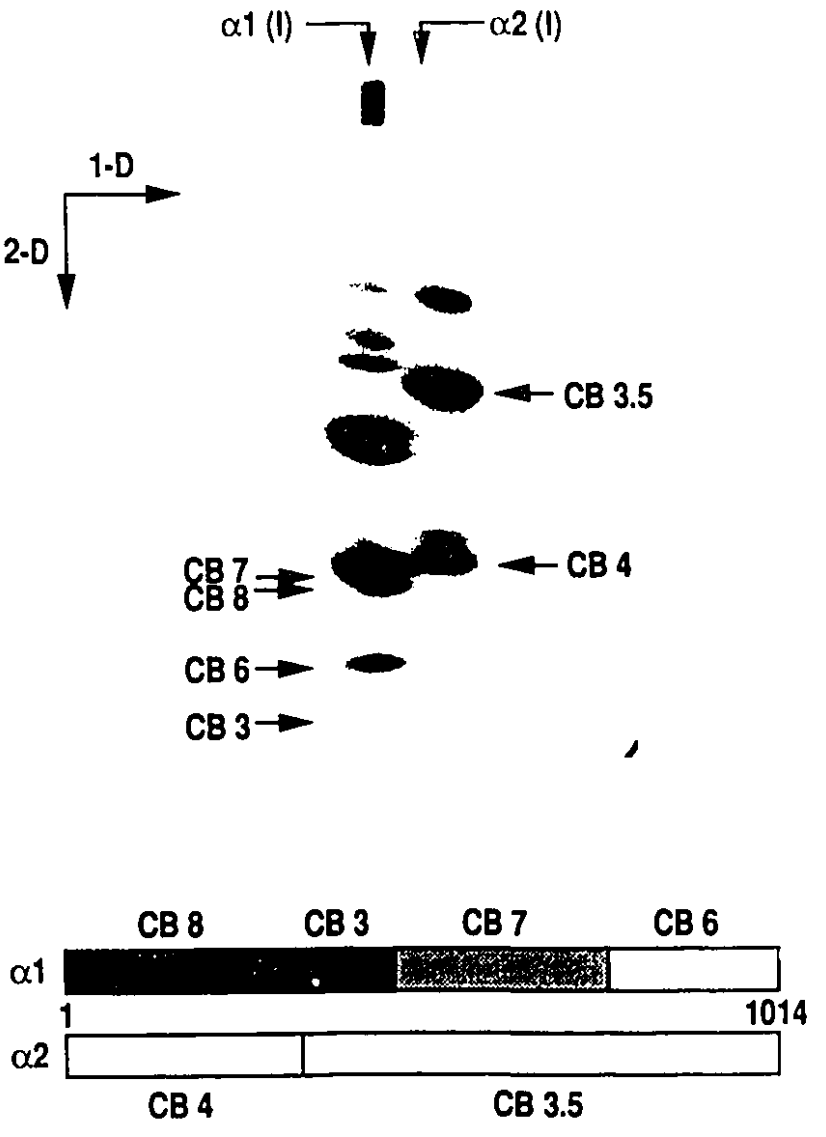


Fig. 29. Cyanogen bromide peptide analysis of collagen derived from OI cell line DP266. ^{3}H-proline-labelled collagen chains derived from DP266 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

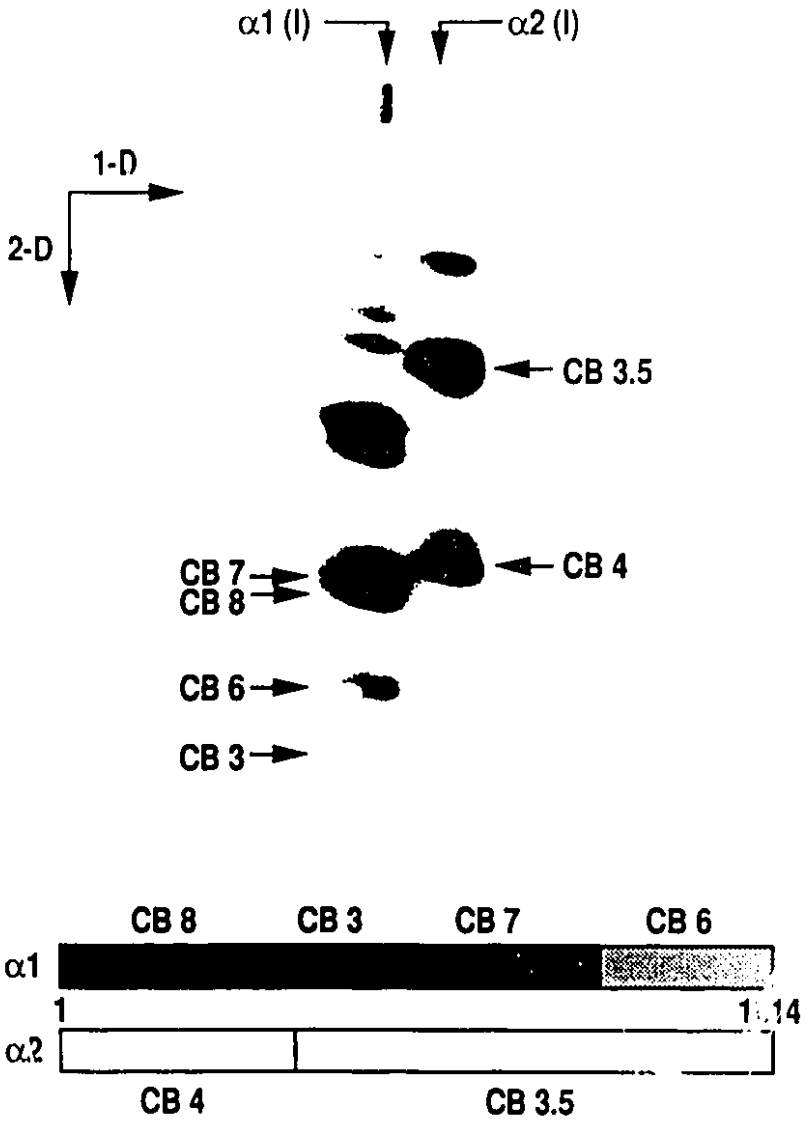


Fig. 30. Cyanogen bromide peptide analysis of collagen derived from OI cell line SL211. ^{3}H-proline-labelled collagen chains derived from SL211 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

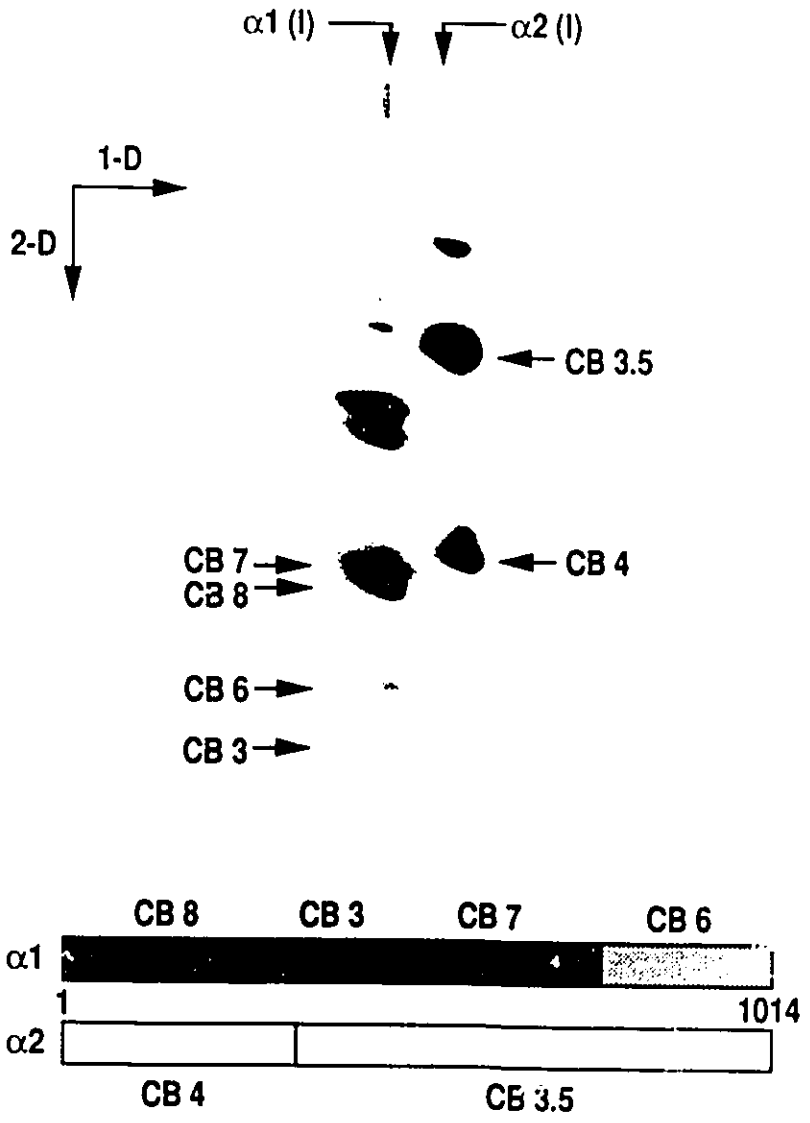


Fig. 31. Cyanogen bromide peptide analysis of collagen derived from OI cell line ML226. ^{3}H-proline-labelled collagen chains derived from ML226 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

that the daughter possesses the same defect exhibited by her father. Analysis of collagen derived from JN245 fibroblasts revealed the most extensively modified chains of all the cell lines studied, with all major α1(I) peptides exhibiting abnormal electrophoretic mobilities (Fig. 32). Such a pattern indicates the presence of a defect near the carboxy terminus of the triple helical domain of either chain.

In general, the results obtained from cyanogen bromide peptide mapping were consistent with the one-dimensional electrophoretic analysis of collagen chains. Collagen chains from Group 1 cell lines exhibited apparently normal electrophoretic mobility, and produced peptide maps which were similar to those obtained for normal collagen chains derived from control cell lines. This suggests the absence of α1(I) and α2(I) helical structural defects in these cell lines, although the presence of a defect at the extreme amino terminus of the α1(I) or α2(I) chain cannot yet be excluded. Collagen chains from Group 2 cell lines exhibited retarded electrophoretic mobility, and produced maps which exhibited varying patterns of peptide overmodification compared to those obtained for normal collagen chains derived from control cell lines. There also appeared to be a correlation between the extent of retarded electrophoretic mobility of α1(I) and α2(I) chains, and the carboxy-terminal extent of the overmodification as indicated by peptide mapping. The only inconsistency observed was for cell line DP266, which was demonstrated by peptide mapping to possess a defect in the central region of either the α1(I) or α2(I) chain, yet which by one-dimensional analysis of collagen chains revealed an electrophoretic pattern characteristic of a defect near the carboxy terminus.

Chemical Cleavage Detection of Mutations

The cyanogen bromide peptide mapping data obtained were indicative of the locations of the molecular defects in the cell lines from Group 2, but did not provide accurate information about the precise sites affected. Identification of the corresponding mutations,

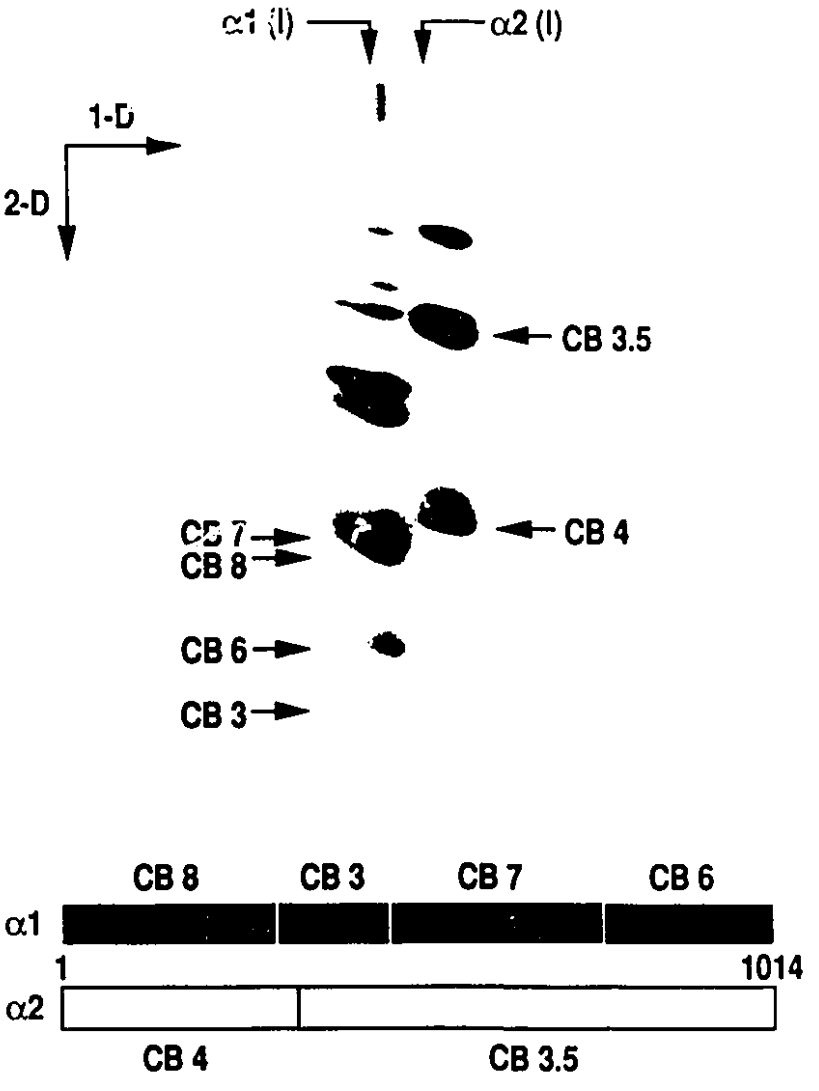


Fig. 32. Cyanogen bromide peptide analysis of collagen derived from OI cell line JN245. ^{3}H-proline-labelled collagen chains derived from JN245 fibroblasts were separated by first-dimension SDS-PAGE under non-reducing conditions. Gel strips were incubated with cyanogen bromide, and the resultant peptides were separated by second-dimension SDS-PAGE and visualized by fluorography. An untreated first-dimension gel strip is located above the two-dimensional gel. The bar diagrams below the gel indicate the patterns of overmodification of the α1(I) and α2(I) peptides. Dark shading indicates apparently complete overmodification, grey shading indicates partial overmodification, and no shading indicates the absence of detectable overmodification.

based on this data alone, would still necessitate the cloning and sequencing of regions between 500 bp and 1 kbp in length from each of the prepro-α1(I) and prepro-α2(I) cDNAs. In addition, peptide mapping was unable to detect significant overmodification in the cell lines comprising Group 1. Nonetheless, it is still possible that these cell lines harbour mutations within the regions of approximately 500 bp of the cDNA sequences covering the extreme amino-terminal ends of the α1(I) and α2(I) triple helical domains. Such mutations could give rise to minimal overmodification that may not be detected by cyanogen bromide peptide mapping. In order to examine this possibility, and to provide more accurate information on the positions of the mutations in the cell lines exhibiting chain overmodification, chemical cleavage mismatch analysis of total RNA from four Group 1 cell lines (GD199, JL227, MB252 and MK345) and seven Group 2 cell lines (SL211, ML2^6, PP238, JN245, DP266, NC323 and SS333) was performed. The cyanogen bromide mapping data were informative in this analysis, allowing the selection of the two chemical cleavage cDNA probes (one for the prepro-α1(I) chain and one for the prepro-α2(I) chain) spanning the predicted site of the defect in each Group 2 cell line. However, chemical cleavage analysis was also performed using all of the other probes in order to verify whether other mismatches could be detected. Total RNA (25 μg) was annealed with the six cDNA probes listed in Table 6 (20 ng), and the resulting heteroduplexes were treated with HA or OT. HA is a particularly useful reagent for the analysis of OI cell lines, since it is able to detect mutations affecting glycine codons (GGN), which always result in mismatched C bases within the noncoding strand of the cDNA probe. Some of the cell lines were also analyzed with OT in order to examine the possibility of mutations resulting in T mismatches.

Group 1 cell lines GD199, JL227 and MB252 did not exhibit HA-reactive mismatches with any of the six cDNA probes utilized. The analysis of cell line JL227 with OT likewise failed to demonstrate any mismatches. In fact, the only mismatch detected during the

analysis of the Group 1 cell lines was observed for MK345. This HA-reactive mismatch was detected with the α1XB probe, and resulted in a cleavage product of approximately 180 bases (Fig. 33), suggesting the presence of a sequence change near position 293 of the prepro-α1(I) cDNA sequence. However, examination of the prepro-α1(I) cDNA sequence reveals that nucleotide positions 292, 295 and 296 are the only ones within close proximity (±5 bp) of the projected cleavage site, which could give rise to a mismatched C base in the mRNA:cDNA heteroduplex during chemical cleavage analysis (78; Fig. 34). Furthermore, the C chemical sequencing ladder, produced by random cleavage of the α1XB probe at all C bases, could be read until the position of the cleavage product. This further pinpointed the sequence change to position 295 or 296, which encode arginine residue 59 of the prepro-α1(I) polypeptide chain.

The chemical cleavage analysis of Group 2 cell lines SL211, ML226 and DP266 failed to detect any HA-reactive mismatches, with the exception of the polymorphism in ML226 (described in section 3.3), although their collagen exhibited extensive overmodification. The analysis of SL211 and ML226 RNA with OT likewise did not result in the detection of any mismatches other than the ML226 polymorphism (described in section 3.3). The only cell lines from Group 2 which exhibited HA-reactive mismatches were PP238, JN245, NC323 and SS333.

Analysis of RNA isolated from PP238 fibroblasts identified a single HA-reactive mismatch which was detected using the α2XX probe (Fig. 35). This C-base mismatch resulted in a cleavage product of approximately 760 bases, suggesting the presence of a sequence change near position 2060 of the prepro-α2(I) cDNA sequence. The mutation would be projected to affect the protein sequence at approximately amino acid residue 640 of the prepro-α2(I) chain, or residue 550 of the α2(I) triple helical domain. This result is consistent with the cyanogen bromide mapping data which predicted a defect in the central region of the α1(I) or the α2(I) triple helical domain.

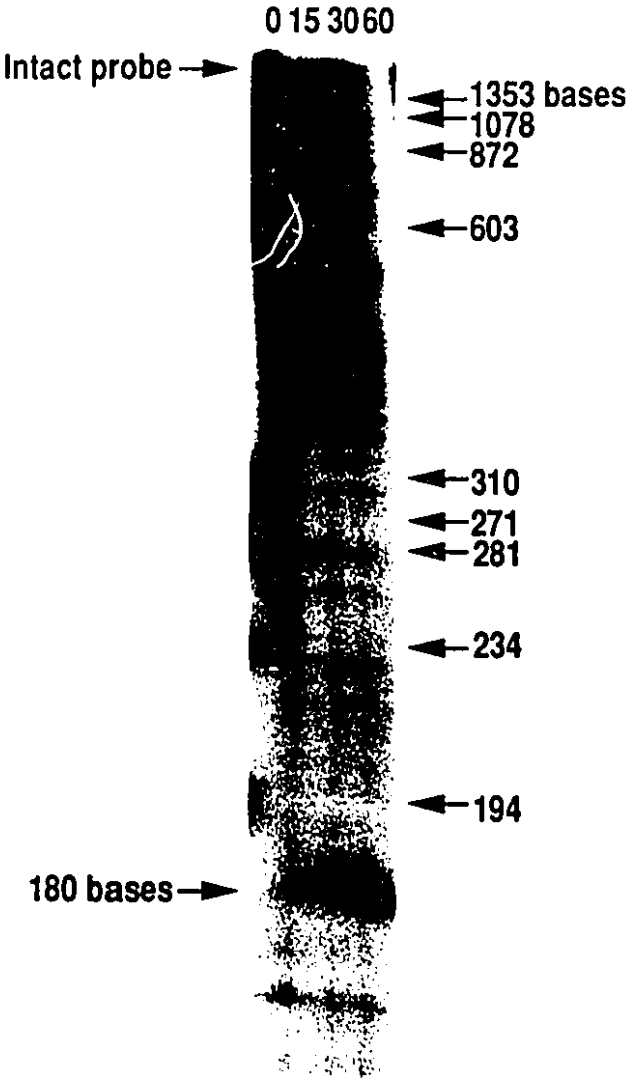


Fig. 33. Chemical cleavage detection of a HA-reactive mismatch in OI cell line MK345. Total RNA (25 μg) from cell line MK345 was annealed with the α1XB probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 1 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

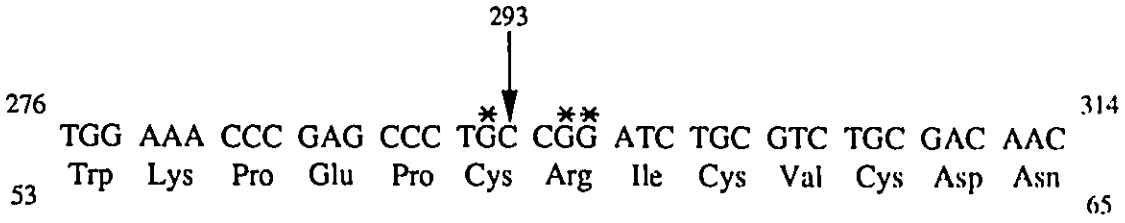


Fig. 34. The prepro-α1(I) cDNA sequence near the predicted site of the HA-reactive mismatch in OI cell line MK345. The published cDNA (78) sequence and the deduced amino acid sequence are shown, with the predicted site of the mismatch indicated by an arrow. The prediction is based solely on the approximate size of the cleavage product observed in Fig. 33 (180 bases). The asterisks indicate the only bases within close proximity (± 5 bp) which could produce C-base mismatches when changed by point mutation.

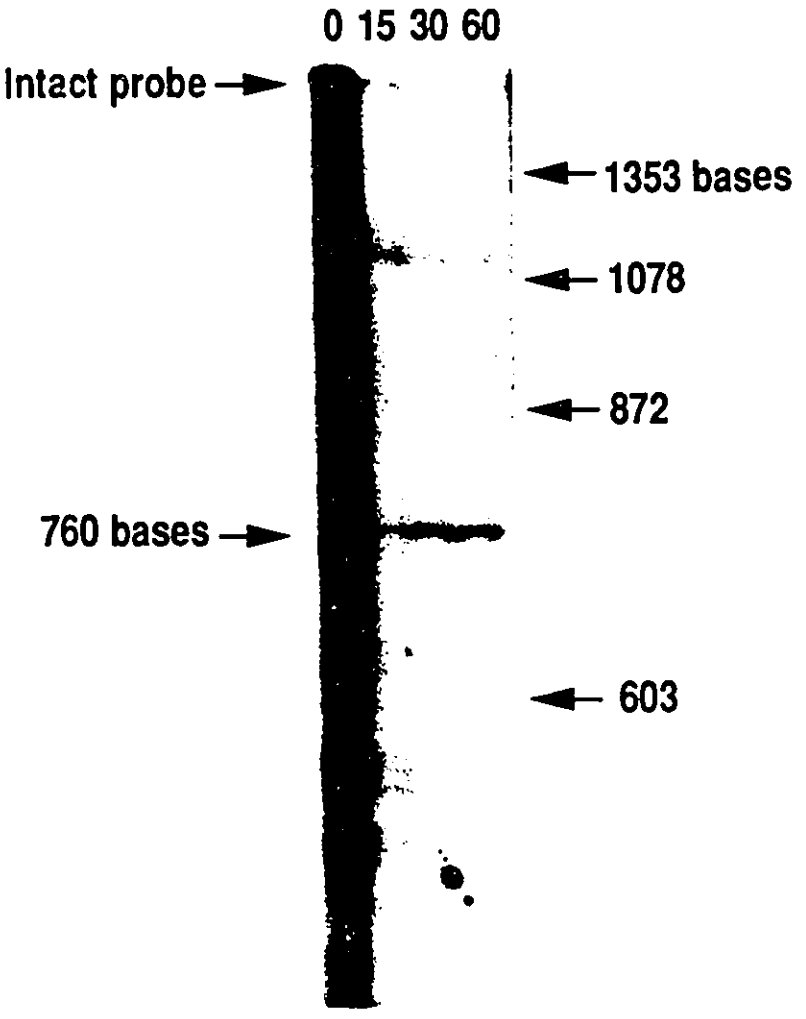


Fig. 35. Chemical cleavage detection of a HA-reactive mismatch in OI cell line PP238. Total RNA (25 μg) from cell line PP238 was annealed with the α2XX probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

Chemical cleavage analysis of RNA from JN245 fibroblasts identified two HA-reactive mismatches. The mismatch detected with the α2XX probe yielded a 530-base fragment (Fig. 36A) indicating a sequence change at approximately position 1830 of the prepro-α2(I) cDNA sequence. This is within close proximity of the HA-reactive polymorphism demonstrated in ML226 and BD268 fibroblasts (described in section 3.3). The mismatch detected with the α2XE probe resulted in the appearance of a cleavage product of approximately 380 bases (Fig. 36B), suggesting the presence of another sequence change near position 3150 of the prepro-α2(I) cDNA. Such a sequence change would be projected to affect the amino acid sequence near residue 1010 of the prepro-α2(I) chain, or residue 920 of the α2(I) triple helical domain. This result is compatible with the cyanogen bromide mapping data which predicted a defect near the carboxy terminus of the α1(I) or α2(I) triple helical domain.

Chemical cleavage analysis of RNA isolated from NC323 fibroblasts identified a single HA-reactive mismatch detected with the α2XX probe (Fig. 37). This C-base mismatch resulted in a cleavage product of approximately 400 bases, suggesting the presence of a sequence change near position 1700 of the prepro-α2(I) cDNA. Such a sequence alteration would affect the amino acid sequence near residue 520 of the prepro-α2(I) chain, or residue 430 of the α2(I) triple helical domain. This result is compatible with the cyanogen bromide mapping data which predicted a defect within the central region of the α1(I) or α2(I) triple helical domain.

RNA isolated from SS333 fibroblasts also resulted in a single mismatch when examined by chemical cleavage. This HA-reactive mismatch, detected with the α1XB probe, resulted in a cleavage product of approximately 1170 bases in size (Fig. 38). This indicated the presence of a sequence change near position 1290 of the prepro-α1(I) cDNA sequence. This would affect the protein sequence near amino acid residue 388 of the prepro-α1(I) chain, or residue 210 of the α1(I) triple helical domain. This result is

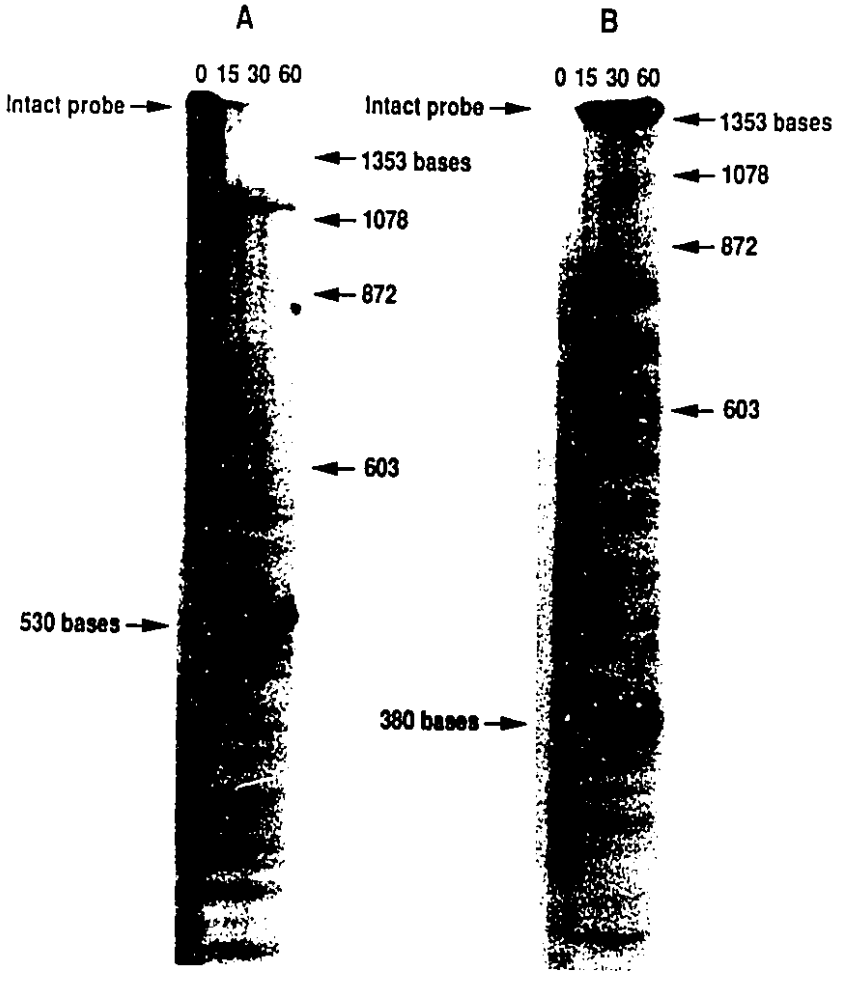


Fig. 36. Chemical cleavage detection of two HA-reactive mismatches in OI cell line JN245. Total RNA (25 µg) from cell line JN245 was annealed with α2XX (A) or α2XE (B) probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

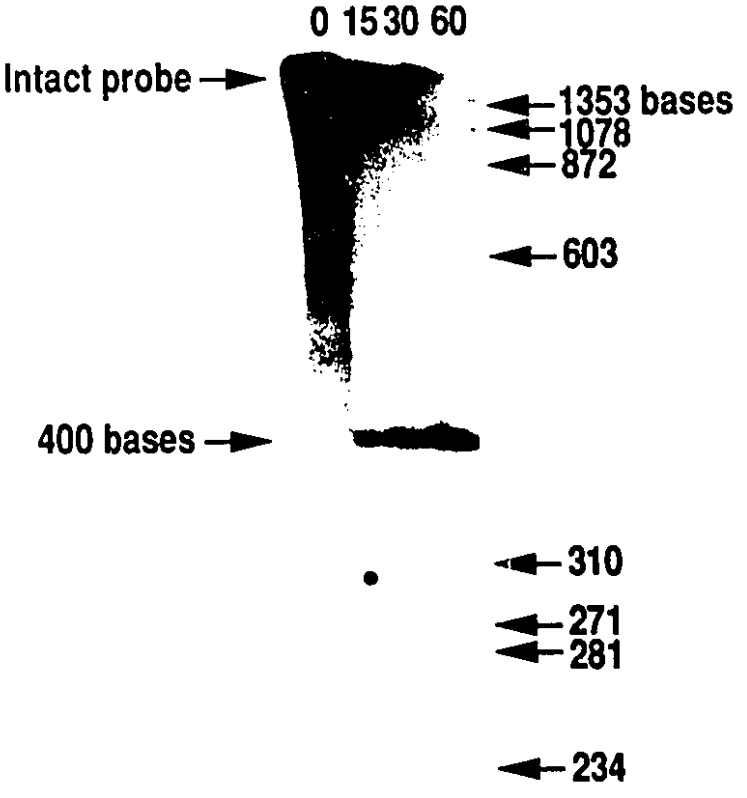


Fig. 37. Chemical cleavage detection of a HA-reactive mismatch in OI cell line NC323. Total RNA (25 μg) from cell line NC323 was annealed with the α2XX probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 1 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

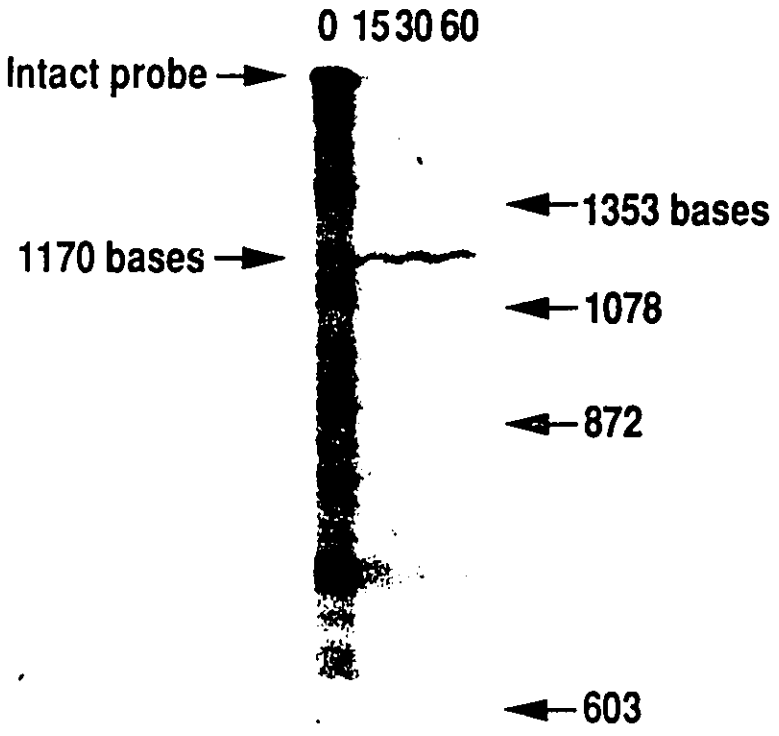


Fig. 38. Chemical cleavage detection of a HA-reactive mismatch in OI cell line SS333. Total RNA (25 μg) from cell line SS333 was annealed with the α1XB probe (20 ng). Heteroduplexes were treated with HA for 0, 15, 30 and 60 min, cleaved with piperidine, and electrophoresed for 4 h on a denaturing 5% polyacrylamide gel. Size markers were end-labelled HaeIII-digested ϕX174 DNA.

compatible with the cyanogen bromide peptide mapping data, which suggested the presence of a defect near the amino terminus of the α1(I) or α2(I) triple helical domain.

