NEW TOOLS FOR THE STUDY OF AN OLD COLLAGEN

Characterization of the human COL9A1, COL9A2 and COL9A3 genes and production of human type IX collagen as a recombinant protein

TERO PIHLAJAMAA

Collagen Research Unit,
Biocenter Oulu and
Department of Medical Biochemistry

OULU 2000

TERO PIHLAJAMAA

NEW TOOLS FOR THE STUDY OF AN OLD COLLAGEN

Characterization of the human COL9A1, COL9A2 and COL9A3 genes and production of human type IX collagen as a recombinant protein

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in the Auditorium of the Department of Medical Biochemistry, on September 15th, 2000, at 1 p.m.

OULUN YLIOPISTO, OULU 2000

Manuscript received 14 August 2000
Accepted 18 August 2000

Communicated by
Professor Raili Myllylä
Professor Aleksander L. Sieron

ISBN 951-42-5735-9

ALSO AVAILABLE IN PRINTED FORMAT

ISBN 951-42-5734-0
ISSN 0355-3221 (URL: http://herkules.oulu.fi/issn03553221/)

OULU UNIVERSITY LIBRARY
OULU 2000

Pihlajamaa, Tero, New tools for the study of an old collagen Characterization of the human COL9A1, COL9A2 and COL9A3 genes and production of human type IX collagen as a recombinant protein
Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry, University of Oulu, FIN-90220 Oulu, Finland
2000
Oulu, Finland
(Manuscript received 14 August 2000)

Abstract

Type IX collagen is a quantitatively minor component of cartilage collagen fibrils. Although a few mutations have been associated with multiple epiphyseal dysplasia, recent evidence suggests involvement of type IX collagen in a wider spectrum of phenotypes. The functional role of this molecule remains undetermined, in part due to difficulties in obtaining high amounts of intact protein.

To facilitate more efficient mutation screening and comparison of the genomic organization of the human genes encoding the α1(IX), α2(IX) and α3(IX) polypeptides, their genomic structures were characterized. Complete nucleotide sequences were determined for the COL9A2 and COL9A3 genes along with sequences for all the exon boundaries in the COL9A1 gene. Putative transcription control elements were identified and the alternative promoter region was characterized in the human and mouse COL9A1 genes. Mutation screening was performed for the COL9A3 gene and two apparently neutral 9-bp deletions within the COL1 domain were identified. These are the first deletions within a triple-helical domain of any collagen that are not associated with a disease phenotype.

An insect cell expression system with an exogenous source of prolyl 4-hydroxylase was used to produce heterotrimeric human type IX collagen. The recombinant protein consisted of the three a chains in a 1:1:1 ratio and showed correct folding and high thermal stability. Up to 10 mg of secreted protein could be purified from a litre of culture medium. The expression system was used to analyze the chain association of type IX collagen *in cellulo*. Although the chains are capable of homotrimerization, a preference for heterotrimer formation was noted.

The neutral deletion was characterized further using the insect cell system. Mutant α3(IX) chains carrying a deletion of one Gly-X-Y triplet were shown to form correctly folded heterotrimers with the wild-type α1(IX) and α2(IX) chains. The results suggest a function for the NC2 domain in neutralizing the effect of the deletion.

This work provides a novel means for the analysis of type IX collagen mutations and their protein-level effects, and should enable future studies to be made of the structure-function relationship in type IX collagen.

Keywords: baculovirus, cartilage, mutation detection

Acknowledgements

This work was carried out at the Department of Medical Biochemistry, University of Oulu, and the Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, USA, during the years 1994-2000.

I wish to express my warmest gratitude to my supervisor, Docent Leena Ala-Kokko, for her continuous guidance and encouragement over the years. Almost any hardship has appeared minor with her unlimited optimism in the background. The numerous extracurricular activities organized by her and her husband, Dr. James Hyland, have been a key element in preserving the mental stability of this young scientist.

I wish to express my sincere gratitude to Professors Kari I. Kivirikko and Taina Pihlajaniemi for creating a supportive and enthusiastic atmosphere in the top-level Collagen Research Unit. Professor Ilmo Hassinen is gratefully acknowledged for his help in computer matters. I wish to express my admiration for Professor Darwin J. Prockop and his brilliance in science and to thank him sincerely for the enjoyable and fruitful time that I spent at Jefferson.

I wish to thank Professor Raili Myllylä and Assistant Professor Aleksander L. Sieron for their swift and thorough preliminary examination of this thesis, and also Malcolm Hicks, M.A., for his careful revision of the language of the manuscript. I am indebted to all my collaborators for sharing their expertise, and to Docent Ilkka Kilpeläinen for letting me write this thesis along-side work on other projects.

My dear colleagues at the departments in Oulu and Philadelphia deserve my warmest thanks. Especially Johanna Myllyharju, Minna Nokelainen and Hongmin Tu are acknowledged for all the assistance and advice they have given me. I will cherish the hours spent in and out of the lab with Susanna Räinä (neé Annunen), Miia Melkoniemi, Merja Välkkilä, Merja Perälä, Constance Yuan, Jaana Lohiniva, Mirka Vuoristo, Jussi Vuoristo, Petteri Paassilta and Jarmo Körkkö as some of my most enjoyable memories. I also thank my volleyball mates in Oulu and Philadelphia for helping me relieve my aggressions on and off the court, and all my friends for the times spent together.

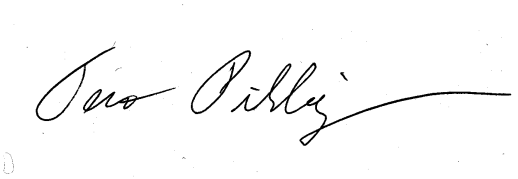
I wish to thank Helena Lindqvist, Aira Harju, Aila Jokinen, Rohini Dhulipala and Robert Hnatuk for their expert technical assistance, Pertti Vuokila for logistics and relaxing chats and Seppo Lähdesmäki for help with the technical equipment. I also wish to thank Ari-Pekka Kvist and Juha Näpänkangas for their help with the computers and

Marja-Leena Kivelä, Auli Kinnunen and Marja Leena Karjalainen for their friendly and efficient secretarial services.

I am very grateful to my parents for supporting me in so many ways, and to my sister Terhi for her friendship and cheerful spirits. Finally, I am most grateful for the love, support and understanding of my fiancée Katriina. Her sense of humour and the wonderful antics of our cats help me survive the disheartening moments in the work.

This research was supported financially by the Finnish Centre of Excellence Programme (2000-2005) of the Academy of Finland (44843), Fibrogen Inc., the Finnish Cultural Foundation and the Farmos Research and Science Foundation.

Helsinki, July 25, 2000



Abbreviations

4Hyp	4-hydroxyproline
bp	base pair(s)
C-	carboxy-
CD	circular dichroism
cDNA	complementary DNA
COL	collagenous
COL*y*A*x*	human gene for the α(*x*) chain of type *y* collagen
Col*y*a*x*	mouse gene for the α(*x*) chain of type *y* collagen
COMP	cartilage oligomeric matrix protein
CSGE	conformation-sensitive gel electrophoresis
Cys	cysteine
E-64	trans-epoxysuccinyl-L-leucylamido(4-guanidino)butane
EDTA	ethylenediaminetetracetic acid
EDM*z*	multiple epiphyseal dysplasia (type *z*)
ECM	extracellular matrix
ER	endoplasmic reticulum
FACIT	fibril-associated collagens with interrupted triple helices
GAG	glycosaminoglycan
Gly	glycine
HB-GAM	heparin-binding growth-associated molecule
HMW	high molecular weight
kb	kilobase(s)
kDa	kilodalton
LMW	low molecular weight
MED	multiple epiphyseal dysplasia
MMP	matrix metalloproteinase
mRNA	messenger RNA
N-	amino-
NC	noncollagenous
PAC	P1 artificial chromosome
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PDI	protein disulphide isomerase
PMSF	phenylmethylsulphonyl fluoride
PRELP	proline arginine-rich end leucine-rich repeat protein
PSACH	pseudoachondroplasia
Pro	proline
RT	reverse transcriptase/transcription
SDS	sodium dodecyl sulphate
Ser	serine
T_m	midpoint temperature of thermal denaturation
X	any amino acid
Y	any amino acid

List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I Pihlajamaa T, Vuoristo MM, Annunen S, Perälä M, Prockop DJ & Ala-Kokko L (1998) Human COL9A1 and COL9A2 genes. Two genes of 90 and 15 kb code for similar polypeptides of the same collagen molecule. Matrix Biol 17: 237-241.

II Pihlajamaa T, Perälä M, Vuoristo MM, Nokelainen M, Bodo M, Schulthess T, Vuorio E, Timpl R, Engel J & Ala-Kokko L (1999) Characterization of recombinant human type IX collagen. Association of α chains into homotrimeric and heterotrimeric molecules. J Biol Chem 274: 22464-22468.

III Paassilta P, Pihlajamaa T, Annunen S, Brewton RG, Wood BR, Johnson CC, Liu J, Gong Y, Warman ML, Prockop DJ, Mayne R & Ala-Kokko L (1999) Complete sequence of 23 kb human COL9A3 gene. Detection of Gly-X-Y triplet deletions which represent neutral variants. J Biol Chem 274: 22469-22475.

In addition, some unpublished data are presented.

Contents

1 Introduction

Although the assigning of a molecule to a certain ultrastructural location within a given tissue may give important clues about the function of the protein, more detailed studies are required to validate the assumptions and suggestions made. Where *in vivo* evidence is unobtainable for technical or ethical reasons, *in vitro* studies with a few isolated components offer an alternative route. With the advent of recombinant DNA technology, the prospects for biomolecular science have become unlimited. Proteins with altered structures can now be produced in recombinant systems, for example, and transgenic organisms can be created for scientific and commercial purposes. With the new technologies currenly available the impossible is suddenly becoming possible.

Cartilage is an example of a complex tissue with inherent obstacles limiting any increase in our scientific knowledge about it. It is an essential component of the human body, although almost completely consisting of extracellular matrix (ECM). As the function of cartilage strongly relies on structural intactness, samples of human cartilage tissue are rarely obtainable for research purposes. Also, several components of the cartilage are present at such low levels that their isolation in large quantities is difficult.

The collagens are a family of structurally related proteins serving in a multitude of functional roles within the extracellular matrix of tissues. A number of collagens are also present in cartilage ECM. Type II collagen is by far the most abundant of these, forming a supportive meshwork of collagen fibrils. Another molecule, type IX collagen, is a minor component of these fibrils, representing a few percent of the total amount. Despite intense research, the suggested functions of type IX collagen have not been confirmed. This is due in part to the unavailability of high quantities of the human protein for studying its function and interactions. The identification of genetic defects causing inherited diseases is a highly useful approach for defining structure-function relationships, but it has not provided any clear insights into the function of type IX collagen, possibly because the information on gene structures required for efficient mutation screening has not been available.

The genomic organizations of the genes encoding human type IX collagen were elucidated in the present work, and the respective protein was produced in a recombinant expression system. The usefulness of the expression system for the analysis of genetic

variations was demonstrated by identifying an unusual adaptation mechanism neutralizing the effect of a short deletion within type IX collagen.

2 Review of the literature

2.1 Collagens

Collagens are a family of structurally related proteins that are present in high amounts in the ECM of numerous tissues. These proteins form a variety of structurally and functionally important supramolecular assemblies. Collagens are characterized by a triple helix consisting of three identical or different polypeptides, called α chains. A characteristic repetitive amino acid triplet -Gly-X-Y- is a structural prerequisite for the formation of the triple helix. Another typical feature is about a 20% occurrence of the planar, rigid imino acid proline, which stabilizes the rod-like α chain. The 19 types of collagen characterized so far are divided into two major structural groups by reference to their ability to associate into macromolecular fibrils, *i.e.* fibrillar and non-fibrillar collagens. The recognition of certain structural or functional similarities between non-fibrillar collagens has led to a further division of these into six subgroups. (For reviews, see Prockop & Kivirikko 1995, Brodsky & Ramshaw 1997.)

2.1.1 Biosynthesis of collagens

The biosynthesis of collagens is a complex process involving both intracellular and extracellular processing events. In addition to the α chains, numerous processing enzymes are required, all of which have to be expressed in a proper temporal and spatial manner. Many of the enzymes serve the sole purpose of collagen biosynthesis. (See Kivirikko & Myllylä 1985.)

2.1.1.1 Intracellular events

The bulk of our information on collagen biosynthesis comes from studies of the fibrillar collagens. The intracellular chain of events starts with transcription of an appropriate collagen gene into an hnRNA molecule, which is transported after processing into the cytosol. In the ribosomes of the rough endoplasmic reticulum the mature mRNA molecule is translated into a polypeptide, which is then co-translationally translocated into the lumen of the endoplasmic reticulum (ER) while simultaneously undergoing modification. The fully translated polypeptide is released into the lumen for post-translational processing after enzymatic cleavage of a signal peptide. Fibrillar collagen chains are synthesized as precursor molecules, proα chains, containing large non-collagenous N and C-propeptides, which are not present in the non-fibrillar collagens. The hydroxylation of most prolyl residues in the Y positions of the -X-Y-Gly- triplets is carried out by prolyl 4-hydroxylase, a tetrameric enzyme consisting of two catalytic α subunits and two protein disulphide isomerase (PDI) chains, or β subunits (see Kivirikko & Myllyharju 1998). PDI is a widely distributed enzyme in animals and assists in the formation of disulphide bonds (Bulleid & Freedman 1988). Prolyl 4-hydroxylation is of major importance, since it is known to stabilize the collagen triple helix by increasing hydrogen bonding (Berg & Prockop 1973, Bella *et al.* 1995). In addition, a few prolyl residues in -Pro-4Hyp-Gly- triplets are hydroxylated by prolyl 3-hydroxylase, but the biological significance of this modification is not known (see Kivirikko & Myllylä 1985). Lysyl hydroxylase catalyzes the hydroxylation of some of the lysines in the Y positions, which confers two important functions on the polypeptide: the possibility to attach O-linked carbohydrates enzymatically and the ability to form intramolecular and intermolecular covalent cross-links. Certain asparagine residues are also targets of N-glycosylation in the propeptides of several collagens, while serine residues in a few collagen types are acceptors for glycosaminoglycan attachment. The extents of the enzymatic modifications, other than proline 4-hydroxylation, show collagen-type-specific and tissue-specific variation. (See Kivirikko & Myllylä 1985, Kielty *et al.* 1993.)

The mechanism of chain selection and association has been mainly studied in fibrillar collagens. The process starts by folding of the N and C-propeptides, which are stabilized by intrachain disulphide bonds (Bächinger *et al.* 1981, Doege & Fessler 1986). The folded C-propeptides then interact in a non-covalent fashion creating a nucleation site for triple helix formation (see McLaughlin & Bulleid 1998). The complex is further stabilized by formation of interchain disulphide bonds (Olsen *et al.* 1976), which is accelerated by PDI (Koivu & Myllylä 1987). The zipper-like folding of the helical α chains into a right-handed triple helix then proceeds in a C to N direction and the process is completed by association of the N-propeptides (Bächinger *et al.* 1980). In general, the folding of the triple helix requires that every third amino acid in each α chain should be glycine, the side chain of which is small enough to fit into the centre of the triple helix. Specific chain recognition is facilitated by internal recognition sequences in the C-propeptides (Lees *et al.* 1997), although these propeptides as such are not necessary for the nucleation or correct alignment of the triple helix. Instead, a minimum of two -Gly-X-4Hyp- triplets in the C-terminus of the collagenous region is required for nucleation (Bulleid *et al.* 1997). The correctly folded procollagen molecules are

transferred by a mechanism that is not yet completely understood into the Golgi complex, where N-linked carbohydrates are further processed. Incorrectly folded proteins are retained in the ER until they are degraded or folded properly (see Hammond & Helenius 1995). Recent evidence has verified the role of prolyl 4-hydroxylase as a molecular chaperone binding incorrectly folded fibrillar procollagens in the ER (Walmsley *et al.* 1999), as suggested earlier (Chessler & Byers 1992). In addition, immunoglobulin heavy chain-binding protein (BiP or GRP78)(Chessler & Byers 1993), heat shock protein HSP47 (see Nagata 1996) and PDI (Kellokumpu *et al.* 1997, Wilson *et al.* 1998) have been implicated as parts of the quality control system for collagen folding, at least in certain situations (Walmsley *et al.* 1999). It has recently been shown (Bonfanti *et al.* 1998) that procollagens undergo cisternal maturation (see Glick & Malhotra 1998) and form large aggregates in the Golgi before deposition into the extracellular space as secretory granules.

2.1.1.2 Extracellular events

The extracellular processing of fibrillar procollagens starts by removal of the propeptides with distinct procollagen N and C-proteinases. As a result, the solubility of the collagen molecules decreases dramatically and the molecules spontaneously associate, forming elongated fibrils. The telopeptides, the short non-collagenous segments remaining at the ends of the triple helix after cleavage of the propeptides, are crucial for fibril formation. Since non-fibrillar collagens that form supramolecular aggregates do not undergo proteolytic removal of their non-collagenous N and C-termini, their conversion from soluble into insoluble structures must involve another, unknown mechanism, *e.g.* formation of insoluble heterotypic molecules with other constituents of the matrix. (See Prockop & Kivirikko 1995, Kadler *et al.* 1996.)

The formation of collagen fibrils is stabilized by the introduction of covalent cross-links within and between the molecules. The cysteine residues form disulphide bonds, while lysyl oxidase initiates reactions between certain lysine and hydroxylysine residues, providing high tensile strength and mechanical stability. (See Kielty *et al.* 1993, Prockop & Kivirikko 1995.)

2.1.2 Fibrillar collagens

The group of fibrillar, or fibril-forming, collagens consists of types I, II, III, V and XI. These molecules are characterized by possession of a triple helix approximately 1000 amino acid residues long flanked by short N and C-telopeptides. The monomers self-associate into long, cross-linked fibrils that provide the structural backbone for a number of tissues. The fibrils are quarter-staggered, *i.e.* the adjacent collagen molecules overlap longitudinally by about one quarter of their length, creating the characteristic banding pattern seen in electron microscopy. The diameter of the fibril varies depending on the

collagen type, and also depending on the tissue concerned, the ultrastructural location and the developmental stage. (See Eyre 1991, Kielty *et al.* 1993, Prockop & Kivirikko 1995, Kadler *et al.* 1996.)

Collagen types I, II and III are referred to as the major fibrillar collagens, implying abundance in a number of tissues. Type I collagen is expressed in most connective tissues and it is the most abundant type of all, being the major structural component of skin, bone, tendon and ligaments. Type I collagen is mostly present in the form of heterotrimers of two α1(I) chains and one α2(I) chain, encoded by the COL1A1 and COL1A2 genes, respectively. Type II is the major collagenous component of cartilage, the vitreous humour and the intervertebral disc, and is also detected in the inner ear and transiently in numerous other tissues during development. It is a homotrimer of three α1(II) chains encoded by the COL2A1 gene. Type III collagen is also a homotrimer, consisting of α1(III) chains. It is expressed in most tissues that contain type I collagen, but not in bone or tendon. Type III is an abundant component of elastic tissues, including the skin, blood vessels, gut and lung, and can be assembled into heterotypic fibrils with type I collagen. (See van der Rest & Garrone 1991, Prockop & Kivirikko 1995.)

All the genes encoding components of the major fibrillar collagens share a similar structure, with only a few differences. This basic structure consists of 52 exons with analogous exon sizes between the genes and species. The exons encoding the triple helix always begin and end with sequences encoding a complete -Gly-X-Y- triplet (the 9-bp rule) and the exons are typically of 54 bp, multiples of 54 bp, or combinations of 45 and 54 bp in length. (See Chu & Prockop 1993, Prockop & Kivirikko 1995.)

Collagen types V and XI are referred to as minor fibrillar collagens, in view of their presence in tissues in lower amounts than types I, II and III. Like the major fibrillar collagens, types V and XI consist of a triple helix spanning about 1000 amino acids in length flanked by large N and C-terminal propeptides. Their N-propeptides are partially retained in mature molecules, however. Also, the common genomic organization of genes producing major fibrillar collagens is not seen in those encoding types V and XI, but instead they share a different exon-intron pattern indicating a common origin distinct from the major fibrillar collagens (Takahara *et al.* 1995, Vuoristo *et al.* 1995, Annunen *et al.* 1999a). Type V is composed of genetically distinct α1(V), α2(V) and α3(V) chains. The most common trimeric form is $[\alpha1(V)]_2\alpha_2(V)$, but other homotrimeric and heterotrimeric forms also exist. Type V is in general expressed in tissues that also contain types I and III, and has been shown to form heterotypic fibrils with type I collagen. (See Fichard *et al.* 1994.) The most abundant form of type XI collagen is a heterotrimer of α1(XI), α2(XI) and α3(XI) chains. The first two chains are distinct products of the COL11A1 and COL11A2 genes, while the α3(XI) chain is an overglycosylated product of the COL2A1 gene. Type XI is mainly present in cartilaginous tissues and in the vitreous body of the eye, and it is a component of fibrils containing type II collagen. The α2(XI) chain is not present in the vitreous body, where it is replaced by the homologous α2(V) chain, resulting in a heterotypic collagen molecule. Also, the α1(XI) chain has been shown to replace the α1(V) chain in bone increasingly with age, and by analogy, the α1(V) chains can replace some of the α1(XI) chains in maturing cartilage. In view of the structural and apparent functional similarities between collagen types V and XI, a common designation as type V/XI collagen has been proposed. (See Eyre & Wu 1987, Fichard *et al.* 1994; for more details on type XI collagen, see section 2.3.3.)