The absence of detected HA-reactive mismatches in Group 1 cell lines GD199, JL227 and MB252 indicates the absence of mutations involving G bases within the prepro-α1(I) and prepro-α2(I) cDNA sequences covered by these probes. Since this would include the codons for all triple helical glycine residues of each chain, these results are consistent with the lack of significant chain overmodification observed for these cell lines. Thus, cell line JL227 may not harbour a type I collagen mutation, while OI cell lines GD199 and MB252 may possess mutations resulting in quantitative defects in type I procollagen biosynthesis. The absence of HA-reactive mismatches in Group 2 cell lines SL211, ML226 and DP266, despite the presence of collagen overmodification, suggests that these individuals possess non-glycine defects, within the α1(I) or α2(I) helical domains, which affect the stability and folding of the type I procollagen triple helix. In addition, since no OT-reactive mismatches were observed in OI cell line SL211, nor in cell lines ML226 and JL227 (derived from the asymptomatic parents), the apparent collagen defects in the former two cell lines probably do not result from mutations producing T mismatches.

The HA-reactive mismatches detected in cell lines PP238, JN245, NC323, SS333 and MK345 are summarized in Table 13, and their projected locations within the prepro-α1(I) and prepro-α2(I) cDNA sequences are illustrated in Figs. 39 and 40. Each of the cell lines was found to exhibit a HA-reactive mismatch which was not observed during the chemical cleavage analysis of normal fibroblasts. This suggested that the mismatches resulted from sequence changes which may be responsible for the OI phenotypes expressed in these individuals. Furthermore, the locations of the mismatches detected in PP238, JN245, NC323 and SS333 suggest that the mutations in these OI patients likely result in glycine substitutions within the triple helical domains of the α1(I) and α2(I) chains. However, the location of the mismatch detected in cell line MK345, within the cDNA region encoding the

Table 13. HA-reactive mismatches detected in OI cell lines PP238, JN245, NC323 , SS333 and MK345.

Cell Line	cDNA Probe	Size of Cleavage Product (bases)	Projected Location of Sequence Change [1]
PP238	α2XX	760	2060
JN245	α2XX	530	1830
	α2XE	380	3150
NC323	α2XX	400	1700
SS333	α1XB	1170	1290
MK345	α1XB	180	293

[1] Numbers represent nucleotide positions within the prepro-α1(I) or prepro-α2(I) cDNA sequences relative to the transcription start sites (78, 80).

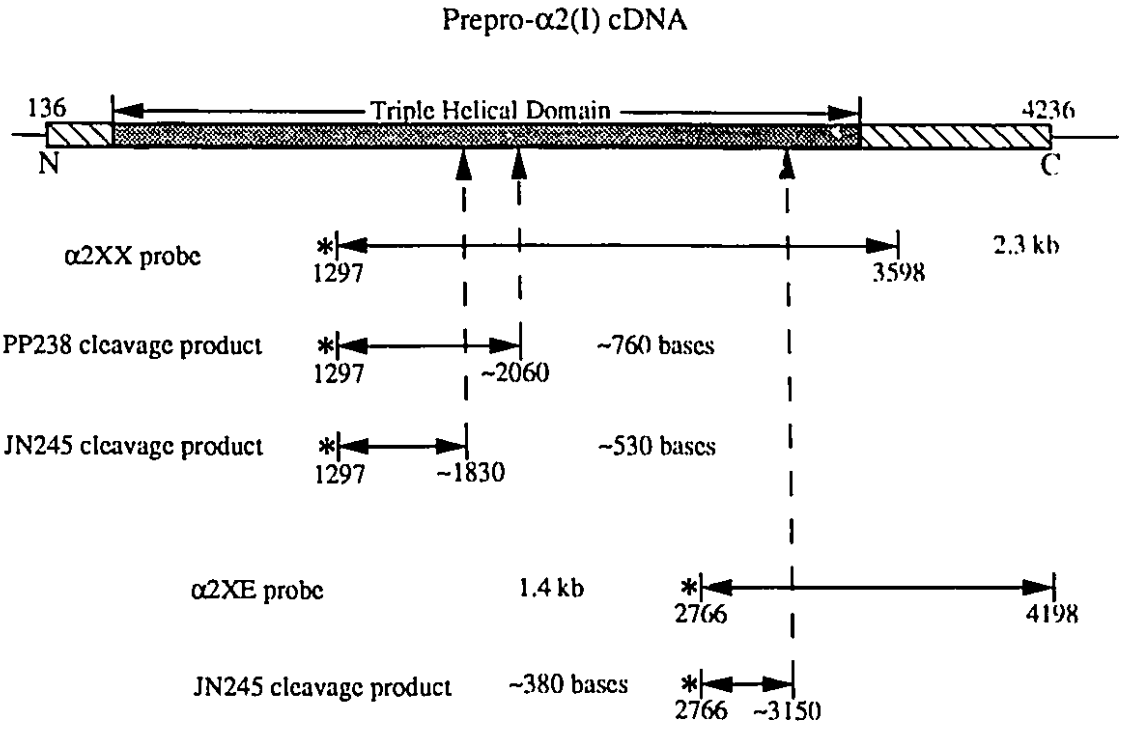


Fig. 39. Schematic diagram of the locations of the HA-reactive mismatches detected in OI cell lines PP238 and JN245. The cDNA probes and HA cleavage products are represented by double-headed arrows below the bar diagram of the full-length prepro-α2(I) cDNA. cDNA coding regions are indicated by bars, and untranslated regions are represented by plain lines. The regions encoding the main triple helical domain (grey shading) and the terminal domains (diagonal shading) are also indicated. cDNA sequence positions are relative to the transcription start site (80). The asterisks denote the locations of the radioactive labels, and the vertical arrows indicate the locations of the detected mismatches within the cDNA sequence.

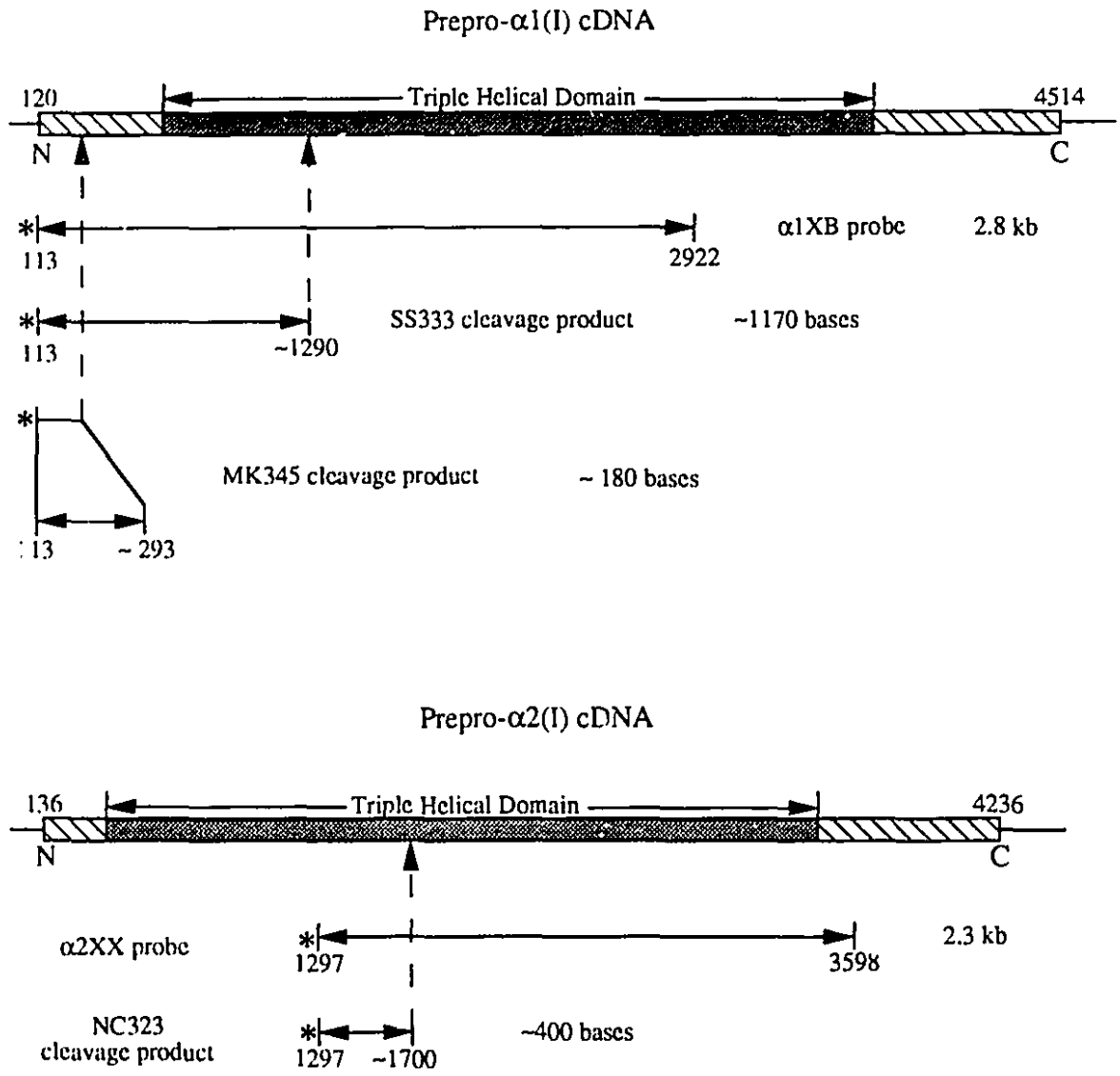


Fig. 40. Schematic diagram of the locations of the HA-reactive mismatches detected in OI cell lines NC323, SS333 and MK345. The cDNA probes and HA cleavage products are represented by double-headed arrows below the bar diagrams of the full-length cDNAs. cDNA coding regions are indicated by bars, and untranslated regions are represented by plain lines. The regions encoding the main triple helical domains (grey shading) and the terminal domains (diagonal shading) are also indicated. cDNA sequence positions are relative to the transcription start sites (78, 80). The asterisks denote the locations of the radioactive labels, and the vertical arrows indicate the locations of the detected mismatches within the cDNA sequences.

amino-terminal propeptide of the prepro-α1(I) chain, was intriguing in suggesting the possibility that defects within this region of the procollagen molecule could give rise to OI, although no such defects have yet been described.

Identification of the Mutations by DNA Sequencing

In order to identify the exact nature of the sequence changes producing the mismatches detected in OI cell lines PP238, JN245, NC323, SS333 and MK345, the sequence analysis of prepro-α1(I) and prepro-α2(I) cDNAs was performed. Total RNA isolated from the corresponding cell lines was used for the oligo-dT or random hexamer-primed synthesis of first strand cDNA. The PCR was then performed to amplify short regions (<400 bp) of the prepro-α1(I) and prepro-α2(I) cDNAs bearing the sequence changes detected by chemical cleavage analysis. The PCR primer pairs COLL-9/COLL-10, COLL-11/COLL-12, COLL-1/COLL-2, COLL-7/COLL-8 and COLL-15/COLL-16 were utilized in the amplification of cDNA derived from PP238, JN245, NC323, SS333 and MK345 fibroblasts, respectively. The PCR products were cloned into the M13mp18 vector for the production of ssDNA sequencing templates. The sequence data are summarized in Table 14.

Analysis of the cDNAs cloned from PP238 fibroblasts identified two alleles differing in sequence at a single base position (Fig. 41). One allele, identical to the published cDNA sequence, had a G at position 2036 (within exon 32), as the second base of the GGT triplet coding for glycine residue 634 of the prepro-α2(I) chain, or residue 544 of the α2(I) triple helical domain. However, the other allele had a T at this position, thereby converting the glycine codon to a GTT triplet coding for a valine residue. This G → T single base pair substitution results in a HA-reactive T:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and is located within 30 bp of the site predicted by chemical cleavage analysis. The sequencing data is also consistent with the cyanogen bromide peptide mapping studies which indicated that the defect in this OI patient was located in the central

Table 14. DNA sequence analysis of the HA-reactive mismatches detected in OI cell lines PP238, JN245, NC323, SS333 and MK345.

Cell Line	Affected cDNA Sequence Position[1]	Observed Sequence	Codon/Amino Acid Residue[2]	Representation Among Clones
PP238	Prepro-α2(I) 2036	G	GGT Gly_{634}	2/11
		T	GTT Val_{634}	9/11
JN245	Prepro-α2(I) 3169	G	GGT Gly_{1012}	2/12
		A	AGT Ser_{1012}	10/12
NC323	Prepro-α2(I) 1684	G	GGT Gly_{517}	7/12
		C	CGT Arg_{517}	5/12
SS333	Prepro-α1(I) 1284	G	GGT Gly_{389}	5/12
		C	CGT Arg_{389}	7/12
MK345	Prepro-α1(I) 296	G	CGG Arg_{57}	8/11
		T	CGT Arg_{57}	3/11

[1] cDNA sequence positions are relative to the corresponding transcription start sites (78, 80).

[2] The amino acid residue numbers refer to positions within the corresponding prepro-α1(I) or prepro-α2(I) chain (77-80).

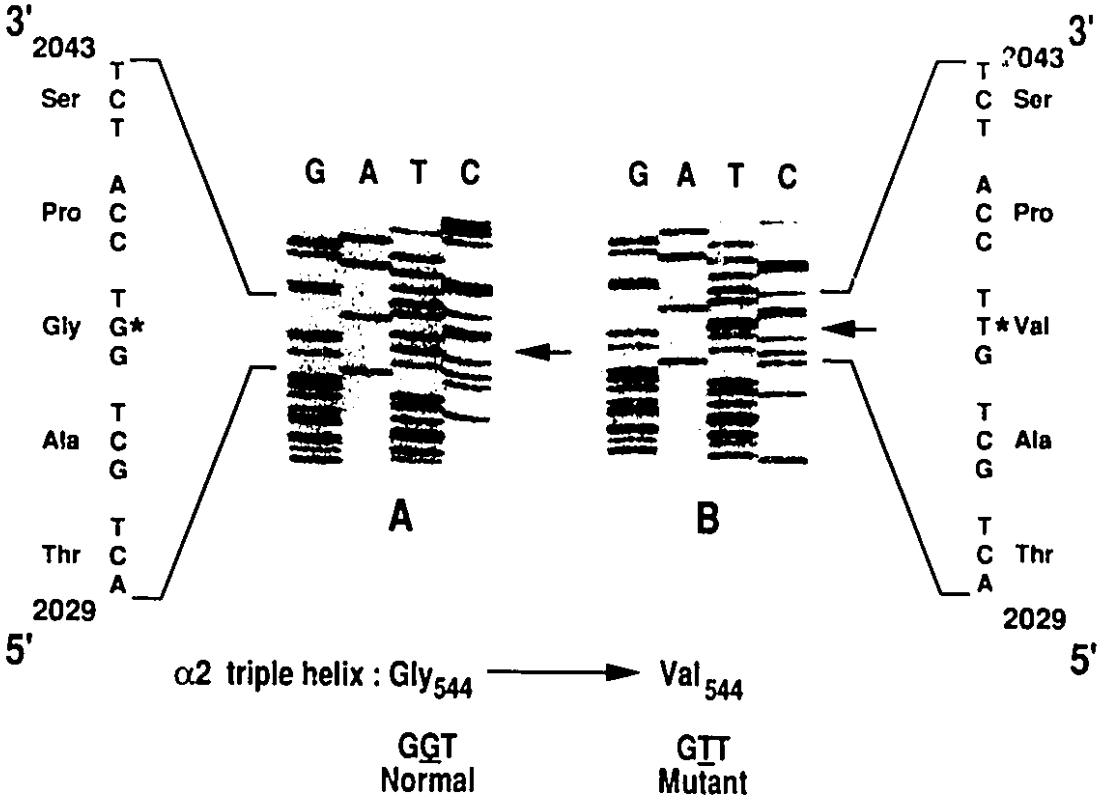


Fig. 41. Sequence analysis of the HA-reactive mismatch detected in OI cell line PP238. PP238 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-9 and COLL-10 to amplify a region of the prepro-α2(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the normal (A) and mutant (B) alleles differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (80). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α2(I) triple helical domain.

region of the α1(I) or α2(I) triple helical domain.

Sequencing of cDNAs cloned from JN245 fibroblasts identified two alleles differing in sequence at a single position (Fig. 42). One allele, identical to the published cDNA sequence, had a G at position 3169 (within exon 46), as the first base of the GGT triplet coding for glycine residue 1012 of the prepro α2(I) chain, or 922 of the α2(I) triple helix. However, the other allele had an A at this position, thereby converting the glycine codon to an AGT triplet coding for a serine residue. This G → A single base pair substitution results in a HA-reactive A:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and was observed within 30 bp of the location predicted by chemical cleavage analysis. This sequence result is also compatible with the cyanogen bromide mapping data, which indicated that the defect in this patient was located near the carboxy terminus of the triple helical domain of either the α1(I) or α2(I) chain. The substitution, in fact, is located only 92 amino acid residues from the end of the triple helical domain.

Sequencing of cDNAs cloned from NC323 fibroblasts identified two alleles differing in sequence at positions 1581 and 1684 of the prepro-α2(I) cDNA sequence. At position 1684 (within exon 26), one of the alleles, exhibiting the published cDNA sequence, had a G as the first base of the GGT triplet coding for glycine residue 517 of the prepro-α2(I) chain, or residue 427 of the α2(I) triple helical domain (Fig. 43). However, the other allele had a C at this position, thereby converting the glycine codon to a CGT triplet coding for an arginine residue. This G → C single base substitution results in a HA-reactive C:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and is located within 20 bp of the site predicted by chemical cleavage analysis. This sequence change is also consistent with the cyanogen bromide mapping data which suggested the presence of a defect in this patient within the central region of either the α1(I) or α2(I) triple helical domain. At position 1581 (within exon 25), one allele had a C as the wobble position of the CCC triplet coding for proline residue 482 of the prepro-α2(I) chain, or residue 392 of the α2(I) triple helical

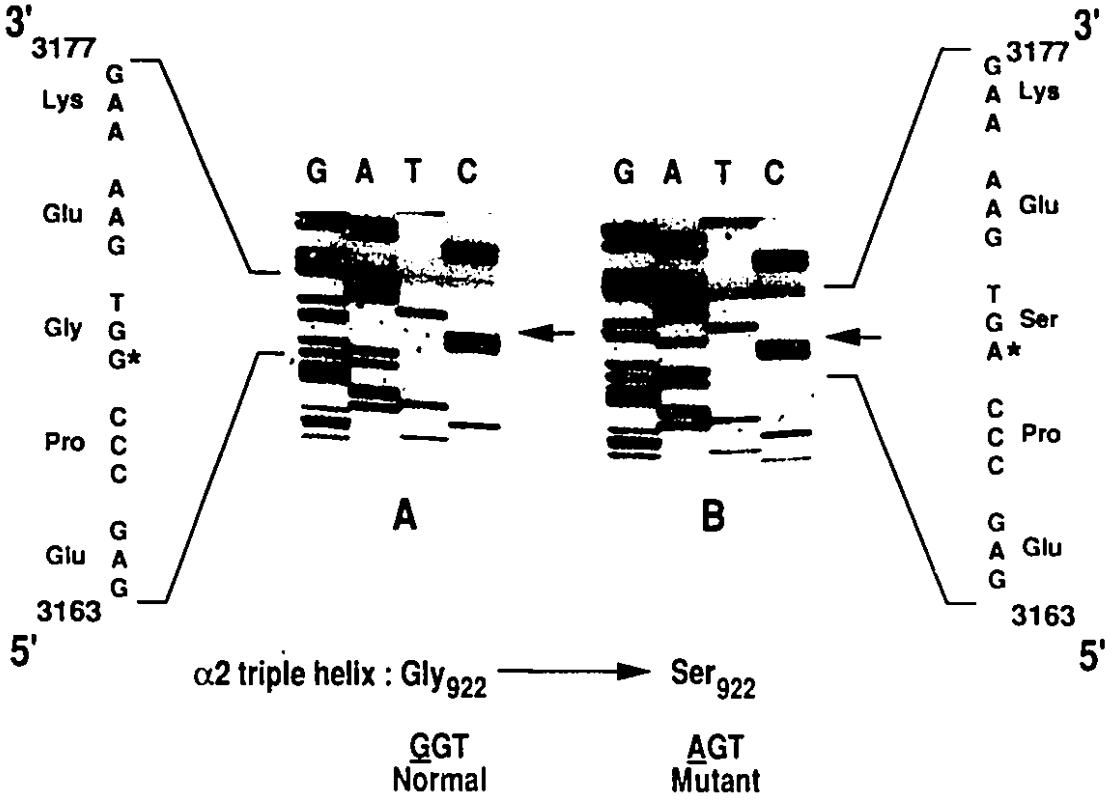


Fig. 42. Sequence analysis of a HA-reactive mismatch detected in OI cell line JN245. JN245 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-11 and COLL-12 to amplify a region of the prepro-α2(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the normal (A) and mutant (B) alleles differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (80). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α2(I) triple helical domain.

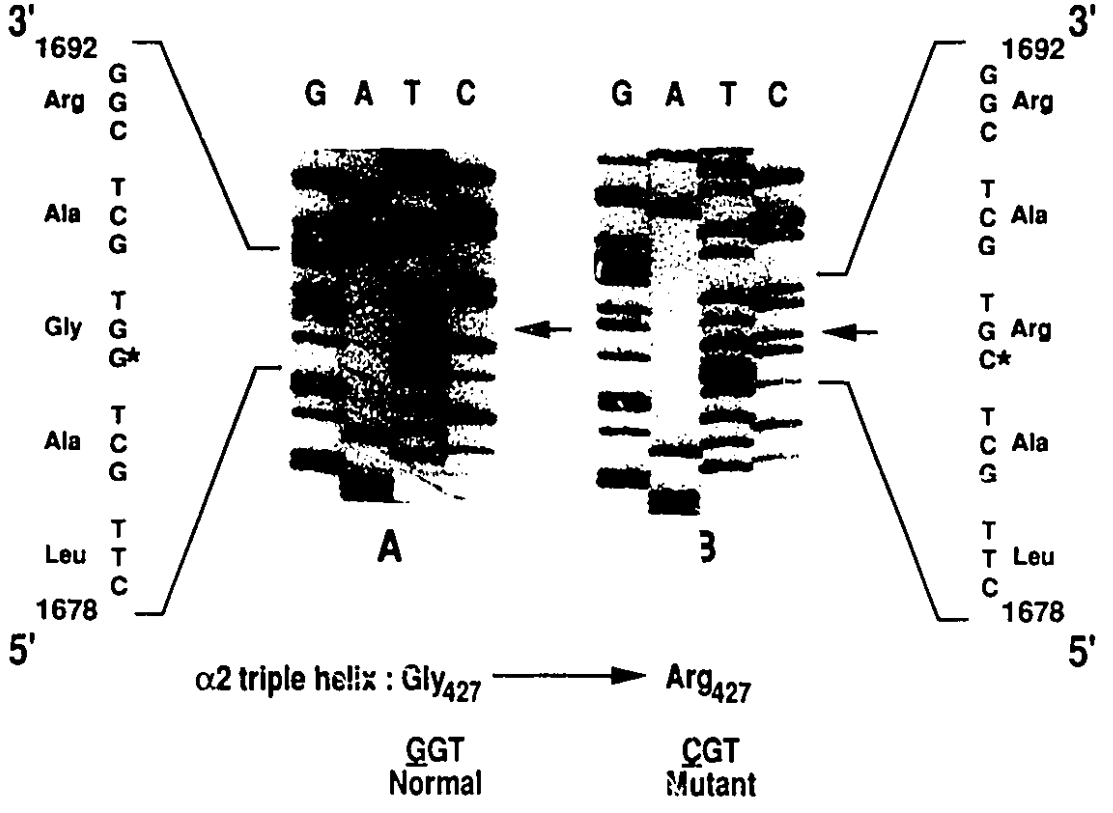


Fig. 43. Sequence analysis of the HA-reactive mismatch detected in OI cell line NC323. NC323 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-1 and COLL-2 to amplify a region of the prepro-α2(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the normal (A) and mutant (B) alleles differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (80). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α2(I) triple helical domain.

domain (data not shown). The other allele had an A at this position resulting in a CCA codon also encoding a proline residue. This corresponds to the OT-reactive PvuII RFLP (351) demonstrated in ML226 and BD268 fibroblasts (described in section 3.3).

Sequencing of cDNAs cloned from SS333 fibroblasts similarly identified two alleles differing in sequence at a single position (Fig. 44). One allele, corresponding to the published cDNA sequence, had a G at position 1284 (exon 18), as the first base of the GGT triplet coding for glycine residue 389 of the prepro-α1(I) chain, or residue 211 of the α1(I) triple helix. However, the other allele had an C at this position, thereby converting the glycine codon to a CGT triplet coding for an arginine residue. This G → C single base pair substitution results in a HA-reactive C:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and was observed within 10 bp of the location predicted by chemical cleavage analysis. This sequence result is also compatible with the cyanogen bromide mapping data, which indicated that the defect in this patient was located near the amino terminus of the triple helical domain of either the α1(I) or α2(I) chain.

The HA-reactive mismatch detected in RNA from MK345 fibroblasts was also identified by cDNA sequence analysis. Since chemical cleavage analysis indicated that the sequence change was located at position 295 or 296, approximately 4.5 kbp 5' of the poly-adenylate tail, it was necessary to synthesize first strand cDNA using random hexamer primers rather than oligo-dT, since previous attempts to amplify cDNA regions more than approximately 3 kbp 5' of the poly-adenylate tail were unsuccessful. PCR amplification was performed using the COLL-15/COLL-16 primer pair, with subsequent purification of the PCR product from a 2.0% agarose gel. The PCR product was cloned into M13mp18 and analyzed by DNA sequencing. Sequence analysis identified two alleles differing in sequence at a single position (Fig. 45). One allele, corresponding to the published cDNA sequence, had a G at position 296 (exon 2), as the wobble position of the CGG triplet coding for arginine residue 59 of the prepro-α1(I) polypeptide chain. The other

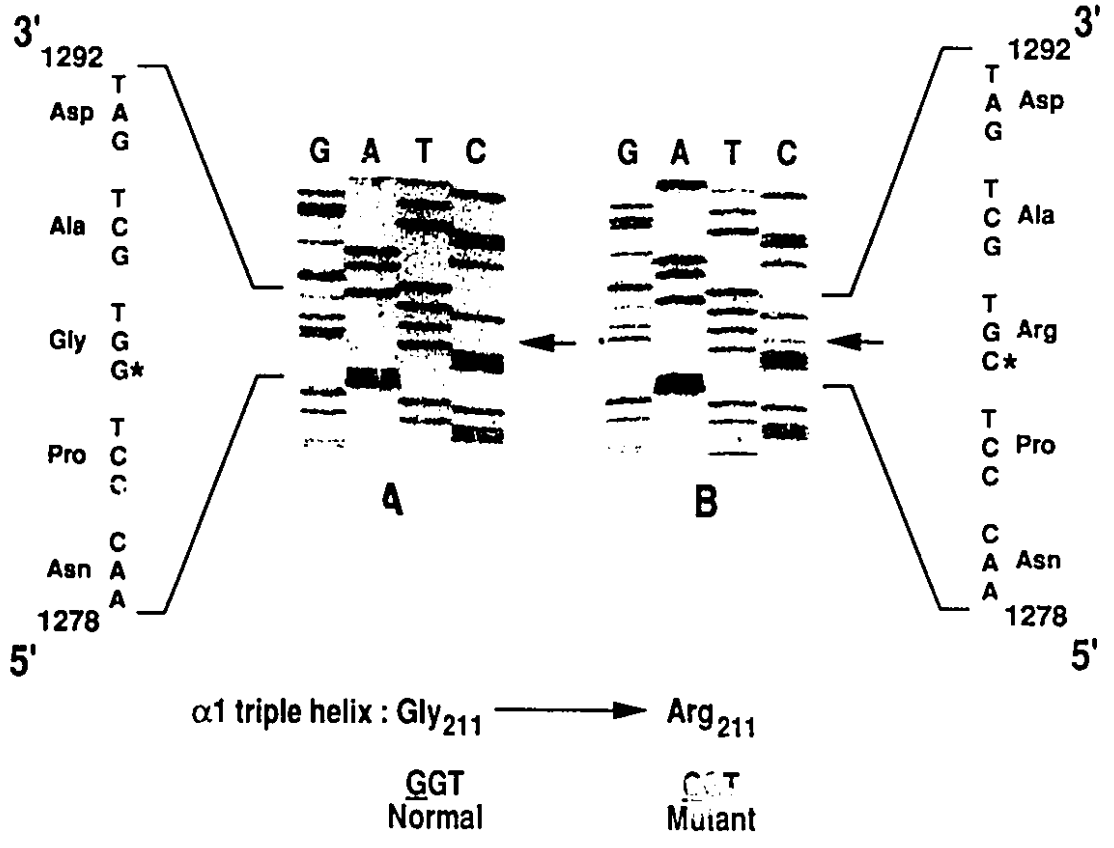


Fig. 44. Sequence analysis of the HA-reactive mismatch detected in OI cell line SS333. SS333 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-7 and COLL-8 to amplify a region of the prepro-α1(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the normal (A) and mutant (B) alleles differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (78). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the α1(I) triple helical domain.