2.1.3 Non-fibrillar collagens

Instead of forming fibrils, the non-fibrillar collagens serve various other functions. Characteristically their triple helix is divided into several segments on account of non-collagenous interruptions. The non-fibrillar collagens may be divided into six subgroups in terms of their structure or function. (See Prockop & Kivirikko 1995.)

The largest subgroup is known as FACIT (fibril-associated collagens with interrupted triple-helices) compraising types IX, XII, XIV, XVI and XIX. The molecules of this subgroup share sequence homologies in certain domains and a conserved pattern of cysteine residues at the junction of their extreme C-terminal collagenous domain and the non-collagenous NC1 domain. (See Shaw & Olsen 1991, Mayne & Brewton 1993; for more details on FACITs, see section 2.1.4.)

The network-forming collagens include types IV, VIII and X. Type IV collagen networks are present in all basement membranes at numerous locations in the body. In most locations this supporting and controlling network contains α1(IV) and α2(IV) chains, but certain basement membranes, *e.g.* glomerular membranes, also include α3(IV), α4(IV), α5(IV) and α6(IV) chains. The collagenous region of these α chains is about 1400 amino acids long with numerous short interruptions. The genes encoding these polypeptides form an interesting exception among the generally randomly located collagen genes as they are located pairwise on three chromosomes in a head-to-head fashion. (See Timpl & Brown 1996, Sado *et al.* 1998.)

The two other network-forming collagens, types VIII and X, are mutually highly similar, but structurally different from type IV. Type VIII collagen forms the hexagonal lattices in Descemet's membrane of the eye, but is also expressed by various other tissues, such as the skin and glomeruli. It may have a function in promoting the motility of endothelial and smooth muscle cells. The suggested chain composition of type VIII collagen is $[\alpha1(VIII)]_2\alpha2(VIII)$. In contrast, type X is a homotrimer of α1(X) chains with expression restricted to hypertrophic chondrocytes in the deep-calcifying zone of cartilage. The hexagonal lattice formed by type X collagen may be involved in endochondral ossification and mineralization, and possibly also in angiogenesis. (See Shuttleworth 1997, Sutmuller *et al.* 1997.)

Collagen types XIII and XVII contain a transmembrane domain near their N-terminus and are non-secretable proteins (Hägg *et al.* 1998, see also Pihlajaniemi & Rehn 1995). While these two molecules are not otherwise similar in structure, they contain an extracellular collagenous domain with several interruptions and have been called MACITs (membrane-associated collagens with interrupted triple-helices). The collagenous region of type XIII collagen consists of three separate domains reaching into the ECM. Analysis of recombinant type XIII collagen suggests that a homotrimeric chain composition is likely (Snellman *et al.* 2000). Type XIII is expressed in a wide variety of tissues and undergoes extensive alternative splicing (see Pihlajaniemi & Rehn 1995). In the skin it is localized at cell-cell and cell-matrix contacts, and may be a component of adherens-type junctions (Peltonen *et al.* 1999). The homotrimeric type XVII collagen, initially characterized as the 180-kDa bullous pemphigoid antigen, or BPAG2, is an important component of the hemidesmosomes of the skin and cornea, organelles that

attach the epithelial cells to the underlying basement membrane. (See Pihlajaniemi & Rehn 1995, Pulkkinen & Uitto 1999.)

Collagen types XV and XVIII are structurally similar proteins with a large globular N-terminus, a highly interrupted triple helix and a large non-helical C-terminus. These proteins have been assigned the name MULTIPLEXINs (for proteins with multiple triple helix domains and interruptions) (Oh *et al.* 1994). The α1(XV) and α1(XVIII) chains have been characterized, but the exact chain composition of these collagens remains to be determined. (See Pihlajaniemi & Rehn 1995.) Type XV is expressed in the basement membrane zones of most capillaries and several other tissues, while type XVIII has a partially overlapping expression pattern with its highest levels in the liver (Myers *et al.* 1996, Saarela *et al.* 1998). The functions of types XV and XVIII are unknown, but the C-terminal fragment of type XVIII, named endostatin, prevents angiogenesis by inhibiting endothelial cell proliferation and migration (O'Reilly *et al.* 1997). The homologous C-terminal region of type XV collagen, termed restin, has recently been shown to inhibit endothelial cell migration (Ramchandran *et al.* 1999).

Type VI collagen is a heterotrimer of α1(VI), α2(VI) and α3(VI) chains consisting of a short triple helix flanked by large globular N and C-terminal domains. It is the only collagen aggregating into beaded microfilaments, which are to be found on the cell surface and around collagen fibres in most connective tissues and may serve to anchor the cells to the macromolecular framework of the ECM. (See Bruckner & van der Rest 1994, Prockop & Kivirikko 1995, Kielty & Shuttleworth 1997.)

Type VII collagen forms anchoring fibrils upon dimerization and lateral association of homotrimeric $\alpha1(VII)_3$ molecules. These fibrils link the epithelial basement membrane to the underlying ECM in skin, cornea and several other epithelial tissues. The highly interrupted triple helix of type VII is the longest among all the collagens, and the gene COL7A1 encoding it has the highest number of exons of the known genes, *i.e.* 108. (See Prockop & Kivirikko 1995, Uitto & Pulkkinen 1996.)

2.1.4 FACIT collagens

As mentioned above, the FACIT group comprises collagen types IX, XII, XIV, XVI and XIX. These molecules are incapable of forming collagen fibrils by themselves, but types IX, XII and XIV are believed to interact with existing fibrils by lateral association of one or more of their triple-helical domains. Type IX, the prototype member of the FACIT group, is formed of α1(IX), α2(IX) and α3(IX) chains and consists of three relatively short triple-helical domains (COL1-COL3 numbered from the C-terminus) interspersed and flanked by four non-collagenous domains (NC1-NC4), of which the NC4 domain consists mainly of the α1(IX) chain. In general, type IX collagen is expressed in the same tissues as type II (see section 2.2.3) and has been shown to associate with fibrils containing type II via lysine-derived cross-links. The location demonstrated for the NC4 domain outside the fibril body, indicates a possible role for this domain as a bridging molecule between various ECM components. Interestingly, this domain shows homology to the N-terminal heparin binding domain of thrombospondin-1, and a homologous

region is found in all FACITs, as also in the N-terminal PARP region of collagen types V and XI and in the N-terminal region of types XV and XVIII (see Bork 1992, Fichard *et al.* 1994, Pihlajaniemi & Rehn 1995). The residues believed to be responsible for heparin binding are not conserved in any of these collagens, and the significance of this domain within the collagens is an enigma (see Pihlajaniemi & Rehn 1995). (See Shaw & Olsen 1991, Bruckner & van der Rest 1994; for more details on type IX collagen, see section 2.2.)

2.1.4.1 Collagen types XII and XIV

Type XII collagen was originally characterized from a chicken tendon fibroblast library as a cDNA resembling α1(IX) and encoding a multidomain protein containing two collagenous domains, COL1 and COL2, and three NC domains (NC1-NC3). The C-terminal COL1 domain was found to be homologous to the COL1 domain of the α1(IX) chain and the N-terminal NC3 domain had a region homologous to the NC4 domain of α1(IX). (Dublet & van der Rest 1987, Gordon *et al.* 1987, 1989.) The NC3 domain also contained multiple fibronectin type III repeats and units homologous to the von Willebrand factor A domain (Yamagata *et al.* 1991). These motifs are likewise present in collagen types VI and VII (see Bork 1992). The protein was shown to be a homotrimer of 220-kDa α1(XII) chains, and appeared in rotary shadowing to consist of a triple-helical tail connected by a central globule to three finger-like extensions representing the NC3 domains (Dublet *et al.* 1989, Lunstrum *et al.* 1991). Characterization of a longer cDNA for α1(XII) from chicken fibroblasts (Yamagata *et al.* 1991) was followed by identification of a corresponding alternative form of type XII collagen in several species (Koch *et al.* 1992, Lunstrum *et al.* 1992, Watt *et al.* 1992, Oh *et al.* 1993). The shorter form of the protein, called XIIB, results from alternative splicing of several exons that are present in the mRNA of the longer form, XIIA (Trueb & Trueb 1992a). This splicing removes a region of NC3 that is capable of binding a GAG side chain (Koch *et al.* 1992, Watt *et al.* 1992) and also contains a heparin binding site (Koch *et al.* 1995). More detailed studies show that the two variants have different spatial and temporal expression patterns, suggesting different functions (Bohme *et al.* 1995, Berthod *et al.* 1997, Gerecke *et al.* 1997). The existence of two C-terminal variants resulting from alternative splicing was recently demonstrated in the mouse and rat, and differences in temporo-spatial expression patterns have also been reported between these isoforms (Kania *et al.* 1999). Type XII is present in dense connective tissues containing type I collagen, such as tendons, ligaments, bone and skin, but also in cartilage, which is devoid of type I collagen (Gordon *et al.* 1987, Sugrue *et al.* 1989, Yamagata *et al.* 1991, Watt *et al.* 1992).

Type XIV collagen was originally isolated from foetal bovine tendon and skin as a homotrimeric molecule structurally highly similar to type XII collagen and as a cDNA clone and a pepsin-resistant fragment from chicken skin (see Mayne & Brewton 1993). Its appearance in electron microscopy was similar to that of type XII collagen described above (Lunstrum *et al.* 1991, Aubert-Foucher *et al.* 1992), and the structural motifs present in the NC3 domain of type XII were also present in the NC3 domain of type XIV,

with differences in the numbers of repeats (Gerecke *et al.* 1993, Wälchli *et al.* 1993). In general, type XIV is present in the same tissues containing type I as is type XII, but the temporo-spatial expression patterns are quite dissimilar, suggesting a difference in function (Castagnola *et al.* 1992, Wälchli *et al.* 1994, see Garrone *et al.* 1997). Developmentally regulated alternative splicing results in two forms of the NC3 domain, possibly differing in their interaction properties due to alternative conformations of a fibronectin type III repeat (Imhof & Trueb 1998). Alternative splicing also yields additional mRNA variants, resulting in differences in the NC1 domain (Wälchli *et al.* 1993) and variants with different 5'-untranslated regions (Gerecke *et al.* 1993), but the expression patterns of these forms are unknown. Type XIV collagen shows partial similarity to a non-collagenous ECM glycoprotein called undulin (Schuppan *et al.* 1990, Trueb & Trueb 1992b). In view of their sequence similarities (Trueb & Trueb 1992b, Brown *et al.* 1993, Gerecke *et al.* 1993, Wälchli *et al.* 1993, Bauer *et al.* 1997) and the assignment of their genes to the same region on human chromosome 8 (Schnittger *et al.* 1995, Imhof & Trueb 1999), these two are believed to be different splice variants encoded by same gene.

The functions of collagen types XII and XIV are currently unknown. *In vivo*, these molecules are found on the surface of type I collagen fibrils with the NC3 domain located outside the fibril body (Schuppan *et al.* 1990, Keene *et al.* 1991, Niyibizi *et al.* 1995). This association may occur via the collagenous domains (Koch *et al.* 1995), analogously with the type IX/type II association, suggesting that these FACITs may mediate interactions between collagen fibrils and other ECM macromolecules or cells. Unlike type IX collagen, types XII and XIV are not covalently cross-linked to the fibril surface (Dublet *et al.* 1989, Lunstrum *et al.* 1991, Aubert-Foucher *et al.* 1992). Type XIV has been shown to bind several cell types via a chondroitin/dermatan sulphate proteoglycan (Ehnis *et al.* 1996). Binding occurs at the N-terminal fibronectin type III repeat of the NC3 domain (Ehnis *et al.* 1998). A small leucine-rich proteoglycan, decorin, is known to compete for this binding, suggesting a role for decorin as a modulator of cellular interactions with type XIV (Font *et al.* 1993, Ehnis *et al.* 1997). Since decorin and another small proteoglycan, fibromodulin, are known to bind the homologous type XII collagen, they may also mediate or modulate its cellular interactions (Font *et al.* 1996). Finally, *in vitro* studies suggest that the NC3 domains of types XII and XIV are involved in mediating the interactions occurring between collagen fibrils (Nishiyama *et al.* 1994, Akutsu *et al.* 1999).

2.1.4.2 Collagen types XVI and XIX

By comparison with the other FACIT molecules, relatively little is known about collagen types XVI and XIX. Type XVI was first isolated as cDNA clones from human skin and placenta. The sequences predicted a protein with 10 collagenous domains interspersed and flanked by cysteine-rich non-collagenous domains, of which only the N-terminal NC11 domain was notably large. Identification of the FACIT hallmarks, a Cys-X-X-X-X-Cys -motif at the NC1/COL1 junction and a region with limited homology to the

thrombospondin-like NC4 domain of the α1(IX) chain, led to the inclusion of type XVI collagen in the FACIT subgroup. (See Mayne & Brewton 1993.) Transfection experiments with α1(XVI) constructs in kidney cells demonstrated a globular appearance for the NC11 domain and suggested a homotrimeric molecular composition (Tillet *et al.* 1995), a finding that was subsequently verified by immunoprecipitation of type XVI from fibroblast and smooth muscle cell cultures (Grassel *et al.* 1996, Lai & Chu 1996). Type XVI is widely expressed late in mouse embryonic development and in adult mice, with strong expression in the heart, kidney, intestine, eye and arterial walls, for example (Lai & Chu 1996). A high concentration of type XVI was recently detected in the subepidermal zone of the skin, where it occurred in close proximity to the type VII collagen of anchoring fibrils and co-localized with the fibrillin1 of microfibrils. This suggests an interaction with components of the anchoring complexes and the microfibrillar apparatus and a possible function in stabilizing the dermo-epidermal interface (Grassel *et al.* 1999). Unlike collagen types IX, XII and XIV, type XVI appears not to be associated with major collagen fibrils in tissues (Grassel *et al.* 1996).

Partial cDNA clones for type XIX collagen, initially called the α1(Y) chain, were identified in a human rhabdomyosarcoma cell cDNA library through homology with type V collagen (Yoshioka *et al.* 1992, Myers *et al.* 1993). Characterization of additional cDNA clones verified the protein as a distinct type of collagen consisting of five collagenous domains flanked by six NC domains (Myers *et al.* 1994, Inoguchi *et al.* 1995). Northern blotting identified an mRNA of over 10 kb with a surprisingly large 3'-untranslated region, where unusual alternative splicing was demonstrated (Myers *et al.* 1993, 1994, Inoguchi *et al.* 1995). Since type XIX appeared to share with the other FACITs the presence of the C-terminal cysteines, two interruptions in COL1 domain and limited homology to the α1(IX) NC4 domain, it was included in the FACIT group (Yoshioka *et al.* 1992, Myers *et al.* 1993, 1994, Inoguchi *et al.* 1995). Type XIX mRNA was found to be extremely extensively present in mouse embryonic tissues, but was restricted mainly to the brain, eye and testis in adult tissues, while immunochemical methods detected the protein in mouse brain (Sumiyoshi *et al.* 1997) and in the basement membrane zone of various human tissues (Myers *et al.* 1997). Although a recent study suggests a role for type XIX in the differentiation of skeletal muscle cells (Myers *et al.* 1999), its exact functions and interactions remain undetermined.

2.2 Type IX collagen

Evidence for the existence of type IX collagen was first obtained in the early 80's, when extraction of homogenized and pepsin-treated mammalian cartilage and intervertebral disc resulted in the isolation of disulphide-bonded collagenous fragments, initially designated as type M collagen and CF2 (Shimokomaki *et al.* 1980, 1981), C-PS1 and C-PS2 collagen (Ayad *et al.* 1981, 1982) or X_1-X_7 (Ricard-Blum *et al.* 1982). Fragments of slightly different size were obtained from chicken sternal cartilage by similar methods and were designated HMW and LMW (Reese & Mayne 1981, Reese *et al.* 1982) or M1 and M2 (von der Mark *et al.* 1982). The intact parental α chains of these proteolytic

fragments were first described as H and J polypeptides secreted by cultured chicken sternal chondrocytes (Gibson *et al.* 1983). Subsequently, the intact molecule containing HMW and LMW, or M1 and M2, was isolated as p-HMW-collagen or pM collagen from chicken sternal cartilage cultures (Bruckner *et al.* 1983, von der Mark *et al.* 1984), and from chicken epiphyseal cartilage as a proteoglycan, PG-Lt (Noro *et al.* 1983). Rotary shadowing experiments with a similar molecule extracted from rat chondrosarcoma showed that it consisted of a short and a long arm connected by a flexible hinge region (Duance *et al.* 1984), a structure earlier described for the HMW/M1 fragment (von der Mark *et al.* 1982, Reese *et al.* 1982). Characterization of a partial cDNA for this novel collagen from chicken sternum revealed a striking multidomain composition (Ninomiya & Olsen 1984), while detailed analysis of the LMW and HMW indicated a molecular composition of three distinct chains (Mayne *et al.* 1985a, van der Rest *et al.* 1985). Eventually the molecule was designated type IX collagen (van der Rest *et al.* 1985).

2.2.1 Structure of type IX collagen

Type IX collagen consists of α1(IX), α2(IX) and α3(IX) chains (van der Rest & Mayne 1987). Complete cDNA sequences have been published for the human α1(IX) and α3(IX) chains, while those for the α2(IX) chain lack the 5'-half of the signal peptide. These sequences show that the collagenous region is divided into COL1, COL2 and COL3 domains, numbered from the C-terminus (Table 1), flanked by four NC domains (NC1-4, numbered from the C-terminus). Additional short interruptions are present, two in the COL1 domain of each chain and one in the COL3 domain. Unlike the situation in the fibrillar collagens, the C-terminal NC1 domains are very short. The size of the NC2 region is 30 residues in the α1(IX) and α2(IX) chains, but 31 in the α3(IX) chain. The size of the NC3 region is different in all three chains, which explains the kink observed in electron microscopy (von der Mark *et al.* 1982, Reese *et al.* 1982, Duance *et al.* 1984). Two cysteines present in each chain, at the NC1/COL1 junction and in the NC3 domain,

Table 1. Sizes of the individual domains of human type IX collagen

	Size (amino acid residues)		
Domain	α1(IX)	α2(IX)	α3(IX)
NC4	268[a]	nd	28[a]
COL3	137	137	137
NC3	12	17	15
COL2	339	339	339
NC2	30	30	31
COL1	115	115	112
NC1	20	25	22

[a] includes a signal peptide
nd not determined

are involved in the formation of stabilizing interchain disulphide bonds. It appears that the N-terminal NC4 region spans only 3 residues in the α2(IX) and α3(IX) chains, although the exact location of the signal peptide cleavage sites has not been determined. In contrast, the NC4 region of the α1(IX) chain is 245 residues long, excluding a putative 23-residue signal peptide, and in view of its calculated pI of 10.4, carries a strong positive charge at physiological pH. The NC4 domain contains four Cys residues that are likely to participate in intrachain disulphide bonding and stabilize the globular conformation demonstrated for this domain by electron microscopy (Irwin *et al.* 1985, Mayne *et al.* 1985b, Bruckner *et al.*1988, Vaughan *et al.* 1988). Where type IX collagen molecules are associated with type II collagen fibrils in cartilage, the short arm formed by the COL3 and NC4 domains projects away from the fibril surface, while the rest of the molecule, the long arm, is oriented parallel to the fibril surface (van der Rest & Mayne 1988, Vaughan *et al.* 1988; for further details, see section 2.4.1). The higher thermal stability demonstrated for the triple helix of the short arm than for that of the long arm (Bruckner *et al.* 1983, Miles *et al.* 1998) is explainable by its high 4-hydroxyproline content (Reese *et al.* 1982). The differing stabilities match the molecular arrangement, in that a lower thermal stability is acceptable for the long arm due to the stabilizing association with the fibril. (Kimura *et al.* 1989, Muragaki *et al.* 1990a,b, 1996, Perälä *et al.* 1993, 1997, Rokos *et al.* 1994, Warman *et al.* 1994, Brewton *et al.* 1995, see also Ninomiya *et al.* 1990.)

The overall domain structure is highly conserved between species, so that the sizes of the COL1-3, NC2 and NC3 regions in each chain are exactly the same in the chicken, for example, as in the human, with minor size differences in the NC1 and NC4 domains. An additional difference is the absence in the human protein of the short interruption seen in the COL2 domain of the chicken α3(IX) chain. The human NC4 domain has a functional attachment site for an N-linked oligosaccharide (Muragaki *et al.* 1990a, Warman *et al.* 1993a). Chicken type IX collagen is also a target for glycosylation, but apparently at a different location (Bruckner *et al.* 1985). (Ninomiya & Olsen 1984, Ninomiya *et al.* 1985, McCormick *et al.* 1987, Vasios *et al.* 1988, Nishimura *et al.* 1989, Brewton *et al.* 1992, Har-El *et al.* 1992.)

2.2.1.1 Type IX collagen as a proteoglycan

The existence of chicken type IX collagen in proteoglycan form was revealed upon the demonstration of identity with the previously isolated PG-Lt (Noro *et al.* 1983, Bruckner *et al.* 1985, Vaughan *et al.* 1985). A sulphated glycosaminoglycan (GAG) side chain (chondroitin/dermatan sulphate) is attached to a Ser residue within the NC3 domain of the α2(IX) chain (Huber *et al.* 1986, Konomi *et al.* 1986, McCormick et al. 1987, Huber *et al.* 1988). Subsequently the presence of a GAG chain was demonstrated in type IX collagen in humans and other mammals (Bruckner *et al.* 1988, Ayad *et al.* 1989, 1991, Arai *et al.* 1992, Bishop *et al.* 1992).

The presence of a GAG chain is not a uniform feature of all type IX molecules, but shows species and tissue dependence. Both proteoglycan and non-proteoglycan forms are synthesized by human, bovine and chicken cartilage in organ culture (Bruckner *et al.*

1988, Ayad *et al.* 1991, Yada *et al.* 1992). Most of the type IX collagen extractable from human foetal cartilage contains a chondroitin/dermatan sulphate GAG chain (Bruckner *et al.* 1988), while virtually all of that extracted from foetal bovine articular cartilage is in a non-proteoglycan form (Ayad *et al.* 1989). Type IX collagen in the vitreous body of the chicken contains a GAG chain 10 times larger than in cartilage, forming a thick coat around the fibrils (Wright & Mayne 1988) and presumably serving as the major structural vitreous proteoglycan instead of the hyaluronan, which fulfils this role in mammals. All the extractable type IX in chicken and bovine vitreous body carries a GAG chain. (Yada *et al.* 1990, Bishop *et al.* 1992, 1994.)

2.2.2 Genes encoding type IX collagen

The human α1(IX), α2(IX) and α3(IX) chains are encoded by the COL9A1, COL9A2 and COL9A3 genes, respectively. These are located on three chromosomes: COL9A1 at 6q12-q13, COL9A2 at 1p32.3-p33 and COL9A3 at 20q13.3 (Kimura *et al.* 1989, Warman *et al.* 1993b, 1994, Brewton *et al.* 1995, Tiller *et al.* 1998).

The type IX collagen genes have been characterized to variable extents in the chicken, mouse and human. The complete exon structure of the chicken α2(IX) gene has been reported (Lozano *et al.* 1985, McCormick *et al.* 1987, see Ninomiya *et al.* 1990), as has a partial structure for the chicken α1(IX) gene (Lozano *et al.* 1985, Vasios *et al.* 1988, Nishimura *et al.* 1989, see Ninomiya *et al.* 1990). A complete sequence for the mouse Col9a2 gene has been published (Perälä *et al.* 1994), along with structures of a few exons at the 5' end of the mouse Col9a1 gene (Muragaki *et al.* 1990b), but in the human case only the structures of four exons in the 5' region of COL9A1 have been reported (Muragaki *et al.* 1990b). Comparison of the structures of the mouse and chicken α2(IX) genes revealed an almost identical exon structure, with 32 exons. Most of the exons coding for the COL2 and COL3 domains followed the "9-bp rule" of the fibrillar collagens, whereas the exons for the COL1 domain and parts of the COL2 domain did not (Perälä *et al.* 1994). The known exon structures of the chicken and mouse α1(IX) genes are also similar to those of the α2(IX) and α3(IX) genes (Muragaki *et al.* 1990b, see Ninomiya *et al.* 1990).

The atypical exon sizes and the presence of six junctional exons encoding both COL and NC regions indicate that the genes for type IX collagen are evolutionarily distinct from those encoding the fibrillar collagens (see Ninomiya *et al.* 1990). They do, however, show some similarity to other FACIT genes. Analysis of the structure of the human COL19A1 gene, encoding the α1(XIX) chain, revealed similarity to the exon structures of the α1(IX) and α2(IX) genes, implying a common ancestral origin (Khaleduzzaman *et al.* 1997). Also, the parts of the type XII collagen gene that encode for the NC1/COL1 region show some similarity to the type IX collagen genes (Gordon *et al.* 1989). The human COL9A1, COL12A1 and COL19A1 genes all map to same locus, 6q12-q13 (Oh *et al.* 1992, Yoshioka *et al.* 1992, Warman *et al.* 1993b, Gerecke *et al.* 1997, Khaleduzzaman *et al.* 1997).

2.2.2.1 Alternative promoters

The NC4 domain of the α1(IX) chain has two alternative forms. While the long form described above contains a large globular NC4 domain and is mainly present in cartilage, a short form lacking this large NC4 domain predominates in the primary corneal stroma and the vitreous body of the eye (Svoboda *et al.* 1988, Yada *et al.* 1990). This results from the use of two different promoters, located about 20 kb apart in the chicken gene for α1(IX). The downstream promoter, giving rise to the short form, is located within the sixth intron of the gene. The transcript of this downstream promoter starts with sequences encoded by alternative exon 1 (or exon 1*) that are directly spliced to sequences encoded by exon 8, thus skipping exon 7. (Nishimura *et al.* 1989, Muragaki *et al.* 1990a,b.) Assuming a signal peptide of 23 residues, the NC4 region of the human short α1(IX) variant contains only two amino acid residues, in contrast to the 245 residues of the long form (Muragaki *et al.* 1990a). The short form of type IX collagen also associates with the surface of collagen fibrils (Wright & Mayne 1988, Ren *et al.* 1991) and is likely to be involved in molecular interactions distinct from those of the long form (Nishimura *et al.* 1989).

2.2.3 Tissue expression of type IX collagen

Type IX collagen is in general expressed in the same tissues as types II and XI, but the temporo-spatial expression patterns of the respective mRNAs, including the alternative forms, are not identical. Type IX is expressed as a quantitatively minor component in all hyaline cartilages, and the developmental expression patterns of the α1(IX), α2(IX) and α3(IX) mRNAs in these tissues appear to be co-ordinated (Perälä *et al.* 1997, Savontaus *et al.* 1998), with expression of the long form of α1(IX) dominating over that of the short form, except in the early stages of development (Hayashi *et al.* 1992, Swiderski & Solursh 1992).

Although it was initially described in cartilaginous tissues, the definition of type IX collagen as ‘a cartilage-specific collagen’ has been shown to be an understatement, since its expression has been detected in numerous extra-cartilaginous tissues. That in ocular tissues has been extensively studied, and expression has been found in the embryonal primary corneal stroma and the embryonal and adult vitreous body of the chicken (Fitch *et al.* 1988, 1995). The presence of type IX collagen has also been demonstrated in the bovine and human vitreous body (Bishop *et al.* 1992, 1994, Warman *et al.* 1993a), and in the embryonal and adult mouse eye (Liu *et al.* 1993). These observations showed that the α1(IX) chain is primarily expressed in its short form in ocular structures.

Type IX collagen is present in all the layers of intervertebral disc, accounting for about 2% or less of its total collagen. It is the short form that is predominant in the central gel-like region, the nucleus pulposus, a situation similar to that in another gelatinous matrix, the vitreous body. In the outer layer, the annulus fibrosus, both isoforms of type IX collagen are present. The cartilage endplates located between the discs and the vertebrae consist of hyaline cartilage reminiscent of articular cartilage and contain only long form

of type IX collagen. A certain proportion of the type IX collagen in all layers of the disc contains a GAG side chain. (Newall & Ayad 1995, see also Humzah & Soames 1988, Buckwalter 1995.)

Type IX collagen mRNAs have been reported in several other non-cartilaginous tissues, such as the heart, brain, skin and kidney, but it is not known whether these mRNAs are translated into protein. (Liu *et al.* 1993, Perälä *et al.* 1997.) In the inner ear type IX collagen associates with thick type II collagen fibrils (Slepecky *et al.* 1992).

2.2.4 Biosynthesis and degradation of type IX collagen

The intracellular steps in the biosynthesis of type IX collagen are mostly similar to those for other collagens. Since no large, cleavable propeptides are present at the termini of type IX collagen or the other FACITs, some differences can be expected relative to the fibrillar collagens. The chain recognition and assembly of type IX collagen was first studied by means of *in vitro* reconstitution experiments with pepsinized LMW fragments containing the COL1 domain and the cysteine-bearing segment of the NC1 domain (Labourdette & van der Rest 1993). The results showed that the simultaneous presence of fragments from all three α chains favoured the formation of α1α2α3 heterotrimers. The chains isolated were shown to be capable of homotrimerization, although only at very low levels in case of the α3(IX) chain. The authors concluded that the fragments contained at least part of the information required for selective chain association. Evidence for the importance of the COL1 domain in the assembly of the FACIT collagens was also obtained by expressing a type XII collagen minigene in HeLa cells (Mazzorana *et al.* 1993, 1995) and insect cells (Mazzorana *et al.* 1996). In these cases the synthesis of disulphide-bonded trimers was seen only if triple helix formation had not been prevented. A model for trimer assembly in the FACITs was proposed, stating that the information for chain selection is contained within the COL1/NC1 junction and that the last triplets of the COL1 domain must fold into a trimer stabilized by prolyl 4-hydroxylation before the assembly is fixed by the formation of disulphide bonds (Lesage *et al.* 1996, Mazzorana *et al.* 1996). A study with hydroxylated synthetic peptides representing the NC1 domains and the five extreme C-terminal tripeptide units of the type IX collagen α chains but lacking the capacity for triple helix formation resulted in an alternative explanation for the apparent importance of prolyl 4-hydroxylase suggesting that the observed generation of disulphide-bonded multimers in the absence of any triple helix formation could be explained by a contribution from a few C-terminal 4-hydroxyprolines to chain recognition and assembly, but not from any non-hydroxylated prolines (Mechling *et al.* 1996). In the assembly of fibrillar collagens hydroxylation is required for the initial nucleation of triple helix folding, and the assembly does not depend on the formation of interchain disulphide bonds (see McLaughlin & Bulleid 1998). In conclusion, the large C-propeptides of fibrillar collagens but not their telopeptides are crucial for the selective association of the α chains, whereas in the FACITs a few amino acids at the COL1/NC1 junction appear to serve in this role (Labourdette & van der Rest 1993, Mazzorana *et al.* 1995, see McLaughlin & Bulleid 1998).

Matrix metalloproteinases (MMPs) degrade a wide variety of ECM components and actively participate in the remodelling of cartilage in endochondral ossification and normal maintenance, as well as in certain pathological conditions such as arthritis (see Birkedal-Hansen *et al.* 1993). MMP-3 (also known as stromelysin-1) produced by chondrocytes and synovial fibroblasts is capable of degrading aggrecan, link protein and collagen types II, IX, X and XI (Wu *et al.* 1991), and MMP-3 may therefore be able to remove type IX collagen from the fibril surface, giving access to additional MMPs and other proteinases. Type IX collagen is also cleaved by MMP-9 (Bollen & Eyre 1993), MMP-2 (Brown *et al.* 1996), cathepsins B and L (Maciewicz & Wotton 1991) and neutrophil elastase (Gadher *et al.* 1988).

2.3 Cartilage

Cartilage is a specialized form of connective tissue found in numerous locations throughout the body and serving different functions. It consists of cartilage-specific cells, chondrocytes, and an abundant extracellular matrix surrounding them, the main constituents of which are water, type II collagen and proteoglycans. Cartilage can be divided into three major forms. The hyaline form is mainly found in articular cartilage, epiphyseal cartilage and in the cartilaginous anlage of bones, the elastic form is characterized by a network of elastin and is located in the external ear, the epiglottis and parts of the larynx, and fibrocartilage, further strengthened by type I collagen fibres, is found in intervertebral discs and at insertion-sites of tendons and ligaments, for example. (See McCarty 1989, Muir 1995.)

2.3.1 Development of cartilage

Cartilage development is initiated by the condensation of mesenchymal precursor cells. As these cells secrete a cartilage ECM around them, they stop producing non-cartilaginous proteins. This differentiation is a complex process and our knowledge of the growth and differentiation factors controlling it is relatively poor. The maturing tissue grows both by interstitial growth resulting from ECM production by mitotically dividing chondrocytes, and by appositional growth arising from differentiation of progenitor cells into chondrocytes. The chondrocytes in mature cartilage are located within small lacunae sparsely distributed throughout an abundant ECM. (See McCarty 1989, Erlebacher *et al.* 1995, Mundlos & Olsen 1997a.)

Cartilage formation is also a crucial precursor to the development of long bones by endochondral ossification, in which cartilage serves as a model, or anlage, which is slowly replaced by uncalcified bone. The sequence of events consists of the formation of hypertrophic chondrocytes, the subsequent death and disappearance of these cells, invasion of the ECM by blood vessels, and deposition of bone matrix by invading osteoblasts. This process is involved in both longitudinal bone growth at the epiphyseal

growth plates and the substitution of bone for epiphyseal cartilage in secondary ossification centres. Cartilage is present in a mature long bone only in a specialized articular form at joint surfaces. (See Erlebacher *et al.* 1995, Mundlos & Olsen 1997a.)

2.3.2 Structure and function of articular cartilage

Articular cartilage is a form of hyaline cartilage covering the epiphyses of long bones at joint surfaces. Like other types of cartilage it is hyperhydrated as a result of an extensive network of hydrophilic sulphated proteoglycans. This creates a swelling pressure, which is counterbalanced by a network of high tensile strength collagen fibres. The fibres at the joint surface run parallel to the surface, while those in the deeper layers are mostly perpendicular to the surface. The articular cartilage provides the joint with shock-absorbing and load-spreading potential. Another function is generation of low-friction properties for joint movement together with the synovial fluid. All cartilages are avascular and rely on passive diffusion for their nutrient supply and metabolite exchange. Partly for this reason the regeneration capacity of cartilage is poor. (See McCarty 1989, Muir 1995, Buckwalter & Mankin 1998.)

2.3.3 Collagenous components of cartilage

The collagen fibrils in cartilage are heteropolymers of collagen types II, IX and XI (Vaughan *et al.* 1988, Mendler *et al.* 1989). Type II, representing about 95% of total collagen in adult cartilage (see Eyre 1991), forms the body of the fibril, with adjacent molecules overlapping by one quarter of their length, or a D-period, resulting in a banded appearance (Fig. 1). The assembly is stabilized by covalent cross-links. (See Eyre *et al.* 1992.)

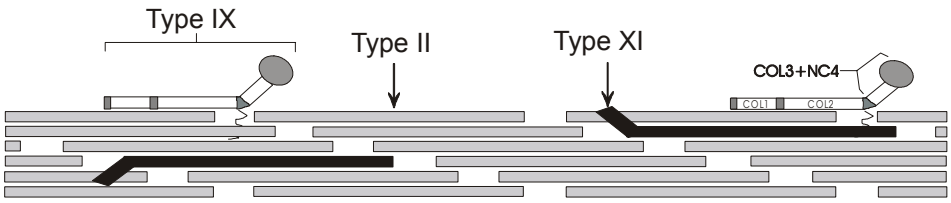


Figure 1. Schematic representation of the arrangement of collagens in a cartilage collagen fibril. The fibril body is formed by type II collagen (grey bars) with minor amounts of type XI (black bars) embedded within it. Type IX molecules are located at the fibril surface, with their COL3 and NC4 domains projecting away from the fibril body. The GAG chain of type IX collagen is located in the gap zone of the fibril.

Collagen types IX and XI are quantitatively minor components of the fibril. Since type XI is immunologically masked in intact fibrils, it is believed to reside in the interior of the fibril (Mendler *et al.* 1989)(Fig.1), in a similar manner to type V collagen within type I fibrils (see Fichard *et al.* 1994). Furthermore, the type XI collagen molecules in cartilage collagen fibrils are primarily cross-linked to each other, suggesting formation of a homopolymeric type XI core fibril around which the type II collagen is assembled (Wu & Eyre 1995). Another suggested role for type XI collagen is regulation of the fibril diameter, perhaps via the retained N-propeptides (see Bruckner & van der Rest 1994, Fichard *et al.* 1994). This is supported by the presence of type XI collagen only on thin cartilage collagen fibrils (Keene *et al.* 1995) and by the presence of abnormally large fibrils in mice that lack a functional α1(XI) chain (Li *et al.* 1995).

Several other collagens are present in cartilage. Type X collagen is restricted to hypertrophic cartilage of the growth plate and appears to function in endochondral ossification. Ultrastructurally, it is found in filamentous mats, probably representing a hexagonal network of type X molecules, and is also associated with collagen fibrils. (See Bruckner & van der Rest 1994, Sutmuller *et al.* 1997.) Type VI collagen is also found in cartilage, where it forms beaded filaments stabilized by interaction with hyaluronan and may participate in the attachment of chondrocytes to ECM. (See Bruckner & van der Rest 1994, Buckwalter & Mankin 1998.)

The splice variants XIIA and XIIB of type XII collagen are both present in cartilage, where XIIA contains a GAG chain in its NC3 domain (Watt *et al.* 1992). Similarly, type XIV is present in cartilage, with a portion of the molecules carrying a GAG chain in the NC3 domain (Castagnola *et al.* 1992, Watt *et al.* 1992). It is not known, however, whether these FACIT molecules associate with the cartilage collagen fibrils. The function of type XIII collagen expressed in cartilage also remains to be demonstrated (Juvonen *et al.* 1992).

2.3.4 Non-collagenous components of cartilage

Entrapped within the collagen network of cartilage are large proteoglycan aggregates formed by interaction of the largest cartilage proteoglycan, aggrecan, with hyaluronic acid. This interaction is stabilized by a non-collagenous protein, the link protein. The aggregates contain up to thousands of sulphated GAG chains, providing the cartilage with its osmotic properties. (See Roughley & Lee 1994.) Another large proteoglycan, perlecan, appears to have an important function in the development or preservation of the integrity of cartilage, but the exact nature of this remains as yet unknown (Arikawa-Hirasawa *et al.* 1999).

Several members of the family of the small leucine-rich proteoglycans are also present in hyaline cartilage, the core proteins of these molecules being capable of interacting with collagen fibrils. Within this group, epiphycan and PRELP are mainly expressed in cartilage, while decorin, biglycan, fibromodulin and lumican are also abundant in non-cartilaginous tissues. Experiments *in vitro* and *in vivo* indicate that decorin, fibromodulin and lumican are associated with the surface of the collagen fibrils and regulate collagen fibrillogenesis, possibly preventing lateral fusion of the fibrils. (See Roughley & Lee 1994, Iozzo 1997.) Decorin resides in the gap-regions of the collagen fibrils in articular cartilage *in vivo*, and is absent from the thinnest fibrils. The location of decorin in areas of cartilage rich in parallel fibril arrays suggests a role in maintaining proper fibril spacing. (Hagg *et al.* 1998.) In contrast, the pericellular location of biglycan suggests a different function, perhaps via interaction with type VI collagen (see Roughley & Lee 1994).

Numerous non-collagenous proteins without a GAG side chain are present in significant amounts in cartilage. Among these the trimeric multi-domain proteins matrilin-1 (also known as cartilage matrix protein or CMP) and matrilin-3 are abundant in nearly all cartilages. Matrilin-1 is a component of type II collagen fibrils, but also associates with aggrecan and forms collagen-independent fibrils by self-assembly, suggesting an important function as a bridging molecule in cartilage ECM. (See Deak *et al.* 1999.) Matrilin-1 knock-out mice nevertheless show little or no change in cartilage ultrastructure (Aszódi *et al.* 1999). Another non-collagenous molecule known to associate with collagen is cartilage oligomeric matrix protein (COMP). This pentameric ECM glycoprotein binds collagen types I and II in a zinc-dependent manner via its C-terminal domain (Rosenberg *et al.* 1998). COMP also binds to chondrocytes, while a homologous protein extractable from articular cartilage, thrombospondin 1, does not (DiCesare *et al.* 1994). (See Neame *et al.* 1999.)