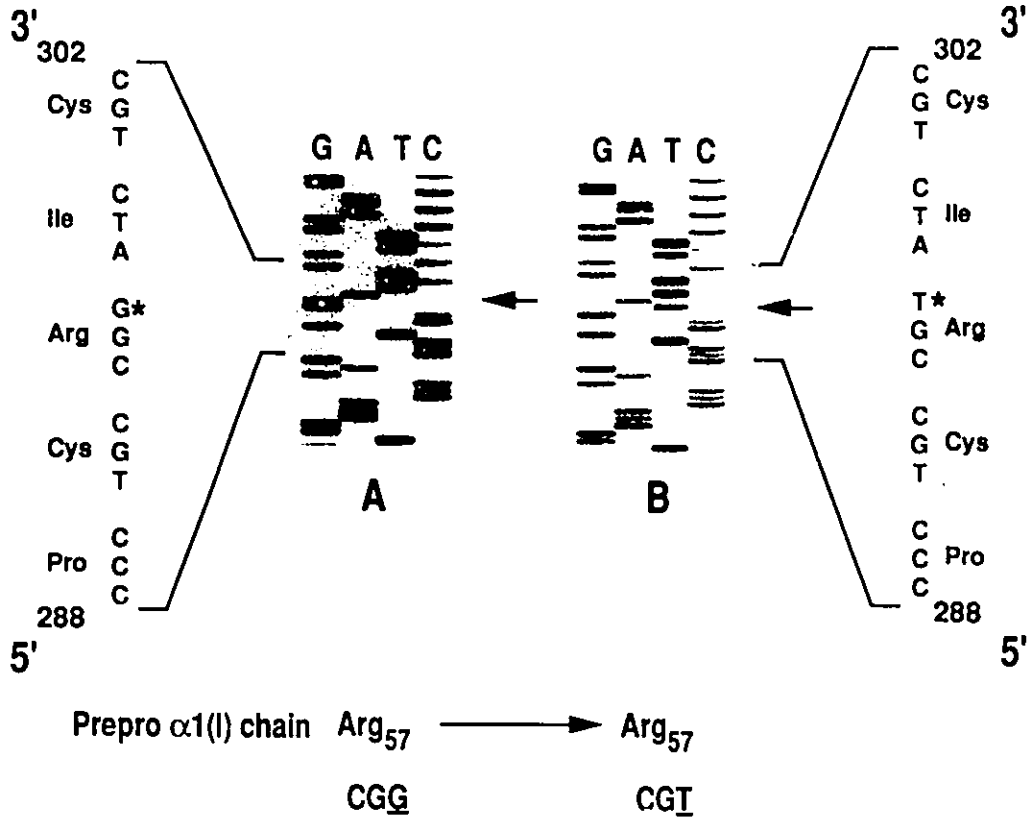


Fig. 45. Sequence analysis of the HA-reactive mismatch detected in OI cell line MK345. MK345 total fibroblast RNA was utilized to synthesize first strand cDNA. The PCR was then performed with primers COLL-15 and COLL-16 to amplify a region of the prepro-α1(I) cDNA containing the HA-reactive sequence change. The PCR products were cloned into M13mp18 and analyzed by DNA sequencing. The arrows and asterisks indicate the position at which the two sequenced alleles (A and B) differ. The numbers above and below the partial DNA sequences are cDNA nucleotide positions relative to the transcription start site (78). The effect of the sequence change on the polypeptide sequence is shown below the sequencing gels, with the affected base underlined. Amino acid residue numbers refer to the position within the prepro-α1(I) chain.

sequence had a T at this position, thereby converting the CGG codon to a CGT triplet, although still encoding an arginine residue. The G → T single base substitution results in a HA-reactive T:C base pair mismatch in the mRNA:cDNA probe heteroduplex, and is located within 5 bp of the site predicted by chemical cleavage analysis. Arginine residue 59 is located within the globular subdomain of the amino-terminal propeptide, and is only 37 amino acid residues carboxy-terminal of the signal peptidase cleavage site. This sequence change represents a polymorphism, since it involves a codon wobble position which does not alter the primary structure of the prepro-α1(I) polypeptide chain. The novel sequence variation results in an XhoII (BstYI) RFLP which may be a useful marker for the COL1A1 gene.

In summary, the sequence data demonstrate that four of the OI patients in this study (PP238, JN245, NC323 and SS333) possess mutations giving rise to four different glycine substitutions within the triple helical domain of the α1(I) or α2(I) chain. The representations of the sequenced alleles among the clones analyzed indicated that these patients are heterozygous for the detected sequence changes. The resulting substitutions likely give rise to the delayed formation of the type I procollagen triple helix, allowing the excessive post-translational modifications which were detected by altered electrophoretic mobility. The positions of the glycine substitutions within triple helical domain were consistent with the degrees of collagen chain overmodification observed for these cell lines. JN245, with the most carboxy-terminal substitution, exhibited the most extensive overmodification, while SS333, with the most amino-terminal substitution, exhibited the least extensive overmodification. Cell lines NC323 and PP238, with substitutions within the central region of the triple helix, displayed intermediate levels of chain overmodification. The sequence data also demonstrated that the mismatch detected in OI cell line MK345 results from a sequence polymorphism which was not observed in any of the other cell lines. This indicates that the polymorphism is rare within the caucasian

population, or that it is the product of a sporadic mutation in this individual.

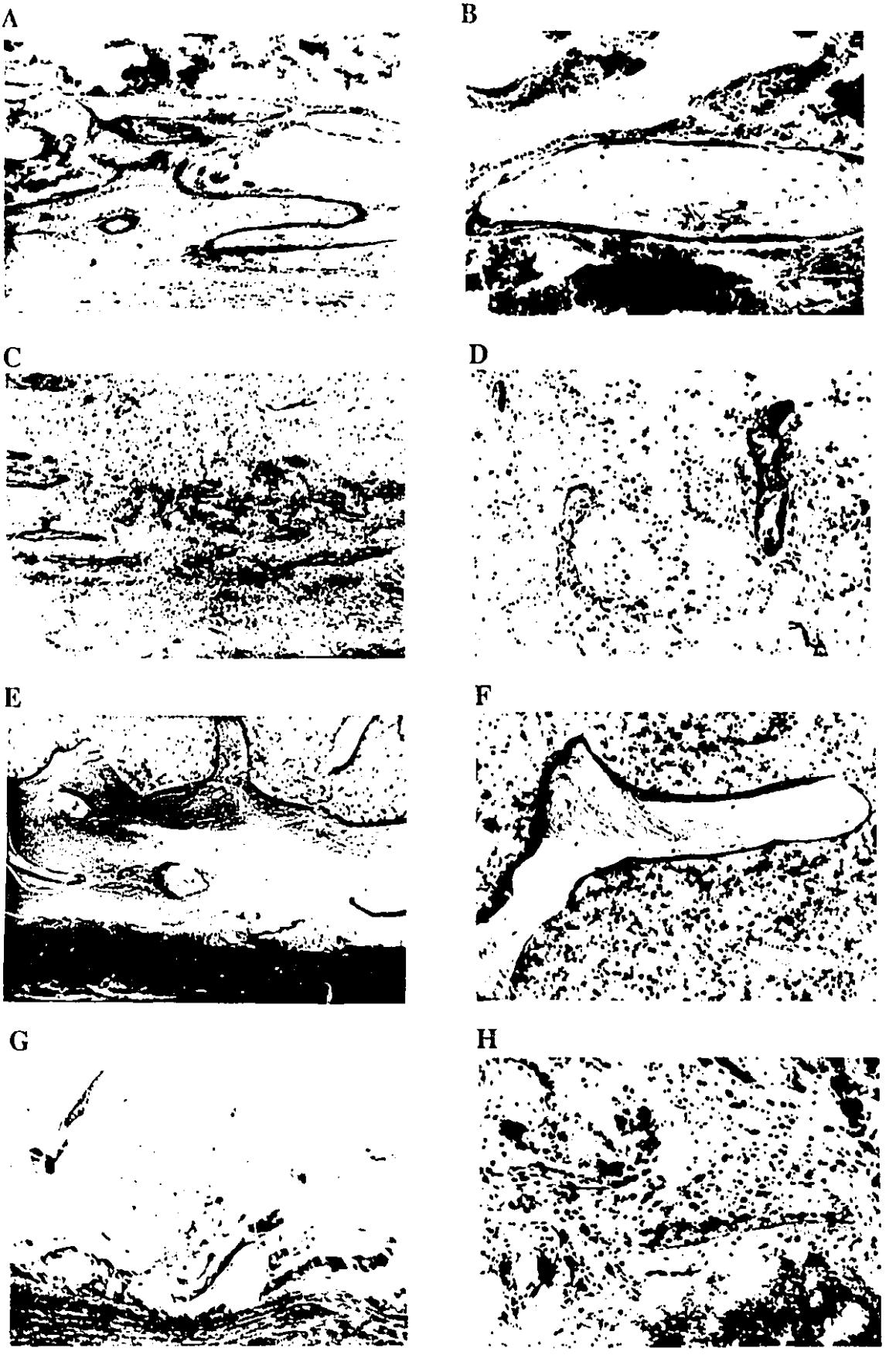
3.5 Analysis of Bone Architectural Changes in OI

Although the relationship between phenotypic severity and specific OI mutations is currently being thoroughly examined, much less work has been done to explore their differential effects at the tissue level. It is presently unclear whether the extensive heterogeneity in the OI phenotypes and in the underlying molecular defects is reflected by a similar heterogeneity at the level of bone architecture. Therefore, in order to assess the effects of the mutations in patients NC323 (type III OI) and SS333 (type II OI) on bone architecture, biopsies were collected from the iliac crest or femur and compared to age- and site-matched controls.

Bone from a normal 24-week fetus and from a normal two year-old exhibited well-organized cortical and trabecular bone structures (Fig. 46 A, B, E, F). However, there were significant age-related differences between the two controls. Bone from the two year-old control contained a thicker and more uniform cortex than bone from the 24 week-old fetus. In addition, bone from the two year-old control contained a decreased amount of cartilaginous matrix within the cores of mineralized cortical and cancellous bone structures, and a more extensive lamellar organization (particularly within the cortex). These features are indicators of the more advanced stage of maturation of the bone from the two year-old control.

Examination of the bone biopsies obtained from the two OI patients demonstrated striking abnormalities in the quantity and organization of mineralized bone, relative to their respective age-matched controls (Fig. 46 C, D, G, H). The severe osteopenia in these patients was reflected by the paucity of mineralized bone structures, particularly in the

Fig. 46. Histological analysis of bone biopsies from OI patients NC323 and SS333. Femoral bone from a 24-week fetal control (A, B) or from patient SS333 (C, D), and iliac crest bone from a 2 year-old control (E, F) or from patient NC323 (G, H) were cut into 5-μm sections and stained with Goldner's stain. The fields photographed are of areas of cortical (A, C, E, G; 18X magnification) or cancellous (B, D, F, H; 45X magnification) bone. The photographs of the cortical regions are oriented with the periosteal surfaces towards the bottom. Cartilaginous structures appear white, surrounded by mineralized bone with a greyish appearance. Osteoid surfaces appear as a dark periphery around the mineralized bone structures.


A
B
C
D
E
F
G
H

cortical regions which were very thin and discontinuous. Furthermore, the structures which were observed exhibited osteoid surfaces which were very thin compared to their normal counterparts, suggesting an impairment in the rate at which new bone matrix is synthesized. In addition, the lamellar organization observed for the normal controls, particularly in the 2 year-old, was largely absent in the two OI patients, further suggesting an abnormality in the process by which woven bone is replaced with mature lamellar bone. These results indicate that the abnormalities in type I collagen structure in these patients give rise to severe deficiencies in bone formation and organization.

The significant age differences between these OI patients prevented a direct comparison of their bone biopsies. However, the abnormalities in bone architecture displayed by these patients, relative to their controls, were similar in that they consisted of quantitative and organizational components. It is interesting to note that two patients with different phenotypic severities, exhibit similar changes in bone architecture. Thus, in the case of these two patients, the differences in OI phenotype do not appear to be reflected at the tissue level.

Chapter 4

Discussion

4.1 Evaluation of the Protein Methods for the Detection of Collagen Defects

A major aim of this thesis was to develop a strategy for the detection and characterization of type I collagen defects in OI. The strategy which was used included two electrophoretic techniques intended to determine whether collagen synthesized by a group of OI cell lines exhibited signs of post-translational overmodification, relative to collagen from normal cell lines. This feature is common to most OI cell lines, and is the result of structural defects affecting triple helix folding and stability.

One of the techniques utilized was a one-dimensional SDS-PAGE analysis of α1(I) and α2(I) chains produced by pepsin digestion of secreted type I procollagen. The second technique was a two-dimensional electrophoretic analysis in which α1(I) and α2(I) chains were separated by first dimension electrophoresis, and subsequently cleaved with cyanogen bromide within the gel. The resulting peptides were then separated in the second dimension electrophoresis. Although these two techniques have definite advantages in their abilities to detect collagen defects and have been used in numerous laboratories investigating OI, they also possess disadvantages.

Justification for the Use of Skin Fibroblasts in the Study of OI

Since the major tissue affected in OI is bone, the ideal cell type to examine for collagen defects is the osteoblast. However, since bone biopsies are much less readily available for study than are skin biopsies, the more commonly utilized cell type is the skin fibroblast. Although, collagen biosynthesis may not be identical in osteoblasts and skin fibroblasts, it is likely that type I collagen mutations would have similar consequences on processes such as post-translational modification, degradation and secretion in these two cell types, although the degree to which these events occur may vary. The description of somatic

mosaicism in a number of cases of OI would suggest the possibility of having a collagen mutation expressed in a proportion of osteoblasts without significant involvement of skin fibroblasts. However, both of these cell types are of mesenchymal origin, and in order to have representation of the mutation in a large enough percentage of osteoblasts to give rise to bone fragility, the mutation would be expected to occur early in embryonal development. Therefore, mosaic individuals in which the mutation is expressed in the osteoblast cell lineage, should also express the mutation in their skin fibroblasts, at least to some extent.

Evaluation of the One-Dimensional Electrophoretic Analysis

The advantages of this method are primarily related to the rapidity and ease with which it can be performed. Once fibroblast cell lines have been established in culture, the cell labelling and electrophoretic analysis can usually be performed within a period of approximately ten days. In addition, the simplicity of the technique allows the simultaneous processing of many samples.

A disadvantage of this technique is that there appears to be some variability in the degree of collagen overmodification observed for an individual cell line between different experiments. Although the precise reasons for these variations are unclear, they may have arisen from slight changes in the culture conditions during cell labelling. These variations in collagen overmodification may give rise to false positives and negatives during the analysis of OI cell lines.

The possibility of significant variability in the extent of collagen post-translational modification between different normal fibroblast cell lines also exists. Little is known concerning how the extent of post-translational modification may be affected by variations in cell passage number, age of the individual at time of biopsy and the specific site of skin biopsy. The cell line to be used as a normal control must be carefully chosen, since it is on comparison with collagen from this cell line that the assessment of post-translational

overmodification will be based. Several potential controls should be compared, and a cell line yielding a representative degree of post-translational modification should be chosen. The improper selection of a control cell line may also lead to false negatives and positives in the course of analyzing OI cell lines.

Another drawback of this technique is the possibility of a lack of sufficient sensitivity to detect certain collagen defects. For example, defects near the amino terminus of the triple helix, often associated with cases of mild OI, may go undetected due to the presence of minimal extents of overmodification. In addition, collagen defects which result in extensive intracellular degradation may elude detection, since this would result in the presence of only small quantities of the mutant chains in the culture medium.

The amount of information which this method can provide about the positions of detected defects is very limited. Since a defect in either the α1(I) or α2(I) chain results in the overmodification of both of them, this method cannot determine the affected chain. In addition, despite a general correlation between the extent of collagen overmodification and the proximity of the defect to the carboxy terminus of the triple helical domain, the accuracy with which defect position can be predicted is insufficient to permit immediate cDNA sequence analysis. The possibility of obtaining such positional information is further hampered by the variability in collagen modification between experiments, as discussed above.

It has been clearly established that structural mutations interfering with triple helix folding result in increased levels of collagen post-translational modification. However, it is unclear whether the relative levels and activities of the modification enzymes themselves may also affect the degree of modification. Since the rate of helix folding is dependent on the extent of proline 4-hydroxylation, it is feasible that changes in the activities of peptidyl proline 4-hydroxylase, relative to those of peptidyl lysine hydroxylase and the glycosyl transferases, could significantly affect the extent of chain modification. Thus, factors

resulting in fluctuations in the activities of these enzymes, relative to each other, may give rise to false positives and negatives.

In summary, the primary use for this technique is in the rapid screening of large groups of OI cell lines for collagen overmodification, indicative of helix structural defects. However, negative findings do not exclude the possibility of the presence of collagen defects resulting in minimal levels of overmodification or in extensive intracellular degradation.

Evaluation of the Two-Dimensional Electrophoretic Analysis

An advantage of this technique, relative to the one-dimensional electrophoretic analysis, is its increased sensitivity for detecting defects near the amino terminus of the triple helical domain of the α1(I) or α2(I) chain, since the increase in molecular weight associated with overmodification is more significant for cyanogen bromide peptides than for whole α chains. In addition, the distribution of the overmodification among the major α1(I) CB peptides provides information about the positions of the defects within the triple helical domain of either the α1(I) or α2(I) chain.

This method is more time-consuming and labor-intensive than the one-dimensional analysis, such that the number of samples that can be simultaneously handled is limited. In addition, a certain amount of experience is required for the analysis of the peptide maps, since these can sometimes be difficult to interpret.

Some of the other disadvantages of this method in detecting collagen defects, are identical to those discussed for the one-dimensional technique, such as the possibility of variability in the extent of collagen modifications, and the choice of controls. Also, this method cannot distinguish between defects in the α1(I) and α2(I) chains.

Despite the increased sensitivity of this method for detecting minimal extents of overmodification, the possibility of failing to detect defects near the amino terminus of the helix exist, for two reasons. Firstly, the four major α1(I) CB peptides which are analyzed

by this method, cover all but the amino-terminal 123 amino acids of the triple helical domain. This region consists of peptides CB2, CB4 and CB5, which individually are too small to be observed by this gel system. In addition, since the region covered by these peptides is the last to fold into a triple helix, these peptides are likely always completely modified. Thus, the modification status of these peptides, even if it could be analyzed by electrophoresis, is likely refractory to any delays in the formation of the triple helix due to carboxy-terminal defects. Secondly, defects occurring near the amino terminus of the α1(I) CB8 peptide, the most amino-terminal of the major CB peptides, or the corresponding region of the α2(I) chain, may result in imperceptible changes in its extent of modification. Thus, defects in the region of approximately 150 amino acids at the amino terminus of the α1(I) chain, or the corresponding region of the α2(I) chain, are likely undetectable.

Although this method is able to provide some information on defect position within the triple helical domain, the precision of the localization is limited by several factors. Firstly, chain overmodification may in some cases occur not only on the amino-terminal side of the defect, but also to some extent on the carboxy-terminal side. This may be the result of the destabilization of the helix on both sides of the defect, due to the introduction of a bulky amino acid. Secondly, the large size of the α1(I) CB peptides may only allow the region of interest, in each chain, to be narrowed down to between approximately 200 and 300 amino acid residues. This would still require the cloning and sequencing of 500 to 1000 base pairs from each of prepro-α1(I) and prepro-α2(I) cDNAs. Thirdly, the α1(I) CB peptides may not all be equally sensitive to changes in modification, due to differences in the number of unhydroxylated and unglycosylated lysine residues present. Therefore, some CB peptides may be able to undergo extensive changes in electrophoretic mobility, allowing distinctions to be made between complete and partial peptide overmodification, while others may only exhibit modest changes.

In summary, this method can be utilized to analyze a group of OI cell lines exhibiting collagen overmodification, in order to determine which of these should be further pursued by chemical cleavage mismatch analysis. It can also be used as a more sensitive assay for cell lines in which no collagen overmodification could be detected by one-dimensional electrophoresis. However, similarly to the one-dimensional technique, the absence of detectable collagen overmodification does not exclude the possibility of a defect near the amino terminus of the α1(I) or α2(I) triple helical domain.

The information provided by this technique about defect position is not sufficiently precise to allow a focused strategy for sequencing the prepro-α1(I) and prepro-α2(I) cDNA regions affected. However, the data is useful in determining the two cDNA probes to utilize in chemical cleavage analysis (ie. one for each chain), thereby saving time and work by obviating the need to cover all six of them. Knowledge of the pattern of overmodification of the α chains also provides supporting evidence for a role of the mutation, identified by subsequent DNA sequence analysis, in the collagen abnormality in the cell line being studied.

Evaluation of HPLC Tryptic Peptide Mapping

HPLC tryptic peptide mapping is another approach which was evaluated for its ability to detect and identify collagen defects in OI. This approach was based on the principle that defects as subtle as single amino acid substitutions would be able to alter the chromatographic profile of trypsin-digested collagen isolated from fibroblast cultures. Any tryptic peptides with altered retention times, on reversed-phase HPLC, could then be subjected to automated peptide microsequencing for the precise identification of the underlying structural change. In theory, only those tryptic peptides with altered structure (sequence, length, hydroxylation, glycosylation) should exhibit changes in chromatographic behavior. Thus, by comparing the tryptic maps of collagen from an OI

cell line to those of collagen from a control cell line, it should be possible to identify those peptides with such structural changes.

HPLC tryptic mapping was performed for collagen obtained from OI cell line SL211 and from control cell line YL278. Initial experiments were performed using whole α1(I) and α2(I) chains which had been purified by HPLC. The resulting tryptic peptide profiles were found to be very reproducible, in terms of the positions and relative heights of the peaks. However, the profiles were very complex, consisting of approximately 75 to 100 peaks. In addition, the chromatography of individual peaks using another buffer system often revealed the presence of several distinct peptides. Thus, the number of peptides produced was considerably greater than the maximum of 80 to 90 which could be expected from each α chain, based on the number of lysine and arginine residues per chain (77-80). The unexpectedly large number of tryptic peptides in the profiles indicated that the proteolytic digestions had not gone to completion. This complexity did not allow any differences to be perceived between the profiles obtained for cell lines SL211 and YL278.

In order to simplify the analysis of the tryptic peptide profiles, subsequent experiments were performed using individual α1(I) CB peptides (CB8, CB3, CB7 and CB6), which had been purified by reversed-phase HPLC. Preliminary experiments with these α1(I) CB peptides from cell line YL278 resulted in simpler profiles containing many fewer peaks than those derived fron whole α chains.

Despite the significant improvements in the resolution of the tryptic peptide profiles, by the use of purified CB peptides, it was felt that this approach was too time-consuming and labor-intensive to be useful in the study of collagen defects. The amount of time required to produce the α1(I) CB peptide profiles for one cell line was between approximately 3 and 6 months, including cell culture, medium collection, and HPLC procedures. This is in contrast to the period of approximately 1 month required to identify the collagen defects in four OI cell lines using the electrophoretic techniques discussed previously, in conjunction

with the molecular biology techniques discussed later. It is nevertheless possible, given a more rapid and discriminating detection technique (eg. on-line mass spectrometry) for identifying the peptides in the α chain profile peaks, that this approach could be of considerable value. However, this strategy has not yet been applied to the detection and identification of collagen defects in OI.

4.2 Evaluation of the Chemical Cleavage Technique for the Detection of Mutations

Advantages of the Chemical Cleavage Technique

The chemical cleavage method was an important component of the approach used in this study, for the detection and identification of OI mutations, and possesses several advantageous technical features. Firstly, the technique is generally applicable since it can be applied to the analysis of any mRNA (or DNA) species, for which probes are available. However, since each mRNA (or DNA) has different characteristics, certain experimental parameters must be adapted in order to analyze them. This is discussed later.

Another advantage of this method is that it appears to be able to detect almost all point mutations, which may account for a large proportion of gene defects. This is particularly important in the study of OI, in which many point mutations resulting in glycine substitutions have been described. The analysis of mRNA:cDNA probe heteroduplexes with HA and OT allows the detection of approximately 50% of point mutations. The remaining point mutations can be detected by analyzing cDNA:DNA probe heteroduplexes, with these same reagents. First strand cDNA is synthesized by the reverse transcription of fibroblast mRNA. The PCR is then utilized to amplify the cDNA region to be analyzed by chemical cleavage, and to selectively label the noncoding strand. This cDNA strand is then hybridized to the coding strand of the probe. In theory, the analysis of RNA:DNA and

DNA:DNA heteroduplexes as described should allow the detection of all point mutations. However, Theophilus et al (354) report that this technique is unable to detect certain point mutations, which are otherwise detected by other methods. The exact proportion of point mutations which remain undetected by this technique is yet to be established.

The chemical cleavage method is advantageous, since it also permits the analysis of large regions of mRNA (or cDNA) at a time, since large probes can be used. This study utilized probes ranging from 1.4 to 2.8 kb in size. This feature is important for the analysis of the type I collagen mRNAs, which together contain over 8 kb of coding sequence. The use of smaller probes would necessitate the preparation of a larger number of probes, and the performance of more chemical cleavage experiments.

The method provides information on the positions of detected mutations within the region analyzed. This allows a focused approach for the identification of the mutation by subsequent cloning and sequencing. This is critical in the analysis of the type I collagen mRNAs, due to the large size of the probes used. If the technique did not provide positional information, large regions of cDNA would have to be cloned and sequenced. In fact, the accuracy of the positional information, ± 20 bp even for cleavage products over 1 kb in size, only requires the sequencing of a cDNA region of approximately 300 bp.

The chemical cleavage method appears to be especially suited for the detection of OI mutations, since virtually all of those described to date affect G bases within the coding sequence. These include point mutations affecting glycine codons (GGX), deletions and insertions, which result in mismatched C bases in mRNA:cDNA hybrids, either by base mispairing or by the looping out of one strand in the hybrid. Thus, the analysis of only one possible base change, using HA treatment of RNA:DNA heteroduplexes, is able to detect virtually all OI mutations. However, the exclusive analysis of G-base changes only, imposes a bias on the types of OI mutations detected. Therefore, novel types of mutations, such as those affecting X- or Y-position residues, may remain undetected.

Limitations of the Chemical Cleavage Technique

The major limitation of the technique is whether the quantity of RNA required to perform all of the analyses is available. This requisite amount of RNA, in turn, depends on the abundance of the messages of interest in the preparation of total RNA, and on the number of cDNA probes to be used in annealing.

Since the successful detection of mismatches requires satisfactory protection of the cDNA probe, messages present in low abundance may require the use of such large amounts of total RNA that the ribosomal RNA may interfere with the annealing process. In cases where the source of RNA is not limiting (eg. cultured cell lines), this problem can be overcome by the isolation of poly-A^+ RNA, in order to enrich the preparation in the mRNA of interest. In cases where the source of RNA is limited (eg. small tissue samples), it may not be possible to isolate sufficient poly-A^+ RNA to perform chemical cleavage analyses with all the desired probes. In these cases, the chemical cleavage analysis may be performed using DNA:DNA heteroduplexes, as described earlier. However, the analyses should be performed using PCR products from duplicate reactions, to distinguish between legitimate mismatches and artifactual ones resulting from Taq DNA polymerase replication errors.

The amount of RNA required also depends on the number of cDNA probes to be used. Thus, the analysis of long mRNAs, necessitating a large number of probes, requires the availability of greater quantities of total RNA. This problem may be alleviated by the design of probes which are greater than 3 kbp in length. However, this is done at the expense of the accuracy with which the positions of detected mismatches are determined.

The limitations described above are not problematic in the study of OI mutations, since the messages encoding the type I collagen chains are expressed at relatively high levels in skin fibroblasts. In addition, there is usually no problem in growing up the number of cells needed to provide the requisite amount of RNA.

Another limitation is that mismatches located near the ends of the probe strand may be difficult to visualize. Mismatches located near the unlabelled 5' end of the labelled probe strand result in small changes in length, and may therefore result in imperceptible changes in electrophoretic mobility. In contrast, those located within approximately 100 bases of the labelled 3' end may be obscured by the background resulting from random probe cleavage. Since large cleavage products are further cleaved into smaller ones, the background intensity generally increases with proximity to the bottom of the gel, such that it is not possible to clearly identify mismatches resulting in products of less than 100 bases in size. These problems can be overcome by designing probes which overlap by at least 200 bp.

Comparison With Other Mutation Detection Methods

In recent years, many new mutation detection methods have been described. Thus, before a specific choice is made, it is critical to outline the criteria which the technique must satisfy. In selecting a method for the detection of gene mutations, such as in OI, four criteria should be met. Most importantly, the method should be capable of scanning cDNA or mRNA sequences, such that there is no requirement for prior knowledge of the nature and position of the mutation. A number of techniques satisfying this criterion, and their abilities to satisfy the following criteria, are listed in Table 15. Additional criteria for evaluating these methods are: 1) the ability to detect all or most point mutations, since these may account for a large proportion of the gene defects in various disorders; 2) the ability to provide information on the position of detected mutations, within the cDNA or mRNA region scanned; and 3) the ability to scan large regions of mRNA (or cDNA) sequence at a time.

Table 15. Features of scanning mutation detection techniques.

Detection Technique	Most Point Mutations Detected?	Positional Data Provided? [1]	Size of Scan Region [2]	References
Southern blotting/RFLP	No	Rarely	Large	308-9
Northern blotting	No	No	Large	355
Polymerase chain reaction	No	No	Small	302-3,356
Solution melting assay	No	No	Small	357
Denaturing gradient gel electrophoresis	Yes	No	Small	310,312 315,358
Single-strand conformation polymorphism analysis	Yes	No	Small	359-60
Carbodiimide/gel mobility	Yes	No	Small	361
Nucleotide analog incorporation	Yes	No	Small	362
DNA sequencing	Yes	Yes	Small [3]	345,363
Carbodiimide/ABC nuclease	Yes	Yes	Small	321
Enzymatic heteroduplex cleavage	Yes [4]	Yes	Large	316-20
Carbodiimide/immunomicroscopy	Yes	Yes [5]	Large	364
Chemical heteroduplex cleavage	Yes	Yes	Large	322-3

[1] This column indicates whether the techniques can provide information on the positions of detected mutations, within the scan region.

[2] The term "small" refers to scan regions of approximately 1 kbp or less; "large" refers to those above this size.

[3] The limitations associated with the small scan region can be overcome, since the steps involved in this approach can be automated.

[4] The proportion of point mutations detected by this method has been estimated at between one and two-thirds (311, 365).

[5] The accuracy of the positional data provided by this method is limited to approximately 200 bp.

Based on these criteria, three of the techniques listed in Table 15 appear to be most applicable. These consist of the enzymatic or chemical cleavage analysis of heteroduplexes and the immunomicroscopic analysis of carbodiimide-modified heteroduplexes. However, two of these methods possess disadvantages which limit their application. The methods involving the enzymatic cleavage of RNA:DNA or RNA:RNA heteroduplexes are hampered by their ability to detect only one to two-thirds of point mutations (311, 365). Although the chemical cleavage method may also be unable to detect all point mutations, it nevertheless appears capable of detecting a larger proportion than the enzymatic cleavage methods. The carbodiimide/immunomicroscopic technique is limited by the accuracy with which mutations can be localized (± 200 bp), in comparison to the chemical cleavage analysis (±20 bp).

The use of DNA sequencing as a method for detecting and identifying OI mutations did not satisfy the specified criteria, due to the small size of the scan regions analyzed (ie. the large number of individual exons in the COL1A1 and COL1A2 genes). However, the procedures commonly utilized in this approach, namely the PCR amplification of genomic DNA and direct sequencing by PCR, can now largely be automated. Thus, under certain circumstances (ie. the availability of the required equipment and funding), this approach may constitute a competitive alternative to the chemical cleavage method for detecting OI mutations.

The chemical cleavage analysis therefore appears to be the method which most adequately satisfies the criteria for mutation detection, described above. However, it is important to remember that these criteria were tailored for the detection of OI mutations. Thus, the analysis of other genes may involve different criteria, which may be adequately satisfied by alternative detection methods.

Important Parameters of the Chemical Cleavage Technique

An important feature of the chemical cleavage method is its general applicability, as described above. This study has examined the effects of various experimental parameters on the detection of a specific mismatch in the prepro-α2(I) mRNA. In doing so, it has improved the general applicability of the chemical cleavage method, by identifying parameters which must be considered when adapting it to the analysis of other mRNAs.