Epiphyseal cartilage is a rich source of potential growth factors or modulators of the cellular phenotype, such as chondromodulin I and II and pleiotrophin/HB-GAM. Anchorin CII, chondroadherin, tenascin, fibronectin and chondronectin, along with several members of the β1 family of integrins (Durr *et al.* 1993), mediate or modify the interactions of the chondrocytes with ECM components. Several other proteins with no known function, such as a typical basement membrane component, laminin (Durr *et al.* 1996), are also abundant in cartilage. (See Buckwalter & Mankin 1998, Neame *et al.* 1999.)

2.4 The function of type IX collagen

In view of its structure and its location in cartilage, several functions have been proposed for type IX collagen in this tissue, although no incontrovertible evidence exists for these. Also, we have no knowledge of the function of the GAG chain in this collagen type, except in the chicken vitreous body, where the large GAG is apparently the major structural proteoglycan (Wright & Mayne 1988).

2.4.1 Type IX collagen in cartilage collagen fibrils

Type IX collagen is a component of most cartilage collagen fibrils, residing at the fibril surface in an antiparallel orientation relative to the type II collagen molecules (Mendler *et al.* 1989, Wu *et al.* 1992)(Fig.1). The interaction is stabilized by lysine-derived cross-links between the N-terminal region of the COL2 domain of each type IX collagen α chain and the N-telopeptide of the α1(II) chain, and also between the central region of the COL2 domain within the α3(IX) chain and the C-telopeptide of the α1(II) chain (Eyre *et al.* 1987, van der Rest & Mayne 1988, Wu & Eyre 1989, Wu *et al.* 1992, Diab *et al.* 1996). This surface location suggests a role for type IX collagen, and perhaps for its GAG chain, in the regulation of fibril diameter. This is supported by the finding that the proportion of type IX collagen decreases upon cartilage maturation, together with a decreasing presence of thin fibrils (Eyre *et al.* 1992), and by the localization of type IX to regions of cartilage where thin fibrils predominate (Irwin *et al.* 1985, Müller-Glauser *et al.* 1986, Wotton *et al.* 1988, Hagg *et al.* 1998). Further support comes from an *in vitro* fibrillogenesis study, in which collagen types II, IX and XI were required in the ratio 8:1:1 to produce thin fibrils (Eikenberry *et al.* 1992). Type IX molecules are also present in a few of the thicker collagen fibrils in cartilage, however (Bruckner *et al.* 1988, Keene *et al.* 1995, Hagg *et al.* 1998).

Type IX molecules in cartilage have been shown to associate with each other both via sites within the long arm of the molecule (Wu *et al.* 1992) and via interaction of the NC4 domains of two separate molecules (Douglas *et al.* 1998). Together with a reported preferential location of type IX at intersections between two collagen fibrils (Müller-Glauser *et al.* 1986), this suggests a role in stabilization of the collagen fibril network. Another proposed function of type IX is modulation of the surface properties of collagen fibrils. The high positive charge of the protruding NC4 domain (Vasios *et al.* 1988, Muragaki *et al.* 1990a) may facilitate binding to adjacent proteoglycans or other ECM macromolecules, but this has not been demonstrated experimentally. Theoretically, these interaction-partners could act as a bridge providing a mechanism for stabilization of the fibril network (Smith & Brandt 1992). A similar bridging function has been suggested for the homologous collagen types XII and XIV (Nishiyama *et al.* 1994, Akutsu *et al.* 1999).

2.4.2 Experiments with transgenic mice

Transgenic animal models have been created to evaluate the functional significance of type IX collagen, *e.g.* a mouse line lacking the α1(IX) chain has been generated (Fässler *et al.* 1994). These mice were normal at birth, but subsequently developed a progressive degenerative joint disease resembling human osteoarthritis. Further analysis indicated an essential role for the α1(IX) chain in the assembly of type IX collagen, since its absence resulted in a functional knock-out of all three type IX collagen polypeptides (Hagg *et al.* 1997). Surprisingly, the cartilage collagen fibrils of the knock-out mice are structurally normal, contradicting the suggested role of type IX in the control of fibril diameter. Thus, type IX collagen appears not to be crucial for fibrillogenesis or skeletal development. It does appear, however, to be important for maintenance of the long-term structural integrity of the ECM. This conclusion is further validated by a study of a mouse line expressing α1(IX) chains with a large in-frame deletion of the complete NC3 domain and most of the COL2 and COL3 domains (Nakata *et al.* 1993). Both homozygous and heterozygous mice developed degenerative changes in their knee cartilage that showed progressive severity with age. In addition, the homozygotes developed mild chondrodysplasia with slight dwarfism and eye abnormalities. The collagen fibrils of the transgenic animals appeared to be abnormally thin, and the mice also developed spinal changes with age, including intervertebral disc degeneration and herniation (Nakata *et al.* 1993, Kimura *et al.* 1996). It was suggested that the age-related cartilage degeneration seen in mice that overexpressed the NC4 domain of type IX collagen was further proof of the role of this collagen type in the maintenance of cartilage integrity (Haimes *et al.* 1995).

The surface location of type IX suggests a role in determining the spacing of parallel fibrils and the prevention of lateral fusion. In support of this, the collagen fibrils produced by immortalized chondrocytes derived from the α1(IX) knock-out mice were of normal diameter but seemed to fuse laterally (Mallein-Gerin *et al.* 1995).

2.4.3 Involvement in human diseases

Mutations in cartilage collagen genes are known to result in mild to severe diseases of bone and cartilage, commonly referred to as osteochondrodysplasias (see Mundlos & Olsen 1997b). Experiments with transgenic mice suggest that alterations in the structure or amount of type IX collagen in humans may cause mild cartilage disorders such as osteoarthritis. This common disease is characterized by slowly progressing degeneration of the articular cartilage. In addition to numerous environmental risk factors, a genetic contribution to the pathogenesis of osteoarthritis has been demonstrated (Spector *et al.* 1996, Felson *et al.* 1998). While mutations in the COL2A1 gene are responsible for a subset of the disease cases (see Prockop 1998), other loci, such as the genes encoding type IX collagen, may be more commonly involved (see Mundlos & Olsen 1997b).

Multiple epiphyseal dysplasia (MED) is a clinically heterogeneous, autosomal dominant osteochondrodysplasia resulting in mildly short stature, epiphyseal

abnormalities and early onset osteoarthritis (see Spranger 1976, International Working Group on Constitutional Diseases of Bone 1998). Some forms of the disease are known to result from mutations in the COL9A2 gene (EDM2 locus) that lead to the skipping of exon 3 in mRNA splicing and to an in-frame deletion of 12 amino acid residues in the COL3 domain of the α2(IX) chain (Muragaki *et al.* 1996, van Mourik *et al.* 1998a, Holden *et al.* 1999). Recent identification of mutations in the COL9A3 gene (EDM3 locus) leading to a deletion of 12 amino acids at a homologous site in the α3(IX) chain verify the importance of type IX collagen for the pathogenesis of MED (Paassilta *et al.* 1999, Bonnemann *et al.* 2000, Lohiniva *et al.* 2000). The exact mechanism concerned in the molecular interactions leading to the disease is not understood. Interestingly, MED is also caused by mutations in the COMP gene (EDM1 locus), which affect either the areas encoding the Ca^{2+}-binding calmodulin repeats (Briggs *et al.* 1995, Ikegawa *et al.* 1998, Délot *et al.* 1999) or the part of the gene that encodes the C-terminal collagen-binding domain of COMP (Briggs *et al.* 1998). Mutations in the same regions also cause a similar but more severe disease called pseudoachondroplasia, PSACH (Briggs *et al.* 1995, 1998, Hecht *et al.* 1995). The mutations found in both PSACH and MED result in accumulation of COMP and type IX collagen in the rough ER of the chondrocytes (Maddox *et al.* 1997, Délot *et al.* 1998). This dilatation of the ER was not seen in MED patients with a mutation in the COL9A2 gene, however (van Mourik *et al.* 1998b). A recent report has demonstrated involvement of a fourth locus, the diastrophic dysplasia sulphate transporter gene (DTDST), in the pathogenesis of an autosomal recessive form of MED, further complicating the phenotypic and genotypic spectra of this disease (Superti-Furga *et al.* 1999).

Genetic factors are likely to be involved in the pathogenesis of human intervertebral disc disease (Battie *et al.* 1995). The changes in the intervertebral discs of α1(IX) transgenic mice (Nakata *et al.* 1993, Kimura *et al.* 1996), together with the known presence of type IX collagen in the intervertebral disc (Newall & Ayad 1995), implicate mutations in the genes encoding type IX collagen as possible contributors. Recently, a putative disease-causing variation in the COL9A2 gene resulting in substitution of tryptophan for arginine was identified (Annunen *et al.* 1999b). This change showed a clear association with intervertebral disc disease in four families, while being completely absent in the unaffected control population.

2.5 Production of proteins in heterologous expression systems

Production of human proteins in high amounts is easy to achieve in prokaryotes, but the recombinant protein produced is often incorrectly folded and forms insoluble aggregates. By contrast, eukaryotic expression systems in general offer access to correctly folded recombinant proteins. While mammalian cells are the choice when authentic post-translational modifications are required, demands for high amounts of the product are more readily met with other expression systems, *e.g.* with yeast cells or insect cells. (See Geisse 1996, Crossen & Gruenwald 1997, Hollenberg & Gellissen 1997.)

2.5.1 Insect cell expression system

Baculoviruses are double-stranded DNA viruses that specifically infect certain insect cells. After infection the viral DNA replicates in the nucleus of the host cell, creating viral particles encapsulated within occlusion bodies. These are an important part of the baculovirus life cycle *in vivo*, but are not needed under cell culture conditions. Thus, a strong promoter driving the expression of polyhedrin, the main protein of the occlusion bodies, can be used to produce high levels of heterologous proteins in insect cells. The recombinant baculoviruses are produced within the cells by co-transfection of viral DNA, most commonly the DNA of *Autographa californica* nuclear polyhedrosis virus, with a transfer vector containing the DNA fragment to be expressed. Homologous recombination integrates the foreign DNA into the viral genome, after which the recombinant viruses are selected, amplified and used to infect large amounts of insect cells to produce the desired heterologous protein. (See Kidd & Emery 1993, Crossen & Gruenwald 1997, Merrington *et al.* 1997.)

The advantages of the baculovirus expression vector system are a high level of recombinant protein production, correct subcellular localization or secretion of the product, the synthesis of a correctly folded, biologically active product and ease of scaling up due to the adaptability of the cells to suspension culture conditions. Insect cells also properly carry out most of the post-translational modifications occurring in mammalian cells. (See Kidd & Emery 1993, Geisse *et al.* 1996, Crossen & Gruenwald 1997.)

2.5.2 Production of collagens as recombinant proteins

Recombinant human type II collagen was first produced successfully in stably transfected mammalian cells (Ala-Kokko *et al.* 1991, Fertala *et al.* 1994). This system was subsequently used to demonstrate local variations in the thermal stability of the collagen triple helix (Arnold *et al.* 1998). Other full-length collagens present in cartilage, namely types V and X, have also been produced in mammalian cells (Fichard *et al.* 1997, Frischholz *et al.* 1998). When studying type XVI collagen the system was used to indicate a homotrimeric chain composition (Tillet *et al.* 1995).

The first studies on the expression of a full-length collagen by baculovirus infection in insect cells indicated that the presence of exogenous prolyl 4-hydroxylase was required to achieve normal thermal stability, and that this also increased the level of recombinant collagen production (Tomita *et al.* 1995, Lamberg *et al.* 1996). In subsequent experiments to produce human type I and II collagens a double promoter virus for the α and β subunits of the human enzyme was successfully employed (Myllyharju *et al.* 1997, Nokelainen *et al.* 1998). Despite the presence of a signal peptide, none of the stable trimeric collagens was efficiently secreted into the culture medium. Also, the recombinant type II and III collagens produced in insect cells showed a slight underhydroxylation and underglycosylation of lysine residues (Lamberg *et al.* 1996, Nokelainen *et al.* 1998). Using the recombinant type II collagen, lysine hydroxylation and hydroxylysine

glycosylation were shown to play a role in fibrillogenesis (Notbohm *et al.* 1999). Of the non-fibrillar collagens, types XIII and XV have been produced in insect cells (Hägg *et al.* 1997, Snellman *et al.* 2000).

Recombinant collagens have also been produced in a few other expression systems. A methylotrophic yeast, *Pichia pastoris,* has been used to produce type III collagen (Vuorela *et al.* 1997). In these cells, and perhaps in eukaryotes in general, the expression of collagen appears to stabilize the prolyl 4-hydroxylase by increasing the half-life of the tetramer (Vuorela *et al.* 1997, 1999). In addition to the full-length collagens, short fragments of numerous collagens have been produced as recombinant proteins in various expression systems, including production of the NC4 domain of chicken type IX collagen in *Escherichia coli* (Douglas *et al.* 1998).

3 Outlines of the present research

Determination of the molecular mechanism underlying a hereditary disease has proved to be a valuable tool for defining a functional role of a protein. Analysis of mRNA is often a straightforward method for identifying a causative genetic variation. This is not the method of choice in the case of cartilage disorders, however, since samples of the patient's chondrocytes are not usually available. The use of mRNA isolated from chondrocytes or other cell types may also entail certain other difficulties. In contrast, the use of genomic DNA is reliable and relatively easy. A necessary requirement is that the genomic organization of the gene under investigation should be known.

When this work began, the genomic structures of the type IX collagen genes were not known and a genetic defect had only been identified in the α2(IX) chain in one phenotype, MED. To facilitate more efficient mutation screening, we decided to characterize the human type IX collagen genes. This would also allow us to compare their genomic organizations and to identify structural or functional similarities and differences.

It was recognized that a possibility for producing human type IX collagen as a recombinant protein would have numerous scientific applications. First, it would allow us to produce large quantities of protein for studying molecular interactions. Secondly, structurally or functionally important regions of the molecule could be identified by producing altered protein, and thirdly, the effects of possible disease-causing genetic variations in type IX collagen could be investigated at the protein level.

The following aims were therefore set out:

1. to characterize the genomic organizations of the human COL9A1, COL9A2 and COL9A3 genes,
2. to develop a rapid screening method for analyzing sequence variations in the COL9A3 gene, and
3. to produce human type IX collagen as a recombinant protein in an insect cell expression system.

4 Materials and methods

4.1 Isolation of genomic clones for type IX collagen (I,III)

To isolate genomic clones for the human COL9A1, COL9A2 and COL9A3 genes, genomic P1 and PAC libraries (Genome systems, Inc., St. Louis, USA) were screened by PCR. Primers for the COL9A1 gene were designed based on the cDNA and genomic sequences for the human α1(IX) chain (Muragaki *et al.* 1990a). The screening identified two positive P1 clones, P1-A (clone address DMPC-HFF#1-1378-C8) and P1-B (DMPC-HFF#1-837-D10). Primers for the COL9A2 gene were designed based on the human α2(IX) cDNA sequences and the genomic structure of the mouse Col9a2 gene (Perälä *et al.* 1993, 1994). A human PAC clone, PAC-1 (clone address PAC-39-22G), containing the COL9A2 gene, was isolated. Genomic clones for the COL9A3 gene were initially obtained by screening of an EMBL3 PS6/T7 genomic phage library with two ^{32}P-radiolabelled probes. Probe p1061 was prepared by RT-PCR as described earlier (Brewton *et al.* 1995) and probe p342 was generated by PCR with a primer pair designed on the basis of the 3' region of the α3(IX) cDNA (Brewton *et al.* 1995). Two phage clones, gRB2B1 and gRB5B1, were isolated and shown not to contain the entire COL9A3 coding region. Additional primer pairs were designed based on the human α3(IX) sequences (Brewton *et al.* 1995) and the mouse Col9a2 gene structure (Perälä *et al.* 1994). These were used for screening the human P1 library by PCR, which yielded three positive P1 clones, P1-C93A, P1-C93B and P1-C93C (addresses DMPC-HFF#1-270-C3, DMPC-HFF#1-753-B10 and DMPC-HFF#1-1082-B5, respectively).

4.2 Characterization of type IX collagen genes (I,III)

All the P1 clones were transferred from the *Escherichia coli* NS3529 strain to the NS3516 strain, which was known to give a higher yield in DNA isolation, via transduction as suggested by the company (Genome Systems, Inc). The P1 and PAC DNA

was isolated by the method of Pierce and Sternberg (1992). Nucleotide sequencing of the isolated DNA was carried out by cycle sequencing (dsDNA Cycle Sequencing System, Life Technologies, Inc.; Cycle Sequencing Kit, Amersham Pharmacia Biotech; Dye Terminator Cycle Sequencing Ready Reaction Kit, PerkinElmer) using primers designed on the basis of published human cDNA sequences (Muragaki *et al.* 1990a, Perälä *et al.* 1993, Brewton *et al.* 1995). Additional primers were designed on the basis of the newly obtained sequences. The sizes of several introns of the COL9A1 gene were estimated by PCR (Expand Long Template PCR System, Roche Molecular Biochemicals), as was the size of intron 26 of the COL9A3 gene. Some intronic sequences within all three genes were determined by sequencing of the PCR products (T7 Sequencing Kit, Amersham Pharmacia Biotech) after purification (QIAEX II Gel Extraction Kit, Qiagen) and cloning into the pUC18 vector (Sure Clone Ligation Kit, Amersham Pharmacia Biotech). The sequencing reactions were analyzed on 6% denaturing polyacrylamide gels. The resulting sequences were stored and analyzed using the Wisconsin Package Version 10.0-UNIX (Genetics Computer Group).

The transcription start sites of the COL9A1 and COL9A2 genes were determined (5'/3'-RACE Kit, Roche Molecular Biochemicals). The 3' end of the human α1(IX) mRNA was analyzed as described previously (Vuoristo *et al.* 1995), using human foetal cartilage as a template. The intragenic tetranucleotide repeat within intron 12 of the COL9A1 gene was characterized by analysis of ^{32}P-radiolabelled PCR products on a 6% denaturing polyacrylamide gel.

Genomic DNA was isolated from mouse liver and used to amplify the alternative promoter region of the mouse Col9a1 gene by PCR with primers designed from sequences corresponding to exons 6 and 7 of the gene (Muragaki *et al.* 1990b). The PCR product was cloned into a plasmid vector and sequenced as above.

4.3 Characterization of genomic variations in the COL9A3 gene (III)

A family with MED was referred to us by Dr. R. Knowlton of Jefferson Medical College, Philadelphia, U.S.A., and a healthy individual with a neutral deletion in the COL9A3 gene by Dr. M. Warman of Case Western Reserve University, Cleveland, U.S.A. Genomic DNA was isolated (Ausübel *et al.* 1989) from venous blood samples from these subjects and control individuals of Finnish origin. These DNA samples were used as templates in PCR amplification of each exon of the COL9A3 gene. Primer pairs for PCR were designed from the intronic sequences for amplification of the exons with at least 70 bp of the flanking sequences. The PCR consisted of 30 cycles of 94°C for 45 s, 60 to 64°C for 45 s and 72°C for 60s followed by a final extension at 72°C for 10 min. The formation of heteroduplexes was enhanced by incubation at 95°C for 5 min followed by 30 min at 68°C. The concentrations of the PCR products were estimated on 1.5% agarose gels.

Heteroduplex analysis of the PCR products was performed by conformation-sensitive gel electrophoresis, CSGE (Ganguly *et al.* 1993, Körkkö *et al.* 1998). The gel, of thickness 1 mm, consisted of 15% formamide (Gibco), 10% acrylamide, 10% ethylene glycol (Sigma), 1,4-bis-acryloylpiperazine (Fluka) in a 1:99 ratio to the acrylamide, 0.1%

ammonium persulphate and 0.07% TEMED in 0.5xTTE buffer (44.4 mM Tris – 14.25 mM Taurine – 0.1 mM EDTA). The gel was pre-electrophoresed for 15 min before loading, and the PCR products were mixed with a loading dye (10x stock solution of 30% glycerol, 0.25% xylene cyanole FF and 0.25% bromphenol blue). The electrophoresis was performed on a standard sequencing apparatus at 45 W for 5 h using 0.5xTTE as the running buffer. After the electrophoresis the samples were visualized by ethidium bromide staining. PCR products that contained heteroduplexes in CSGE were analyzed by direct sequencing (T7 Sequenase PCR Product Sequencing Kit, USB) or by sequencing after cloning into a plasmid vector (see section 4.2).

4.4 Preparation of baculoviruses expressing recombinant human type IX collagen α chains (II)

Total RNA was extracted from human foetal cartilage and used to generate cDNAs with an oligo(dT) primer and Moloney murine leukaemia virus reverse transcriptase (Life Technologies, Inc.). Aliquots of the cDNAs were used for a single step amplification of the α1(IX), α2(IX) and α3(IX) cDNAs by PCR (Expand Long Template PCR System, Roche Molecular Biochemicals) with primers designed for the 5' and 3' ends of each cDNA based on the published sequences (Muragaki *et al.* 1990a, Rokos *et al.* 1994, Perälä *et al.* 1993, Brewton *et al.* 1995) and the sequence of the COL9A2 gene (I). The forward and reverse primers for the amplification of the α1(IX) and α2(IX) cDNAs contained engineered *NotI* and *EagI* cleavage sites, respectively. Both primers used for amplifying the α3(IX) cDNA contained engineered *XbaI* cleavage sites. The PCR products were digested with the appropriate enzymes and ligated into the pVL1392 vector. The constructs were completely sequenced (Sequenase Reagent Kit, Amersham Pharmacia Biotech) using cDNA-specific primers. The three recombinant constructs were co-transfected into *Spodoptera frugiperda* (Sf9; Invitrogen) insect cells with a modified *Autographa californica* nuclear polyhedrosis virus DNA (BaculoGold Transfection Kit, Pharmingen), and the resultant viral pools were collected, amplified and plaque-purified (Crossen & Gruenwald 1997).

4.5 Cell culture and optimization of recombinant protein production (II)

Sf9 or *Trichoplusia ni* (High Five; Invitrogen) insect cells were cultured in monolayers in TNM-FH medium (Sigma) supplemented with 10% foetal bovine serum (Bioclear) at 27°C. The recombinant baculoviruses for the type IX collagen α chains were used to infect cells seeded at a density of 5-6 x 10^6 cells/ml together with a double promoter virus 4PHαβ coding for the α and β subunits of human prolyl 4-hydroxylase (Nokelainen *et al.* 1998). Various combinations of the multiplicities of infection for each virus were tested to

define the conditions yielding the highest level of recombinant protein production. High Five cells were used in the subsequent infections, and the multiplicities of infection were 2:2:2:4 for the α1(IX), α2(IX), α3(IX) and 4PHαβ viruses, respectively. High Five cells were also cultured in suspension in Sf-900 II SFM medium (Life Technologies, Inc.). The initial cell density in suspension at the beginning of an infection was set at 1 x 10^6 cells/ml. L-ascorbic acid phosphate (Wako) was added to the culture medium daily at a final concentration of 80 µg/ml during all the infections. Prolyl 4-hydroxylase activity was determined after the infections by analyzing the cell homogenates as described previously (Kivirikko & Myllylä 1982).

4.6 Isolation of intracellular recombinant human type IX collagen (II)

After 72 h of infection, the High Five cells were detached from the culture plates by pipetting and harvested by centrifugation at 1000 x g for 5 min. The cells cultured in suspension were also harvested by centrifugation. To extract the intracellular proteins, the cells were homogenized in 0.27 M NaCl, 0.2% Triton X-100 and 0.07 M Tris-HCl buffer, pH 7.4, as described earlier (Lamberg *et al.* 1996). The Triton-insoluble pellet was suspended in 0.1% SDS and incubated at room temperature for 2 h, after which the insoluble remains were discarded. As an alternative, the cells were suspended in 0.75 M NaCl and 0.5 M acetic acid, pH 2 (7.5 x 10^6 cells/ml), and homogenized on ice for 30 s using a glass-Teflon homogenizer. The insoluble material was pelleted by centrifugation at 12 000 x g for 20 min at 4°C, and proteins were precipitated from the supernatant by increasing the NaCl concentration to 3 M and mixing at 4°C for 12–16 h (Duance *et al.* 1984). The precipitate was collected by centrifugation and dissolved in 50 mM acetic acid. The homogenization was performed either without protease inhibitors or in the presence of various mixtures of 10 mM EDTA, 1 mM PMSF, 10 µM E-64, 1 µM leupeptin, 1 µM pepstatin and 75 nM aprotinin.

4.7 Purification of secreted recombinant human type IX collagen (II)

The culture medium from High Five cells infected with the four recombinant viruses was analyzed for the presence of type IX collagen by dialysis into 50 mM acetic acid followed by SDS-PAGE and Western blotting with the monoclonal antibody 95D1A. This antibody was generated against recombinant type XIII collagen, but was shown to recognize various denatured collagens (types I, II, III and XIII) via a collagenous epitope containing a Lys-Gly-Glu sequence (Snellman *et al.* 2000). To purify type IX collagen, the proteins were precipitated from the culture medium by adding solid ammonium sulphate to a 25% saturation and placing the mixture on ice for 1 h (Gibson *et al.* 1983). The precipitated proteins were collected by centrifugation at 12 000 x g for 20 min at 4°C and dissolved in 0.5 M urea, 0.2 M NaCl, 0.05 M Tris-HCl buffer, pH 7.4, at 4°C overnight to a concentration of about 1mg/ml. The dissolved recombinant protein was purified by gel

filtration through a Sephacryl S-300HR column (Amersham Pharmacia Biotech) in the same buffer. Further purification was achieved by cation exchange with CM-Sepharose fast flow matrix (Amersham Pharmacia Biotech) in a buffer of 2 M urea, 50 mM PIPES, and 20 mM NaCl, pH 6.5, eluting with an increasing NaCl concentration gradient (0.02–1 M NaCl).

To study type IX collagen chain association, adherent High Five cells were infected with the viruses encoding the α chains individually and in all possible combinations together with the 4PHαβ virus. The resulting recombinant molecules were characterized by SDS-PAGE under non-reducing conditions and by Western blotting with the 95D1A antibody.

4.8 Characterization of recombinant type IX collagen (II)

The purified recombinant protein was characterized by SDS-PAGE followed by Coomassie Brilliant Blue staining or Western blotting with the monoclonal antibody 95D1A. For amino acid analysis, the material was dialyzed against 50 mM acetic acid, hydrolyzed in 6 M HCl at 110°C for 16 h and analyzed in an Applied Biosystems 421 apparatus. The thermal stability of the material was determined by circular dichroism (CD) analysis at a fixed wavelength (221 nm), raising the temperature linearly at a rate of 60°C/h (Gaill *et al.* 1991). For N-terminal sequencing, purified recombinant type IX collagen was electrophoresed under reducing conditions and transferred to ProBlott polyvinylidene difluoride-type membrane, after which the excised bands were subjected to Edman degradation with a 477/120A liquid-phase protein/peptide sequencer (Applied Biosystems).

4.9 Production and analysis of recombinant human type IX collagen containing an internal Gly-Pro-Pro deletion in the α3(IX) chain (III)

Based on the published cDNA sequence for the α3(IX) chain (Brewton *et al.* 1995), two primers containing *CspI* cleavage sites were designed for the region encoded by exon 30, which contained the naturally occurring 9-bp deletion. These primers, together with primers from the 5' and 3'-non-coding regions, were used to amplify the α3(IX) coding region in two fragments by PCR. The template in the PCR was cDNA that had been transcribed from total RNA extracted from human foetal cartilage. The 1729 bp and 700 bp PCR products were digested with *CspI*, purified after agarose gel electrophoresis (QIAEX II Gel Extraction Kit, Qiagen) and ligated to the pVL1392 vector. The nucleotide sequence of the construct was checked by automated sequencing using cDNA-specific primers (ABI PRISM model 377 sequencer, PerkinElmer; Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS, PerkinElmer).

The construct carrying the 9-bp deletion was used to generate a recombinant baculovirus by co-transfection into Sf9 cells as above (section 4.4). The viral pool was collected, amplified and plaque-purified (Crossen & Gruenwald 1997). High Five cells were infected with the recombinant virus for the α3(IX) chain containing the Gly-Pro-Pro deletion and viruses for the α1(IX) and α2(IX) chains (II; section 4.4) together with the 4PHαβ virus (Nokelainen *et al.* 1998). Expressions were performed in suspensions as described above (II; section 4.5). For production of the wild-type collagen, the mutant α3(IX) virus was replaced by the wild-type α3(IX) virus (II; section 4.4). Mutant and wild-type recombinant collagens were precipitated from the culture medium as indicated above (II; section 4.7) and dissolved overnight at 4°C in 0.5 M urea, 0.2 M NaCl and 0.05 M Tris-HCl buffer, pH 7.4. For pepsin digestion, the samples were adjusted to pH 2.0 and the digestion was performed at room temperature for 4 hours, after which the pH was adjusted to 7.5. Negative controls were obtained by incubation at room temperature without pepsin. A set of samples were denatured by incubation at 60°C for 5 min prior to pepsin digestion. The samples were analyzed by 8% SDS-PAGE under non-reducing conditions followed by Coomassie staining. The recombinant collagen with the Gly-Pro-Pro deletion was further purified by cation exchange, as described above (II, section 4.7).

5 Results

5.1 Characterization of the human genes encoding type IX collagen (I,III)

Nucleotide sequencing of two genomic P1 clones, P1-A and P1-B, showed that they overlapped and contained the entire COL9A1 gene, while a PAC clone, PAC-1, was shown to contain the entire coding sequence for the COL9A2 gene (Fig. 1 in I). All three P1 clones obtained for the COL9A3 gene were shown by PCR and sequencing to contain the entire coding region, and clone P1-C93A was selected for detailed characterization of the gene. The genomic phage clones obtained here (gRB2B1 and gRB5B1) and the clone gRB35 isolated previously (Brewton *et al.* 1995) failed to cover the entire COL9A3 coding region, and were not characterized further.

Nucleotide sequencing defined over 25 kb of genomic sequences for the COL9A1 gene, including all 38 exons and a minimum of about 160 bp of flanking intronic sequences (Table I in I). Due to large size or difficulties caused by certain segments, *e.g.* Alu repeats, some introns were not sequenced completely, and instead their sizes were estimated by PCR amplifications with two sets of primers per intron. Thus, the overall size of the COL9A1 gene was estimated to be about 90 kb. For the COL9A2 gene, all the exon and intron sequences were determined by nucleotide sequencing. The gene was shown to span only 15 kb of genomic DNA, divided into 32 exons (Table I in I). More than 26 kb of nucleotide sequences were determined for the COL9A3 gene, which is about 23 kb in size and consists of 32 exons (Fig. 2 in III). The size of intron 26 was estimated by PCR. The nucleotide sequences obtained here were submitted to GenBank (accession numbers AF036110-AF036130 for the COL9A1 gene, AF019406 for the COL9A2 gene, and AF026801 and AF026802 for the COL9A3 gene).

Comparison of the exon sizes in the three genes (Table 2) shows almost perfect conservation, excluding the exons of the COL9A1 gene encoding the NC4 domain, which are absent from the COL9A2 and COL9A3 genes. An expected size difference is seen in the exons coding for the NC3 domain, in that exon 16 in COL9A1 is 33 bp in length while the corresponding exon 10 in the COL9A2 and COL9A3 genes spans 48 and 42 bp,

respectively. This size difference explains the flexible kink seen in type IX collagen. Exon 2 of the COL9A3 gene, the junction exon encoding the NC4 and COL3 domains, codes for one collagen triplet less than exon 2 in COL9A2 or the corresponding exon 8 of the COL9A1 gene. The length of the COL3 domain is similar in all three polypeptides, however, because a complementing insertion is present in exon 4 of the COL9A3 gene.

*Table 2. Comparison of exon and intron sizes in the homologous regions of the human type IX collagen genes**

Exon size (bp)			Intron size (bp)		
COL9A1	COL9A2	COL9A3	COL9A1	COL9A2	COL9A3
75 (8)	75 (2)	69 (2)	349	1195	867
36	36	36	129	84	661
63	63	72	~5800[a]	1419	631
54	54	54	~600	121	1194
36	36	36	~2000	93	288
24	24	24	356	343	226
54	54	54	~1300	341	297
54	54	54	648	114	425
33	48	42	~700	213	1817
57	57	57	~2000	83	467
54	54	54	~3300	325	697
54	54	54	~2500	396	333
54	54	54	~3700	99	509
54	54	54	~1350	119	420
54	54	54	133	1690	626
54	54	54	133	435	787
54	54	54	433	203	105
54	54	54	~950	1260	477
45	45	45	~1000	373	127
54	54	54	88	863	84
54	54	54	~9000	264	539
54	54	54	626	84	102
72	72	72	511	220	1421
36	36	36	684	91	880
45	45	45	~1350	84	~2200
33	33	33	~4200	130	229
147	147	147	200	343	144
55	55	55	1770	319	547
189	189	183	~10000	731	1419
78 (37)	78 (31)	78 (31)	~10000	430	1773

* the numbers of the first and last exons listed, as counted from the 5' end, are indicated in parentheses

[a] intron sizes marked with (~) were estimated by PCR

In contrast to the exons, the introns differ markedly in size between the three genes (Table 2). The average intron size of the COL9A1 gene is several times higher than that of the other two genes. Also, the largest introns of the COL9A1 and COL9A3 genes are mostly located in the 3' region, while the 5' region in the COL9A2 gene contains several large introns. Intron 10 of the COL9A3 gene was found to contain 12 tandem repeats of a 31-bp sequence showing no homology to any known sequences.

A tetranucleotide repeat was identified in intron 12 during the nucleotide sequencing of the COL9A1 gene. Five alleles of six to ten repeats were found in 55 Caucasians. The allele frequencies were 0.01, 0.01, 0.05, 0.51 and 0.42.

5.1.1 Identification of putative transcription control elements within the human type IX collagen genes (I,III)

The transcription start sites of the COL9A1 and COL9A2 genes were located by 5' RACE analysis to distances of 131 and 88 bp from the translation start site, respectively (Table I in I). The transcription start site was not determined for COL9A3.

A computer search for the presence of consensus recognition sequences of various transcription factors was performed for the promoter region and first intron of each gene. A CCAAT or TATA box was not found in any of the type IX collagen promoters used with cartilaginous tissues. Analysis of sequences up to 3 kb upstream from the translation start site of the COL9A3 gene identified 30 consensus recognition sites for Sp1, of which 11 were concentrated between the positions –361 and –35. Several Sp1 sites are also found within the promoters of the human COL9A2 and mouse Col9a2 genes (Perälä *et al.* 1994). Interestingly, a perfectly conserved sequence of 14 bp (CTCTCAGGTGACAG) was found located between two Sp1 sites in these two genes (Fig. 2). A similar sequence was also identified in the promoter of the human COL9A1 gene, but no Sp1 sites were present. In addition, the promoter of the human COL2A1 gene was found to contain a comparable conserved sequence that is also located between Sp1 sites. In the COL9A2, Col9a2 (Perälä *et al.* 1994) and COL2A1 (Vikkula *et al.* 1992) genes this sequence is located closely upstream of the major transcription start site. Recent reports implicate SOX9 as a transcription factor involved in chondrogenesis (see Lefebvre & de Crombrugghe 1998). Two consensus binding sites for the SRY/SOX protein motif were identified in reverse orientation within the COL9A3 promoter, one located far upstream and the other within intron 1. Sequences matching this consensus are also present in the promoter region of COL9A1, and in the first introns of the COL9A1 and COL9A2 genes (T.Pihlajamaa, unpublished). No other putative cartilage-specific motifs were identified in any of the upstream regions analyzed.

The alternative transcript of the COL9A1 gene is initiated in intron 6 (Table I in I). This intron spans 663 and 820 bp in the chicken (Nishimura *et al.* 1989) and human genes, respectively. To extend the comparison, we cloned and sequenced intron 6 of the mouse gene (GenBank accession number AF020297), determining a size of 799 bp for it. In addition to the sizes, the sequences from the 5' end of intron 6 to the start of

Figure 2. The 5' regions of the human COL9A2 and mouse Col9a2 (Perälä *et al.* 1994) genes. The transcription start site of the human gene is indicated by (▼). The major (♦) and minor (◇) transcription start sites are shown for the mouse gene (Perälä *et al.* 1994). The conserved 14-bp sequence is indicated by an open box and the Sp1 recognition sites are marked with horizontal lines. The translation start sites are indicated (filled box).

translation of the alternative exon 1* were also well conserved in these genes (Fig. 3). The transcription start site of the cornea-specific transcript in the chicken is located 13 nucleotides downstream of a TATA box and 41 nucleotides downstream of a CCAAT box (Nishimura *et al.* 1989), while the human and mouse genes were found here to have a TATA and a CCAAT box at corresponding locations (Fig. 3). This promoter in the human gene also contains two SRY/SOX consensus sites seen in the type IX collagen promoters that are active in cartilage, but lacks the 14-bp conserved sequence described above (T.Pihlajamaa, unpublished). The distances between the alternative exon 1* and exon 7 were found to be short: 62 bp in the human gene, 58 in the mouse gene, 57 in the rat gene (Ting *et al.* 1993) and 48 in the chick gene (Nishimura *et al.* 1989).

5.2 Characterization of recombinant human type IX collagen (II)

To produce recombinant human type IX collagen, three recombinant viruses were generated, each coding for one of the three α chains of human type IX collagen. These were used to infect insect cells together with a double promoter virus, 4PHαβ (Nokelainen *et al.* 1998), coding for the α and β subunits of human prolyl 4-hydroxylase. High Five cells were used as host cells, since they yielded higher expression levels than Sf9 cells. After a 72 h infection, the cells were harvested and homogenized in a buffer containing Triton X-100. Although the same buffer has been succesfully used to isolate other collagens from insect cells (Lamberg *et al.* 1996, Myllyharju *et al.* 1997), no soluble type IX could be obtained by homogenizing the cells in Triton X-100 buffer. Instead, the individual α chains were detected in the insoluble fraction. Selective salt

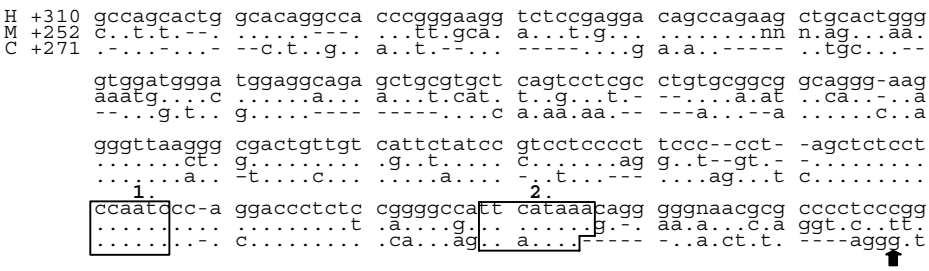


Figure 3. The regions of highest homology within the sixth intron of the human (H), mouse (M) and chicken (C) (Nishimura *et al.* 1989) α1(IX) genes. Box 1 indicates the location of the CCAAT box and box 2 denotes the apparent TATA box. The sequences are numbered from the 5' end of the intron. The transcription start site of the alternative exon 1* in the chicken gene is indicated by an arrow. Gaps created for alignment are marked with dashes and sequence identity with the human sequence is indicated by dots.

precipitation, previously used to extract type IX collagen from tissues (Duance *et al.* 1984), was therefore used here to isolate the intracellular collagen. This method yielded intracellular type IX at a level of 4-8 mg/litre of culture, as determined by comparison with known amounts of Coomassie-stained recombinant human type II collagen after SDS-PAGE. Analysis by Western blotting with the monoclonal antibody 95D1A nevertheless revealed that the isolated material was partially degraded, and its quality could not be markedly improved despite the use of various protease inhibitors.

SDS-PAGE and Western blotting with the 95D1A antibody were used to demonstrate the presence of type IX collagen in dialyzed culture medium (Fig. 1 in II). By comparison with the intracellular material isolated by acetic acid/NaCl extraction and precipitation with 3 M NaCl, the recombinant type IX collagen present in the medium appeared to be significantly less degraded (Fig. 1 in II). Also, the intracellular material appeared to migrate slightly faster in SDS-PAGE under non-reducing conditions (not shown). Therefore, a protocol was developed for isolating and purifying type IX collagen from the culture medium. The addition of solid ammonium sulphate to 25% saturation was found to result in almost quantitative precipitation of type IX collagen with minimal co-precipitation of other proteins (Fig. 2 in II). The precipitating type IX collagen was in the form of disulphide-bonded trimers, which appeared upon visual examination after SDS-PAGE analysis to consist of three different α chains in a 1:1:1 ratio (Fig. 2 in II). Up to 10 mg of type IX collagen could be obtained from 1 litre of culture medium. The precipitated protein was dissolved in a buffer containing 0.5 M urea to enhance its solubilization, and purified by gel filtration on a Sephacryl S-300HR column. This purification step was efficient in removing low molecular weight contaminants, *e.g.* the remaining bovine serum albumin (Fig. 3 in II). As a final purification step, the recombinant protein was subjected to cation exchange chromatography on a CM-Sepharose fast flow column.