Perhaps the most critical parameter to consider is the RNA:probe molar ratio utilized during the annealing period, which must be empirically determined for each specific mRNA and source (tissue or cell line). This study has demonstrated that variations in the RNA:probe molar ratio, achieved by fixing either component while varying the other, have significant effects on the intensity of the detected mismatch. Thus, when initially establishing the annealing conditions for the analysis of the mRNA of interest, the quantity of total or poly-A^+ RNA yielding satisfactory mismatch intensity should be determined. In addition, the quantity of probe used should be varied to determine the maximum amount which can be utilized without adversely affecting background cleavage. These preliminary experiments will establish conditions yielding satisfactory mismatch signals (relative to background intensity) and film exposure times.

Another parameter which may also be varied is the annealing temperature. Although the variation of this parameter was not examined in this study, it is possible that this could affect mismatch detection by resulting in an altered equilibrium between free and hybridized probe. When first establishing annealing conditions, the temperature can be estimated as described by Dean (337), and subsequent minor adjustments can be made to improve mismatch detection, if necessary.

The source of RNA used should be carefully chosen in order to ensure that the abundance of the message is sufficient to yield satisfactory probe protection, and that enough RNA can be isolated to perform all the required analyses. The relative abundance

of a specific mRNA in different tissues or cell lines should be considered, as well as the manner in which expression may be augmented by varying culture conditions prior to RNA isolation. For example, in this study, ascorbate-stimulated fibroblasts were utilized, since this treatment is known to increase type I collagen mRNA levels (325). The augmented levels resulted in a significant improvement in the intensity of the detected mismatch. Ascorbate stimulation also improved the yields of total RNA, allowing more chemical cleavage experiments to be performed from the same number of fibroblast cells.

The manner in which the probes are designed and labelled is also important to consider. Probes should be designed to overlap by approximately 200 bp, and should not exceed approximately 3 kbp in length, since this will adversely affect the accuracy with which the mutation position is determined. The restriction sites used to prepare the probes should be carefully selected so as to yield fragments whose noncoding strands have recessed 3' ends and blunt or recessed 5' ends. This allows the selective labelling of this strand at its 3' end by filling-in using Klenow fragment and a single specific ^{32}P-labelled dNTP. If an overhanging 5' end is the only feasible option, the restriction site used should create a recess whose first base is distinct from that at the 3' end of the noncoding strand. This will permit the selective labelling of the 3' end of the non-coding strand, which was shown to significantly improve mismatch detection by decreasing the intensity of background probe cleavage.

In summary, a careful consideration of these parameters will assist in establishing the chemical cleavage technique for the analysis of various mRNA (or cDNA) species. If these parameters are not considered, chemical cleavage experiments may give rise to unreliable mismatch detection.

4.3 Evaluation of the Cloning and Sequencing PCR Products

The PCR is a method which is commonly used to generate specific cDNA and genomic DNA sequences for procedures such as sequence analysis and *in vitro* transcription, or for use in an expression system. These procedures often require that the PCR product initially be cloned into a plasmid vector. However, the cloning of PCR products often proves problematic (348). One of the generally applicable approaches utilized is blunt-end cloning. An aim of this thesis was to examine the importance of various parameters, relating to the preparation of PCR product insert DNA on the efficiency of blunt-end cloning into the M13 mp18 sequencing vector. The various methods of insert DNA preparation differed by: a) whether Taq DNA polymerase was extracted immediately following amplification, b) whether blunt-ending was performed, c) the choice of blunt-ending enzyme between T4 DNA polymerase and Klenow fragment, and d) the order in which blunt-ending and gel purification were performed. The results demonstrated that in order to maximize cloning efficiencies, PCR amplification should be followed by phenol:chloroform extraction of Taq DNA polymerase, blunt-ending with Klenow fragment, and subsequently by DEAE-cellulose gel purification.

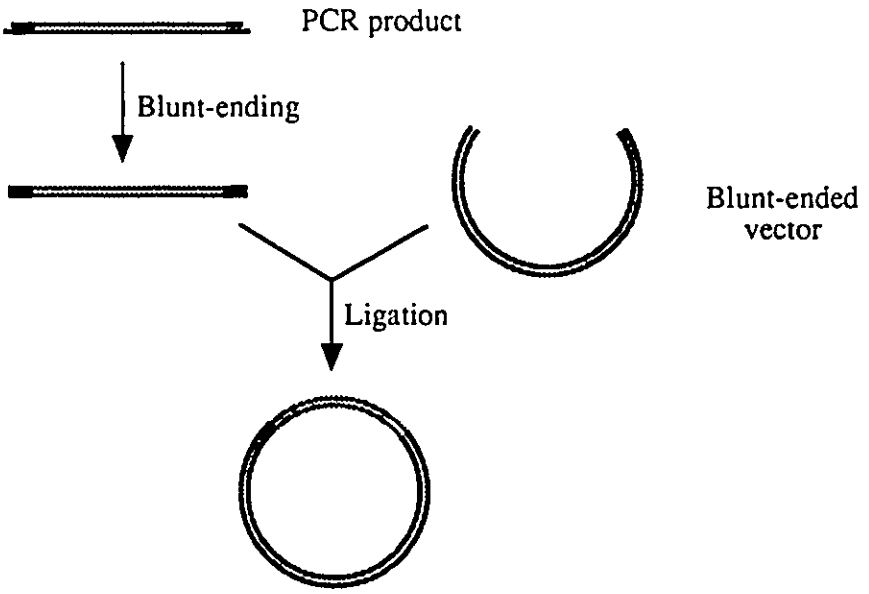
Although blunt-end cloning may appear to be the obvious approach to use to introduce PCR products into plasmid vectors, other techniques have also been described. The strengths and weaknesses of blunt-end cloning and of these alternative techniques, diagramatically represented in Fig. 47, are discussed and compared.

Evaluation of Blunt-End Cloning of PCR Products

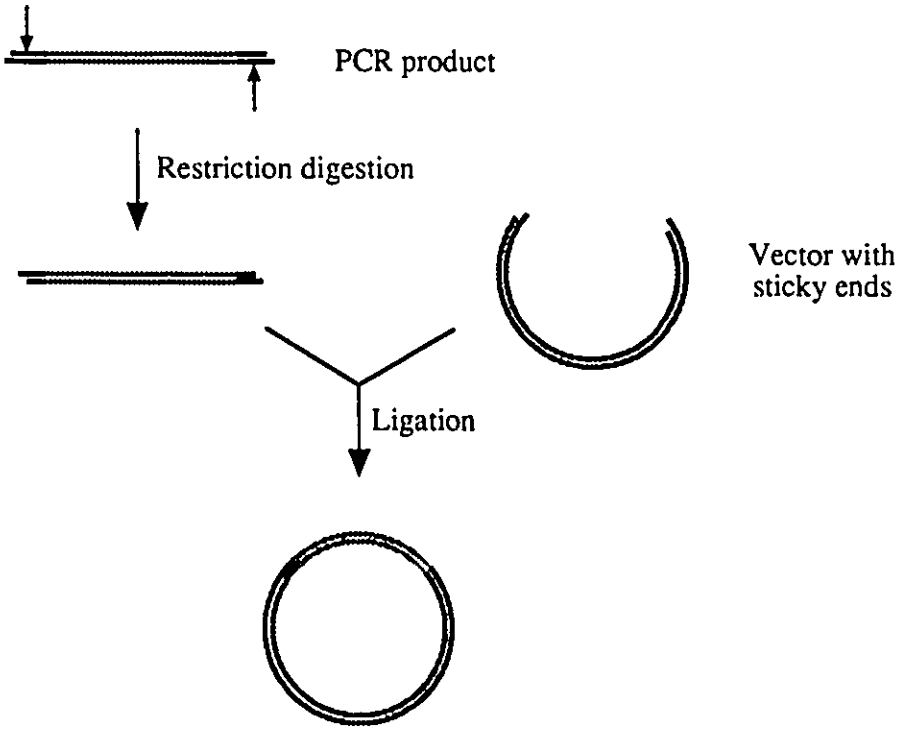
The major advantage of blunt-end cloning (Fig. 47A) is its simplicity and general applicability; any fragment of DNA possessing blunt ends can be cloned into any blunt-ended vector. In addition, no constraints are placed on PCR primer design, since the only features required for ligation and cloning, blunt ends, are sequence-independent.

Fig. 47. Approaches for the cloning of PCR products. The diagram illustrates different approaches for the cloning of PCR products. These consist of the cloning of blunt-ended PCR products (A), cloning by the use of external (B) or internal (C) restriction sites, and cloning using a specialized vector possessing 3' T-base overhangs. The regions of the amplified PCR product which are encoded by the primers are illustrated in black. The arrows pointing at the PCR products indicate the sites of endonucleolytic cleavage utilized in approaches (B) and (C).

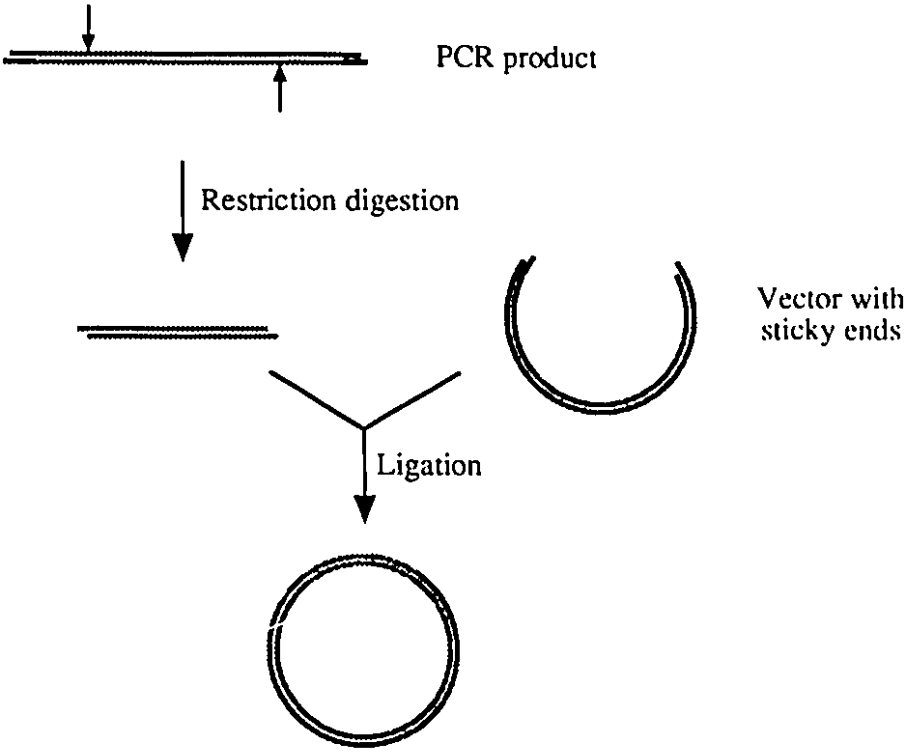
A) Blunt-end cloning



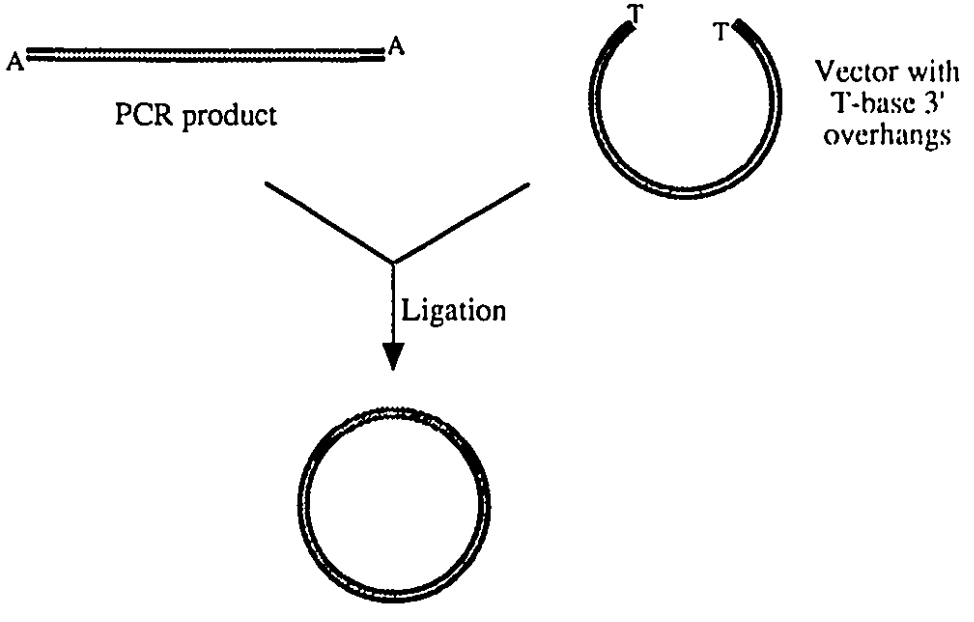
B) Cloning using external restriction sites



C) Cloning using internal restriction sites



D) Cloning using vector with T-base overhangs



A disadvantage of using blunt-end cloning is the relative inefficiency of blunt-end ligations, which may give rise to poor cloning efficiencies, in comparison with those involving cohesive ends. Moreover, it is not possible to perform directional cloning, to obtain a specific orientation, since the ends of the DNA insert are equivalent. Thirdly, as a result of a terminal transferase-like activity associated with Taq DNA polymerase, PCR products appear to contain single A-base 3' overhangs which interfere with direct blunt-end ligation. As demonstrated in this report, this requires the use of an additional enzymatic treatment, with a blunt-ending enzyme such as Klenow fragment.

Evaluation of Cloning of PCR Products Containing External Restriction Sites

This approach, shown in Fig. 47B, can be generally applicable if the same restriction enzyme recognition sites are routinely incorporated into all PCR primers utilized. In this case, only one vector, digested with the appropriate enzyme(s), needs to be prepared. The use of the recognition sequences of different restriction enzymes into the sense and antisense PCR primers allows the PCR product to be cloned in a directional manner. Another advantage of this method over blunt-end cloning, is that ligation occurs more efficiently due to the involvement of cohesive ends. In addition, since the ends involved in ligation are generated following PCR amplification, the success of this approach is unaffected by the addition of 3' A-base overhangs.

The most problematic feature of this approach is the poor efficiency with which many commonly used restriction endonucleases cleave sites near DNA termini (349). Since a large proportion of the DNA termini may remain uncleaved, this method can result in low cloning efficiencies (348). This problem can be alleviated somewhat by the use of longer incubation times and higher concentrations of the restriction endonuclease, although this may increase the chances of illicit cleavage by contaminating nucleases. An alternate option is the introduction of additional bases to the PCR primers such that the restriction sites are

further from the DNA termini. It is clear that the number of enzymes which can be successfully utilized in this approach is limited. A second disadvantage of this approach is that the presence of the endonuclease recognition sequence, which is not complementary to the template, may interfere with the sequence specificity of the primer. In addition, the endonuclease recognition sequence may place constraints on primer design by being able to form unwanted secondary structure with potential template-specific sequences or by participating in "primer dimerization".

Evaluation of Cloning of PCR Products Containing Internal Restriction Sites

Most of the advantages of this approach (Fig. 47C) are similar to those discussed above for the use of primer-encoded restriction sites. However, an additional advantage is that the use of restriction sites located within the amplified DNA fragment, distant from the termini, eliminates the problems associated with restriction enzyme cleavage efficiency.

Despite the numerous strengths of this approach, this method is usually the most restricted in its applicability, since it requires the presence of infrequently occurring restriction sites within the amplified product, which are compatible with sites in the cloning vector used. Moreover, since different templates have distinct restriction maps, it may be necessary to prepare a large number of cloning vectors, each cleaved with a specific pair of endonucleases. This approach may severely limit primer design, depending on the locations of useful restriction sites relative to the region of interest.

Evaluation of Cloning of PCR Products into a Vector Containing T-base Overhangs

The recent commercial availability of a vector containing single T-base 3' overhangs, created by the restriction endonuclease HphI, has provided an alternative method for cloning PCR products (Fig. 47D). The technique is mediated by the complementarity between the overhangs present on the PCR product and those on the vector. In contrast to

blunt-end cloning, whose efficiency is hampered by the presence of A-base overhangs, this method takes advantage of them.

This approach is unique among the cloning methods discussed, since it obviates the need for any enzymatic steps prior to ligation to the vector. In addition, it is generally applicable, and no constraints are placed on primer design, since the only features required for ligation are the A-base overhangs added during amplification. The ligation of PCR products to this specialized vector may be more efficient than blunt-end ligation, despite the cohesive ends consisting of only one base each.

The extent of the addition of A-base overhangs to PCR products, and the conditions which may affect it, are unclear. Since this event is critical for the success of this cloning method, factors which may affect its rate, such as the concentration of dATP, may also affect cloning efficiency. It is possible that this non-template dependent enzymatic process occurs only under specific conditions existing at a particular stage of the amplification. Consequently, this may result in the selective cloning of PCR products synthesized under these permissive conditions. Until the enzyme activity responsible for the A-base addition has been more extensively characterized, these questions will remain unanswered. Another disadvantage of this method is the cost of purchasing the specialized vector which is required.

Criteria for Selecting Blunt-End Cloning

One of the most important criteria for selecting the blunt-end cloning method was its general applicability. Of the methods discussed above, only that utilizing internal restriction sites fails to satisfy this criterion. Due to the repetitive sequence of the type I collagen cDNAs, within the regions encoding the triple helical domains, restriction sites which would be useful in cloning are scarce.

Reliability was another desired feature of the cloning method chosen. The blunt-end

cloning approach has proven to be very reliable, despite the observation of variability in efficiency between experiments. However, this variability appears to depend primarily on the yield of the PCR product obtained. The method utilizing the vector with T-base overhangs was not considered, due to its cost and its reliance on an enzymatic modification which is incompletely understood. The method utilizing external restriction sites was excluded, since initial experiments resulted in extremely poor cloning efficiencies, presumably due to inefficient endonucleolytic cleavage. However, it is possible that further experiments examining the effects of different cleavage conditions, or of other combinations of restriction enzyme sites, would have resolved these technical problems.

Taq DNA Polymerase Artifacts

One of the problems associated with PCR amplification is the high frequency of replication errors committed by Taq DNA polymerase, relative to DNA polymerases possessing editing functions, such as Klenow fragment. The rate of base substitutions has been reported to be approximately 1.1-2.8 x 10^{-4} per nucleotide (307, 366, 367), and appears to depend on the number of copies of DNA template at the start of PCR. Although these errors are not problematic when generating DNA to be used as probes in Northern or Southern blotting experiments, they must be seriously considered when amplifying DNA for subsequent sequence characterization. It is particularly critical to assess the possibility of Taq DNA polymerase replication errors when attempting to attribute single base substitutions to sequence polymorphisms and disease mutations.

The accuracy of the information provided by chemical cleavage analysis, concerning the positions of detected mismatches, allowed the easy identification of the underlying sequence changes in some of the clones analyzed. All of the individuals examined by cDNA sequencing were found to be heterozygous for the sequence changes detected by chemical cleavage analysis, whether they were polymorphisms or OI mutations. Therefore, in all cloning experiments, the M13 clones analyzed could be divided into two

groups based on which allele they represented. In most cases, the alleles were approximately equally represented among the 10 to 12 clones sequenced. Any observed sequence changes, exhibiting clone representations which were distinct from those of the identified alleles, were considered to be the result of Taq DNA polymerase replication errors. In fact, these replication errors were always present in single clones out of those analyzed. In some experiments, replication errors were identified by sequencing clones derived from duplicate PCR reactions. This is the definitive manner to determine whether observed sequence changes are replication errors, due to the extremely low probability of having the same error occur in two independent PCR reactions.

In the course of sequencing various cDNA regions from a number of OI cell lines, 29 Taq DNA polymerase replication errors were identified based on their representation among the clones examined. In general, the replication errors, which were all single base substitutions, tended to occur in clusters such that clones derived from some PCR reactions contained numerous errors, while those derived from other reactions contained none or few. This suggests that the occurrence of replication errors may depend on the presence of specific conditions within the PCR reaction mixture, such as the concentration of dNTPs. Thus errors might be expected to occur in later cycles when dNTPs become depleted. This is supported by the observation that replication errors were always represented by single clones among those sequenced. If the errors had occurred early in amplification, a larger proportion of clones would have exhibited the sequence changes.

Among the 29 replication errors observed during sequencing, there appeared to be a definite preference in the base pair change (Table 16). The most common base pair change observed was A:T → G:C which accounted for 59% of the errors. This is in agreement with a report demonstrating that A:T → G:C transitions are predominant among the replication errors committed by Taq DNA polymerase (367). The reverse substitution, G:C → A:T, was the second most preferred change, representing 24% of the errors. The

Table 16. Taq DNA polymerase replication errors identified by DNA sequence analysis.

Base Pair Change	Frequency [1]
A:T → G:C	17/29 (59%)
G:C → A:T	7/29 (24%)
C:G → A:T	3/29 (10%)
G:C → C:G	1/29 (3.5%)
T:A → A:T	1/29 (3.5%)

[1] The 29 replication errors listed in the table were identified in the course of sequencing approximately 32 kbp of various cDNA clones. These were judged to be errors based on their representation in single clones among the 10-12 which were sequenced for each cDNA insert.

exhibited preference for these base pair changes suggests that G:T and/or A:C base pairs are the mismatches which are most commonly created by Taq DNA polymerase. The more likely mismatch intermediate is a G:T base pair, since this has been shown to be more thermally stable than an A:C base pair (368).

4.4 Polymorphisms in the COL1A1 and COL1A2 Coding Regions

The frequency and distribution of polymorphisms in the type I collagen cDNAs have been examined by chemical cleavage analysis with HA and OT. The three polymorphisms which were detected by this method are summarized in Table 17.

One polymorphism, detected by reactivity with HA, was observed in the prepro-α1(I) cDNA. The polymorphism was demonstrated to affect the wobble position of the exon 2 triplet encoding arginine residue 57 of the prepro-α1(I) chain. The allele detected by HA reactivity, containing the CGT codon, was present in only one of the thirty-two chromosomes examined with this reagent.

Two polymorphisms were observed in the prepro-α2(I) cDNA. One of these was detected using OT, and affected the wobble position of the exon 25 triplet encoding proline residue 482 of the prepro-α2(I) chain. The allele detected by reactivity with OT, containing the CCC codon, was present in three of the eighteen chromosomes examined by chemical cleavage analysis with this reagent. The second prepro-α2(I) cDNA polymorphism was detected using HA, and affected the 5' position of the exon 28 triplet encoding residue 549 of the prepro-α2(I) chain. At the level of the amino acid sequence, this results in an alanine/proline dimorphism at Y-position residue 459 of the triple helical domain of this chain. The allele detected by reactivity with HA, containing the CCT proline codon, was detected in six of the sixteen individuals examined by chemical cleavage analysis with this

Table 17. Polymorphisms identified in the COL1A1 and COL1A2 coding regions.

Gene	Polymorphism [1]			Probe [2]	Reagent	Frequency [3]
COL1A1	CGG/CGT	Exon 2	Arg_{57}	α1XB	HA	1/32
COL1A2	CCA/CCC	Exon 25	Pro_{482}	α2XX	OT	3/18
	GCT/CCT	Exon 28	Ala/Pro_{549}	α2XX	HA	6-10/32 [4]

[1] The underlined codon sequence is that which is detected by chemical cleavage analysis with the specified reagent. The amino acid residue numbers are relative to the first residues in the prepro-α1(I) and prepro-α2(I) chains (78, 80).

[2] The abbreviations refer to the probes listed in Table 6.

[3] The frequency reported is that of the HA- or OT-reactive allele and is calculated as a fraction of the number of chromosomes analyzed (32 with HA, 16 with OT).

[4] The frequency of this HA-reactive allele is reported as a range, since homozygosity or heterozygosity for the cleaved allele was not determined for 4 of the 6 individuals exhibiting the polymorphism.

reagent. In two of these six individuals, heterozygosity for this sequence polymorphism was confirmed by cDNA sequence analysis. Due to the possibility of homozygosity in the other four individuals, the HA-reactive allele may be present in from six to ten of the thirty-two chromosomes analyzed with this reagent.

Frequency and Distribution of Collagen Polymorphism

This study was not intended as a comprehensive determination of all the sequence variations in these cDNAs, since the group of individuals studied was small (eight with OT and sixteen with HA), and only those sequence changes resulting in reactive C and T bases could be detected. The goal of the study was rather to provide some indication as to the frequency with which polymorphisms occur, and to determine whether some coding regions were more frequently affected than others.

The small number of polymorphisms detected indicates that sequence variation in the prepro-α1(I) and prepro-α2(I) cDNAs is rare. The low frequency of polymorphisms is unexpected considering the large size of these cDNAs (4.0-4.5 kbp). In addition, there would not be expected to exist any selective pressure for conserving certain nucleotide positions, such as codon wobble bases which do not result in amino acid changes. The data, however, suggest that this is not the case, since only two such polymorphisms, affecting the wobble positions of arginine and proline codons, were detected by chemical cleavage analysis. This would indicate that there is considerable selective pressure to conserve wobble positions. This conclusion is supported by the remarkable preference of specific bases in the wobble positions of codons encoding glycine, proline, and alanine residues in the triple helical domains of the α1(I) and α2(I) chains (78, 80). In general, the base preference, in order of decreasing usage, is T, C, A and G. This preference is particularly striking for Y-position proline codons in the prepro-α2(I) cDNA, 84% and 13% of which contained T and C in the wobble position, respectively. The absence of

selective pressure would be expected to result in approximately equal usage of all four bases in this position. A possible explanation for this conservation is that during DNA replication, the ability of the cell to synthesize all four dNTPs may not be equal. The sequence of the genome may therefore have evolved to reflect these limitations. An alternative explanation is that the four base pairs, A:T, T:A, G:C and C:G, may undergo mutation at different rates, perhaps due to inherent properties of the DNA polymerase, the DNA repair enzymes and/or the distinct chemical properties of the bases. For example, the unequal representation of glycine codons suggests that G → T mutations may be relatively common.

Additional cDNA positions which might be predicted to be tolerant of sequence change are the triplets encoding X- and Y-position residues of the α1(I) and α2(I) triple helical domains. The functions of the majority of X- and Y-position residues are to stabilize the triple helix and to participate in fibril formation. However, since these functions are performed as a cooperative effort involving many X- and Y-position amino acids, it is possible that any such individual residue may be dispensable. Thus, although a critical number of X- and Y-position proline residues per chain are required for triple helix formation to occur, it is possible that the substitution of any individual proline residue may be tolerated. In addition, it could be expected that the substitution of some charged and hydrophobic X- and Y-position residues, whose primary roles may be to participate in intermolecular interactions during fibrillogenesis, would also be tolerated. The use of HA and OT in chemical cleavage analysis did not permit the identification of sequence changes resulting in proline substitutions, since these would give rise to mismatched G bases in RNA:DNA heteroduplexes. However, the observation of a single such polymorphism, affecting a Y-position residue within the α2(I) triple helical domain, suggests that non-proline X- and Y-position residues are generally conserved in identity. A possible explanation is that the substitution of one such residue may be sufficient to disrupt the

packing of collagen molecules in a fibril, particularly if the replacing amino acid has different chemical properties from that which was replaced. For example, the substitution of an arginine residue by a leucine may not be tolerated.

A number of COL1A1 and COL1A2 polymorphisms have been described in the literature, and are summarized in Table 18. Only three have been described in the COL1A1 gene, while fifteen have been observed in COL1A2. Considering the large size of these genes, these numbers are quite low, even though RFLPs, which constitute the majority of the polymorphisms described, can only detect a fraction of all the sequence variation present. In addition, the large sizes of the introns present in COL1A1 and COL1A2 would suggest that polymorphisms should be more frequent. The greater number of polymorphisms described in the COL1A2 gene may be a reflection of the larger size of this gene (32 kbp), relative to COL1A1 (18 kbp). In fact COL1A2 appears to be the most polymorphic of the collagen genes studied to date (378, Table 18). However, the reasons for this are unclear, since the differences in polymorphism are probably not exclusively a result of variations in gene size.

The higher frequency of reported polymorphisms in the COL1A2 gene, in comparison with COL1A1, is also reflected by the difference in interspecies sequence conservation exhibited by the two chains. For example, the pro-α1(I) chain amino acid sequence exhibits 90% conservation between avian and mammalian species (77, 78), in contrast to the 83% sequence conservation displayed by the pro-α2(I) chain (80). Thus, during evolution, the sequence of the α1(I) chain has been more tightly conserved, indicating that it is less tolerant of change than that of the α2(I) chain. In turn, this would suggest that polymorphisms, resulting in changes at the protein level, would be more common in the α2(I) chain.

Eleven of the eighteen polymorphisms identified in the COL1A1 and COL1A2 genes are RFLPs due to sequence changes which have not yet been characterized. In addition,

Table 18. Previously described polymorphisms in the human COL1A1 and COL1A2 genes.

Gene	Polymorphism(s)	Location(s)	Reference(s)
COL1A1	RsaI RFLP	ND [1]	170
	Sequence polymorphism	3'UT region	369
	HaeIII RFLP	Intron 43	370
COL1A2	EcoRI RFLPs (2)	ND	371, 372
	PstI RFLP	ND	371
	MspI RFLP	ND	373
	BglII RFLPs (5)	ND	372, 374
	StuI RFLP	ND	375
	PvuII RFLP	Exon 25	351
	RsaI RFLP/ 38-bp deletion	Unspecified intron	376
	Sequence polymorphism	Exon 6	377
	Sequence polymorphism	Exon 47	377
	Sequence polymorphism/ MaeIII RFLP	Intron 33	377

[1] Abbreviations used: ND, not determined; UT, untranslated.

whether these sequence variations are located within coding or noncoding sequences has not been established. It is probable that they are located in noncoding regions, since this appears to be where most sequence diversity is distributed (379). In fact, it has been estimated that polymorphisms may occur with a frequency of approximately one per every hundred base pairs in noncoding sequences, such as introns and untranslated regions (380). Assuming this frequency, the COL1A1 and COL1A2 genes would perhaps be expected to contain between approximately 100 and 300 polymorphisms, of which only a fraction could be detected as RFLPs.

To date only seven described COL1A1 and COL1A2 polymorphisms have had their underlying sequence changes precisely identified. Four of these were located within noncoding sequences, with three being present within introns, and the fourth within a 3'-untranslated region. Thus far, only three recently published polymorphisms have been demonstrated to be located within coding regions. These polymorphisms were shown to affect the wobble positions of the triplets encoding residues Asp_{82} (exon 6), Pro_{482} (exon 25) and Gly_{1045} (exon 47) of the prepro-α2(I) chain. The sequence variation affecting the proline codon results in a PvuII RFLP (351), and is identical to the polymorphism detected by chemical cleavage analysis using OT, as described in this study. Constantinou et al reported the frequency of the allele containing the CCC triplet (indicated by the absence of PvuII cleavage) to be present in 40% of the 68 chromosomes analyzed (351). This is in contrast to the present study, in which this allele, detected by OT reactivity was present in only 16.7% of the 18 chromosomes analyzed. This discrepancy can perhaps be explained by considering that the number of chromosomes analyzed may be too limited to allow meaningful comparison. An alternative explanation is that the frequencies of the two alleles may differ depending on the racial origins of the individuals studied. All but one of the individuals examined in this study were caucasians, whereas the races of the subjects examined by Constantinou et al were not specified in their report.

Effects of the Alanine/Proline Polymorphism on Collagen Structure and Function

The identified sequence variation resulting in an alanine/proline polymorphism at position 549 of the prepro-α2(I) chain constitutes only the fourth coding sequence polymorphism described to date, and the first resulting in an alteration in the amino acid sequence. This is the first report of a proline codon at this position, since de Wet et al have observed an alanine codon at this position (381). However, it appears that this position can also be occupied by aspartic acid, as indicated by the cDNA sequence reported by Kuivaniemi et al (80). Thus, variation of the amino acid residue in this position does not seem to result in deleterious effects on the structure or function of type I collagen.