Amino acid analysis of the purified material yielded results that corresponded well with the calculated values for human type IX collagen (Table I in II). Based on these

results and on the appearance of the material on Coomassie-stained SDS-PAGE gels, a purity of over 90% was estimated for the recombinant type IX collagen. The triple helicity of the recombinant protein was verified by analysis of its thermal denaturation properties by CD. The melting profile was biphasic, with about 2/3 of the transition centring at T_m = 37.5°C and about 1/3 of the transition centring at T_m = 46.0°C (Fig. 4 in II).

In an attempt to identify the signal peptide cleavage sites of each recombinant α chain, amino acid sequencing of the N-terminal amino acids was performed as described in section 4.8. The sequence present at the N-terminus of the α1(IX) chain was AVKRRPR, corresponding to the predicted signal peptide cleavage site (Muragaki *et al.* 1990a). The sequences for the α2(IX) and α3(IX) chains could not be determined, apparently because of N-terminal blocking possibly on account of the presence of glutamine (I, Brewton *et al.* 1995) at the cleavage site (Blombäck 1967).

5.2.1 Association of recombinant type IX collagen α chains into disulphide-bonded molecules (II)

The formation of disulphide-bonded molecules by the recombinant type IX collagen α chains was studied by expressing each chain individually and in all possible combinations in High Five cells in the presence of exogenous prolyl 4-hydroxylase. Samples of the culture media were analyzed by SDS-PAGE followed by Western blotting with the 95D1A antibody. The results show that all three α chains of human type IX collagen appear to be capable of forming disulphide-bonded homodimers (Fig. 5, lanes 2-4 in II). The α1(IX) chains also readily form homotrimers (Fig. 5, lane 2 in II), whereas homotrimers of the α2(IX) and α3(IX) chains were only just detectable (Fig. 5, lanes 3 & 4 in II). When all three chains are expressed simultaneously, the heterotrimer α1(IX)α2(IX)α3(IX) is the predominant disulphide-bonded molecule formed (Fig. 5, lane 8 in II). Other disulphide-bonded trimers may be present, but in quantities that are below the detection limit. Co-expression of any two α chains resulted in the formation of detectable amounts of disulphide-bonded heterodimeric molecules only when the chains concerned were α1(IX) and α3(IX) (Fig. 5, lane 6 in II). Any other combination of two α chains appeared to result only in the formation of disulphide-bonded homomeric molecules (Fig. 5, lanes 5 & 7 in II). Monomers were also abundant in all the samples (Fig. 5, lanes 2-8 in II).

5.3 Identification and characterization of Gly-Pro-Pro triplet deletions in the α3(IX) chain (III)

A procedure was developed for the analysis of genetic variations in the COL9A3 gene, consisting of PCR amplification of the exons and their flanking regions followed by

heteroduplex analysis by CSGE. Analysis of a proband diagnosed with MED led to the identification of several neutral polymorphisms and one potential disease-causing mutation, a 9-bp deletion in exon 30, resulting in removal of a Gly-Pro-Pro triplet at the 5' end of the COL1 domain (Fig. 5 in III). Surprisingly, analyses of other family members showed the presence of this deletion in both affected and unaffected family members, while two affected individuals did not have it (Fig. 4 in III). Hence, the deletion did not co-segregate with the disease phenotype. A similar 9-bp deletion was also found in a healthy member of a second family during the analysis of the COL9A3 gene in the control population. This deletion removed a Gly-Pro-Pro triplet located adjacent to that removed in the first family (Fig. 5 in III). Seven individuals within the second family had the deletion, but there was no associated phenotype. Reverse transcriptase-PCR amplification with RNA extracted from cultured skin fibroblasts of one family member was used to demonstrate that the normal allele and that containing the deletion were expressed equally. The two deletions were not found in 350 additional chromosomes, indicating that they are not common polymorphisms.

The insect cell expression method for the production of recombinant type IX collagen (II) was used to analyze the ability of the α3(IX) chain carrying the deletion to participate in formation of a collagen triple helix. A baculovirus expressing an α3(IX) chain with the Gly-Pro-Pro deletion was used to infect insect cells together with the wild-type α1(IX) and α2(IX) viruses described elsewhere (II). The recombinant protein secreted into the culture medium was isolated, purified and analyzed by SDS-PAGE (Fig. 6 in III). A band similar in size to that of a wild-type type IX collagen heterotrimer was detected. Upon reduction the mutant recombinant protein showed the presence of three different α chains in a 1:1:1 ratio, indicating that the α3(IX) chain carrying the deletion was able to participate in the formation of secretable, heterotrimeric type IX collagen (Fig. 6 in III). Pepsin digestion with or without a preceding denaturation step was used to analyze the triple-helical conformation of partially purified, mutant recombinant protein. When the material was not denatured prior to pepsin digestion, pepsin-resistant fragments characteristic of triple-helical type IX collagen were detectable by SDS-PAGE analysis (Fig. 7, lanes 3 & 7 in III). A denaturation step before the digestion resulted in disappearance of the pepsin resistance (Fig. 7, lanes 5 & 9 in III). The results indicate that despite the absence of one Gly-X-Y unit, the α3(IX) chain was assembled into a type IX collagen heterotrimer that was secretable and triple-helical.

6 Discussion

When the present work was begun, the primary structures of the human type IX collagen α chains had been characterized except for the extreme 5' region of the α2(IX) chain. In contrast, the genomic organizations of the corresponding human genes were practically undetermined, with only the structures of a few exons of the COL9A1 gene having been reported. The results obtained here fill in both the primary and the genomic structures of human type IX collagen.

The genomic organizations of the human genes encoding type IX collagen are remarkably similar. The exon sizes of the COL9A2 and COL9A3 genes are almost identical and both consist of 32 exons. The corresponding region of the COL9A1 gene also shows practically identical exon sizes, although this region of the COL9A1 gene spans nearly 70 kb of genomic DNA, whereas those in COL9A2 and COL9A3 are of size 15 and 23 kb, respectively. The intron sizes of the three genes are thus notably different. The locations of the longest introns also appear to differ among the type IX collagen genes. The evolutionary significance of these findings is at present unknown.

An interesting deviation regarding the conservation of exon sizes is seen in the area encoding the COL3 domain, for where exon 2 in the COL9A3 gene codes for one Gly-X-Y triplet less than the exon 2 of the COL9A1 and COL9A2 genes, a compensating insertion of 9 bp in exon 4 of the COL9A3 gene preserves a domain size identical with the other two α chains (Fig. 3 in III). This implies a strong evolutional pressure towards maintaining the correct structure of the COL3 domain. Interestingly, the only disease-causing mutations identified in the type IX collagen genes so far lead to an in-frame deletion of 12 amino acids from the COL3 domain (Muragaki *et al.* 1996, Holden *et al.* 1999, Paassilta *et al.* 1999, Bonnemann *et al.* 2000, Lohiniva *et al.* 2000). It is not understood why these deletions result in development of the MED phenotype. Perhaps a uniform length of the COL3 region in each type IX collagen α chain is required to facilitate proper folding of the positively charged NC4 domain, which is suspected to interact with other components of the cartilage ECM. In support of this, patients with EDM2 or EDM3 have not been reported to have ocular symptoms, which may simply result from the demonstrated absence of the NC4 domain in type IX collagen expressed in ocular tissues.

Although the transciption factors controlling the expression of cartilage collagens remain relatively poorly characterized (see Lefebvre & de Crombrugghe 1998), the present homology search for published consensus recognition sites of transcription factors, such as SOX9, identified a few such elements in the promoters of the human type IX collagen genes. Of special interest was a 14-bp sequence that was found to be perfectly conserved between the human and mouse genes for the α2(IX) chain, while also showing similarity to a sequence in the promoters of the human COL2A1 and COL9A1 genes. Although the mere presence of a short consensus sequence is by no means proof of the importance of the respective region for actual transcriptional control, the 14-bp sequence in the COL9A2, Col9a2 and COL2A1 genes also shows a conserved location between two Sp1 binding sites adjacent to the major transcription start site of each gene. With the genomic sequences published in this work, it is now possible to carry out more detailed studies of the human type IX collagen gene promoter and enhancer regions to analyze the functional importance of the elements described.

The analysis of the alternative promoter region driving expression of the short form of α1(IX) indicated that intron 6, separating the alternative exon 1* from exon 7, spanned less than 70 bp in four species. This is below the minimum distance required for correct splicing of mRNA, as demonstrated in a mammalian experimental system (Wieringa *et al.* 1984). Thus, the absence of exon 7 from the α1(IX) short transcript appears to arise from a mechanistic limitation of the splicing machinery rather than controlled alternative splicing.

Since type IX collagen is a quantitatively minor component of the collagen fibrils in cartilage and other tissues and becomes covalently cross-linked to type II collagen, it cannot be obtained in high quantities from human sources. Also, extraction from tissues by pepsin digestion removes non-collagenous regions of possible functional importance, thereby limiting the usefulness of such material in subsequent studies. To generate a source of high amounts of full-length human type IX collagen, a protocol for the production and purification of a recombinant protein in insect cells was developed. Expression of each type IX collagen α chain was achieved by infection with three baculoviruses encoding the polypeptides with their native signal sequences. Since an exogenous source of prolyl 4-hydroxylase is required in insect cells to achieve high thermal stability of the triple helix in the production of collagens (Lamberg *et al.* 1996), a double promoter virus for the α and β subunits of this enzyme (Nokelainen *et al.* 1998) was used together with the viruses for the α chains of type IX collagen.

The initial attempts to isolate intracellular recombinant type IX collagen were unsuccessful because the protein was insoluble in a neutral buffer containing Triton X-100 and appeared to be degraded after acid extraction and selective salt precipitation. By contrast, the material present in the culture medium did not show evidence of degradation, as judged by SDS-PAGE. Based on a previous double-immunostaining experiment (Lamberg *et al.* 1996), it is likely that the four recombinant viruses used for infection purposes do not infect all insect cells equally. This may lead to intracellular accumulation of underhydroxylated and improperly folded molecules that are insoluble or susceptible to degradation. In support of this, SDS-PAGE analysis showed that the intracellular material appeared to migrate slightly faster than the secreted collagen (results not shown). Thus, secretion into the medium may serve as an inherent means of quality control in the purification of recombinant collagen.

Disulphide-bonded type IX collagen heterotrimers could be easily obtained from the culture medium with an apparent purity of over 90% and with only minor quantitative losses. Amino acid analysis of the purified recombinant type IX collagen indicated that the composition was in agreement with that expected in view of the deduced amino acid sequence of the cDNA. About 17% of the lysines in the recombinant protein had been hydroxylated, but the extent of lysine hydroxylation of tissue-derived type IX collagen is not known. The 4-hydroxylation level of the prolines in the recombinant protein was shown to be about 40%, which translates to about 80% of the theoretical maximum. Analysis by CD demonstrated a triple-helical conformation which showed a biphasic transition profile. Previous analyses of the melting properties of tissue-derived type IX collagen have demonstrated the presence of several regions of differing thermal stability in the molecule (Bruckner *et al.* 1983, 1985), while a more recent study has reported T_m values of about 40°C for the pepsinized COL1 and COL2 domains, but 49°C for the COL3 domain (Miles *et al.* 1998). Since these values are only slightly higher than those obtained for the recombinant type IX collagen, an adequate level of prolyl 4-hydroxylase activity was evidently achieved with the expression system. The level of prolyl 4-hydroxylation of type IX collagen *in vivo* has not been reported, but levels analogous to those obtained here were seen in type IX produced by cultured chondrocytes (Gibson *et al.* 1983). A small difference in melting behaviour of the recombinant protein compared with the pepsin-derived material is to be expected due to the effect of the neighbouring domains within the intact molecule. There appears, however, to be some discrepancy between the thermal properties of the intact recombinant collagen and intact type IX from rat chondrosarcoma (Miles *et al.* 1998). The significance of this finding is difficult to evaluate, since no data were presented on the isolation and characterization of the rat protein.

Somewhat surprisingly, the trimeric recombinant type IX collagen was efficiently secreted into the culture medium. Previous attempts at producing other secretable collagens in insect cells have mainly led to very inefficient secretion of the recombinant protein (Tomita *et al.* 1995, 1997, Lamberg *et al.* 1996, Myllyharju *et al.* 1997, Nokelainen *et al.* 1998). The difference may simply indicate that the type IX collagen signal peptides are more efficient than those of the other collagens studied, or alternatively, it may reflect a limitation on the capacity of the secretory pathway, possibly because the expression level of type IX collagen was several times lower than those of the other collagens. The expression of large amounts of secretory heterologous proteins sometimes results in inefficient secretion (see Ailor & Betenbaugh 1999). As with other eukaryotic cells, the participation of molecular chaperones in events along the secretory pathway is evident in insect cells but remains somewhat poorly characterized (see Ailor & Betenbaugh 1999). However, a baculovirus infection is known to interfere with the synthesis of endogenous proteins (Jarvis & Summers 1989), and therefore insect cell chaperones are unlikely to contribute significantly to the synthesis of heterologous proteins. It was found here that correctly folded recombinant type IX collagen was secreted when exogenous prolyl 4-hydroxylase was present, in addition to which, the expression level of the collagen appeared to correlate directly with the level of this enzyme activity. The presence of active prolyl 4-hydroxylase has been reported previously to dramatically enhance trimer formation in other collagens when produced in insect and yeast cells (Lamberg *et al.* 1996, Vuorela *et al.* 1997, Snellman *et al.* 2000),

verifying the essential role of this enzyme in the assembly of the collagen triple helix (Mazzorana *et al.* 1996, Mechling *et al.* 1996, see McLaughlin & Bulleid 1998). The efficient synthesis and secretion of the recombinant type IX collagen achieved here can be interpreted to indicate a negligible role for chaperones in the synthesis of type IX collagen, perhaps with the exception of prolyl 4-hydroxylase or its β subunit (PDI) (Wilson *et al.* 1998, Walmsley *et al.* 1999). Alternatively, the level of endogenous chaperones may remain high enough for the synthesis and secretion of type IX collagen while being insufficient in the case of the other collagens expressed at higher levels, resulting in their inefficient secretion. This latter view is supported by a recent report that expression of exogenous HSP47 enhances the secretion of heterotrimeric type I collagen in insect cells (Tomita *et al.* 1999).

Assembly of the type IX collagen chains has previously been studied with C-terminal, pepsin-resistant LMW fragments containing the COL1 domain and the cysteine region of the NC1 domain (Labourdette & van der Rest 1993), or with synthetic peptides containing six C-terminal Gly-X-Y triplets with 4-hydroxyproline and the complete NC1 region of each chain (Mechling *et al.* 1996). By contrast, we used recombinant baculoviruses to study the assembly and secretion of type IX collagen in insect cells by identifying the various molecular species from the culture medium by Western blotting. The results show that all recombinant type IX collagen α chains are able to form disulphide-bonded homodimers, whereas only the α1(IX) chains show clear homotrimer formation. These findings support previous observations that the C-terminal fragments of α1(IX) chains exhibit the highest potential for homotrimer formation (Mechling *et al.* 1996). The origin of the monomers and disulphide-bonded dimers in the culture medium in our experiment is somewhat unclear. Since collagens are in general secreted as trimers, at least a portion of the monomers and dimers may originate from trimers that were secreted but not fully disulphide-bonded. Some of the molecules may also represent immature molecules that were released into the medium due to the inevitable lysis of a small fraction of the cells during infection. True secretion of monomeric and dimeric molecules is also a possibility, however, since such a phenomenon has been documented both in insect cells and in mammalian cells (Fukuda *et al.* 1997, Tomita *et al.* 1997). The apparent abundance of monomers and dimers must also be in part an artifact arising because the efficiency of electroblotting is inversely related to the size of the blotted molecules.

When all three α chains were expressed simultaneously, a heterotrimer α1α2α3 was the predominant product formed, as expected from previous results (Labourdette & van der Rest 1993, Mechling *et al.* 1996). However, the use of fragments lacking in the C-terminal region of the NC1 domain also resulted in formation of homotrimers (Labourdette & van der Rest 1993), and it is therefore possible that the C-termini of the NC1 domains enhance formation of the heterotrimer in preference to homotrimers. It should be noted that homotrimers may have been formed in our experiment but failed to be secreted efficiently or were not detected due to the low quantities present.

Disulphide bonds form in the ER of eukaryotic cells provided that the reacting cysteinyl residues are close enough to each other (Creighton 1984). Two Cys residues at the COL1/NC1 junction in the three type IX collagen α chains are involved in interchain disulphide bond formation, and their functionality in connecting any two of the α chains has been demonstrated by Labourdette and van der Rest (1993), who were able to detect

all possible dimeric combinations in minor quantities. In our study, however, α1α3 was the only heterodimer detected. It is possible that the other two disulphide-bonded heterodimeric species were also formed but in quantities that were below the detection limit of the experimental system. It is also possible that these molecules were present in the reassociation experiment with the synthetic peptides (Mechling *et al.* 1996) but remained undetected. Even though our results were obtained with a semiquantitative analysis, it appears that co-expression of the α1(IX) and α3(IX) chains results in formation of the heterodimer α1α3, which dominates over the respective homotrimers. This suggests that formation of the α1α2α3 heterotrimer may involve a preferred route in which the α2(IX) chain can only associate with a pre-existing α1α3 dimer to yield the heterotrimer. Such a mechanism might be an intrinsic property of the NC1 domains, or else result from the action of chaperones and foldases. More precise experimentation would be required, however, to study the details of type IX collagen chain association *in cellulo*.

As discussed above, an important benefit to arise from this research was that the genomic organization of the type IX collagen genes can now be utilized to set up fast, reliable mutation screening protocols for the identification of genetic variations causing MED and similar disorders. Indeed, this has already resulted in interesting discoveries with regard to the molecular aetiology of MED and intervertebral disc disease (Annunen *et al.* 1999b, Paassilta *et al.* 1999, Lohiniva *et al.* 2000). Also, the allele frequencies of the tetranucleotide repeat detected here show that this intragenic polymorphism may be useful in future studies of the COL9A1 gene. A protocol was set up here for screening of the COL9A3 gene using CSGE, which has recently been characterized as a highly sensitive and reliable method (Körkkö *et al.* 1998), and its efficiency was demonstrated by the identification of several neutral polymorphisms in the COL9A3 gene. Also, a 9-bp deletion resulting in removal of three amino acids from the COL1 domain was identified in a patient with MED. Similar deletions in the COL1A1 gene result in lethal osteogenesis imperfecta, apparently by causing a shift in the phase of the chains in the triple helix, which may interfere with cross-link formation or lateral interaction of the molecules (Hawkins *et al.* 1991, Wallis *et al.* 1992). Surprisingly, the MED phenotype of the family members did not co-segregate with the 9-bp deletion. In another family a different 9-bp deletion affecting the same domain was identified, again with no segregating skeletal manifestations. These deletions must therefore be rare neutral variants of the COL9A3 gene, and thereby represent the first deletions within the triple-helical domain of any collagen that are not associated with a disease phenotype.

Assembly of the type IX collagen trimer is initiated at the C-terminus of the molecule. A few amino acids at the cysteine-containing COL1/NC1 junction appear to be important for chain assembly (Mechling *et al.* 1996). Once assembled, the three chains fold into a triple helix by a zipper-like process operating in a C to N direction (see McLaughlin & Bulleid 1998). This means that the α3(IX) chains carrying the deletion would be expected to participate in trimer assembly. To study this, we expressed the α3(IX) chain that was lacking in one Gly-Pro-Pro unit from the COL1 domain in insect cells together with wild-type α1(IX) and α2(IX) chains. As expected, the mutant α3(IX) chain participated in the formation of disulphide-bonded trimers that were secreted into the culture medium. As the deletion might affect or prevent the folding of the COL2 and COL3 domains into a triple helix, we performed a pepsin digestion to analyze whether or not a triple helix had

been formed. The digestion products of the mutant type IX collagen were identical to those originating from the wild-type recombinant, which has been shown by another method to contain thermally stable triple helices in all three collagenous domains (see section 5.2). Digestion of denatured collagen by pepsin has been shown to proceed quite slowly, and thus the usefulness of pepsin as a probe for a triple-helical conformation is limited (Bruckner & Prockop 1981, Bulleid *et al.* 1998). However, our results were obtained using an extended exposure to a large molar excess of pepsin, which should reduce the likelihood of false positive results (Bruckner & Prockop 1981). Also, a heat denaturation step was used as a positive control. In the light of the comparison with digested wild-type collagen, it can therefore be stated that the COL2 and COL3 domains of the mutant collagen were triple-helical. The results thus suggest, that the NC2 domain may be able to act as a "secondary nucleation site" to align the COL2/NC2 junction in the correct register despite a deletion in the α3(IX) COL1 domain (Fig. 4, panel B). Hydroxylatable proline residues are present at the extreme C-terminal end of the COL2 domain in the α1(IX) and α3(IX) chains, and only a few residues further towards the N-terminus in the α2(IX) chain. Studies of the FACIT COL1/NC1 junction (Mechling *et al.* 1996) and the nucleation mechanism in fibrillar collagens (Bulleid *et al.* 1997) suggest that these prolines may well participate in recreating the correct alignment of the triple helix. Also, a Cys residue is present at positions 21 and 22 in the NC2 domain (counting from the C-terminus of the domain) of the α1(IX) and α3(IX) chains, respectively, but absent from the α2(IX) chain. The location of these cysteines is conserved in the chicken α1(IX) and α3(IX) chains (van der Rest *et al.* 1985, Brewton *et al.* 1992, Har-El *et al.* 1992), suggesting that they may be important in forming an interchain disulphide bond. In wild-type type IX collagen, the COL1 domain of the α3(IX) chain is one Gly-X-Y unit shorter than that of the other two chains. It is possible that the conserved cysteines are among the key residues that help adaptation to this size difference in the wild-type protein and participate in the formation of a nucleation site for the folding of the COL2 domain (Fig. 4, panel A). The re-alignment of the chains of the mutant protein could thus simply represent an extended performance of the existing adaptation mechanism within the NC2 domain. The NC3 domain is known to re-align the α chains and form a nucleation site for the folding of the COL3 domain, which introduces a flexible kink into the molecule at the NC3 domain. A similar mechanism may perhaps be involved in the nucleation event in both the NC2 and the NC3 domains. This is supported by electron microscopy of human type IX collagen, which has shown that the NC2 domain may also adopt various conformations (Bruckner *et al.* 1988).