The polymorphism affects a Y-position residue within the sequence -Gly-Pro-Y-Gly-Pro-Pro-Gly-. The high content of imino acids in the X and Y positions in these two consecutive Gly-X-Y triplets would suggest that a tight triple helical structure may be advantageous in this region. However, the indication that one of these Y positions can be filled with proline, alanine or aspartic acid, residues with very different chemical properties, without apparent effects on collagen function suggests that the increased helix stability provided by an additional proline residue is not required. Whether this position of the α2(I) chain can be occupied by any amino acid, without deleterious effects on collagen function, is unknown.

The possible presence of amino acids with diverse chemical and physical properties, such as alanine, proline and aspartic acid, at the same position within a protein would be expected to have very different effects on its structure. For example, alanine and aspartic acid, due to their differences in hydropathy, may introduce different protein conformations as a result of their preferences for hydrophobic and hydrophilic surroundings, respectively. Proline residues may produce distinct conformations due to their ability to introduce sharp kinks in the three-dimensional folding of a polypeptide chain. Thus, varying the identity of a single residue, between amino acids such as alanine, proline and aspartic acid, would

perhaps be expected to result in deleterious effects on the structures and functions of many proteins.

The unique properties of collagen structure appear to provide some flexibility in the type of amino acids which can be tolerated in X and Y positions. Due to the special role of imino acids in stabilizing the triple helical structure of collagen, the substitution of X and Y-position residues for proline and hydroxyproline, respectively, may be readily tolerated in most regions of the helix. However, such polymorphisms involving proline and certain amino acids such as alanine and aspartic acid, may not be tolerated equally well in all regions of the helix. For example, the substitution of a proline residue in a region requiring greater thermal stability, such as the helical region immediately adjacent to the carboxy-terminal propeptide, may not be tolerated. Conversely, the substitution to proline of a non-proline X- or Y-position residue in a region requiring greater flexibility, may not be tolerated. This may be the case in the regions containing the collagenase cleavage and helix cross-linking sites, which have low levels of X- and Y-position imino acids (77-80). Such substitutions could potentially hide these sites, which are important in collagen turnover and stabilization, respectively, by resulting in a tighter triple helical conformation.

Although no substitutions involving X- or Y-position residues have yet been demonstrated to result in OI, it is possible that such defects could give rise to bone fragility by interfering with fibrillogenesis and/or mineralization, depending on the identities of the amino acid residues involved, and on the helical region affected. For example, a substitution within the helical region opposite the hole zone of the collagen fibril may prevent normal mineral deposition. The description of the alanine/proline polymorphism in this study is important, since it indicates that such variations in the identities of X- and Y-position amino acid residues occur within the normal population, and are not necessarily involved in the manifestation of a connective tissue disorder such as OI. This conclusion must be considered before proposing a role of X- and Y-position substitutions in producing

OI. However, this does not exclude the possibility that substitutions affecting X- and Y-position residues, located in different regions of the α1(I) and α2(I) triple helical domains and involving various replacing amino acids, could result in OI.

Importance of Identifying Polymorphisms

One of the most important reasons for identifying polymorphisms is that they serve as convenient gene markers in the linkage analyses of families exhibiting inherited disorders. The polymorphisms act as tags for tracing the inheritance of specific gene alleles throughout a pedigree. However, a problem which is often encountered in such analyses is the absence of informative polymorphisms. Thus, the identification of a large number of intragenic polymorphisms is of great value.

This study has identified two novel polymorphisms which can be used in linkage analyses involving disorders associated with mutations in the COL1A1 and COL1A2 genes, as well as genes in closely linked loci. The identified COL1A1 polymorphism results in an XhoII (BstYI) RFLP, and is of particular importance since very few markers have been described for this gene. A possible disadvantage of this marker is that one of the alleles may be relatively rare, having been detected in only one out of the 32 chromosomes analyzed. However, due to the small number of chromosomes examined, this frequency may not be representative of larger populations.

The COL1A2 polymorphism detected by HA reactivity results in an EcoRII RFLP which may be a useful marker. However, due to the repetitive nature of the cDNA sequence encoding the triple helical domain, this restriction endonuclease is a frequent cutter. Therefore, in order for this polymorphism to be a useful RFLP, the method of detection should not employ large probes, since this would result in the visualization of a large number of restriction fragments. An alternative approach to RFLP analysis for allele detection, is the use of allele-specific oligonucleotide hybridization, which is capable of differentiating between sequences differing by a single base substitution (382).

The characterization of HA-reactive prepro-α1(I) and prepro-α2(I) cDNA polymorphisms will also assist in the identification of OI mutations by chemical cleavage analysis. This is due to the ability to disregard mismatches which have been detected in normal individuals, thereby allowing the investigator to focus on mismatches which are more likely to have arisen from a disease-related mutation. In addition, the identification of a sequence change in an OI patient, which is distinct from the sequence variation exhibited by normal individuals, provides additional evidence supporting its role in producing the OI phenotype.

Polymorphisms are also informative in identifying coding regions which are tolerant of sequence change, and therefore less critical for the normal function of the encoded protein. Although few polymorphisms could be detected in the type I collagen cDNAs, it is likely that the examination of a larger group of individuals, and the chemical cleavage analysis of the noncoding cDNA strand will reveal further sequence variation. Such experiments will provide additional information on the sequence conservation in these cDNAs within the human population, and as a consequence, a more detailed assessment of the structural elements of type I collagen prepro-α chains which are required or dispensable for normal function.

4.5 Identification of Type I Collagen Defects in OI Patients

This study consisted of a group of ten OI patients exhibiting a range of phenotypic severity, as well as the parents of type IV OI patient SL211. Fibroblast cell lines derived from these individuals were examined to determine whether they synthesized type I collagen exhibiting post-translational overmodification. In addition, chemical cleavage analysis of fibroblast RNA was performed in order to detect and locate potential OI

mutations. Based on these experimental data, the twelve cell lines could be divided into three groups, as summarized in Table 19.

Cell Lines in Group A

The cell lines in this group, GD199, JL227, PM244, MB252 and MK345, did not exhibit any detectable overmodification of secreted type I procollagen, nor any detectable C-base mismatches in chemical cleavage analysis. These findings are most easily reconciled with the presence of mutations which are not associated with structural defects within the triple helical domains of the α1(I) and α2(I) chains. However, the data do not exclude the possibility of mutations altering the structure of other domains of these chains.

The experimental data indicate that type I OI cell lines GD199, PM244 and MB252 likely possess mutations resulting in quantitative collagen defects. These findings are therefore in agreement with the literature on this form of OI. The mutations in these patients may give rise to nonfunctional COL1A1 alleles (174), abnormal prepro-α1(I) mRNA processing and transport (177), or the synthesis of structurally abnormal chains which cannot associate with normal chains (178). The three type I OI cell lines in Group A are presently being examined by RNA slot blot analysis to determine whether any of them exhibit deviations from the 2:1 ratio in the steady-state levels of the type I collagen messages. This will reveal if any of the cell lines possess mutations resulting in non-functional COL1A1 alleles or abnormal RNA processing.

JL227 is the clinically normal mother of type IV OI patient SL211. The experimental data suggest that JL227 does not harbour any type I collagen defects. However, the paradoxical biochemical and phenotypic features exhibited by her spouse (ML226) and daughter, suggest that JL227 may in fact possess a recessive mutation. This is discussed further below.

Table 19. Results of the biochemical analyses of the cell lines from the OI patients and parents.

Group	Results of Biochemical Analyses	Cell Line	OI Type
A	No apparent collagen overmodification	GD199	IA
	No mismatches detected	JL227	N[1]
		PM244	I
		MB252	IA
		MK345	III
B	Collagen overmodification detected	SL211	IVB
	No mismatches detected	ML226	N
		DP266	III
C	Collagen overmodification detected	PP238	IV
	Mismatches detected	JN245	IV
	Mutations identified	NC323	III
		SS333	IIA

[1] Abbreviation used: N, normal phenotype (parents of patient SL211).

MK345 is the second affected child born to clinically normal parents. The results for this patient, indicating the absence of collagen overmodification and detectable C-base mismatches, were somewhat surprising, since a large proportion of type III OI cases have been demonstrated to possess structural defects similar to those in other forms of OI. Once again, the possibility of a recessive mutation could be considered to explain the experimental data and the recurrence of OI in this family, particularly in a relatively isolated population, such as native Indians. However, it is difficult to imagine a type I collagen mutation which could give rise to such a severe phenotype in a homozygous state, but which could be asymptomatic in a heterozygous individual.

A more plausible explanation may be the presence of a dominant mutation in MK345, and germ-line (and perhaps somatic) mosaicism in one parent to explain the recurrence of OI in this family. One possibility would be a mutation affecting the pro-α1(I) carboxy-terminal propeptide which allows the abnormal chain to associate with its normal counterparts, but which prevents helix formation. The abnormal procollagen molecules assembled would be rapidly degraded, resulting in the exclusive secretion of normal procollagen. Such a mutation has been described by Cole et al in a patient with type IIB OI (205). If the defect is located within the carboxy-terminal propeptide region covered by the chemical cleavage probes, it is likely a mutation not affecting a G base, since no C-base mismatches were detected. The defect could also be located within the extreme carboxy-terminal end of the propeptide encoded by cDNA regions not covered by the chemical cleavage probes. This patient is currently being studied by cDNA sequence analysis to determine whether these propeptide regions contain such defects.

Although the phenotypes, biochemical data, and apparent modes of inheritance of the patients in this group can be rationalized by the presence of various types of collagen mutations, the possibility exists that one or more of them may harbour non-collagen mutations. Such mutations would explain the apparent absence of overmodified type I

collagen secreted by their fibroblasts and the inability to detect mismatches by chemical cleavage analysis. A small number of OI families have been demonstrated to exhibit a lack of linkage to COL1A1 or COL1A2 (158, 172). However, no specific mutations have yet been identified in such families.

Cell Lines in Group B

The cell lines in this group, SL211, ML226 and DP2o6, exhibited significant collagen overmodification, but no detectable HA-reactive mismatches in chemical cleavage analysis. The data suggest the presence of α1(I) or α2(I) helical defects resulting in delayed helix formation. However, the absence of detected C mismatches indicates that the defects do not involve the mutation of G bases within the coding sequence. This would exclude the possibility of mutations giving rise to glycine substitutions, deletions and insertions, and suggests that the collagen abnormalities exhibited by these cell lines may involve X- and Y-position residues, defects which have not yet been described in OI. Although X- and Y-position residues have their side chains directed away from the surface of the triple helix, it is nevertheless possible that a very bulky amino acid, such as tryptophan, could be disruptive. An alternative possibility is that the chemical cleavage method is not capable of detecting all types of C-base mismatches under the conditions used in this study. Theophilus et al have reported that certain mismatches exhibit a resistance to chemical cleavage (354). If this is the case with the cell lines in this group, glycine substitutions would then be the most likely explanation for the observed patterns of α1(I) and α2(I) chain overmodification.

The observation of indistinguishable patterns of collagen chain overmodification exhibited by type IV OI patient SL211 and her asymptomatic father, ML226, suggests that they possess the same collagen structural defect. Furthermore, it indicates, that despite the differences in phenotype, the defect has been genetically transmitted from the father.

The existence of differences in phenotype between SL211 and ML226 has several possible explanations. Since most collagen defects are dominantly expressed, it may be suggested that the mutation resulting in overmodification in patient SL211 is responsible and sufficient for producing a type IV OI phenotype. The father's normal phenotype could then be explained by considering that type IV OI is the form of the disease exhibiting the most extensive intrafamilial variability, which can be attributed to differences in genetic background and to environmental factors. The normal phenotype exhibited by ML226 could also be due to somatic and germ-line mosaicism, whereby the presence of the mutation in mature sperm cells would result in its transmission to the offspring. The number of offspring affected would depend on the percentage of mature gametes carrying the mutation. In this family, SL211 is the only affected child, of three offspring.

The observation that skin fibroblasts from ML226 exhibit the collagen abnormality indicates the presence of the mutation in at least some somatic tissues, in addition to the germ-line. The distribution of the mutation among various somatic tissues and the percentage of cells affected would depend on the precise stage of development at which the error occurred. In bone, an osteoblast harbouring the mutation would synthesize and secrete abnormal type I procollagen, which would only be incorporated into the fibrils being assembled by that cell. An osteoblast possessing two normal alleles would secrete normal procollagen and assemble only structurally normal fibrils. Thus, bone would consist of discrete regions in which the ECM organization is either normal or abnormal. In order for bone fragility to result, the proportion of regions containing abnormal matrix must reach a certain threshold. The normal phenotype of ML226 suggests that the percentage of osteoblasts expressing the mutant allele is insufficient to attain this threshold. This is despite the observation that a significant proportion of his skin fibroblasts appear to synthesize abnormal type I procollagen. Thus, although skin fibroblasts and osteoblasts are of common mesenchymal origin, the degree to which these cell lineages are affected in

ML226 may be significantly different.

A different explanation for the observed biochemical data is that the mutation resulting in the collagen overmodification in ML226 is recessive, and that SL211 exhibits compound heterozygosity, a situation which has already been described in one case of type III OI (224). The second mutation, contributing to the type IV OI phenotype in SL211, may then have been inherited from the asymptomatic mother, JL227. In this case, the mother may be heterozygous for a nonfunctional COL1A2 allele, or a germ-line mosaic for a quantitative or structural defect. On the other hand, JL227 may be normal, and the second mutation present in SL211 may have occurred sporadically early in her gestational development.

The rarity of compound heterozygosity suggests the likelihood of germ-line and somatic mosaicism of ML226 for a dominant collagen mutation. Although no mismatches were detected for cell lines SL211, ML226 and DP266, they are presently being examined by direct sequencing of the potentially-affected prepro-α1(I) and prepro-α2(I) cDNA regions, which have been identified by the two-dimensional electrophoretic technique.

Cell Lines in Group C

The cell lines in this group, PP238, JN245, NC323 and SS333, exhibited significant collagen overmodification, and the presence of HA-sensitive mismatches in chemical cleavage analysis, features which were indicative of α1(I) or α2(I) helical glycine substitutions. Indeed, subsequent cDNA sequence analysis demonstrated that these patients harboured point mutations resulting in four novel glycine substitutions within the triple helical domains of the α1(I) and α2(I) chains. Patients SS333 and NC323 exhibited Gly → Arg substitutions at positions 211 and 427 of the α1(I) and α2(I) triple helical domains, respectively. Patients PP238 and JN245 exhibited Gly → Val and Gly → Ser substitutions at positions 544 and 922 of the α2(I) triple helical domain, respectively.

The glycine substitutions identified in these patients, although involving different α chains, amino acid positions and/or replacing residues, and producing distinct OI

phenotypes, likely have similar effects on the biosynthesis and structure of type I collagen. The defects probably give rise to the decreased stability of procollagen molecules incorporating abnormal chains, thereby resulting in enhanced intracellular degradation and poor secretion. Consequently, the tissues of these patients likely contain insufficient quantities of type I collagen. However, some of the abnormal procollagen molecules are probably secreted. The incorporation of these structurally altered collagens into the ECM may then result in the formation of abnormal fibrils.

The "phenotypic severity rules" have been described in order to correlate OI phenotypes and glycine substitutions. However, the identification of numerous exceptions to the "rules" indicates that factors other than the position and type of the substitution are important in determining phenotypic severity. The occurrence of such exceptions underlines the fact that little is known concerning the precise manner in which distinct quantitative and qualitative abnormalities, associated with different structural collagen defects, contribute to bone architectural changes and to phenotype determination. These points are discussed in detail in sections 4.6 and 4.7.

Biochemical Heterogeneity in Type III OI

The distribution of the type III OI patients in this study among groups A, B and C is indicative of the genetic and biochemical heterogeneity exhibited by this form of the disease. The patient in Group A, MK345 is an example of an individual who may possess a non-collagen mutation, or a dominant type I collagen mutation inherited from a parent exhibiting germ-line mosaicism. Patient DP266, in Group B, however, is an individual who may possess a defect involving an X- or Y-position residue within the triple helical domain of the α1(I) or α2(I) chain, if indeed the chemical cleavage method is capable of detecting all C-base mismatches. In contrast, the patient in Group C, NC323, was demonstrated to harbour a dominant mutation similar to those identified in the other forms

of OI (ie. resulting in a glycine substitution within the triple helix). The possibility of such diverse mutations in type III OI is consistent with previous studies which have demonstrated different modes of inheritance in these patients.

Examination of the Frequencies of Different Glycine Substitutions

To date, at least 63 different glycine substitutions have been reported, including those described in this study. The distribution of these substitutions according to the identity of the replacing amino acid (or stop codon) is summarized in Table 20.

There are 676 glycine residues in the α1(I) and α2(I) triple helical domains, excluding the small number located in the X and Y positions. Taking into account the non-random base usage in the wobble position of glycine codons, the expected frequencies of the different possible glycine substitutions have been calculated (Table 20).

Most types of glycine substitutions appear to have been observed at close to the frequencies expected. However, Gly → Cys and Gly → Ala substitutions have been observed at frequencies which are significantly higher and lower, respectively, than those expected. An explanation for the unexpectedly large number of observed Gly → Cys substitutions, particularly in the α1(I) chain, is the manner in which these are often detected. These defects are easily identified, since they are accompanied by the appearance of disulfide-bonded α1(I) chains upon electrophoretic collagen analysis.

The unexpectedly low number of observed Gly → Ala substitutions may be due to the absence of a change in electrostatic charge or disulfide-bonding associated with these defects, thereby rendering their detection more difficult. However, there may be other reasons to explain the underrepresentation of Gly → Ala defects. For example, it is possible that different bases undergo mutation at different rates, such that certain glycine substitutions occur less frequently than anticipated. Alternatively, Gly → Ala substitutions may result in mild phenotypes in comparison with those involving other replacing amino

Table 20. Theoretical and observed frequencies of the replacing amino acids involved in glycine substitutions.

Amino Acid	Number Observed [1]	Observed Frequency (%)	Expected Frequency(%)
Cys	21	33.3	12.7
Arg	12	19.0	20.7
Ser	12	19.0	12.7
Asp	10	15.9	12.7
Val	7	11.1	16.7
Ala	1	1.6	16.7
Glu	0	0.0	4.0
STOP	0	0.0	3.3
Trp	0	0.0	0.7
	63	100	100

[1] A small number of glycine substitutions, involving the same replacing amino acid, have been observed more than once at the same position. However, these duplicate substitutions have only been counted once.

acids, such as arginine or aspartic acid. Of the amino acids which can arise from glycine codon point mutations, alanine possesses the smallest side group (ie. methyl). Thus, Gly → Ala substitutions may often be asymptomatic, depending on the precise polypeptide chain position affected.

The absence of observed glycine codon mutations resulting in tryptophan, glutamic acid and stop codons is not surprising, since these are expected to be relatively rare. The description of additional OI mutations will likely eventually uncover a small number of these defects, thereby allowing an assessment of the phenotypic severities associated with them.

4.6 Examination of the Phenotypic Severity Rules

This study has identified mutations in four OI patients. All four mutations resulted in glycine substitutions within the α1(I) or α2(I) triple helical domain. The comparison of these novel collagen defects with those that have been previously reported allows the evaluation of the rules which have been described to correlate phenotypic severity with type I collagen defects. The phenotypic gradient rule can be examined by comparing these defects to those involving the same chain and replacing amino acid. This helps to determine how phenotypic severity is affected by defect position for substitutions involving each type of amino acid. The chain rule can be investigated by comparing the phenotypic gradients in the α1(I) and α2(I) chains, for substitutions involving the same replacing amino acid. This examines whether all replacing amino acids result in more severe phenotypes when in the α1(I) chain than at an equivalent position in the α2(I) chain. In addition, the phenotypic gradients exhibited by substitutions involving different replacing amino acids can be compared. This aids in assessing the relative effects of different

replacing amino acid residues at the same position in either α chain on phenotypic severity.

Gly → Arg Substitutions

Patients SS333 (type II OI) and NC323 (type III OI) were shown to be heterozygous for point mutations converting glycine residues 211 of the α1(I) helix and 427 of the α2(I) helix, respectively, to arginine residues. Analysis of previously described OI mutations (reviewed by Kuivaniemi et al (173) and summarized in Table 21) identified seven resulting in Gly → Arg substitutions in the α1(I) chain. These substitutions are uniformly distributed throughout the α1(I) triple helical domain, being demonstrated at positions 85, 154, 391, 550, 667, 847 and 976. This has allowed a fairly informative assessment of the correlation between phenotypic severity and defect position for Gly → Arg substitutions in the α1(I) chain. The five defects at positions 391, 550, 667, 847, and 976, were identified in patients with type II OI, while those at positions 154 and 85 resulted in type III and mild OI phenotypes, respectively. Gly → Arg substitutions in the α1(I) chain seem to adhere well to the phenotypic gradient rule, since the resultant OI phenotype becomes milder with proximity to the amino terminus. The pattern in the phenotypic severity associated with these substitutions indicates that defects within the region of approximately 600 amino acids at the carboxy-terminal end of the α1(I) triple helical domain result in a lethal phenotype. Furthermore, defects within the region of approximately 150 amino acids at the amino terminus result in non-lethal phenotypes, which decrease in severity with proximity to the telopeptide. However, the phenotypes resulting from substitutions within the region of phenotypic transition, from type II to type III OI, between positions 154 and 391 were unknown. Consequently, the amino-terminal extent of the lethal phenotype associated with Gly → Arg substitutions in α1(I) had not been identified.

The defect identified in patient SS333 in this report is the most amino-terminal α1(I) Gly → Arg substitution described to date, which results in a lethal phenotype. It is

Table 21. Summary of OI mutations.

A. Glycine Substitutions

	Chain	Position [1]	OI Phenotype [2]	References [3]
1. Gly → Cys	α1(I)	19	OP/hypermobility [4]	173
		43	I	208
		94	I	383
		175	I	384
		178	I	385
		205	IV	214
		223	IV	214
		244	II	217
		382	IV	214
		415	III/IV	386
		526	III	383
		691	II	387
		718	II	383
		748	II	388
		904	II	209
		988	II	389
		1017	I	390
	α2(I)	259	Moderate	391
		472	II	173
		646	IV	391
		787	II	217

[1] The amino acid positions are numbered relative to the first glycine residue in the major triple helical domain.

[2] Whenever possible, the OI phenotype is reported according to the Sillence classification system. In the cases where Sillence classifications are not provided, the OI phenotypes are reported using the descriptive terms utilized by the authors.

[3] The references are numbered as in the main text, and are listed in the bibliography.

[4] Abbreviations used: OP, osteoporosis; ND, not determined.

	Chain	Position	OI Phenotype	References
2. Gly → Arg	α1(I)	85	Mild	392
		154	III	393
		211 [5]	**II**	-
		391	II	394
		550	II	220
		667	II	395
		847	II	186
		976	II	396
	α2(I)	**427**	**III**	-
		457	II	173
		694	II	397
		1012	IV	216
3. Gly → Ser	α1(I)	565	II	173
		598	II	398
		631	II	398
		832	IV	215
		844	III	399
		913	II	400
		964	II	196
		1003	II	393
		1009	severe	400
	α2(I)	661	OP	210
		865	II	396
		922	**IV**	-

[5] The substitutions shown in bold type are those described in this report.

	Chain	Position	OI Phenotype	References
4. Gly → Asp	α1(I)	97	II	214
		541	II	401
		559	II	214
		673	II	173
		883	II	221
	α2(I)	547	II	402
		580	II	214
		805	II	214
		907	II	403
		976	II	214
5. Gly → Val	α1(I)	256	II	404
		637	II	405
		973	II	396
		1006	II	396
		1009	II	400
	α2(I)	**544**	**IV**	-
		586	Moderate	173
6. Gly → Ala	α1(I)	928	II	396

B.In-Frame Deletions

Chain	Deletion of	Mutation Type	OI Phenotype	References
α1(I)	Exon 8	Splicing	Moderate	173
	Exon 14	Splicing (homozygous)	II	226
	Exon 17	ND	I	406
	Exons 23-25	Intron-to-intron genomic deletion	II	194
	Exon 27	Splicing	II	214
	Exon 30	Splicing	III	196
	Exon 44	Splicing	II	214
	Amino acids 874-876 (exon 44)	9-bp genomic deletion	II	196
	Exon 47	Splicing	II	173
α2(I)	Exon 9 or 10	Splicing	OP/hypermobility	173
	Exon 11	Splicing	Mild	247
	Exon 12	ND	Mild	214
	Exon 13	ND	I	407
	Val residue 255 (exon 19)	3-bp genomic deletion	Moderate	173
	Exon 21	Splicing	Moderate	173
	Exon 26	Splicing	IV	197
	Exon 28	Splicing mutation and nonfunctional COL1A2	II [6]	224
	Exon 33	ND	II	408
	Exon 34-40	Intron-to-intron genomic deletion	II	195

[6] This patient was shown to exhibit compound heterozygosity for a mutation resulting in a nonfunctional COL1A2 allele and a splicing defect in the other COL1A2 allele.

C. In-Frame Insertions

Chain	Insertion of	Mutation Type	OI Phenotype	References
α1(I)	Sequences from exons 14-17	exon-to-exon genomic duplication	II	198
α1(I)	75 bp from intron 35	ND	II	200
α1(I)	50-70 amino acids	600-bp genomic duplication	II	199
α2(I)	18 bp from intron 33	ND	Moderate	173

D.Translational Frameshifts

Chain	Frameshift Location	Mutation	OI Phenotype	References
α1(I)	C-propeptide; shortened chain	1-bp insertion in exon 49	II	205
α1(I)	C-propeptide; elongated chain	5-bp deletion in exon 52	I	178
α2(I)	C-propeptide;	4-bp deletion in exon 52 (homozygous)	III	206

nevertheless consistent with the phenotypic gradient rule since it is located on the carboxy-terminal side of the nonlethal substitution at position 154. This defect located at residue 211 of the α1(I) helix therefore extends the lethal phenotype produced by Gly → Arg substitutions by 180 residues from the previous amino-terminal position of 391. In addition, due to its proximity to the nonlethal substitution at position 154, it may approach the extreme amino-terminal extent of the type II OI phenotype associated with Gly → Arg substitutions in the α1(I) chain. Therefore, with the description of this new defect, the region of phenotypic transition from type II to type III OI resulting from Gly → Arg substitutions in the α1(I) helix, has been considerably focused from 237 to 57 residues. Since only one such nonlethal substitution has yet been described in this region, it is unclear whether this transition will consist of a region of phenotypic overlap between OI types II and III, or of a well defined boundary. This point will be resolved with the description of additional Gly → Arg substitutions in this region. The identification of additional Gly → Arg defects in the amino-terminal 150 residues of the α1(I) helical domain will also aid in further describing the phenotype changes from OI type III to yet milder forms with proximity to the amino-terminal extremity.

In contrast to the large number of α1(I) Gly → Arg substitutions described to date, relatively few substitutions have been identified in the α2(I) chain. The three Gly → Arg substitutions which have thus far been demonstrated, were located at positions 457, 694 and 1012 of the α2(I) triple helical domain. The substitutions at positions 457 and 694 resulted in type II OI, while that at position 1012 resulted in a considerably milder phenotype, type IV OI. These data indicate that Gly → Arg substitutions within the central region of the α2(I) triple helical domain result in lethal phenotypes, as observed with those occurring in the α1(I) chain. This suggests that such substitutions occurring in a large part of this chain likely disrupt triple helix stability in a manner similar to their α1(I) counterparts. In contrast to the central helical region, the extreme carboxy-terminal end of

the α2(I) triple helical domain appears to be tolerant of at least some Gly → Arg substitutions. However, the sparsity of α2(I) Gly → Arg substitutions described to date has made it difficult to assess the validity of the phenotypic gradient rule for this chain, and to predict the amino-terminal extent of the type II OI phenotype associated with these defects.

The Gly → Arg substitution at position 427 of the α2(I) triple helical domain identified in type III OI patient NC323 is consistent with the phenotypic gradient rule since it is located amino-terminal to the previously described type II OI substitutions, at positions 457 and 694. In addition, since this defect is located only 30 residues away from the lethal defect at position 457, it may identify the region of the α2(I) triple helical domain in which the phenotype associated with Gly → Arg substitutions undergoes a transition from type II to type III OI. Thus, the phenotypic transition which was expected to occur within the region of approximately 450 amino acids at the amino-terminal end of the α2(I) triple helical domain has been greatly focused to a region of 30 residues bounded by positions 427 and 457. The description of the Gly → Arg substitution in patient NC323 therefore suggests that type II OI phenotypes result from defects in the α2(I) carboxy-terminal region of approximately 550 residue, and milder forms arise from defects which are more amino-terminal. Since the defect in patient NC323 is the most amino-terminal Gly → Arg substitution thus far identified, it is unclear whether the pattern of decreasing phenotypic severity of non-lethal phenotypes observed in the α1(I) chain will be reproduced in the α2(I) chain.

The possibility remains that any of the four defects located near the suggested OI type II/type III transition regions result in phenotypes which are either exceptionally mild or severe. This would result in the α1(I) and α2(I) regions of phenotypic transition being located either more amino- or carboxy-terminal than proposed in this study. However, the majority of the exceptions to the phenotypic gradient rule which have thus far been

demonstrated, have been glycine substitutions involving cysteine or serine residues, as discussed later. It appears that the phenotypic gradients exhibited by substitutions involving more deleterious amino acids, such as arginine, adhere more strongly to the gradient rule. This supports the validity of the transition regions identified in this study.

The locations of the Gly → Arg defects identified in patients NC323 and SS333 were fortuitous in that they allowed a more precise mapping of the regions where phenotypic transition, from type II to type III OI, occurs in both the α1(I) and the α2(I) triple helical domains. This has permitted a direct comparison of the relative severities of substitutions occurring in these chains. The results are in agreement with the chain rule, since they demonstrate that the region of transition is considerably more carboxy-terminal (by approximately 250 residues) in the α1(I) chain than in the α2(I) chain. These regions are located between residues 154 and 211 for the α(I) helix, and between residues 427 and 457 for the α2(I) helix. It therefore appears that even for glycine substitutions involving a very bulky and disruptive replacing amino acid, such as arginine, there are regions of the α2(I) helical domain which are more tolerant than the corresponding regions of the α1(I) helix.

It is at present unclear whether the tolerance of the extreme carboxy-terminal end of the α2(I) triple helix, to certain Gly → Arg substitutions, is also a feature of the analogous region of the α1(I) chain, since no such defects have yet been described. The differences in phenotypic expression of α1(I) and α2(I) Gly → Arg substitutions (ie. as described by the chain rule) are likely to some extent a consequence of the 2:1 ratio of chains in the type I procollagen molecule. Defects in the α1(I) chain probably result in more extensive "protein suicide" than those in the α2(I) chain. They may also result from possible differences in the positions of α1(I) and α2(I) chains relative to the central axis of the collagen triple helix. As a result of the differences in amino acid sequence, α2(I) chains may participate in triple helix formation such that they are further from the central axis. This may allow α2(I) chains to more readily tolerate certain glycine substitutions. The ability of the α2(I) chain

to more readily tolerate such sequence changes, in comparison with the α1(I) chain, is also indicated by its more extensive interspecies sequence divergence, as discussed previously.

The regions of phenotypic transition are also of interest because they may represent areas in which there is phenotypic overlap between OI types II and III. Thus, some Gly → Arg substitutions in these regions may give rise to severe forms of OI which are associated with the radiological features of type II OI, but with variability in the length of post-natal survival, characteristics of patients classified as subtype IIB. Alternatively, such substitutions may give rise to very severe forms of type III OI (by radiological criteria), in which the length of post-natal survival is determined by the magnitude of various environmental factors (ie. physical and nutritional) experienced during gestation and delivery.