However, if the NC2 domain is assumed to be unable to correct the chain alignment, then the results of the pepsin digestion indicate that the deletion was propagated in one Gly-X-Y stagger all the way to the NC4 domain (Fig. 4, panel C). This may not compromise the pepsin resistance of the COL2, COL3 and NC3 domains, and could act in effect as a deletion of three residues from the COL3 domain. Such a change in the phase of the α chains would seem likely to interfere with the nucleation of the COL2 and COL3 triple helices. An in-frame deletion of 12 amino acids within the COL3 domain of either the α2(IX) or the α3(IX) chain is also known to cause MED (Muragaki *et al.* 1996, Holden *et al.* 1999, Paassilta *et al.* 1999, Bonnemann *et al.* 2000, Lohiniva *et al.* 2000), and it therefore appears unlikely that the deletions described in this work would, in a sense, create a three-residue deletion in the same domain but not result in a disease

phenotype. The NC3 domain of each chain contains two cysteine residues that are perfectly conserved at amino acid positions three and seven, counting from the C-terminus of the domain. Pepsin is only able to cleave the α2(IX) chain within this domain in the chicken (Reese *et al.* 1982, van der Rest *et al.* 1985), and the resulting HMW fragment remains as a trimer due to disulphide bonding in the NC3 domain (Mayne *et al.* 1985b). The same apparently holds true for type IX collagen in other species, includingthe human species, although a more heterogenous pattern of digestion products is seen (Duance *et al.* 1984, Bruckner *et al.* 1988). As evidenced by SDS-PAGE under non-reducing conditions, the mutant type IX collagen presents a pattern of pepsin-resistant fragments similar to the wild-type disulphide-bonded HMW fragments (Reese & Mayne 1981, Reese *et al.* 1982, Duance *et al.* 1984, Bruckner *et al.* 1988). Thus the three

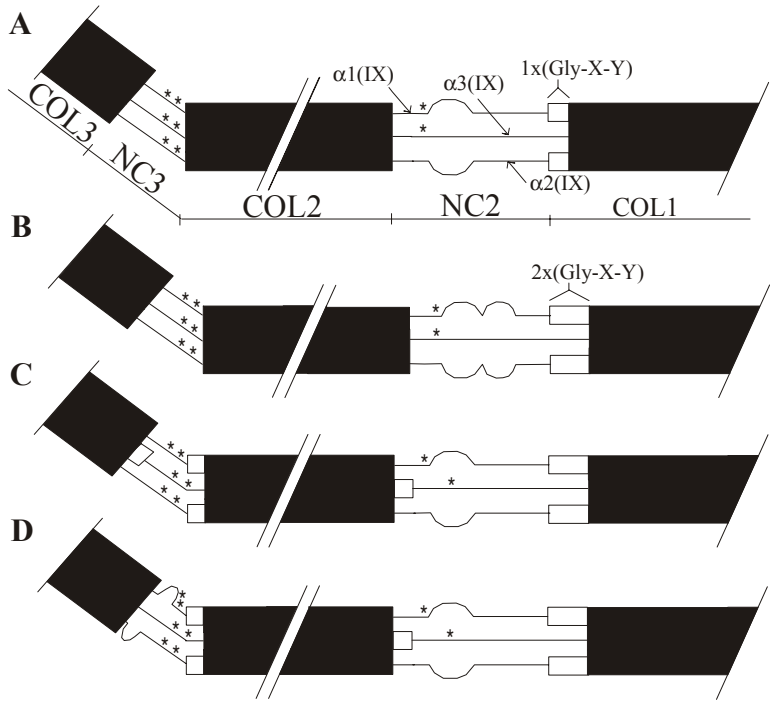


Figure 4. Mechanistic options for neutralization of the Gly-Pro-Pro deletion. Panel A, wild-type type IX collagen. Panel B, deletion neutralized by the NC2 domain. Panel C, deletion propagated to the N-terminus. Panel D, deletion neutralized by the NC3 domain. Only short segments of the triple helix are shown (black boxes). Open bars denote non-helical -Gly-X-Y- regions. The positions of the cysteine residues are indicated by asterisks. Not drawn exactly to scale.

chains of the mutant collagen must also be disulphide-bonded at the NC3 domain. It is uncertain whether or not these interchain disulphide bonds are formed in a similar manner to those in the wild-type collagen, but theoretically a deletion that was propagated to the N-terminus might still allow the formation of disulphide bonds to create a molecule migrating as a trimer.

Finally, it can be speculated that the NC3 domain could also have facilitated a corrective re-alignment of the α chains, perhaps with the assistance of the conserved cysteines (Fig. 4, panel D). Since the NC3 domain is only about 15 residues in length, one half of the size of NC2, such an adaptation would require a considerable conformational change in it, which appears unlikely.

Whatever the molecular mechanism involved, the fact remains that the deletion within the COL9A3 gene as described here represents an unusual neutral variant. The findings also act as a warning that caution should be exercised when evaluating the consequences of DNA mutations in the absence of complementary details at the level of protein chemistry or cell biology. The recombinant expression system created in this work may prove useful for more detailed analysis of the protein-level events occurring in this Gly-Pro-Pro deletion, as also in future studies of other genetic variations in the type IX collagen genes. As a result of this research, a solid set of new tools is now available for the study of type IX collagen.

References

Ailor E & Betenbaugh MJ (1999) Modifying secretion and post-translational processing in insect cells. Curr Opin Biotechnol 10: 142-145.

Akutsu N, Milbury CM, Burgeson RE & Nishiyama T (1999) Effect of type XII or XIV collagen NC-3 domain on the human dermal fibroblast migration into reconstituted collagen gel. Exp Dermatol 8: 17-21.

Ala-Kokko L, Hyland J, Smith C, Kivirikko KI, Jimenez SA & Prockop DJ (1991) Expression of a human cartilage procollagen gene (COL2A1) in mouse 3T3 cells. J Biol Chem 266: 14175-14178.

Annunen S, Körkkö J, Czarny M, Warman ML, Brunner HG, Kääriäinen H, Mulliken JB, Tranebjaerg L, Brooks DG, Cox GF, Cruysberg JR, Curtis MA, Davenport SLH, Friedrich CA, Kaitila I, Krawczynski MR, Latos-Bielenska A, Mukai S, Olsen BR, Shinno N, Somer M, Vikkula M, Zlotogora J, Prockop DJ & Ala-Kokko L (1999a) Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome, but other mutations cause overlapping Marshall/Stickler phenotypes. Am J Hum Genet 65: 974-983.

Annunen S, Paassilta P, Lohiniva J, Perälä M, Pihlajamaa T, Karppinen J, Tervonen O, Kröger H, Lähde S, Vanharanta H, Ryhänen L, Göring HH, Ott J, Prockop DJ & Ala-Kokko L (1999b) An allele of COL9A2 associated with intervertebral disc disease. Science 285: 409-412.

Arai M, Yada T, Suzuki S & Kimata K (1992) Isolation and characterization of type IX collagen-proteoglycan from the Swarm rat chondrosarcoma. Biochim Biophys Acta 1117: 60-70.

Arikawa-Hirasawa E, Watanabe H, Takami H, Hassell JR & Yamada Y (1999) Perlecan is essential for cartilage and cephalic development. Nat Genet 23: 354-358.

Arnold WV, Fertala A, Sieron AL, Hattori H, Mechling D, Bächinger H-P & Prockop DJ (1998) Recombinant procollagen II: Deletion of D period segments identifies sequences that are required for helix stabilization and generates a temperature-sensitive N-proteinase cleavage site. J Biol Chem 273: 31822-31828.

Aszódi A, Bateman JF, Hirsch E, Baranyi M, Hunziker EB, Hauser N, Bösze Z & Fässler R (1999) Normal skeletal development of mice lacking matrilin 1: redundant function of matrilins in cartilage? Mol Cell Biol 19: 7841-7845.

Aubert-Foucher E, Font B, Eichenberger D, Goldschmidt D, Lethias C & van der Rest M (1992) Purification and characterization of native type XIV collagen. J Biol Chem 267: 15759-15764.

Ausübel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG & Struhl K (eds) (1989) Current Protocols in Molecular Biology, Wiley, New York.

Ayad S, Abedin MZ, Grundy SM & Weiss JB (1981) Isolation and characterisation of an unusual collagen from hyaline cartilage and intervertebral disc. FEBS Lett 123: 195-199.

Ayad S, Abedin MZ, Weiss JB & Grundy SM (1982) Characterisation of another short-chain disulphide-bonded collagen from cartilage, vitreous and intervertebral disc. FEBS Lett 139: 300-304.

Ayad S, Marriott A, Brierley VH & Grant ME (1991) Mammalian cartilage synthesizes both proteoglycan and non-proteoglycan forms of type IX collagen. Biochem J 278: 441-445.

Ayad S, Marriott A, Morgan K & Grant ME (1989) Bovine cartilage types VI and IX collagens. Characterization of their forms in vivo. Biochem J 262: 753-761.

Bächinger H-P, Bruckner P, Timpl R, Prockop DJ & Engel J (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. Eur J Biochem 106: 619-632.

Bächinger H-P, Fessler LI, Timpl R & Fessler JH (1981) Chain assembly intermediate in the biosynthesis of type III procollagen in chick embryo blood vessels. J Biol Chem 256: 13193-13199.

Battie MC, Videman T, Gibbons LE, Fisher LD, Manninen H & Gill K (1995) 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. Spine 20: 2601-2612.

Bauer M, Dieterich W, Ehnis T & Schuppan D (1997) Complete primary structure of human collagen type XIV (undulin). Biochim Biophys Acta 1354: 183-188.

Bella J, Brosdky B & Berman HM (1995) Hydration structure of a collagen peptide. Structure 3: 893-906.

Berg RA & Prockop DJ (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem Biophys Res Commun 52: 115-120.

Berthod F, Germain L, Guignard R, Lethias C, Garrone R, Damour O, van der Rest M & Auger FA (1997) Differential expression of collagens XII and XIV in human skin and in reconstructed skin. J Invest Dermatol 108: 737-742.

Birkedal-Hansen H, Moore WGI, Bodden MK, Windsor LJ, Birkedal-Hansen B, DeCarlo A & Engler JA (1993) Matrix metalloproteinases: a review. Crit Rev Oral Biol Med 4: 197-250.

Bishop PN, Crossman MV, McLeod D & Ayad S (1994) Extraction and characterization of the tissue forms of collagen types II and IX from bovine vitreous. Biochem J 299: 497-505.

Bishop P, McLeod D & Ayad S (1992) Extraction and characterisation of the intact form of bovine vitreous type IX collagen. Biochem Biophys Res Commun 185: 392-397.

Blombäck (1967) Methods Enzymol 11: 398-411.

Bohme K, Li Y, Oh PS & Olsen BR (1995) Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development. Dev Dyn 204: 432-445.

Bollen A-M & Eyre DR (1993) Direct extraction of gelatinases from rat bone. Conn Tissue Res 29: 223-230.

Bonfanti L, Mironov AA Jr, Martínez-Menárguez JA, Martella O, Fusella A, Baldassarre M, Buccione R, Geuze HJ, Mironov AA & Luini A (1998) Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. Cell 95: 993-1003.

Bonnemann CG, Cox GF, Shapiro F, Wu J-J, Feener CA, Thompson TG, Anthony DC, Eyre DR, Darras BT & Kunkel LM (2000) A mutation in the α3 chain of type IX collagen causes autosomal dominant multiple epiphyseal dysplasia with mild myopathy. Proc Natl Acad Sci USA 97: 1212-1217.

Bork P (1992) The modular architecture of vertebrate collagens. FEBS Lett 307: 49-54.

Brewton RG, Ouspenskaia MV, van der Rest M & Mayne R (1992) Cloning of the chicken α3(IX) collagen chain completes the primary structure of type IX collagen. Eur J Biochem 205: 443-449.

Brewton RG, Wood BM, Ren Z-X, Gong Y, Tiller GE, Warman ML, Lee B, Horton WA, Olsen BR, Baker JR & Mayne R (1995) Molecular cloning of the α3 chain of human type IX collagen: linkage of the gene COL9A3 to chromosome 20q13.3. Genomics 30: 329-336.

Briggs MD, Hoffman SMG, King LM, Olsen AS, Mohrenweiser H, Leroy JG, Mortier GR, Rimoin DL, Lachman RS, Gaines ES, Cekleniak JA, Knowlton RG & Cohn DH (1995) Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. Nat Genet 10: 330-336.

Briggs MD, Mortier GR, Cole WG, King LM, Golik SS, Bonaventure J, Nuytinck L, De Paepe A, Leroy JG, Biesecker L, Lipson M, Wilcox WR, Lachman RS, Rimoin DL, Knowlton RG &

Cohn DH (1998) Diverse mutations in the gene for cartilage oligomeric matrix protein in the pseudoachondroplasia-multiple epiphyseal dysplasia disease spectrum. Am J Hum Genet 62: 311-319.

Brodsky B & Ramshaw JA (1997) The collagen triple-helix structure. Matrix Biol 15: 545-554.

Brown DJ, Bishop P, Hamdi H & Kenney MC (1996) Cleavage of structural components of mammalian vitreous by endogenous matrix metalloproteinase-2. Curr Eye Res 15: 439-445.

Brown JC, Mann K, Wiedemann H & Timpl R (1993) Structure and binding properties of collagen type XIV isolated from human placenta. J Cell Biol 120: 557-567.

Bruckner P, Mayne R & Tuderman L (1983) p-HMW-collagen, a minor collagen obtained from chick embryo cartilage without proteolytic treatment of the tissue. Eur J Biochem 136: 333-339.

Bruckner P, Mendler M, Steinmann B, Huber S & Winterhalter KH (1988) The structure of human collagen type IX and its organization in fetal and infant cartilage fibrils. J Biol Chem 263: 16911-16917.

Bruckner P & Prockop DJ (1981) Proteolytic enzymes as probes for the triple-helical conformation of procollagen. Anal Biochem 110: 360-368.

Bruckner P & van der Rest M (1994) Structure and function of cartilage collagens. Microsc Res Tech 28: 378-384.

Bruckner P, Vaughan L & Winterhalter KH (1985) Type IX collagen from sternal cartilage of chicken embryo contains covalently bound glycosaminoglycans. Proc Natl Acad Sci USA 82: 2608-2612.

Buckwalter JA (1995) Aging and degeneration of the human intervertebral disc. Spine 20: 1307-1314.

Buckwalter JA & Mankin HJ (1998) Articular cartilage: tissue design and chondrocyte-matrix interactions. Instr Course Lect 47: 477-486.

Bulleid NJ, Dalley JA & Lees JF (1997) The C-propeptide domain of collagen can be replaced with a transmembrane domain without affecting trimer formation or collagen triple helix folding during biosynthesis. EMBO J 16: 6694-6701.

Bulleid NJ & Freedman RB (1988) Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. Nature 335: 649-651.

Bulleid NJ, Wilson RR & Lad U (1998) A cautionary note when using pepsin as a probe for the formation of a collagen triple helix. Matrix Biology 17: 233-236.

Castagnola P, Tavella S, Gerecke DR, Dublet B, Gordon MK, Seyer J, Cancedda R, van der Rest M & Olsen BR (1992) Tissue-specific expression of type XIV collagen - a member of the FACIT class of collagens. Eur J Cell Biol 59: 340-347.

Chessler SD & Byers PH (1992) Defective folding and stable association with protein disulfide isomerase/prolyl hydroxylase of type I procollagen with a deletion in the pro α2(I) chain that preserves the Gly-X-Y repeat pattern. J Biol Chem 267: 7751-7757.

Chessler SD & Byers PH (1993) BiP binds type I procollagen pro α chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. J Biol Chem 268: 18226-18233.

Chu M-L & Prockop DJ (1993) Collagen: gene structure. In: Royce PM & Steinmann B (eds) Connective Tissue and Its Heritable Disorders. Molecular, Genetic and Medical Aspects. Wiley-Liss, New York, p 149-165.

Creighton TE (1984) Disulfide bond formation in proteins. Methods enzymol 107: 305-329.

Crossen R & Gruenwald S (1997) Baculovirus Expression Vector System. Instruction Manual. PharMingen, San Diego.

Deak F, Wagener R, Kiss I & Paulsson M (1999) The matrilins: a novel family of oligomeric extracellular matrix proteins. Matrix Biol 18: 55-64.

Délot E, Brodie SG, King LM, Wilcox WR & Cohn DH (1998) Physiological and pathological secretion of cartilage oligomeric matrix protein by cells in culture. J Biol Chem 273: 26692-26697.

Délot E, King LM, Briggs MD, Wilcox WR & Cohn DH (1999) Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene. Hum Mol Genet 8: 123-128.

Diab M, Wu J-J & Eyre DR (1996) Collagen type IX from human cartilage: a structural profile of intermolecular cross-linking sites. Biochem J 314: 327-332.

DiCesare PE, Mörgelin M, Mann K & Paulsson M (1994) Cartilage oligomeric matrix protein and thrombospondin 1. Purification from articular cartilage, electron microscopic structure, and chondrocyte binding. Eur J Biochem 223: 927-937.

Doege KJ & Fessler JH (1986) Folding of carboxyl domain and assembly of procollagen I. J Biol Chem 261: 8924-8935.

Douglas SP, Jenkins JM & Kadler KE (1998) Collagen IX: evidence for a structural association between NC4 domains in cartilage and a novel cleavage site in the α1(IX) chain. Matrix Biol 16: 497-505.

Duance VC, Wotton SF, Voyle CA & Bailey AJ (1984) Isolation and characterization of the precursor of type M collagen. Biochem J 221: 885-889.

Dublet B, Oh S, Sugrue SP, Gordon MK, Gerecke DR, Olsen BR & van der Rest M (1989) The structure of avian type XII collagen. α1(XII) chains contain 190-kDa non-triple helical amino-terminal domains and form homotrimeric molecules. J Biol Chem 264: 13150-13156.

Dublet B & van der Rest (1987) Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-derived fragments. J Biol Chem 262: 17724-17727.

Durr J, Goodman S, Potocnik A, von der Mark H & von der Mark K (1993) Localization of β1-integrins in human cartilage and their role in chondrocyte adhesion to collagen and fibronectin. Exp Cell Res 207: 235-244.

Durr J, Lammi P, Goodman SL, Aigner T & von der Mark K (1996) Identification and immunolocalization of laminin in cartilage. Exp Cell Res 222: 225-233.

Ehnis T, Dieterich W, Bauer M, Kresse H & Schuppan D (1997) Localization of a binding site for the proteoglycan decorin on collagen XIV (undulin). J Biol Chem 272: 20414-20419.

Ehnis T, Dieterich W, Bauer M, von Lampe B & Schuppan D (1996) A chondroitin/dermatan sulfate form of CD44 is a receptor for collagen XIV (undulin). Exp Cell Res 229: 388-397.

Ehnis T, Dieterich W, Bauer M & Schuppan D (1998) Localization of a cell adhesion site on collagen XIV (undulin). Exp Cell Res 239: 477-480.

Eikenberry EF, Mendler M, Bürgin R, Winterhalter K & Bruckner P (1992) Fibrillar organization in cartilage. In: Kuettner K, Schleyerbach R, Peyron JG & Hascall VC (eds) Articular Cartilage and Osteoarthritis. Raven Press, New York, p 133-148.