Gly → Val Substitutions

Patient PP238, exhibiting the phenotypic features of type IV OI, was demonstrated to possess a mutation resulting in a Gly → Val substitution at position 544 of th α2(I) helical domain. Thus far, only one Gly → Val substitution has been described in the triple helical domain of the α2(I) chain. This defect, located at position 586, resulted in a moderate OI phenotype. The Gly → Val substitution identified in patient PP238 is consistent with this previously described defect, since it is situated within close proximity, and also results in a moderate phenotype (type IV OI). Therefore, these specific defects indicate that Gly → Val substitutions within the central region of the α2(I) helical domain result in non-lethal phenotypes. However, the identification of only two Gly → Val substitutions is not sufficient to predict the phenotypes resulting from analogous defects in other regions of the α2(I) triple helical domain. For example, it is unclear whether Gly → Val substitutions are able to produce type II OI phenotypes when located at more carboxy-terminal positions, and milder phenotypes at more amino-terminal positions. In addition, it has not been established whether the area bounded by residues 544 and 586 represents a segment of a

continous phenotypic gradient, or whether it is a region which is tolerant of Gly → Val substitutions. If these defects constitute part of a continuous phenotypic gradient, this would indicate that Gly → Val substitutions are less severe than Gly → Arg substitutions.

In comparison with the α2(I) chain, a larger number of Gly → Val substitutions have been identified in the α1(I) chain, revealing a pattern in phenotypic severity which is similar to that observed for α1(I) Gly → Arg substitutions. The five α1(I) Gly → Val substitutions described thus far, located at positions 256, 637, 973, 1006 and 1009, all gave rise to type II OI phenotypes and are therefore consistent with the gradient rule. This indicates that valine residues are able to sufficiently alter the structure of the triple helix, thereby resulting in extensive degradation and/or the incorporation of abnormal collagen into the ECM. Since no nonlethal substitutions have yet been identified, the amino-terminal extent of the type II OI phenotype associated with α1(I) Gly → Val substitutions is unknown.

The paucity of described α2(I) Gly → Val substitutions prevents an extensive comparison of the phenotypic severity patterns exhibited by the two chains. However, by analogy with Gly → Arg substitutions, the large region of the α1(I) chain in which Gly → Val substitutions give rise to type II OI (positions 256, 637, 973, 1006 and 1009), suggests that such defects located near the carboxy-terminal end of the α2(I) triple helical domain, could result in similar phenotypic severities. Assuming this, the region of phenotypic transition from OI type II to type III, associated with Gly → Val substitutions in the α1(I) chain, would be predicted to be at least 300 residues more amino-terminal than that for the α2(I) chain, in agreement with the chain rule. The identification of further substitutions will be required to resolve these uncertainties.

Gly → Ser Substitutions

Patient JN245, affected by type IV OI, was demonstrated to possess a mutation resulting in a Gly → Ser substitution at position 922 of the α2(I) helical domain. Thus far, only two other such substitutions have been identified in this chain. One substitution, affecting position 661, was associated with osteoporosis, but with only mild bone fragility. This case can perhaps be considered a mild case of type I OI. The second substitution was observed at position 865 and gave rise to type II OI. The identification of these two defects suggests that substitutions from position 865 towards the carboxy terminus of the triple helical domain result in type II OI, and that substitutions which are more amino-terminal result in progressively milder phenotypes. However, the identification of a Gly → Ser substitution at position 922, resulting in a milder form of OI than that produced by the defect at position 865, indicates that these do not strictly adhere to the phenotypic gradient rule. In addition, this suggests that the region of the α2(I) chain surrounding residue 922 is tolerant of some substitutions, while adjacent regions may be relatively intolerant, resulting in type II OI phenotypes.

To date, nine different Gly → Ser substitutions have been described in the α1(I) chain. Similar to the α2(I) chain, Gly → Ser substitutions in the α1(I) chain do not appear to adhere to the phenotypic gradient rule. Lethal defects have been demonstrated at positions 565, 598, 631, 913, 964 and 1003. However, this region is interrupted by two non-lethal Gly → Ser substitutions at positions 832 (type IV OI) and 844 (type III OI). In addition, the extreme carboxy-terminal end of the α1(I) triple helical domain appears to be tolerant of some Gly → Ser substitutions, since a defect resulting in severe OI was demonstrated at position 1009. These findings suggest that the α1(I) chain contains two regions which may be relatively tolerant of some substitutions, depending on the nature of the replacing amino acid and precise position affected.

Due to the small number of α2(I) defects described to date, it is difficult to determine whether the regions of the α2(I) helical domain which are tolerant of Gly → Ser substitutions coincide with those of the α1(I) chain. Thus, it is unknown whether Gly → Ser substitutions near residues 832 and 844, or near the carboxy terminus of the α2(I) chain would result in milder phenotypes than those in adjacent regions. However, the tolerant region located near the carboxy terminus of the α1(I) chain may be analogous to that in the α2(I) chain, which was identified by the presence of the non-lethal Gly → Arg substitution at position 1014. In addition, since a lethal defect was identified at position 913 of the α1(I) helical domain, this chain does not appear to contain a region of tolerance corresponding to that identified in the α2(I) chain (surrounding residue 922).

The validity of the chain rule is difficult to assess for Gly → Ser substitutions, since the occurrence of exceptions to the gradient rule interferes with the identification of phenotypic transition regions. The presence of the α2(I) Gly → Ser substitution at position 661, associated with a very mild phenotype, suggests that the lethal phenotype associated with some of these defects does not extend amino-terminal to this position. In contrast, the transition region in the α1(I) chain appears to extend at least as amino-terminal as position 565. This then suggests that the OI type II/type III transition region is considerably more carboxy-terminal in the α2(I) chain than in the α1(I) chain, in agreement with the chain rule.

Comparison of Substitutions Involving Different Amino Acids

Comparison of the phenotypic gradients exhibited by glycine substitutions involving different replacing amino acids indicates that Gly → Asp defects are more deleterious than those involving the other residues observed to date, including arginine. This is indicated by the fact that no non-lethal Gly → Asp substitutions have yet been described in either the α1(I) or α2(I) chain. In addition, the Gly → Asp substitution at position 97 of the α1(I)

helical domain is the most amino-terminal defect resulting in a lethal phenotype which has been reported. Since relatively fewer Gly → Asp substitutions have been identified near the amino terminal of the α1(I) helical domain, it is unclear whether these result in a more severe phenotypic gradient than Gly → Arg substitutions, or whether the defect at position 97 affects an exceptionally intolerant region of the triple helix. It is difficult to explain how aspartic acid residues, with their smaller side chains, could result in phenotypes which are more severe than those produced by arginine residues. However, it is possible that the opposite charges of their side chains would have different effects on helix stability, fibril formation and perhaps mineralization due to the proximity to the amino terminus of the α1(I) chain. Alternatively, the region containing residue 97 may be especially intolerant, possibly due to the proximity of the helix lysine residue involved in collagen cross-linking. However, the manner in which this would give rise to bone fragility, rather than joint laxity, is unclear. The description of additional Gly → Asp and Gly → Arg substitutions in this region will provide information to help answer these questions.

Gly → Arg substitutions appear to be the next most deleterious defects observed to date, as indicated by the amino-terminal extent of the lethal phenotype in the α1(I) and α2(I) chains. The deleterious effect of Gly → Arg substitutions is likely to be predominantly due to the bulky nature of the side chain of arginine, in comparison with residues such as cysteine and serine. Thus arginine residues likely disrupt the triple helical conformation to a greater extent.

The amino-terminal extent of the lethal phenotype associated with α1(I) Gly → Val substitutions has not yet been determined. Therefore, it is difficult to assess how these defects compare with Gly → Arg substitutions in terms of phenotypic severity. However, comparison of their respective α2(I) defects indicates that Gly → Val substitutions are the less deleterious of the two. Therefore, it is unlikely the lethal phenotype observed with Gly → Val substitutions would extend as amino-terminal as those involving arginine.

Since only one Gly → Ala substitution has been observed to date, at position 928 of the α1(I) helical domain, it is not possible to determine how alanine compares to other amino acids in its ability to disrupt the triple helix. However, in comparison with valine, an amino acid with similar chemical properties, the smaller size of the alanine side chain suggests that it would perhaps be associated with less severe phenotypes.

Glycine substitutions involving serine or cysteine also appear to be less severe than those involving valine. This is particularly evident when comparing defects which have been reported in the α1(I) chain, in which Gly → Val substitutions result in lethal phenotypes as amino-terminal as position 256. In contrast, the amino-terminal extent of lethal Gly → Cys substitutions appears to be between position 526 and 691. While the amino-terminal extent of lethal Gly → Ser substitutions has not yet been determined, the presence of two unexpectedly mild defects at positions 832 and 844 suggests that these give rise to less severe phenotypes than Gly → Val substitutions.

In summary, the order in which the various replacing amino acids compare in their deleterious effects on collagen stability appears to be: Asp > Arg > Val > Ala, Ser, Cys. This order appears to reflect not only the relative sizes of the amino acid side chains, but also whether they are charged. A more thorough understanding of the relative effects of these substituting amino acids, as well as glutamic acid and tryptophan, on OI phenotypic severity, will be determined with the description of additional OI mutations.

Exceptions to the Phenotypic Gradient Rule

The phenotypic gradient rule predicts that the severity of OI increases with proximity to the carboxy terminus of the α1(I) or α2(I) helical domain. Thus, substitutions near the carboxy terminus result in greater helix destabilization than analogous substitutions near the amino terminus. The rule is based on the fact that the type I procollagen helix is believed to form in a zipper-like fashion in a carboxy- to amino-terminal direction. The rule is also

based on the assumption that the helical conformation is uniform along the entire length of the molecule. However, as discussed above, numerous exceptions have been described which are associated with OI phenotypes which are exceptionally mild. The occurrence of these exceptions can be explained by the following model.

The model proposes that the collagen triple helix, rather than unfolding as a single full-cooperative region, unfolds sequentially through a series of cooperative blocks, each of which unfolds independently of the others (173, 398, 399). In addition, each block has a different ability to destabilize the entire triple helix. In other words, the unfolding of some of these blocks causes the whole helix to become destabilized and to unfold, while others do not. A significant amount of evidence, obtained from the study of collagen defects in OI (398, 399), exists in support of this model. The triple helix has been shown to exhibit sensitivity to proteolytic digestion at specific regions in the helix at pre-denaturation temperatures.

The model also proposes that each glycine position within these blocks has different abilities to tolerate different replacing amino acid residues. Thus, some positions might not tolerate any replacing amino acid, while others would tolerate only amino acids such as serine and cysteine, for example. The ability of different glycine positions to tolerate different amino acids could perhaps be determined by the local concentration of imino acids in the X and Y positions, although this has not been established. The substitution of a specific glycine residue by an amino acid exceeding that position's tolerance level would result in the thermal destabilization of the affected cooperative block. This could then subsequently result in the destabilization of the entire triple helix.

The model is able to explain the observed pattern in the distribution of the exceptions among the different replacing amino acids. Thus, numerous exceptions have been described for amino acids such as cysteine and serine, in comparison with arginine. Cysteine and serine residues have a greater likelihood of being tolerated than arginine, due

to their small size and the absence of a charged side chain.

This model would also predict that exceptions would be less frequent in the α1(I) chain than in the α2(I) chain, due to the 2:1 chain stoichiometry. Substitutions tolerated at certain positions in the α1(I) chain would also be expected to be tolerated at analogous positions in the α2(I) chain. However, the reverse would not be true. Whether this prediction is in fact true is difficult to establish since a much larger number of α1(I) defects have been described, than in the α2(I) chain. The identification of additional glycine substitutions will help to determine whether this model is able to adequately account for the observed patterns in phenotypic severity relative to the nature and positions of the underlying type I collagen defects.

4.7 Effects of Type I Collagen Defects on Bone Architecture

The description of a growing number of OI mutations has been instrumental in examining the correlation between phenotypic severity and specific type I collagen defects, such as glycine substitutions. These studies have contributed to a greater understanding of the consequences of such defects on type I collagen structure and biosynthesis. However, a detailed understanding of the mechanisms by which the main clinical feature of OI, bone fragility, is produced by various type I collagen structural abnormalities remains lacking, as very few studies have examined these patients at the tissue level. In order to investigate the effects of specific collagen defects on bone architecture, bone biopsies were collected and examined by histology. Although mutations were identified in four of the OI patients in this study, bone biopsies were available only from patients NC323 (type III OI) and SS333 (type II OI). Bone from both of these patients exhibited marked abnormalities in architecture relative to their respective age and site-matched normal controls. The changes

in architecture included quantitative deficiencies in ECM production and mineralization. In addition, the persistence of cartilage within bone trabeculae of the diaphysis, and the absence of lamellar bone and mature Haversian systems indicated abnormalities in the bone maturation process by which cartilaginous precursor structures are replaced by woven bone, and subsequently by mature lamellar bone. Despite the distinct severities in their phenotypes, no major differences in the changes in bone architecture exhibited by these patients, relative to their respective controls, were evident. This study underlines the importance of analyzing the bone architecture of OI patients with specific identified defects exhibiting a wide spectrum of phenotypic severities. This is critical in assessing whether the heterogeneity in phenotype and biochemical defects can be correlated with specific quantitative and qualitative changes at the tissue level.

Relatively few OI patients have been examined at the level of bone histology, and fewer still have also had their mutations identified. Cole et al have published three studies of the pathological alterations in bone from six probands with type II OI (192, 193, 207). Two of these studies examined the clinical and pathological features of five infants exhibiting Gly → Arg or Gly → Val substitutions at different positions within the triple helical domain of the pro-α1(I) chain (394-396). The infants exhibited varying phenotypic severities which were reflected in their OI subtyping. Previous biochemical studies had demonstrated that these defects resulted in the synthesis of structurally abnormal type I procollagen molecules exhibiting increased post-translational modification of the constituent chains, decreased thermal stability, and a decreased rate of secretion likely due to elevated intracellular degradation. The decrease in the rate of secretion appeared to correlate with the phenotypic severity. Bone and skin from these infants exhibited diminished quantities of type I collagen. Importantly, the collagen which was incorporated into the ECM consisted of both normal and mutant molecules. The pathological changes decribed in bone from these five patients consisted of deficient endochondral and intramembranous ossification, abnormal matrix production, the absence of lamellar bone and Haversian systems, the

persistence of cartilage within bone trabeculae, and a failure of osteoblasts to mature into osteocytes. The changes in bone architecture were similar between patients, and their magnitudes appeared to correlate with phenotypic severity. These observations indicated clear abnormalities in the processes of bone mineralization and maturation, resulting from the production of insufficient quantities of type I collagen, and the incorporation of structurally abnormal molecules into the ECM.

The type II OI patient in this study, SS333 (α1(I) $Gly_{211} \rightarrow$ Arg), exhibited changes in bone architecture which were similar to some of those described by Cole et al, possibly due to the similar nature of their biochemical defects, Gly $\rightarrow$ Arg substitutions in the α1(I) triple helical domain. The similarity in the nature of the defects suggests that they may produce their effects on bone architecture via the same biochemical mechanisms. That is, the defect in patient SS333 likely results in the secretion of diminished quantities of type I collagen, and in the incorporation of abnormal molecules into the matrix. However, the more amino-terminal location of the Gly $\rightarrow$ Arg substitution in patient SS333 (position 211), in comparison with the infants described by Cole et al (positions 391, 667 and 976), may be expected to result in less extensive intracellular degradation. This, in turn, may be predicted to give rise to less severe changes in bone architecture. However, since this does not appear to be the case, the type II OI phenotype and bone changes observed in patient SS333 may arise more predominantly from the incorporation of larger quantities of abnormal collagen into the matrix, than in the patients studied by Cole et al.

Glycine substitutions affecting different chain positions, but involving the same replacing amino acid, have distinct effects on procollagen thermal stability and secretion, and on phenotypic severity. Thus, it is possible that they may also affect bone architecture to varying degrees, depending on the amount of collagen in the ECM and on the quantity of the structurally abnormal collagen incorporated during fibrillogenesis. The relative importance of these two factors in the pathogenesis in OI, and in determining phenotypic

severity is at present unclear. However, the role of one of these factors, the decreased matrix collagen content, has been addressed in a study of a type IIB infant bearing a 1-bp insertion within exon 49 of the COL1A1 gene (205, 207). This mutation resulted in a translational frameshift within the carboxy-terminal propeptide region of the pro-α1(I) chain, which allowed it to associate with normal pro-α1(I) and pro-α2(I) chains, but prevented the formation of a triple helical structure. Abnormal molecules underwent increased intracellular degradation and inefficient secretion, resulting in tissue collagen contents which were 20% of normal values. A feature which distinguished this frameshift defect from the Gly → Arg and Gly → Val substitutions described above was that no abnormal collagen molecules appeared to be incorporated into the ECM. However, despite the differences in the biochemical effects of these mutations, the changes in bone architecture observed in this patient were similar to those present in the patients bearing glycine substitutions. This suggested that the quantitative deficiency in bone collagen may be sufficient in itself to give rise to the abnormalities in mineralization and maturation, and that the incorporation of abnormal collagens into the ECM is not necessarily required for producing the observed changes in bone architecture and the associated lethal phenotype. Since the observed histological changes constitute more than a quantitative deficiency of mineralized matrix, these observations further suggest that a certain amount of type I collagen is critical in the development of normal bone organization. However, it is nevertheless possible that very small quantities of abnormal collagen are present in the tissues of this patient, and that this may be sufficient to significantly affect bone organization and mineralization, mediated by changes in collagen fibril structure.

It is interesting to note that type I OI patients exhibit decreased quantities of mineralized bone matrix, but an apparently normal lamellar bone organization (Dr. Francis H. Glorieux, personal communication). Since fibroblasts from type I OI patients secrete only half the normal amount of type I procollagen (174, 175), it is presumed that the bone

matrix of these patients would also be deplete of collagen. This indicates that, to some degree, decreased bone collagen content is compatible with normal bone organization. However, below a certain threshold, this will not be the case.

Phenotypic severity likely is not solely determined by the quantity of collagen present in bone ECM, since some OI mutations have been demonstrated to have negligible effects on the thermal stability of the procollagen molecule, while nonetheless being able to give rise to lethal OI phenotypes. For example, Gly → Ser substitutions at positions 631 and 844 of the α1(I) triple helical domain have been identified in patients with type II OI (398, 399). Such defects presumably result in the assembly of abnormal procollagen molecules which are secreted relatively efficiently by skin fibroblasts, in relation to normal procollagen. If osteoblasts process such abnormal procollagens in a manner similar to skin fibroblasts, this would suggest that the lethal phenotypes associated with these types of defects arise predominantly from the incorporation of mutant molecules into the bone ECM. However, it is still possible that such abnormal molecules, although secreted efficiently, may not be efficiently incorporated into fibrils as a result of their conformational abnormalities. These molecules may alternatively be degraded extracellularly by proteases present in the extracellular space, such that the tissues of these patients may be quantitatively deficient in type I collagen. However, the extent to which these abnormal molecules can undergo proteolytic turnover outside the cell is not clear.

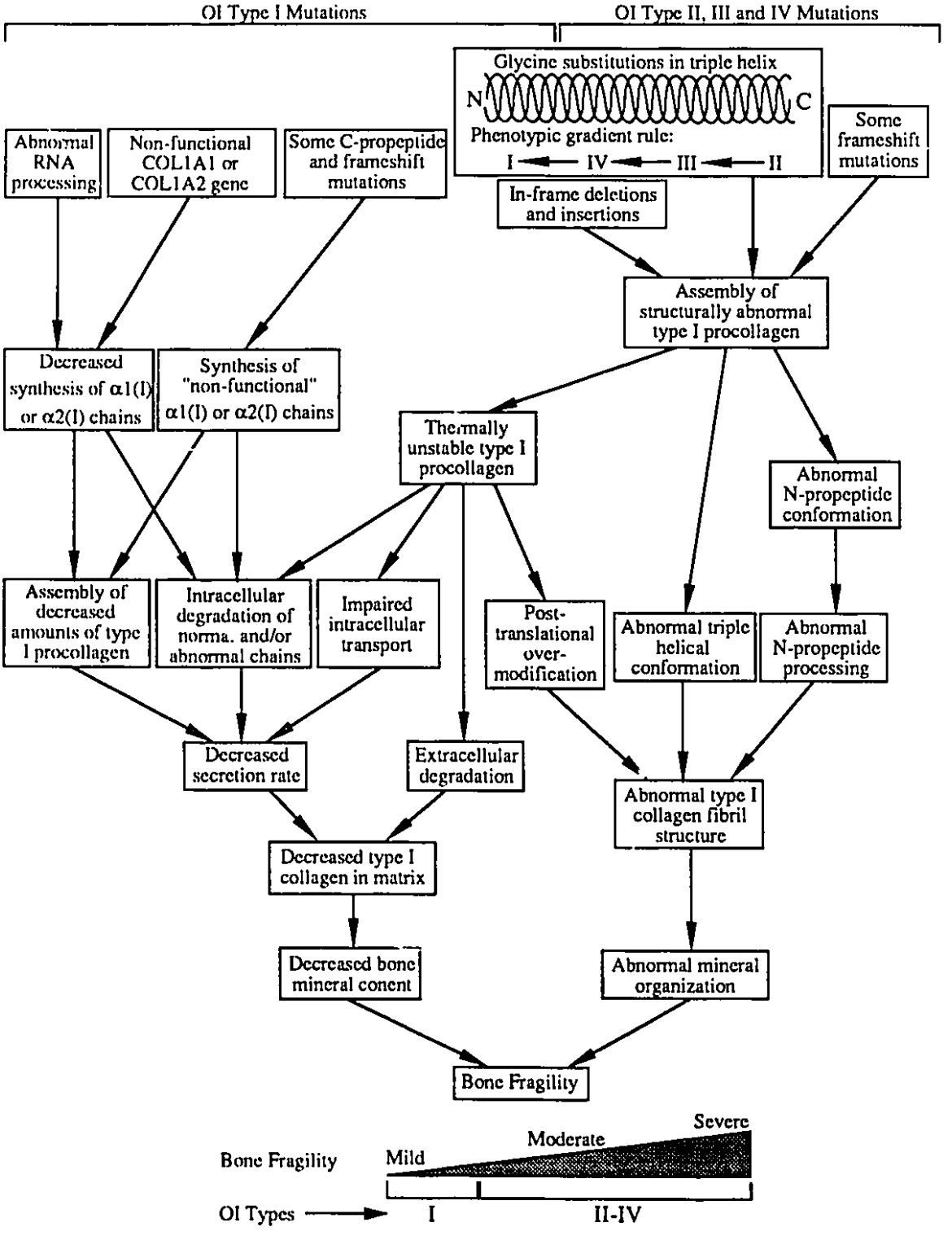
The correlation between OI phenotype, changes in bone architecture and specific collagen defects have not been sufficiently examined. In fact, NC323 is the first reported case of nonlethal OI which has been examined by bone histology, and whose defect has been precisely identified. In spite of their differences in phenotypic severity, there were no major differences in the changes in bone architecture exhibited by patients NC323 and SS333, relative to their respective age-matched controls. A possible explanation for this observation is that patients NC323 and SS333 may be considered more comparable from

the point of view of the biochemical consequences of their mutations, than would otherwise be indicated by their OI classification. The examination of the phenotypic gradients resulting from Gly → Arg substitutions in the α1(I) and α2(I) triple helical domains, and of the defects observed in these patients, indicates that SS333 and NC323 may constitute the mild and severe limits of the type II and type III OI phenotypes, respectively. Thus, their defects, falling within or near the OI type II/OI type III transition regions of the α1(I) and α2(I) chains, may result in only modest differences in bone architecture and bone fragility. Since phenotypic severity depends not only on the biochemical consequences of the mutations, but also to some extent on different environmental factors (mechanical stress, diet etc.), the differential classification of the patients, based on radiological observations and on the length of post-natal survival, may then be the result of differences in the magnitude of various environmental factors experienced during gestation and delivery.

An alternative explanation for the similarities in the changes in bone architecture in patients NC323 and SS333 is that histological analyses are not discriminating enough to distinguish between all of the OI types. For example, it may not be possible to distinguish between OI types II and III. Perhaps only very large differences in phenotypic severity are reflected by differences in appearance on bone histology. More subtle differences in phenotypic severity and bone fragility observed between OI patients may only be perceptible at the ultrastructural level, in terms of the quantity and structure of the mineral phase, which constitutes the main mechanically stabilizing component of bone. The intimate association between type I collagen fibrils and the mineral phase (294) suggests that mutations resulting in different structural defects may result in different alterations in the quantity and structure of the mineral.

Collagen mutations can affect the ECM in two distinct ways. The diverse ways in which various type I collagen defects can give rise to bone fragility are schematically represented in Fig. 48. Firstly, they may result in a decreased number of collagen fibrils

Fig. 48. Flow diagram summarizing the effects of quantitative and structural defects on type I collagen biosynthesis and bone fragility. The various types of mutations which have been described are indicated at the top of the diagram, along with the OI types that they are associated with. The diagram indicates how different types of collagen defects give rise to bone fragility, mediated by their various effects on type I collagen biosynthesis.


OI Type I Mutations
OI Type II, III and IV Mutations
Glycine substitutions in triple helix
N
C
Phenotypic gradient rule:
I ← IV ← III ← II
Abnormal RNA processing
Non-functional COL1A1 or COL1A2 gene
Some C-propeptide and frameshift mutations
In-frame deletions and insertions
Some frameshift mutations
Assembly of structurally abnormal type I procollagen
Decreased synthesis of α1(I) or α2(I) chains
Synthesis of "non-functional" α1(I) or α2(I) chains
Thermally unstable type I procollagen
Abnormal N-propeptide conformation
Assembly of decreased amounts of type I procollagen
Intracellular degradation of norma. and/or abnormal chains
Impaired intracellular transport
Post-translational over-modification
Abnormal triple helical conformation
Abnormal N-propeptide processing
Decreased secretion rate
Extracellular degradation
Abnormal type I collagen fibril structure
Decreased type I collagen in matrix
Decreased bone mineral conent
Abnormal mineral organization
Bone Fragility
Bone Fragility
Mild
Moderate
Severe
OI Types
I
II-IV

due to a decrease in the rate of collagen secretion. The decreased number of fibrils results in a decreased amount of mineral deposited in the matrix, since most of the mineral is normally deposited in the interior of collagen fibrils. Secondly, the mutations may result in the assembly of structurally abnormal fibrils due to the incorporation of collagen molecules with altered conformations. The effects of an altered structure of collagen fibrils on mineral deposition and structure has not been examined. However, it seems likely that the effects could be mediated by alterations in the rate of mineral deposition, as well as alterations in the structure of the mineral phase itself. The rate of mineralization could be affected as a result of changes in the ability of various bone matrix proteins, involved in the process of mineral deposition, to access or interact with the hole zones of the collagen fibril. Abnormalities in the structure of the mineral phase may consist of an altered geometry of the plate-like crystals and/or of the alignment of the crystals relative to each other within the collagen fibril. Changes in the geometry of individual crystals could possibly arise from abnormalities in the shape and size of the hole zone, the arrangement of collagen functional groups perhaps involved in mineral deposition, and the abnormal interactions with proteins involved in mineralization. In contrast, changes in crystal alignment may result if the crystals formed are structurally normal, but are no longer in parallel rows due to the altered packing of collagen molecules within the fibril.

The apparent abnormalities in bone maturation observed in patients NC323 and SS333, and those described by Cole et al, may be explained by the differences in the mechanisms by which woven and lamellar bone are mineralized. Woven bone is mineralized by a vesicle-mediated process which appears to take place predominantly at the surfaces of collagen fibrils (296, 409). However, the mineralization of lamellar bone occurs by a process in which crystals form and grow within the hole zones in the interior of the fibril (294, 296). Since the latter process is likely to be more dependent on the structure of the collagen fibril, it would be expected to be more susceptible to collagen structural defects.

Thus, the presence of structurally abnormal collagen fibrils in OI bone may only permit the mineralization of immature woven bone.

In conclusion, the histological analysis of additional nonlethal cases of OI will be necessary to determine the nature of the relationship between phenotypic severity and the quantitative and organizational defects observed at the tissue level. Also, it appears feasible that additional ultrastructural studies of the structure of the bone mineral in OI patients may be required in order to comprehend how different collagen defects result in a spectrum of bone fragility. These studies will also contribute to our understanding of how different collagen defects can also give rise to similarities in bone architecture.

Conclusions

The present study has evaluated and optimized techniques for the detection and identification of gene mutations. These methods consisted of the chemical cleavage detection of mismatches in mRNA:cDNA heteroduplexes and the cloning of PCR products for subsequent DNA sequence analysis, and were applied to the characterization of polymorphisms and OI mutations in the genes encoding type I collagen.

The chemical cleavage analysis of the prepro-α1(I) and prepro-α2(I) mRNAs from control and OI fibroblasts identified three HA- and OT-reactive polymorphisms, two of which have not yet been described. These novel polymorphisms may serve as useful markers in the linkage analysis of inherited diseases resulting from mutations in the COL1A1 and COL1A2 genes, and genes in closely linked loci. The COL1A1 wobble base polymorphism, although rare, may be of particular use, due to the paucity of markers thus far demonstrated in this gene. The COL1A2 gene sequence variation gives rise to a Y-position proline/alanine dimorphism within the pro-α2(I) triple helical domain. Thus, although glycine residues in the first position of the Gly-X-Y triplet are required for normal triple helix formation, at least some X- and Y-position residues appear to be tolerant of certain substitutions. Such neutral sequence changes must be considered when examining OI patients for type I collagen mutations.

Despite the fact that all four possible base changes were not examined by this analysis, it appears that polymorphisms within the COL1A1 and COL1A2 coding regions are rare, suggesting that these genes exhibit strong evolutionary sequence conservation. The rarity of coding sequence polymorphisms is unexpected, since there would appear to be no significant selective advantage in maintaining the identity of bases in functionally neutral positions (eg. codon wobble positions).

This study has also examined a group of ten patients exhibiting different forms of OI, with the purpose of further defining the correlation between phenotypic severity and specific type I collagen defects. The biochemical data for fibroblasts from three of these patients (GD199, PM244, MB252), exhibiting type I OI, are consistent with the presence of mutations, likely within the COL1A1 gene, resulting in quantitative defects in type I collagen synthesis. Although no specific mutations were identified in these patients, the results are in agreement with the literature on this mild form of the disease.

The severe phenotype in type III OI patient MK345 and the apparent absence of collagen overmodification were compatible with the presence of a defect within the carboxy-terminal propeptide of the pro-α1(I) chain. However, no mutations were detected, and the data do not exclude the possibility of a non-collagen mutation in this patient.

In two of the other patients (SL211, type IV OI; DP266, type III OI) and in ML226, the asymptomatic father of SL211, apparent collagen structural abnormalities were observed, yet no mutations were detected by chemical cleavage analysis. The protein data were consistent with the presence of defects within the triple helical domain of either chain. If the chemical cleavage method is indeed capable of detecting all possible C-base mismatches, the data would suggest that the mutations in these patients may involve X or Y-position residues, defects which have not yet been reported in OI.

In the remaining four OI patients, novel point mutations resulting in the substitution of glycine residues within the triple helical domain of either the pro-α1(I) or pro-α2(I) chain were identified. The defect in type II OI patient SS333 (α1 $Gly_{211} \rightarrow$ Arg) is compatible with previously described Gly → Arg substitutions in the α1(I) chain, being located in a region bounded by lethal defects on the carboxy-terminal side, and non-lethal defects on the amino-terminal side. This substitution is interesting, since it more precisely defines the region of the α1(I) helical domain in which the OI phenotype undergoes a transition

between lethal and non-lethal forms. The defect in type III OI patient NC323 (α2 Gly_{427} $\rightarrow$ Arg) also appears to identify the corresponding phenotypic transition region of the α2(I) chain, due to its proximity to a previously reported lethal defect (α2 Gly_{457} $\rightarrow$ Arg). Therefore, the defect in patient NC323 is important since it identifies the amino-terminal extent of the lethal phenotype associated with Gly $\rightarrow$ Arg substitutions in the α2(I) chain. The defect in type IV OI patient PP238 (α2 Gly_{544} $\rightarrow$ Val) is consistent with the phenotypic gradient rule, being located near a previously reported defect also associated with a moderate phenotype (α2 Gly_{586} $\rightarrow$ Val). In contrast, the defect identified in type IV OI patient JN245 (α2 Gly_{922} $\rightarrow$ Ser) resulted in a phenotype of unexpected mildness. This substitution, although located at a more carboxy-terminal position than a previously described lethal defect (α2 Gly_{865} $\rightarrow$ Ser), resulted in a less severe phenotype. Therefore, the substitution in patient JN245 identifies a region of the α2(I) helical domain which appears to be tolerant of at least some types of defects.

Taken together, the defects demonstrated in three of these OI patients identify regions of the α1(I) and α2(I) helical domains in which the disease severity undergoes position-dependent transitions. The identification of the phenotypic transition regions associated with Gly $\rightarrow$ Arg substitutions is of value since these are likely the most commonly occurring single type of defect in OI. These results contribute to the existing database of OI mutations, and will assist in the construction of more detailed phenotypic maps of the type I collagen chains. In turn, this information will be critical in providing parents with more accurate prognosis of phenotypic severity in affected offspring.

Suggestions for Future Work

Although significant progress has been made in the last five years in the understanding of OI, predominantly through the identification of the mutations in a large number of patients, many questions are yet to be answered. The following work should assist in resolving some of these problems.

1) The mutations in additional OI patients should be identified in order to further define the correlation between specific type I collagen defects and phenotypic severity. This will allow more detailed "phenotypic maps" to be constructed for the different types of defects in the pro-α1(I) and pro-α2(I) chains. As a consequence, more accurate prognosis of phenotypic severity will be possible, thereby allowing prospective parents to make responsible decisions. In addition, the description of additional mutations, and their effects on biochemical parameters, such as thermal stability, efficiency of secretion and collagen fibril structure, will help to determine the manner in which type I collagen defects give rise to bone fragility.

2) In examining the relationship between phenotypic severity and type I collagen defects, it will be necessary to demonstrate that the OI patients studied are heterozygous for the identified defects. This is due to the possible occurrence of somatic cell mosaicism, which has been reported in several cases of OI, and which can result in phenotypes of unexpectedly mild severity. Thus, it will be essential to exclude mosaicism as a possible explanation for phenotypes which are apparently incompatible with the phenotypic severity rules.

3) Whenever possible, OI patients whose type I collagen defects have been identified, should be examined at the level of bone histology. This will be necessary to assess the relationship between various type I collagen defects and their effects on the quantity and organization of mineral in bone of OI patients. Ideally, these analyses should be performed on bone from patients exhibiting a wide range of phenotypic severities, and would reveal whether the heterogeneity in clinical severity is reflected at the tissue level.

4) The effects of different type I collagen defects on the structure of the collagen fibrils and mineral phase present in bone should be investigated. This will help to determine whether the bone fragility in OI results predominantly from insufficient quantities of type I collagen in the matrix, abnormalities in collagen fibril structure, or alterations in the organization or structure of the mineral crystals themselves.

5) The transgenic mouse model of OI should be utilized to examine the effects of specific structural type I collagen defects on phenotype and various biochemical parameters (thermal stability, secretion etc.). This approach may be particularly useful for assessing the consequences of defects which are expected to occur rarely in nature (eg. Gly → Trp substitutions) or which are located at or near specific sites of interest (eg. helical cross-linking sites).

6) Additional individuals should be examined for sequence polymorphisms in the coding regions of the type I collagen genes. This information will be useful in generating additional markers for these genes, as well as providing data about the evolutionary conservation of their coding sequences. These studies will complement the study of OI patients by identifying those regions of the type I collagen chains which are tolerant of sequence variation.

Statements of Original Contributions

1) The work presented in this thesis has optimized the chemical cleavage method for base mismatch detection, a technique which is useful for the detection of gene mutations. It has improved the applicability of this method to the analysis of mRNA in general, by identifying parameters which significantly affect mismatch detection.

2) This study describes the optimization of a method for the reliable and efficient blunt-end cloning of PCR products for subsequent sequence analysis. The work has identified technical parameters which significantly affect cloning efficiencies.

3) This study provides important data on the frequency and distribution of polymorphisms in the coding regions of the COL1A1 and COL1A2 genes. It describes two novel polymorphisms which may serve as useful markers for the linkage analysis of inherited disorders associated with mutations in these genes, and with genes in closely linked loci. This work is also the first to describe a type I collagen gene polymorphism resulting in an alteration at the level of the amino acid sequence.

4) This study has identified the mutations in four OI patients, giving rise to novel glycine substitutions within the triple helical domains of the pro-α1(I) and pro-α2(I) chains. The Gly → Arg substitutions demonstrated in patients SS333 and NC323 allowed the identification of the regions of the pro-α1(I) and pro-α2(I) triple helical domains in which the phenotype associated with these defects exhibits a transition between lethal and non-lethal forms. In addition, the pro-α2(I) Gly → Ser substitution demonstrated in patient JN245 identified a region of the pro-α2(I) triple helical domain in which certain types of defects are tolerated. This data contributes to the existing database of biochemical defects

described in OI, and provides important new information about the relationship between the nature and location of type I collagen defects and the phenotypic severity.

5) This study is the first to report the histological analysis of bone from a non-lethal case of OI for which the biochemical defect has been identified. Such information is critical to understanding how various type I collagen defects, resulting in OI of varying severity, affect bone architecture and result in bone fragility.

Bibliography

1 Hay, E.D., ed. (1991) Cell Biology of Extracellular Matrix, 2nd ed. , Plenum, New York, NY

2 Miller, E.J. (1984) *In*: Extracellular Matrix Biochemistry (Piez, K.A. and Reddi, A.H., Eds.), pp. 41-81, Elsevier, New York, NY

3 Piez, K.A. (1984) *In*: Extracellular Matrix Biochemistry (Piez, K.A. and Reddi, A.H., Eds.), pp. 1-39, Elsevier, New York, NY

4 Solursh, M. (1989) Am. J. Med. Genet. **34**:30-34

5 Hay, E.D. (1989) Am. J. Med. Genet. **34**:14-29

6 Hay, E.D. (1991) *In*: Cell Biology of Extracellular Matrix, 2nd ed. (Hay, E.D., Ed.), pp. 419-462, Plenum, New York, NY

7 van der Rest, M. and Garrone, R. (1991) FASEB J. **5**:2814-2823

8 Reid, K.B.M. (1974) Biochem. J. **141**:189-203

9 Rosenberry, T.L. and Richardson, J.M. (1977) Biochemistry **16**:3550-3558

10 Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) J. Biol. Chem. **261**:6878-6887

11 Ramachandran, G.N. and Kartha, G. (1954) Nature **174**:269-270

12 Rich, A. and Crick, F.H.C. (1955) Nature **176**:915-916

13 van der Rest, M. and Mayne, R. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 195-221, Academic, New York, NY

14 Ninomiya, Y. and Olsen, B.R. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**:3014-3018

15 Sage, H. and Bornstein, P. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 173-194, Academic, New York, NY

16 Yamaguchi, N., Benya, P.D., van der Rest, M. and Ninomiya, Y. (1989) J. Biol. Chem. **264**:16022-16029

17 Fessler, J.H. and Fessler, L.I. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 81-103, Academic, New York, NY

18 Vuorio, E. and de Crombrugghe, B. (1990) Ann. Rev. Biochem. **59**:837-872

19 Petruska, J.A. and Hodge, A.J. (1964) Proc. Natl. Acad. Sci. U.S.A. **51**:871-876

20 Fietzek, P.P. and Kuhn, K. (1976) Int. Rev. Conn. Tissue Res. **7**:1-60

21 Fleischmajer, R., Timpl, R., Tuderman, L., Raisher, L., Wiestner, M., Perlish, J.S. and Graves, P.N. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**:7360-7364

22 Trus, B.L. and Piez, K.A. (1976) J. Mol. Biol. **108**:705-732

23 Hofmann, H., Fietzek, P.P. and Kuhn, K. (1978) J. Mol. Biol. **125**:137-165

24 Hodge, A.J. and Schmitt, F.O. (1960) Proc. Natl. Acad. Sci. U.S.A. **46**:186-197

25 Stark, M., Miller, E.J. and Kuhn, K. (1972) Eur. J. Biochem. **27**:192-196

26 Miller, E.J. (1976) Mol. Cell. Biochem. **13**:165-192

27 Smith, J.W. (1968) Nature **219**:157-158

28 Woodhead-Galloway, J. (1982) *In*: Collagen in Health and Disease (Weiss, J.B. and Jayson, M.I.V., Eds.), pp. 28-48, Churchill Livingstone, New York, NY

29 Miller, E.J. and Gay, S. (1987) Methods Enzymol. **144**:3-41

30 Ala-Kokko, L., Kontusaari, S., Baldwin, C.T., Kuivaniemi, H. and Prockop, D.J. (1989) Biochem. J. **260**:509-516

31 Burgeson, R.E. and Hollister, D.W. (1979) Biochem. Biophys. Res. Commun. **87**:1124-1131

32 Eyre, D. and Wu, J.-J. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 261-281, Academic, New York, NY

33 Niyibizi, C. and Eyre, D.R. (1989) FEBS Lett. **242**:314-318

34 Keene, D.R., Sakai, L.Y., Bachinger, H.P. and Burgeson, R.E. (1987) J. Cell Biol. **105**:2393-2402

35 Mendler, M., Eich-Bender, S.G., Vaughan, L., Winterhalter, K.H. and Bruckner, P. (1989) J. Cell Biol. **108**:191-197

36 Birk, D.E., Fitch, J.M., Babiarz, J.P. and Linsenmayer, T.F. (1988) J. Cell Biol. **106**:999-1008

37 Paulsson, M. (1992) Crit. Rev. Biochem. Mol. Biol. **27**:93-127

38 Glanville, R.W. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 43-79, Academic, New York, NY

39 Hostikka, S.L., Eddy, R.L., Byers, M.G., Hoyhtya, M., Shows, T.B. and Tryggvason, K. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**:1606-1610

40 Yurchenko, P.D. and Schittny, J.C. (1990) FASEB J. **4**:1577-1590

41 Noonan, D.M., Fulle, A., Valente, P., Cai, S., Horigan, E., Sasaki, M., Yamada, Y. and Hassell, J.R. (1991) J. Biol. Chem. **266**:22939-22947

42 Sakai, L.Y., Engvall, E., Hollister, D.W. and Burgeson, R.E. (1982) Am. J. Pathol. **108**:310-318

43 Labermeier, U. and Kenney, M.C. (1983) Biochem. Biophys. Res. Commun. **116**:619-625

44 Kapoor, R., Bornstein, P. and Sage, H. (1985) *In*: Basement Membranes (Shibata, S., Ed.), pp. 105-118, Elsevier, New York, NY

45 Muragaki, Y., Jacenko, O., Apte, S., Mattei, M.-G., Ninomiya, Y. and Olsen, B.R. (1991) J. Biol. Chem. **266**:7721-7727

46 Sawada, H., Konomi, H. and Hirosawa, K. (1990) J. Cell Biol. **110**:219-227

47 Benya, P.D. and Padilla, S.R. (1986) J. Biol. Chem. **261**:4160-4169

48 Schmid, T.M. and Linsenmayer, T.F. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 223-259, Academic, New York, NY

49 Vasios, G., Ninomiya, Y. and Olsen, B.R. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 283-309, Academic, New York, NY

50 Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. and Olsen, B.R. (1986) J. Biol. Chem. **261**:5041-5050

51 Schmid, T.M. and Linsenmayer, T.F. (1990) Dev. Biol. 138:53-62

52 Jander, R., Rauterberg, J. and Glanville, R.W. (1983) Eur. J. Biochem. **133**:39-46

53 Chu, M.-L., Zhang, R.-Z., Pan, T.C., Stokes, D., Conway, D., Kuo, H.-J., Glanville, R.W., Mayer, U., Mann, K., Deutzmann, R. and Timpl, R. (1990) EMBO J. **9**:385-393

54 Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. and Colombatti, A. (1990) Biochemistry **29**:1245-1254

55 Timpl, R. and Engel, J. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 105-143, Academic, New York, NY

56 Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. and Engel, J. (1983) Biochem. J. **211**:303-311

57 Bruns, R.R., Press, W., Engvall, E., Timpl, R. and Gross, J. (1986) J. Cell Biol. **103**:393-404

58 Burgeson, R.E. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 145-172, Academic, New York, NY

59 Sakai, L.Y., Keene, D.R., Morris, N.P. and Burgeson, R.E. (1986) J. Cell Biol. **103**:1577-15860

60 Keene, D.R., Sakai, L.Y., Lunstrum, G.P., Morris, N.P. and Burgeson, R.E. (1987) J. Cell Biol. **104**:611-621

61 van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N.G., Chretien, M. and Olsen, B.R. (1985) J. Biol. Chem. **260**:220-225

62 Huber, S., van der Rest, M., Bruckner, P., Rodriguez, E., Winterhalter, K.H. and Vaughan, L. (1986) J. Biol. Chem. **261**:5965-5968

63 Muragaki, Y., Nishimura, I., Henney, A., Ninomiya, Y. and Olsen, B.R. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**:2400-2404

64 Brewton, R.G., Wright, D.W. and Mayne, R. (1991) J. Biol. Chem. **266**:4752-4757

65 Eyre, D.R., Apon, S., Wu, J.-J., Ericsson, L.H. and Walsh, K.A. (1987) FEBS Lett. **220**:337-341

66 van der Rest, M. and Mayne, R. (1988) J. Biol. Chem. **263**:1615-1618

67 Vasios, G., Nishimura, I., Konomi, H., van der Rest, M., Ninomiya, Y. and Olsen, B.R. (1988) J. Biol. Chem. **263**:2324-2329

68 Gordon, M.K., Gerecke, D.R. and Olsen, B.R. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**:6040-6044

69 Dublet, B., Oh, S., Sugrue, S.P., Gordon, M.K., Gerecke, D.R., Olsen, B.R. and van der Rest, M. (1989) J. Biol. Chem. **264**:13150-13156

70 Dublet, B. and van der Rest, M. (1991) J. Biol. Chem. **266**:6853-6858

71 Kuhn, K. (1987) *In*: Biology of Extracellular Matrix: Structure and Function of Collagen Types (Mayne, R. and Burgeson, R.E., Eds.), pp. 1-42, Academic, New York, NY

72 Kivirikko, K.I. and Myllyla, R. (1984) *In*: Extracellular Matrix Biochemistry (Piez, K.A. and Reddi, A.H., Eds.), pp. 83-118, Elsevier, New York, NY

73 Cheah, K.S.E. and Grant, M.E. (1982) *In*: Collagen in Health and Disease (Weiss, J.B. and Jayson, M.I.V., Eds.), pp. 73-100, Churchill Livingstone, New York, NY

74 de Wet, W.J., Chu, M.-L. and Prockop, D.J. (1983) J. Biol.Chem. **258**:14385-14389

75 Myers, J.C., Chu, M.-L., Faro, S.H., Clark, W.J., Prockop, D.J. and Ramirez, F. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**:3516-3520

76 Chu, M.-L., Myers, J.C., Bernard, M.P., Ding, J.-F. and Ramirez, F. (1982) Nucleic Acids Res. **10**:5925-5934

77 Bernard, M.P., Chu, M.-L., Myers, J.C., Ramirez, F., Eikenberry, E.F. and Prockop, D.J. (1983) Biochemistry **22**: 5213-5223

78 Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C.T., Jaenisch, R. and Prockop, D.J. (1988) Biochem. J. **253**: 919-922

79 Bernard, M.P., Myers, J.C., Chu, M.-L., Ramirez, F., Eikenberry, E.F. and Prockop, D.J. (1983) Biochemistry **22**: 1139-1145

80 Kuivaniemi, H., Tromp, G., Chu, M.-L. and Prockop, D.J. (1988) Biochem. J. **252**: 633-640

81 Baldwin, C.T., Reginato, A.M., Smith, C., Jimenez, S.A. and Prockop, D.J. (1989) Biochem. J. **262**:521-528

82 Cheah, K.S.E., Stoker, N.G., Griffin, J.R., Grosveld, F.G. and Solomon, E. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**:2555-2559

83 Rapoport, T.A. (1986) CRC Crit. Rev. Biochem. **20**:73-137

84 Eyre, D.R. (1984) Ann. Rev. Biochem. **53**:717-748

85 Palmiter, R.D., Davidson, J.M., Gagnon, J., Rowe, D.W. and Bornstein, P. (1979) J. Biol. Chem. **254**:1433-1436

86 Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. **67**:835-851

87 Jackson, R.C. and Blobel, G. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**:5598-5602

88 Kivirikko, K.I. and Myllyla, R. (1980) *In*: The Enzymology of Post-translational Modifications of Protein (Freedman, R.B. and Hawkins, H.C., Eds.), pp. 53-104, Academic, London

89 Kivirikko, K.I., Myllyla, R. and Pihlajaniemi, T. (1989) FASEB J. **3**:1609-1617

90 Prockop, D J., Berg, R.A., Kivirikko, K.I. and Uitto, J. (1976) *In*: Biochemistry of Collagen (Ramachandran, G.N. and Reddi, A.H., Eds.), pp. 163-273, Plenum, New York, NY

91 Fessler, L.I. and Fessler, J.H. (1974) J. Biol. Chem. **249**:7637-7646

92 Kivirikko, K.I. and Myllyla, R. (1979) Int. Rev. Connec. Tiss. Res. **8**:23-72

93 Clark, C.C. (1979) J. Biol. Chem. **254**:10798-10802

94 Snider, M.D. (1984) *In*: Biology of Carbohydrates (Ginsburg, V. and Robbins, P.W., Eds.), **2**:163-198, Wiley & Sons, New York, NY

95 Kornfeld, R. and Kornfeld, S. (1985) Ann. Rev. Biochem. **54**:631-664

96 Housley, T.J., Rowland, F.N., Ledger, P.W., Kaplan, J. and Tanzer, M.L. (1980) J. Biol. Chem. **255**:121-128

97 Bachinger, H.P., Fessler, L.I., Timpl, R. and Fessler, J.H. (1981) J. Biol. Chem. **256**:13193-13199

98 Bachinger, H.P., Bruckner, P., Timpl, R., Prockop, D.J. and Engel, J. (1980) Eur. J. Biochem. **106**:619-632

99 Bruckner, P., Eikenberry, E.F. and Prockop, D.J. (1981) Eur. J. Biochem. **118**:607-613

100 Steinmann, B., Bruckner, P. and Superti-Furga, A. (1991) J. Biol. Chem. **266**:1299-1303

101 Trelstad, R.L. and Hayashi, K. (1979) Dev. Biol. **71**:228-242

102 Hojima, Y., van der Rest, M. and Prockop, D.J. (1985) J. Biol. Chem. **260**:15996-16003

103 Hojima, Y., McKenzie, J., van der Rest, M. and Prockop, D.J. (1989) J. Biol. Chem. **264**:11336-11345

104 Bateman, J F. and Golub, S.B. (1990) Biochem. J. **267**:573-577

105 Gerstenfeld, L., Beldekas, J.C., Sonenshein, G.E. and Franzblau, C. (1984) J. Biol. Chem. **259**:9158-9162

106 Fessler, L.I. and Fessler, J.H. (1979) J. Biol. Chem. **254**:233-239

107 Miyahara, M., Hayashi, K., Berger, J., Tanzawa, K., Njieha, F.K., Trelstad, R.L. and Prockop, D.J. (1984) J. Biol. Chem. **259**:9891-9898

108 Paglia, L., Wilczek, J., de Leon, L.D., Martin, G.R., Horlein, D. and Muller, P. (1979) Biochemistry **18**:5030-5034

109 Williams, B.R., Gelman, R.A., Poppke, D.C. and Piez, K.A. (1978) J. Biol. Chem. **253**:6578-6585

110 Gelman, R.A., Williams, B.R. and Piez, K.A. (1979) J. Biol. Chem. **254**:180-186

111 Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1973) J. Mol. Biol. **79**:137-148

112 Tanzer, M.L. (1973) Science **180**:561-566

113 Pinnell, S.R. and Martin, G.R. (1968) Proc. Natl. Acad. Sci. U.S.A. **61**:708-714

114 Siegel, R.C. (1979) Int. Rev. Conn. Tiss. Res. **8**:73-118

115 Siegel, R.C. and Fu, J.C.C. (1976) J. Biol. Chem. **251**:5779-5785

116 Cronlund, A.L., Smith, B.D. and Kagan, H.M. (1985) Conn. Tissue Res. **14**:109-119

117 Davis, N.R. and Bailey, A.J. (1971) Biochem. Biophys. Res. Comm. **45**:1416-1422

118 Barnes, M.J., Constable, B.J., Morton, L.F. and Kodicek, E. (1971) Biochem. J. **125**:433-437

119 Barnes, M.J., Constable, B.J., Morton, L.F. and Kodicek, E. (1971) Biochem. J. **125**:925-928

120 Fujimoto, D., Akiba, K.-Y. and Nakamura, N. (1977) Biochem. Biophys. Res. Comm. **76**:1124-1129

121 Eyre, D.R. (1981) *In*: The Chemistry and Biology of Mineralized Connective Tissues (Veis, A., Ed.), pp. 51-55, Elsevier, New York, NY

122 Ogawa, T., Ono, T., Tsuda, M. and Kawanishi, Y. (1982) Biochem. Biophys. Res. Comm. **107**:1252-1257

123 Kuypers, R., Tyler, M., Kurth, L.B., Jenkins, I.D. and Horgan, D.J. (1992) Biochem. J. **283**:129-136

124 Yamauchi, M., London, R.E., Guenat, C., Hashimoto, F. and Mechanic, G.L. (1987) J. Biol. Chem. **262**:11428-11434

125 Bornstein, P. and Traub, W. (1979) *In*: The Proteins, 3rd ed. (Neurath, H. and Hill, R.L., Eds.), 4:411-632, Academic, New York, NY

126 Tanzer, M.L. (1965) Int. Rev. Conn. Tiss. Res. 3:91-112

127 Danks, D.M. (1989) *In*: The Metabolic Basis of Inherited Disease, 6th ed. (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., Eds.), pp. 1411-1431, McGraw-Hill, New York, NY

128 Mudd, S.H., Levy, H.L. and Skovby, F. (1989) *In*: The Metabolic Basis of Inherited Disease, 6th ed. (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., Eds.), pp. 693-734, McGraw-Hill, New York, NY

129 D'Alessio, M., Bernard, M., Pretorius, P.J., de Wet, W. and Ramirez, F. (1988) Gene **67**:105-115

130 Westerhausen, A., Constantinou, C.D., Pack, M., Peng, M., Hanning, C., Olsen, A.S. and Prockop, D.J. (1991) Matrix **11**:375-379

131 Yamada, Y., Avvedimento, V.E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. and de Crombrugghe, B. (1980) Cell **22**:887-892

132 Khillan, J.S., Schmidt, A., Overbeek, P.A., de Crombrugghe, B. and Westphal, H. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**:725-729

133 Yamada, Y., Miyashita, T., Savagner, P., Horton, W., Brown, K.S., Abramczuk, J., Hou-Xiang, X., Kohno, K., Bolander, M. and Bruggeman, L. (1990) Ann. N.Y. Acad. Sci. **580**:81-87

134 Karsenty, G., Golumbek, P. and de Crombrugghe, B. (1988) J. Biol. Chem. **263**:13909-13915

135 Brenner, D.A., Rippe, R.A. and Veloz, L. (1989) Nucleic Acids Res. **17**:6055-6064

136 Hatamochi, A., Golumbek, P.T., Van Schaftingen, E. and de Crombrugghe, B. (1988) J. Biol. Chem. **263**:5940-5947

137 Ruteshouser, E.C. and de Crombrugghe, B. (1989) J. Biol. Chem. **264**:13740-13744

138 Ignotz, R.A., Endo, T. and Massague, J. (1987) J. Biol. Chem. **262**:6443-6446

139 Goldring, M.B. and Krane, S.M. (1987) J. Biol. Chem. **262**:16724-16729

140 Solis-Herruzo, J.A., Brenner, D.A. and Chojkier, M. (1988) J. Biol. Chem. **263**:5841-5845

141 Hamalainen, L., Oikarinen, J. and Kivirikko, K.I. (1985) J. Biol. Chem. **260**:720-725

142 Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and de Crombrugghe, B. (1988) Cell **52**:405-414

143 Rossouw, C.M.S., Vergeer, W.P., du Plooy, S.J., Bernard, M.P., Ramirez, F. and de Wet, W.J. (1987) J. Biol. Chem. **262**:15151-15157

144 Rossi, P. and de Crombrugghe, B. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**:5590-5594

145 Horton, W., Miyashita, T., Kohno, K., Hassell, J.R. and Yamada, Y. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**:8864-8868

146 Bornstein, P., McKay, J., Morishima, J.K., Devarayalu, S. and Gelinas, R.E. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**:8869-8873

147 Bornstein, P., McKay, J., Liska, D.J., Apone, S. and Devarayalu, S. (1988) Mol. Cell. Biol. **8**:4851-4857

148 Schniecke, A., Harbers, K. and Jaenisch, R. (1983) Nature **304**:315-320

149 Kratochwil, K., von der Mark, K., Kollar, E.J., Jaenisch, R., Mooslehner, K., Schwarz, M., Haase, K., Gmachl, I. and Harbers, K. (1989) Cell **57**:807-816

150 Tajima, S. and Pinnell, S.R. (1982) Biochem. Biophys. Res. Commun. **106**:632-637

151 Chojkier, M., Houglum K., Solis-Herruzo, J. and Brenner, D.A. (1989) J. Biol. Chem. **264**:16957-16962

152 Hentze, M.W. (1991) Biochim. Biophys. Acta **1090**:281-292

153 Penttinen, R.P., Kobayashi, S. and Bornstein, P. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**:1105-1108

154 Lyons, B.L. and Schwarz, R.I. (1984) Nucleic Acids Res. **12**:2569-2579

155 Yamada, Y., Mudryj, M. and de Crombrugghe, B. (19[illegible] J. Biol. Chem. **258**:14914-14919

156 Sillence, D.O. (1988) Ann. N.Y. Acad. Sci. **543**:1-15

157 Rowe, D.W. and Shapiro, J.R. (1990) *In*: Metabolic Bone Disease and Clinically Related Disorders (Avioli, L.V. and Krane, S.M., Eds.), pp. 659-701, W.B. Saunders, Philadelphia, PA

158 Byers, P.H. (1989) *In*: The Metabolic Basis of Inherited Disease, 6th ed. (Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., Eds.), pp. 2805-2842, McGraw-Hill, New York, NY

159 Komai, T., Kunii, H. and Ozaki, Y. (1956) Am. J. Hum. Genet. **8**:110-119

160 Sillence, D.O (1981) Clin. Orthoped. Rel. Res. **159**:11-25

161 Sillence, D.O., Senn, A. and Danks, D.M. (1979) J. Med. Genet. **16**:101-116

162 Sillence, D.O., Rimoin, D.L. and Danks, D.M. (1979) *In*: Birth Defects: Original Article Series - Penetrance and Variability in Malformation Syndromes (O'Donnell, J.J. and Hall, B.D., Eds.), **15**:113-129, Alan R. Liss, New York, NY

163 Beighton, P., de Paepe, A., Danks, D., Finidori, G., Gedde-Dahl, T., Goodman, R., Hall, J.G., Hollister, D.W., Horton, W., McKusick, V.A., Opitz, J.M., Pope, F.M., Pyeritz, R.E., Rimoin, D.L., Dillence, D., Spranger, J.W., Thompson, E., Tsipouras, P., Viljoen, D., Winship, I. and Young, I. (1988) Am. J. Med. Genet. **29**:581-594

164 Sillence, D.O., Barlow, K.K., Garber, A.P., Hall, J.G. and Rimoin, D.L. (1984) Am. J. Med. Genet. **17**:407-423

165 Byers, P.H., Tsipouras, P., Bonadio, J.F., Starman, B.J. and Schwartz, R.C. (1988) Am. J. Hum. Genet. **42**:237-248

166 Sillence, D.O., Barlow, K.K., Cole, W.G., Dietrich, S., Garber, A.P. and Rimoin, D.L. (1986) Am. J. Med. Genet. **23**:821-832

167 Maroteaux, P., Frézal, J. and Cohen-Solal, L. (1986) Am. J. Med. Genet. **24**:219-230

168 Maroteaux, P., Cohen-Solal, L. and Bonaventure, J. (1988) Ann. N.Y. Acad. Sci. **543**:16-29

169 Wenstrup, R.J., Willing, M.C., Starman, B.J. and Byers, P.H. (1990) Am. J. Hum. Genet. **46**:975-982

170 Sykes, B., Ogilvie, D., Wordsworth, P., Anderson, J. and Jones, N. (1986) Lancet **ii**:69-72

171 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) Am. J. Hum. Genet. **46**:292-307

172 Aitchison, K., Ogilvie, D., Honeyman, M., Thompson, E. and Sykes, B. (1988) Hum. Genet. **78**:233-236

173 Kuivaniemi, H., Tromp, G. and Prockop, D.J. (1991) FASEB J. **5**:2052-2060

174 Barsh, G.S., David, K.E. and Byers, P.H. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**:3838-3842

175 Rowe, D.W., Shapiro, J.R., Poirier, M. and Schlesinger, S. (1985) J. Clin. Invest. 76:604-611

176 Bateman, J.F., Chan, D., Lamande, S., Mascara, T. and Cole, W.G. (1988) Ann. N.Y. Acad. Sci. **543**:95-105

177 Genovese, C. and Rowe, D. (1987) Methods Enzymol. **145**:223-235

178 Willing, M.C., Cohn, D.H. and Byers, P.H. (1990) J. Clin. Invest. **85**:282-290

179 Sykes, B., Francis, M.J.O. and Smith, R. (1977) New Engl. J. Med. **296**:1200-1203

180 Beighton, P., Wallis, G., Viljoen, D. and Versfeld, G. (1988) Ann. N.Y. Acad. Sci. **543**:40-46

181 Dickson, L.A., Pihlajaniemi, T., Deak, S., Pope, F.M., Nicholls, A., Prockop, D.J. and Myers, J.C. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**:4524-4528

182 Bonadio, J. and Byers, P.H. (1985) Nature **316**:363-366

183 Bateman, J.F., Mascara, T., Chan, D. and Cole, W.G. (1984) Biochem. J. **217**:103-115

184 Bateman, J.F., Chan, D., Mascara, T., Rogers, J.G. and Cole, W.G. (1986) Biochem. J. **240**:699-708

185 Steinmann, B., Rao, V.H., Vogel, A., Bruckner, P., Gitzelmann, R. and Byers, P.H. (1984) J. Biol. Chem. 259.11129-11138

186 Wallis, G.A., Starman, B.J., Schwartz, M.F. and Byers, P.H. (1990) J. Biol. Chem. **265**:18628-18633

187 Prockop, D.J. and Kivirikko, K.I. (1984) New Engl. J. Med. **311**:376-386

188 Stacey, A., Bateman, J., Choi, T., Mascara, T., Cole, W. and Jaenisch, R. (1988) Nature **332**:131-136

189 Vogel, B.E., Doelz, R., Kadler, K.E., Hojima, Y., Engel, J. and Prockop, D.J. (1988) J. Biol. Chem. **263**:19249-19255

190 Kadler, K.E., Torre-Blanco, A., Adachi, E., Vogel, B.E., Hojima, Y. and Prockop, D.J. (1991) Biochemistry **30**:5081-5088

191 Torre-Blanco, A., Adachi, E., Romanic, A.M. and Prockop, D.J. (1992) J. Biol. Chem. **267**:4968-4973

192 Cole, W.G., Chow, C.W., Rogers, J.G. and Bateman, J.F. (1990) J. Med. Genet. **27**:228-235

193 Cole, W.G., Patterson, E., Bonadio, J., Campbell, P.E. and Fortune, D.W. (1992) J. Med. Genet. **29**:112-118

194 Chu, M.-L., Gargiulo, V., Williams, C.J. and Ramirez, F. (1985) J. Biol. Chem. **260**:691-694

195 Willing, M.C., Cohn, D.H., Starman, B.J., Holbrook, K.A., Greenberg, C.R. and Byers, P.H. (1988) J. Biol. Chem. **263**:8398-8404

196 Wallis, G.A., Starman, B.J. and Byers, P.H. (1989) Am. J. Hum. Genet. **45**:A228

197 Wenstrup, R., Shrago, A., Phillips, C., Byers, P. and Cohn, D.H. (1990) Ann. N.Y. Acad. Sci. **580**:546-548

198 Cohn, D.H., Pruchno, C.J., Zhuang, X. and Byers, P.H. (1990) Am. J. Hum. Genet. **47**:A110

199 Byers, P.H., Starman, B.J., Cohn, D.H. and Horwitz, A.L. (1988) J. Biol. Chem. **263**:7855-7861

200 Genovese, C., Brufsky, A., Shapiro, J. and Rowe, D. (1989) J. Biol. Chem. **264**:9632-9637

201 Williams, C.J. and Prockop, D.J. (1983) J. Biol. Chem. **258**:5915-5921

202 Barsh, G.S. and Byers, P.H. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**:5142-5146

203 Penttinen, R.P., Lichtenstein, J.R., Martin, G.R. and McKusick, V.A. (1975) Proc. Natl. Acad. Sci. U.S.A. 72:586-589

204 Sippola, M., Kaffe, S. and Prockop, D.J. (1984) J. Biol. Chem. **259**:14094-14100

205 Bateman, J.F., Lamande, S.R., Dahl, H.-H.M., Chan, D., Mascara, T. and Cole, W.G. (1989) J. Biol. Chem. **264**:10960-10964

206 Pihlajaniemi, T., Dickson, L.A., Pope, F.M., Korhonen, V.R., Nicholls, A., Prockop, D.J. and Myers, J.C. (1984) J. Biol. Chem. **259**:12941-12944

207 Cole, W.G., Campbell, P.E., Rogers, J.G. and Bateman, J.F. (1990) J. Med. Genet. **27**:545-551

208 Shapiro, J.R., Stover, M.L., Burn, V.E., McKinstry, M.B., Burshell, A.L., Chipman, S.D. and Rowe, D.W. (1992) J. Clin. Invest. **89**:567-573

209 Constantinou, C.D., Nielsen, K.B. and Prockop, D.J. (1989) J. Clin. Invest. **83**:574-584

210 Spotila, L.D., Constantinou, C.D., Sereda, L., Ganguly, A., Riggs, B.L. and Prockop, D.J. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:5423-5427

211 Shapiro, J.R., Burn, V.E., Chipman, S.D., Velis, K.P. and Bansal, M. (1989) Conn. Tissue Res. **21**:117-124

212 Prockop, D.J. (1988) Arthritis Rheum. **31**:1-8

213 Constantinou, C.D., Pack, M. and Prockop, D.J. (1989) Cytogenet. Cell Genet. **51**:979

214 Byers, P.H. (1990) Trends Genet. **6**:293-300

215 Marini, J.C., Grange, D.K., Gottesman, G.S., Lewis, M.B. and Koeplin, D.A. (1989) J. Biol. Chem. **264**:11893-11900

216 Wenstrup, R.J., Cohn, D.H., Cohen, T. and Byers, P.H. (1988) J. Biol. Chem. **263**:7734-7740

217 Fertala, A., Westerhausen, A., Morris, G.M., Rooney, J.E. and Prockop, D.J. (1990) Am. J. Hum. Genet. **47**:A216

218 Bonadio, J., Saunders, T.L., Tsai, E., Goldstein, S.A., Morris-Wiman, J., Brinkley, L., Dolan, D.F., Altschuler, R.A., Hawkins, J.E. Jr., Bateman, J.F., Mascara, T. and Jaenisch, R. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**:7145-7149

219 Khillan, J.S., Olsen, A.S., Kontusaari, S., Sokolov, B. and Prockop, D.J. (1991) J. Biol. Chem. **266**:23373-23379

220 Wallis, G.A., Starman, B.J., Zinn, A.B. and Byers, P.H. (1990) Am. J. Hum. Genet. **46**:1034-1040

221 Cohn, D.H., Starman, B.J., Blumberg, B. and Byers, P.H. (1990) Am. J. Hum. Genet. **46**:591-601

222 Constantinou, C.D., Pack, M., Young, S.B. and Prockop, D.J. (1990) Am. J. Hum. Genet. **47**:670-679

223 de Wet, W.J., Pihlajaniemi, T., Myers, J., Kelly, T.E. and Prockop, D.J. (1983) J. Biol. Chem. **258**:7721-7728

224 Tromp, G. and Prockop, D.J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**:5254-5258

225 Nicholls, A.C., Pope, F.M. and Schloon, H. (1979) Lancet **i**:1193

226 Bonadio, J., Ramirez, F. and Barr, M. (1990) J. Biol. Chem. **265**:2262-2268

227 Elejalde, B.R. and de Elejalde, M.M. (1983) Am. J. Med. Genet. **14**:353-359

228 Grange, D.K., Lewis, M.B. and Marini, J.C. (1990) Am. J. Med. Genet. **36**:258-264

229 Robinson, L.P., Worthen, N.J., Lachman, R.S., Adomian, G.E. and Rimoin, D.L. (1987) Prenat. Diagn. 7:7-15

230 McKusick, V.A. (1986) Mendelian Inheritance in Man, 7th ed., Johns Hopkins University Press, Baltimore, MD

231 Lichtenstein, J.R., Martin, G.R., Kohn, L.D., Byers, P.H. and McKusick, V.A. (1973) Science **182**:298-300

232 Minor, R.R., Sippola-Thiele, M., McKeon, J., Berger, J. and Prockop, D.J. (1986) J. Biol. Chem. **261**:10006-10014

233 Williams, B., Cranley, R., Doty, S. and Lichtenstein, J.R. (1973) Am. J. Hum. Genet. **25**:86A

234 Lapiere, C.M. and Nusgens, B. (1976) *In*: Biochemistry of Collagen (Ramachandran, G.N. and Reddi, A.H., Eds.), pp. 377-447, Plenum, New York, NY

235 Bailey, A.J. and Lapiere, C.M. (1973) Eur. J. Biochem. **34**:91-96

236 Steinmann, B., Tuderman, L., Peltonen, L., Martin, G.R., McKusick, V.A. and Prockop, D.J. (1980) J. Biol. Chem. **255**:8887-8893

237 Eyre, D.R., Shapiro, F.D. and Aldridge, J.F. (1985) J. Biol. Chem. **260**:11322-11329

238 Cole, W.G., Chan, D., Chambers, G.W., Walker, I.D. and Bateman, J.F. (1986) J. Biol. Chem. **261**:5496-5503

239 Wirtz, M.K., Glanville, R.W., Steinmann, B., Rao, V.H. and Hollister, D.W. (1987) J. Biol. Chem. **262**:16376-16385

240 Weil, D., Bernard, M., Combates, N., Wirtz, M.K., Hollister, D.W., Steinmann, B. and Ramirez, F. (1988) J. Biol. Chem. **263**:8561-8564

241 Weil, D., D'Alessio, M., Ramirez, F., de Wet, W., Cole, W.G., Chan, D. and Bateman, J.F. (1989) EMBO J. **8**:1705-1710

242 Weil, D., D'Alessio, M., Ramirez, F., Steinmann, B., Wirtz, M.K., Glanville, R.W. and Hollister, D.W. (1989) J. Biol. Chem. **264**:16804-16809

243 Weil, D., D'Alessio, M., Ramirez, F. and Eyre, D.R. (1990) J. Biol. Chem. **265**:16007-16011

244 Vasan, N.S., Kuivaniemi, H., Vogel, B.E., Minor, R.R., Wooton, J.A.M., Tromp, G., Weksberg, R. and Prockop, D.J. (1991) Am. J. Hum. Genet. **48**:305-317

245 Wirtz, M.K., Keene, D.R., Hori, H., Glanville, R.W., Steinmann, B., Rao, V.H. and Hollister, D.W. (1990) J. Biol. Chem. **265**:6312-6317

246 de Wet, W., Sippola, M., Tromp, G., Prockop, D.J., Chu, M.-L. and Ramirez, F. (1986) J. Biol. Chem. **261**:3857-3862

247 Kuivaniemi, H., Sabol, C., Tromp, G., Sippola-Thiele, M. and Prockop, D.J. (1988) J. Biol. Chem. **263**:11407-11413

248 Horton, W.A. (1987) *In*: Practice of Pediatrics (Kelley, V.C., Ed.), Vol. 10, Chap. 55, pp. 1-30, Harper & Row, New York, NY

249 Horton, W.A., Machado, M.A., Chou, J.W. and Campbell, D. (1987) Pediatr. Res. **22**:324-329

250 Eyre, D.R., Upton, M.P., Shapiro, F.D., Wilkinson, R.H. and Vawter, G.F. (1986) Am. J. Hum. Genet. **39**:52-67

251 Godfrey, M., Keene, D.R., Blank, E., Hori, H., Sakai, L.Y., Sherwin, L.A. and Hollister, D.W. (1988) Am. J. Hum. Genet. **43**:894-903

252 Borochowitz, Z., Ornoy, A., Lachman, R., Murray, L. and Rimoin, D. (1985) Clin. Res. **33**:129a

253 Godfrey, M. and Hollister, D.W. (1988) Am. J. Hum. Genet. **43**:904-913

254 Murray, L.W., Bautista, J., James, P.L. and Rimoin, D.L. (1989) Am. J. Hum. Genet. **45**:5-15

255 Vissing, H., D'Alessio, M., Lee, B., Ramirez, F., Godfrey, M. and Hollister, D.W. (1989) J. Biol. Chem. **264**:18265-18267

256 Tiller, G.E., Rimoin, D.L., Murray, L.W. and Cohn, D.H. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**:3889-3893

257 Lee, B., Vissing, H., Ramirez, F., Rogers, D. and Rimoin, D.L. (1989) Science **244**:978-980

258 Francomano, C.A., Liberfarb, R.M., Hirose, T., Maumanee, I.H., Streeten, E.A., Meyers, D.A. and Pyeritz, R.E. (1987) Genomics **1**:293-296

259 Knowlton, R.G., Weaver, E.J., Struyk, A.F., Knobloch, W.H., King, R.A., Norris, K., Shamban, A., Uitto, J., Jimenez, S.A. and Prockop, D.J. (1989) Am. J. Hum. Genet. **45**:681-688

260 Francomano, C.A., Rowan, B.G., Liberfarb, R.M., Hirose, T., Maumenee, I.H., Stoll, H.U. and Pyeritz, R.E. (1988) Am. J. Hum. Genet. **43**:A83

261 Schwartz, R.C., Watkins, D., Fryer, A.E., Goldberg, R., Marion, R., Polomeno, R.C., Spallone, A., Upadhyaya, M., Harper, P. and Tsipouras, P. (1989) Am. J. Hum. Genet. **45**:A218

262 Ahmad, N.N., Ala-Kokko, L., Knowlton, R.G., Jimenez, S.A., Weaver, E.J., Maguire, J.I., Tasman, W. and Prockop, D.J. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:6624-6627

263 Vandenberg, P., Khillan, J.S., Prockop, D.J., Helminen, H., Kontusaari, S. and Ala-Kokko, L. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:7640-7644

264 Garofalo, S., Vuorio, E., Metsaranta, M., Rosati, R., Toman, D., Vaughan, J., Lozano, G., Mayne, R., Ellard, J., Horton, W. and de Crombrugghe, B. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:9648-9652

265 Horton, W.A. and Rimoin, D.L. (1970) Pediatr. Res. **13**:1266-1270

266 Poole, A.R., Pidoux, I., Reiner, A., Rosenberg, L., Hollister, D.W., Murray, L. and Rimoin, D.L. (1988) J. Clin. Invest. **81**:579-589

267 Poole, A.R., Pidoux, I., Reiner, A., Choi, H. and Rosenberg, L.C. (1984) J. Cell Biol. **98**:54-65

268 Palotie, A., Vaisanen, P., Ott, J., Ryhanen, L., Elima, K., Vikkula, M., Cheah, K., Vuorio, E. and Peltonen, L. (1989) Lancet **ii**:924-927

269 Knowlton, R.G., Katzenstein, P.L., Moskowitz, R.W., Weaver, E.J., Malemud, C.J., Pathria, M.N., Jimenez, S.A. and Prockop, D.J. (1990) New Engl. J. Med. **322**:526-530

270 Ala-Kokko, L., Baldwin, C.T., Moskowitz, R.W. and Prockop, D.J. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**:6565-6568

271 Pope, F.M., Martin, G.R., Lichtenstein, J.R., Penttinen, R., Gerson, B., Rowe, D.W. and McKusick, V.A. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**:1314-1316

272 Superti-Furga, A. and Steinmann, B. (1989) Lancet **i**:903-904

273 Pope, F.M., Martin, G.R. and McKusick, V.A. (1977) J. Med. Genet. **14**:200-204

274 Vissing, H., D'Alessio, M., Lee, B., Ramirez, F., Byers, P.H., Steinmann, B. and Superti-Furga, A. (1991) J. Biol. Chem. **266**:5244-5248

275 Aumailley, M., Poschl, E., Martin, G.R., Yamada, Y. and Muller, P.K. (1988) Eur. J. Clin. Invest. **18**:207-212

276 Superti-Furga, A. and Steinmann, B. (1988) Biochem. Biophys. Res. Commun. **150**:140-147

277 Byers, P.H., Holbrook, K.A., McGillivray, B., MacLeod, P.M. and Lowry, R.B. (1979) Hum. Genet. **47**:141-150

278 Pope, F.M., Nicholls, A.C., Narcisi, P., Bartlett, J., Neil-Dwyer, G. and Doshi, B. (1981) Lancet **i**:973-975

279 Stolle, C.A., Pyeritz, R.E., Myers, J.C. and Prockop, D.J. (1985) J. Biol. Chem. **260**:1937-1944

280 Byers, P.H., Holbrook, K.A., Barsh, G.S., Smith, L.T. and Bornstein, P. (1981) Lab. Invest. **44**:336-341

281 Clark, J.G., Kuhn, C. III and Uitto, J. (1980) Am. Rev. Respir. Dis. **122**:971-978

282 Pyeritz, R.E., Stolle, C.A., Parfrey, N.A. and Myers, J.C. (1984) Am. J. Med. Genet. **19**:607-622

283 Tsipouras, P., Byers, P.H., Schwartz, R.C., Chu, M.-L., Weil, D., Pepe, G., Cassidy, S.B. and Ramirez, F. (1986) Hum. Genet. **74**:41-46

284 Nicholls, A.C., De Paepe, A., Narcisi, P., Dalgleish, R., De Keyser, F., Matton, M. and Pope, F.M. (1988) Hum. Genet. **78**:276-281

285 Tromp, G., Kuivaniemi, H., Shikata, H. and Prockop, D.J. (1989) J. Biol. Chem. **264**:1349-1352

286 Tromp, G., Kuivaniemi, H., Stolle, C., Pope, F.M. and Prockop, D.J. (1989) J. Biol. Chem. **264**:19313-19317

287 Kontusaari, S., Tromp, G., Kuivaniemi, H., Romanic, A.M., Prockop, D.J. (1990) J. Clin. Invest. **86**:1465-1473

288 Richards, A.J., Lloyd, J.C., Ward, P.N., De Paepe, A., Narcisi, P. and Pope, F.M. (1991) J. Med. Genet. **28**:458-463

289 Lee, B., D'Alessio, M., Vissing, H., Ramirez, F., Steinmann, B. and Superti-Furga, A. (1991) Am. J. Hum. Genet. **48**:511-517

290 Kontusaari, S., Tromp, G., Kuivaniemi, H., Ladda, R.L. and Prockop, D.J. (1990) Am. J. Hum. Genet. **47**:112-120

291 Kuivaniemi, H., Kontusaari, S., Tromp, G., Zhao, M., Sabol, C. and Prockop, D.J. (1990) J. Biol. Chem. **265**:12067-12074

292 Cole, W.G., Chiodo, A.A., Lamande, S.R., Janeczko, R., Ramirez, F., Dahl, H.-H.M., Chan, D. and Bateman, J.F. (1990) J. Biol. Chem. **265**:17070-17077

293 Lee, B., Vitale, E., Superti-Furga, A., Steinmann, B. and Ramirez, F. (1991) J. Biol. Chem. **266**:5256-5259

294 Weiner, S. and Traub, W. (1992) FASEB J. **6**:879-885

295 Jee, W.S.S. (1988) *In*: Cell and Tissue Biology: A textbook of Histology, 6th ed. (Weiss, L., Ed.), pp. 211-254, Urban & Schwarzenberg, Baltimore, MD

296 Termine, J.D. (1990) *In*: Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism, 1st ed., pp. 16-18, American Society for Bone and Mineral Research, Kelseyville, CA

297 Arsenault, A.L. (1991) Calcif. Tissue Int. **48**:56-62

298 Robinson, R.A. (1952) J. Bone Joint Surg. **34A**:389-434

299 Lowenstam, H.A. and Weiner, S. (1989) On Biomineralization, pp. 144-167, Oxford University Press, New York, NY

300 Hodge, A.J. (1989) Conn. Tissue Res. **21**:137-147

301 Traub, W., Arad, T. and Weiner, S. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**:9822-9826

302 Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science **230**: 1350-1354

303 Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science **239**: 487-491

304 Caskey, C.T. (1987) Science **236**: 1223-1229

305 Landegren, U., Kaiser, R., Caskey, C.T. and Hood, L. (1988) Science **242**: 229-237

306 Cotton, R.G.H. (1989) Biochem. J. **263**: 1-10

307 Rossiter, B.J.F. and Caskey, C.T. (1990) J. Biol. Chem. **265**:12753-12756

308 Kan, Y.W. and Dozy, A.M. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**: 5631-5635

309 Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Am. J. Hum. Genet. **32**: 314-331

310 Myers, R.M., Maniatis, T. and Lerman, L.S. (1987) Methods Enzymol. **155**: 501-527

311 Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985) Nature (London) **313**: 495-498

312 Fischer, S.G. and Lerman, L.S. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**: 1579-1583

313 Myers, R.M., Fischer, S.G., Maniatis, T. and Lerman, L.S. (1985) Nucleic Acids Res. **13**: 3111-3130

314 Myers, R.M., Fischer, S.G., Lerman, L.S. and Maniatis, T. (1985) Nucleic Acids Res. **13**: 3131-3145

315 Sheffield, V.C., Cox, D.R., Lerman, L.S. and Myers, R.M. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**: 232-236

316 Shenk, T.E., Rhodes, C., Rigby, P.W.J. and Berg, P. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**: 989-993

317 Dodgson, J.B. and Wells, R.D. (1977) Biochemistry **16**: 2374-2379

318 Freeman, G.J. and Huang, A.S. (1981) J. Gen. Virol. **57**: 103-117

319 Myers, R.M., Larin, Z. and Maniatis, T. (1985) Science **230**: 1242-1246

320 Winter, E., Yamamoto, F., Almoguera, C. and Perucho, M. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**: 7575-7579

321 Thomas, D.C., Kunkel, T.A., Casna, N.J., Ford, J.P. and Sancar, A. (1986) J. Biol. Chem. **261**: 14496-14505

322 Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**: 4397-4401

323 Dahl, H.-H.M., Lamande, S.R., Cotton, R.G.H. and Bateman, J.F. (1989) Anal. Biochem. **183**: 263-268

324 Prockop, D.J. (1992) New Engl. J. Med. **326**:540-546

325 Pinnell, S.R., Murad, S. and Darr, D. (1987) Arch. Dermatol. **123**:1684-1686

326 Barsh, G.S., Peterson, K.E. and Byers, P.H. (1981) Collagen Rel. Res. **1**:543-548

327 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**:156-159

328 Hanahan, D. (1983) J. Mol. Biol. **166**:557-580

329 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., pp. A.2-A.3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

330 Ibid., p. A.9

331 Ibid., pp. 1.38-1.39

332 Ibid., pp. 1.42-1.43

333 Ibid., pp. 1.46

334 Tartof, K.D. and Hobbs, C.A. (1987) Bethesda Res. Lab. Focus **9**:12

335 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., pp. 6.28-6.29, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

336 Ibid., p. E.16

337 Dean, M. (1987) Nucleic Acids Res. **15**:6754

338 Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. **65**:499-560

339 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., pp. F.4-F.5, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

340 Ibid., pp. 6.24-6.27

341 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, 1st ed., p. 124, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

342 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**:103-119

343 Messing, J. (1983) Methods Enzymol. **101**:20-78

344 King, P.V. and Blakesley, R.W. (1986) Bethseda Res. Lab. Focus **8**:1-2

345 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**:5463-5467

346 Schenk, R.K., Olah, A.J. and Herrmann, W. (1984) *In*: Methods of Calcified Tissue Preparation (Dickson, G.R., Ed.), pp. 1-56, Elsevier, New York, NY

347 Wake, S.A. and Mercer, J.F.B. (1985) Biochem J. **228**:425-432

348 Lorens, J.B. (1991) PCR Methods Applications **1**:140-141

349 Kaufman, D.L. and Evans, G.A. (1990) Biotechniques **9**:304-306

350 Clark, J.M. (1988) Nucleic Acids Res. **16**:9677-9686

351 Constantinou, C.D., Spotila, L.D., Zhuang, J., Sereda, L., Hanning, C. and Prockop, D.J. (1990) Nucleic Acids Res. **18**:5577

352 Peltonen, L., Palotie, A., Hayashi, T. and Prockop, D.J. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**:162-166

353 Fontana, A. and Gross, E. (1986) *In*: Practical Protein Chemistry - A Handbook (Darbre, A., Ed.), pp. 67-120, John Wiley & Sons, New York, NY

354 Theophilus, B.D.M., Latham, T., Grabowski, G.A. and Smith, F.I. (1989) Nucleic Acids Res. **17**:1707-1722

355 Yang, T.P., Patel, P.I., Chinault, A.C., Stout, J.T., Jackson, L.G., Hildebrand, B.M. and Caskey, C.T. (1984) **310**:412-414

356 Chamberlain, J.S., Gibbs, R.A., Ranier, J.E. and Caskey, C.T. (1990) *In*: PCR Protocols: A Guide to Mthods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds.), pp. 272-281, Academic Press, New York, NY

357 Smith, F.I., Latham, T.E., Ferrier, J.A. and Palese, P. (1988) Genomics **3**:217-223

358 Riesner, D., Steiger, G., Zimmat, R., Owens, R.A., Wagenhoter, M., Hillen, W., Vollbach, S. and Henco, K. (1989) Electrophoresis **10**:377-389

359 Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Seyika, T. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**:2766-2770

360 Orita, M., Suzuki, Y., Seyika, T. and Hayashi, K. (1989) Genomics **5**:874-879

361 Novack, D.F., Casna, N.J., Fischer, S.G. and Ford, J.P. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**: 586-590

362 Kornher, J.S. and Livak, K.J. (1989) Nucleic Acids Res. **17**:7779-7784

363 Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988) Proc. Natl. Acad. Sci. U.S.A. 85:9431-9440

364 Ganguly, A., Rooney, J.E., Hosomi, S., Zeiger, A.R. and Prockop, D.J. (1989) Genomics **4**:530-538

365 Gibbs, R.A. and Caskey, C.T. (1987) Science **236**:303-305

366 Tindall, K.R. and Kunkel, T.A. (1988) Biochemistry **27**:6008-6013

367 Keohavong, P. and Thilly, W.G. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**:9253-9257

368 Werntges, H., Steger, G., Riesner, D. and Fritz, H.-J. (1986) Nucleic Acids Res. **14**:3773-3790

369 Westerhausen, A.I., Constantinou, C.D. and Prockop, D.J. (1990) Nucleic Acids Res. **18**:4968

370 Mackay, K., Hawkins, J.R., Superti-Furga, A., Steinmann, B. and Dalgleish, R. (1990) Nucleic Acids Res. **18**:5926

371 Tsipouras, P., Myers, J.C., Ramirez, F. and Prockop, D.J. (1983) J. Clin. Invest. **72**:1262-1267

372 Brebner, D.K., Grobler-Rabie, A.F., Bester, A.J., Mathew, C.G. and Boyd, C.D. (1985) Hum. Genet. **70**:25-27

373 Grobler-Rabie, A.F., Brebner, D.K., Vandenplas, S., Wall..., G., Dalgleish, R., Kaufman, R.E., Bester, A.J., Mathew, C.G.P. and Boyd, C.D. (1985) J. Med. Genet. **22**:182-186

374 Ahti, H., Palotie, A., Peltonen, L. and Ramirez, F. (1987) Hum. Genet. **75**:79-80

375 Falk, C.T., Schwartz, R.C., Ramirez, F. and Tsipouras, P. (1986) Am. J. Hum. Genet. 38:269-279

376 Watson, C.J. and Dalgleish, R. (1990) Nucleic Acids Res. **18**:5925

377 Strobel, D., Tsuneyoshi, T., Kuivaniemi, H., Tromp, G., Spotila, L.D., Baldwin, C.T., Constantinou, C.D., Ganguly, A., Sereda, L., Sokolov, B.P. and Prockop, D.J. (1992) Matrix **12**:87-91

378 Borresen, A.-L. (1986) Ann. Clin. Res. **18**:258-263

379 Cavalli-Sforza, L.L., Wilson, A.C., Cantor, C.R., Cook-Deegan, R.M. and King, M.-C. (1991) Genomics **11**:490-491

380 Levitt, R.C. (1991) Genomics **11**:484-489

381 de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M.-L., Dickson, L., Weil, D. and Ramirez, F. (1987) J. Biol. Chem. **262**:16032-16036

382 Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**:278-282

383 Starman, B.J., Eyre, D., Charbonneau, H., Harrylock, M., Weis, M.A., Weiss, L., Graham, J.M. Jr. and Byers, P.H. (1989) J. Clin. Invest. **84**:1206-1214

384 Wirtz, M.K., Rao, V.H., Glanville, R.W., Labhard, M.E., Pretorius, P.J., de Vries, W.N., de Wet, W.J. and Hollister, D.W. (1990) Matrix **10**:A251

385 Valli, M., Mottes, M., Tenni, R., Sangalli, A., Lira, M.G., Rossi, A., Antoniazzi, F., Cetta, G. and Pignatti, P.F. (1991) J. Biol. Chem. 266:1872-1878

386 Nicholls, A.C., Oliver, J., Renouf, D.V., Keston, M. and Pope, F.M. (1991) J. Med. Genet. **28**:757-764

387 Steinmann, B., Westerhausen, A., Constantinou, C.D., Superti-Furga, A. and Prockop, D.J. (1991) Biochem. J. **279**:747-752

388 Vogel, B.E., Minor, R.R., Freund, M. and Prockop, D.J. (1987) J. Biol. Chem. **262**:14737-14744

389 Cohn, D.H., Byers, P.H., Steinmann, B. and Gelinas, R.E. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**:6045-6047

390 Cohn, D.H., Apone, S., Eyre, D.R., Starman, B.J., Andreassen, P., Charbonneau, H., Nicholls, A.C., Pope, F.M. and Byers, P.H. (1988) J. Biol. Chem. **263**:14605-14607

391 Wenstrup, R.J., Shrago-Howe, A.W., Lever, L.W., Phillips, C.L., Byers, P.H. and Cohn, D.H. (1991) J. Biol. Chem. **266**:2590-2594

392 Deak, S.B., Scholz, P.M., Amenta, P.S., Constantinou, C.D., Levi-Minzi, S.A., Gonzalez-Lavin, L. and Mackenzie, J.W. (1991) J. Biol. Chem. **266**:21827-21832

393 Pruchno, C.J., Cohn, D.H., Wallis, G.A., Willing, M.C., Starman, B.J., Zhang, X. and Byers, P.H. (1991) Hum. Genet. **87**:33-40

394 Bateman, J.F., Chan, D., Walker, I.D., Rogers, J.G. and Cole, W.G. (1987) J. Biol. Chem. **262**:7021-7027

395 Bateman, J.F., Lamande, S.R., Dahl, H.-H.M., Chan, D. and Cole, W.G. (1988) J. Biol. Chem. **263**:11627-11630

396 Lamande, S.R., Dahl, H.-H.M., Cole, W.G. and Bateman, J.F. (1989) J. Biol. Chem. 264:15809-15812

397 Tsuneyoshi, T., Constantinou, C.D., Mikkelson, M. and Prockop, D.J. (1990) Am. J. Hum. Genet. **47**:A240

398 Westerhausen, A., Kishi, J. and Prockop, D.J. (1990) J. Biol. Chem. **265**:13995-14000

399 Pack, M., Constantinou, C.D., Kalia, K., Nielsen, K.B. and Prockop, D.J. (1989) J. Biol. Chem. **264**:19694-19699

400 Cohn, D.H., Wallis, G., Zhang, X. and Byers, P.H. (1990) Matrix **10**:A236

401 Zhuang, J., Constantinou, C.D., Ganguly, A. and Prockop, D.J. (1990) Matrix **10**:A252

402 Bonadio, J., Patterson, E. and Smiley, E. (1988) Coll. Rel. Res. **8**:506A

403 Baldwin, C.T., Constantinou, C.D., Dumars, K.W. and Prockop, D.J. (1989) J. Biol. Chem. **264**:3002-3006

404 Patterson, E., Smiley, E. and Bonadio, J. (1989) J. Biol. Chem. **264**:10083-10087

405 Westerhausen, A., Constantinou, C.D. and Prockop, D.J. (1990) Am. J. Hum. Genet. **47**:A242

406 Pruchno, C.J., Willing, M.C., Starman, B.J., Aylesworth, A. and Byers, P.H. (1989) Am. J. Hum. Genet. **45**:A213

407 Zhuang, J., Constantinou, C.D. and Prockop, D.J. (1990) Matrix **10**:A252

408 Baldwin, C.T., Ganguly, A., Rooney, J., Hosomi, S., Conway, D.M. and Prockop, D.J. (1988) Coll. Rel. Res. **8**:A508

409 Bonucci, E. (1971) Clin. Orthop. Rel. Res. **78**:108-139

Manuscripts and Abstracts

The following manuscripts and abstracts were derived from the work presented in this thesis:

1) Sztrolovics, R., van der Rest, M. and Roughley, P.J. (1992) Optimization of chemical cleavage detection of mismatches in RNA:DNA heteroduplexes and of cloning PCR products for sequence analysis: Use in characterization of a novel sequence polymorphism in a human type I collagen gene, Anal. Biochem., manuscript submitted

2) Sztrolovics, R., Glorieux, F.H., Travers, R., van der Rest, M. and Roughley, P.J. (1992) Comparison Between Molecular Defects and Bone Histological Changes in Osteogenesis Imperfecta, manuscript in preparation

3) Sztrolovics, R., Glorieux, F.H., van der Rest, M. and Roughley, P.J. (1992) Identification of a carboxy-terminal glycine → serine substitution in the α2(I) chain which produces an unexpectedly mild OI phenotype, manuscript in preparation

4) Sztrolovics, R., Glorieux, F.H., Travers, R., van der Rest, M. and Roughley, P.J. (1992) Correlation between phenotypic severity and specific collagen mutations, Orthop. Trans., abstract in press