Erlebacher A, Filvaroff EH, Gitelman SE & Derynck R (1995) Toward a molecular understanding of skeletal development. Cell 80: 371-378.

Eyre DR, Apon S, Wu J-J, Ericsson LH & Walsh KA (1987) Collagen type IX: evidence for covalent linkages to type II collagen in cartilage. FEBS Lett 220: 337-341.

Eyre DR (1991) The collagens of articular cartilage. Semin Arthritis Rheum 21: 2-11.

Eyre DR & Wu J-J (1987) Type XI or α1α2α3 collagen. In: Burgeson RE & Mayne R (eds) Structure and Function of Collagen Types, Academic Press, Orlando, p 261-281.

Eyre DR, Wu J-J & Woods P (1992) Cartilage specific collagens. In: Kuettner K, Schleyerbach R, Peyron JG & Hascall VC (eds) Articular Cartilage and Osteoarthritis. Raven Press, New York, p 119-131.

Fässler R, Schnegelsberg PN, Dausman J, Shinya T, Muragaki Y, McCarthy MT, Olsen BR & Jaenisch R (1994) Mice lacking α1(IX) collagen develop noninflammatory degenerative joint disease. Proc Natl Acad Sci USA 91: 5070-5074.

Felson DT, Couropmitree NN, Chaisson CE, Hannan MT, Zhang Y, McAlindon TE, LaValley M, Levy D & Myers RH (1998) Evidence for a Mendelian gene in a segregation analysis of generalized radiographic osteoarthritis: the Framingham Study. Arthritis Rheum 41: 1064-1071.

Fertala A, Sieron AL, Ganguly A, Li S-W, Ala-Kokko L, Anumula KR & Prockop DJ (1994) Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). Biochem J 298: 31-37.

Fichard A, Kleman J-P & Ruggiero F (1994) Another look at collagen V and XI molecules. Matrix Biol 14: 515-531.

Fichard A, Tillet E, Delacoux F, Garrone R & Ruggiero F (1997) Human recombinant α1(V) collagen chain. Homotrimeric assembly and subsequent processing. J Biol Chem 272: 30083-30087.

Fitch JM, Mentzer A, Mayne R & Linsenmayer TF (1988) Acquisition of type IX collagen by the developing avian primary corneal stroma and vitreous. Dev Biol 128: 396-405.

Fitch JM, Gordon MK, Gibney EP & Linsenmayer TF (1995) Analysis of transcriptional isoforms of collagen types IX, II, and I in the developing avian cornea by competitive polymerase chain reaction. Dev Dyn 202: 42-53.

Font B, Aubert-Foucher E, Goldschmidt D, Eichenberger D & van der Rest M (1993) Binding of collagen XIV with the dermatan sulfate side chain of decorin. J Biol Chem 268: 25015-25018.

Font B, Eichenberger D, Rosenberg LM & van der Rest M (1996) Characterization of the interactions of type XII collagen with two small proteoglycans from fetal bovine tendon, decorin and fibromodulin. Matrix Biol 15: 341-348.

Frischholz S, Beier F, Girkontaite I, Wagner K, Pöschl E, Turnay J, Mayer U & von der Mark K (1998) Characterization of human type X procollagen and its NC-1 domain expressed as recombinant proteins in HEK293 cells. J Biol Chem 273: 4547-4555.

Fukuda K, Hori H, Utani A, Burbelo PD & Yamada Y (1997) Formation of recombinant triple-helical $[\alpha1(IV)]_2\alpha2(IV)$ collagen molecules in CHO cells. Biochem Biophys Res Commun 231: 178-182.

Gadher SJ, Eyre DR, Duance VC, Wotton SF, Heck LW, Schmid TM & Woolley DE (1988) Susceptibility of cartilage collagens type II, IX, X, and XI to human synovial collagenase and neutrophil elastase. Eur J Biochem 175: 1-7.

Gaill F, Wiedemann H, Mann K, Kühn K, Timpl R & Engel J (1991) Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents. J Mol Biol 221: 209-223.

Ganguly A, Rock MJ & Prockop DJ (1993) Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90: 10325-10329.

Garrone R, Lethias C & Le Guellec D (1997) Distribution of minor collagens during skin development. Microsc Res Tech 38: 407-412.

Geisse S, Gram H, Kleuser B & Kocher HP (1996) Eukaryotic expression systems: a comparison. Protein Expr Purif 8: 271-282.

Gerecke DR, Foley JW, Castagnola P, Gennari M, Dublet B, Cancedda R, Linsenmayer TF, van der Rest M, Olsen BR & Gordon MK (1993) Type XIV collagen is encoded by alternative transcripts with distinct 5' regions and is a multidomain protein with homologies to von Willebrand's factor, fibronectin, and other matrix proteins. J Biol Chem 268: 12177-12184.

Gerecke DR, Olson PF, Koch M, Knoll JHM, Taylor R, Hudson DL, Champliaud M-F, Olsen BR & Burgeson RE (1997) Complete primary structure of two splice variants of collagen XII, and assignment of α1(XII) collagen (COL12A1), α1(IX) collagen (COL9A1), and α1(XIX) collagen (COL19A1) to human chromosome 6q12-q13. Genomics 41: 236-242.

Gibson GJ, Kielty CM, Garner C, Schor SL & Grant ME (1983) Identification and partial characterization of three low-molecular-weight collagenous polypeptides synthesized by chondrocytes cultured within collagen gels in the absence and in the presence of fibronectin. Biochem J 211: 417-426.

Glick BS & Malhotra V (1998) The curious status of the Golgi apparatus. Cell 95: 883-889.

Gordon MK, Gerecke DR, Dublet B, van der Rest M & Olsen BR (1989) Type XII collagen. A large multidomain molecule with partial homology to type IX collagen. J Biol Chem 264: 19772-19778.

Gordon MK, Gerecke DR & Olsen BR (1987) Type XII collagen: distinct extracellular matrix component discovered by cDNA cloning. Proc Natl Acad Sci USA 84: 6040-6044.

Grassel S, Timpl R, Tan EM & Chu M-L (1996) Biosynthesis and processing of type XVI collagen in human fibroblasts and smooth muscle cells. Eur J Biochem 242: 576-584.

Grassel S, Unsold C, Schacke H, Bruckner-Tuderman L & Bruckner P (1999) Collagen XVI is expressed by human dermal fibroblasts and keratinocytes and is associated with the microfibrillar apparatus in the upper papillary dermis. Matrix Biol 18: 309-317.

Hägg PM, Hägg PO, Peltonen S, Autio-Harmainen H & Pihlajaniemi T (1997) Location of type XV collagen in human tissues and its accumulation in the interstitial matrix of the fibrotic kidney. Am J Pathol 150: 2075-2086.

Hägg P, Rehn M, Huhtala P, Väisänen T, Tamminen M & Pihlajaniemi T (1998) Type XIII collagen is identified as a plasma membrane protein. J Biol Chem 273: 15590-15597.

Hagg R, Bruckner P & Hedbom E (1998) Cartilage fibrils of mammals are biochemically heterogeneous: differential distribution of decorin and collagen IX. J Cell Biol 142: 285-294.

Hagg R, Hedbom E, Mollers U, Aszódi A, Fässler R & Bruckner P (1997) Absence of the α1(IX) chain leads to a functional knock-out of the entire collagen IX protein in mice. J Biol Chem 272: 20650-20654.

Haimes HB, Jimenez PA, Li Y, Shinya T & Olsen BR (1995) Overexpression of the NC4 domain of type IX collagen induces osteoarthritis in mice. Inflamm Res 44 Suppl 2: S127-128.

Hammond C & Helenius A (1995) Quality control in the secretory pathway. Curr Opin Cell Biol 7: 523-529.

Har-el R, Sharma YD, Aguilera A, Ueyama N, Wu J-J, Eyre DR, Juricic L, Chandrasekaran S, Li M, Nah H-D, Upholt WB & Tanzer ML (1992) Cloning and developmental expression of the α3 chain of chicken type IX collagen. J Biol Chem 267: 10070-10076.

Hawkins JR, Superti-Furga A, Steinmann B & Dalgleish R (1991) A 9-base pair deletion in COL1A1 in a lethal variant of osteogenesis imperfecta. J Biol Chem 266: 22370-22374.

Hayashi M, Hayashi K, Iyama K-I, Trelstad RL, Linsenmayer TF & Mayne R (1992) Notochord of chick embryos secretes short-form type IX collagen prior to the onset of vertebral chondrogenesis. Dev Dyn 194: 169-176.

Hecht JT, Nelson LD, Crowder E, Wang Y, Elder FF, Harrison WR, Francomano CA, Prange CK, Lennon GG, Deere M & Lawler J (1995) Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia. Nat Genet 10: 325-329.

Holden P, Canty EG, Mortier GR, Zabel B, Spranger J, Carr A, Grant ME, Loughlin JA & Briggs MD (1999) Identification of novel pro-α2(IX) collagen gene mutations in two families with distinctive oligo-epiphyseal forms of multiple epiphyseal dysplasia. Am J Hum Genet 65: 31-38.

Hollenberg CP & Gellissen G (1997) Production of recombinant proteins by methylotrophic yeasts. Curr Opin Biotechnol 8: 554-560.

Huber S, van der Rest M, Bruckner P, Rodriguez E, Winterhalter KH & Vaughan L (1986) Identification of the type IX collagen polypeptide chains. The α2(IX) polypeptide carries the chondroitin sulfate chain(s). J Biol Chem 261: 5965-5968.

Huber S, Winterhalter KH & Vaughan L (1988) Isolation and sequence analysis of the glycosaminoglycan attachment site of type IX collagen. J Biol Chem 263: 752-756.

Humzah MD & Soames RW (1988) Human intervertebral disc: structure and function. Anat Rec 220: 337-356.

Ikegawa S, Ohashi H, Nishimura G, Kim KC, Sannohe A, Kimizuka M, Fukushima Y, Nagai T & Nakamura Y (1998) Novel and recurrent COMP (cartilage oligomeric matrix protein) mutations in pseudoachondroplasia and multiple epiphyseal dysplasia. Hum Genet 103: 633-638.

Imhof M & Trueb B (1998) An alternative insert of three amino acids is incorporated into collagen XIV in a developmentally regulated fashion. FEBS Lett 438: 325-328.

Imhof M & Trueb B (1999) Comparative cytogenetic mapping of COL14A1, the gene for human and mouse collagen XIV. Cytogenet Cell Genet 84: 217-219.

Inoguchi K, Yoshioka H, Khaleduzzaman M & Ninomiya Y (1995) The mRNA for α1(XIX) collagen chain, a new member of FACITs, contains a long unusual 3' untranslated region and displays many unique splicing variants. J Biochem (Tokyo) 117: 137-146.

International Working Group on Constitutional Diseases of Bone (1998) International nomenclature and classification of the osteochondrodysplasias (1997). International Working Group on Constitutional Diseases of Bone. Am J Med Genet 79: 376-382.

Iozzo RV (1997) The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. Crit Rev Biochem Mol Biol 32: 141-174.

Irwin MH, Silvers SH & Mayne R (1985) Monoclonal antibody against chicken type IX collagen: preparation, characterization, and recognition of the intact form of type IX collagen secreted by chondrocytes. J Cell Biol 101: 814-823.

Jarvis DL & Summers MD (1989) Glycosylation and secretion of human tissue plasminogen activator in recombinant baculovirus-infected insect cells. Mol Cell Biol 9: 214-223.

Juvonen M, Sandberg M & Pihlajaniemi T (1992) Patterns of expression of the six alternatively spliced exons affecting the structures of the COL1 and NC2 domains of the α1(XIII) collagen chain in human tissues and cell lines. J Biol Chem 267: 24700-24707.

Kadler KE, Holmes DF, Trotter JA & Chapman JA (1996) Collagen fibril formation. Biochem J 316: 1-11.

Kania AM, Reichenberger E, Baur ST, Karimbux NY, Taylor RW, Olsen BR & Nishimura I (1999) Structural variation of type XII collagen at its carboxyl-terminal NC1 domain generated by tissue-specific alternative splicing. J Biol Chem 274: 22053-22059.

Keene DR, Lunstrum GP, Morris NP, Stoddard DW & Burgeson RE (1991) Two type XII-like collagens localize to the surface of banded collagen fibrils. J Cell Biol 113: 971-978.

Keene DR, Oxford JT & Morris NP (1995) Ultrastructural localization of collagen types II, IX, and XI in the growth plate of human rib and fetal bovine epiphyseal cartilage: type XI collagen is restricted to thin fibrils. J Histochem Cytochem 43: 967-979.

Kellokumpu S, Suokas M, Risteli L & Myllylä R (1997) Protein disulfide isomerase and newly synthesized procollagen chains form higher-order structures in the lumen of the endoplasmic reticulum. J Biol Chem 272: 2770-2777.

Khaleduzzaman M, Sumiyoshi H, Ueki Y, Inoguchi K, Ninomiya Y & Yoshioka H (1997) Structure of the human type XIX collagen (COL19A1) gene, which suggests it has arisen from an ancestor gene of the FACIT family. Genomics 45: 304-312.

Kidd IM & Emery VC (1993) The use of baculoviruses as expression vectors. Appl Biochem Biotechnol 42: 137-159.

Kielty CM, Hopkinson I & Grant ME (1993) Collagen: the collagen family: structure, assembly, and organization in the extracellular matrix. In: Royce PM & Steinmann B (eds) Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects. Wiley-Liss, New York, p 103-147.

Kielty CM & Shuttleworth CA (1997) Microfibrillar elements of the dermal matrix. Microsc Res Tech 38: 413-427.

Kimura T, Mattei M-G, Stevens JW, Goldring MB, Ninomiya Y & Olsen BR (1989) Molecular cloning of rat and human type IX collagen cDNA and localization of the α1(IX) gene on the human chromosome 6. Eur J Biochem 179: 71-78.

Kimura T, Nakata K, Tsumaki N, Miyamoto S, Matsui Y, Ebara S & Ochi T (1996) Progressive degeneration of articular cartilage and intervertebral discs. An experimental study in transgenic mice bearing a type IX collagen mutation. Int Orthop 20:177-181.

Kivirikko KI & Myllyharju J (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. Matrix Biol 16: 357-368.

Kivirikko KI & Myllylä R (1982) Post-translational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol 82 Pt A: 245-304.

Kivirikko KI & Myllylä R (1985) Post-translational processing of collagens. Ann NY Acad Sci 460: 187-201.

Koch M, Bernasconi C & Chiquet M (1992) A major oligomeric fibroblast proteoglycan identified as a novel large form of type-XII collagen. Eur J Biochem 207: 847-856.

Koch M, Bohrmann B, Matthison M, Hagios C, Trueb B & Chiquet M (1995) Large and small splice variants of collagen XII: differential expression and ligand binding. J Cell Biol 130: 1005-1014.

Koivu J & Myllylä R (1987) Interchain disulfide bond formation in types I and II procollagen. Evidence for a protein disulphide isomerase catalyzing bond formation. J Biol Chem 262: 6159-6164.

Konomi H, Seyer JM, Ninomiya Y & Olsen BR (1986) Peptide-specific antibodies identify the α2 chain as the proteoglycan subunit of type IX collagen. J Biol Chem 261: 6742-6746.

Körkkö J, Annunen S, Pihlajamaa T, Prockop DJ & Ala-Kokko L (1998) Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. Proc Natl Acad Sci USA 95: 1681-1685.

Labourdette L & van der Rest M (1993) Analysis of the role of the COL1 domain and its adjacent cysteine-containing sequence in the chain assembly of type IX collagen. FEBS Lett 320: 211-214.

Lai CH & Chu M-L (1996) Tissue distribution and developmental expression of type XVI collagen in the mouse. Tissue Cell 28: 155-164.

Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T & Kivirikko KI (1996) Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. J Biol Chem 271: 11988-11995.

Lees JF, Tasab M & Bulleid NJ (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J 16: 908-916.

Lefebvre V & de Crombrugghe B (1998) Toward understanding SOX9 function in chondrocyte differentiation. Matrix Biol 16: 529-540.

Lesage A, Penin F, Geourjon C, Marion D & van der Rest M (1996) Trimeric assembly and three-dimensional structure model of the FACIT collagen COL1-NC1 junction from CD and NMR analysis. Biochemistry 35: 9647-9660.

Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, Oxford JT, Morris NP, Andrikopoulos K, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE & Olsen BR (1995) A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. Cell 80: 423-430.

Liu C-Y, Olsen BR & Kao WW-Y (1993) Developmental patterns of two α1(IX) collagen mRNA isoforms in mouse. Dev Dyn 198: 150-157.

Lohiniva J, Paassilta P, Seppänen U, Vierimaa O, Kivirikko S & Ala-Kokko L (2000) Splicing mutations in the COL3 domain of collagen IX cause multiple epiphyseal dysplasia. Am J Med Genet 90: 216-222.

Lozano G, Ninomiya Y, Thompson H & Olsen BR (1985) A distinct class of vertebrate collagen genes encodes chicken type IX collagen polypeptides. Proc Natl Acad Sci USA 82: 4050-4054.

Lunstrum GP, McDonough AM, Marinkovich MP, Keene DR, Morris NP & Burgeson RE (1992) Identification and partial purification of a large, variant form of type XII collagen. J Biol Chem 267: 20087-20092.

Lunstrum GP, Morris NP, McDonough AM, Keene DR & Burgeson RE (1991) Identification and partial characterization of two type XII-like collagen molecules. J Cell Biol 113: 963-969.

Maciewicz RA & Wotton SF (1991) Degradation of cartilage matrix components by the cysteine proteinases, cathepsins B and L. Biomed Biochim Acta 50: 561-564.

Maddox BK, Keene DR, Sakai LY, Charbonneau NL, Morris NP, Ridgway CC, Boswell BA, Sussman MD, Horton WA, Bächinger H-P & Hecht JT (1997) The fate of cartilage oligomeric matrix protein is determined by the cell type in the case of a novel mutation in pseudoachondroplasia. J Biol Chem 272: 30993-30997.

Mallein-Gerin F, Ruggiero F, Quinn TM, Bard F, Grodzinsky AJ, Olsen BR & van der Rest M (1995) Analysis of collagen synthesis and assembly in culture by immortalized mouse chondrocytes in the presence or absence of α1(IX) collagen chains. Exp Cell Res 219: 257-265.

von der Mark K, van Menxel M & Wiedemann H (1982) Isolation and characterization of new collagens from chick cartilage. Eur J Biochem 124: 57-62.

von der Mark K, van Menxel M & Wiedemann H (1984) Isolation and characterization of a precursor form of M collagen from embryonic chicken cartilage. Eur J Biochem 138: 629-633.

Mayne R & Brewton RG (1993) New members of the collagen superfamily. Curr Opin Cell Biol 5: 883-890.

Mayne R, van der Rest M, Ninomiya Y & Olsen BR (1985b) The structure of type IX collagen. Ann NY Acad Sci 460: 38-46.

Mayne R, van der Rest M, Weaver DC & Butler WT (1985a) The structure of a small collagenous fragment isolated form chicken hyaline cartilage. J Cell Biochem 27: 133-141.

Mazzorana M, Giry-Lozinguez C & van der Rest M (1995) Trimeric assembly of collagen XII: effect of deletion of the C-terminal part of the molecule. Matrix Biology 14: 583-588.

Mazzorana M, Gruffat H, Sergeant A & van der Rest M (1993) Mechanisms of collagen trimer formation. Construction and expression of a recombinant minigene in HeLa cells reveals a

direct effect of prolyl hydroxylation on chain assembly of type XII collagen. J Biol Chem 268: 3029-3032.

Mazzorana M, Snellman A, Kivirikko KI, van der Rest M & Pihlajaniemi T (1996) Involvement of prolyl 4-hydroxylase in the assembly of trimeric minicollagen XII. Study in a baculovirus expression system. J Biol Chem 271: 29003-29008.

McCarty DJ (1989) Arthritis and Allied Conditions. Lea & Febiger, Philadelphia.

McCormick D, van der Rest M, Goodship J, Lozano G, Ninomiya Y & Olsen BR (1987) Structure of the glycosaminoglycan domain in the type IX collagen-proteoglycan. Proc Natl Acad Sci USA 84: 4044-4048.

McLaughlin SH & Bulleid NJ (1998) Molecular recognition in procollagen chain assembly. Matrix Biol 16: 369-377.

Mechling DE, Gambee JE, Morris NP, Sakai LY, Keene DR, Mayne R & Bächinger H-P (1996) Type IX collagen NC1 domain peptides can trimerize *in vitro* without forming a triple helix. J Biol Chem 271: 13781-13785.

Mendler M, Eich-Bender SG, Vaughan L, Winterhalter KH & Bruckner P (1989) Cartilage contains mixed fibrils of collagen types II, IX and XI. J Cell Biol 108: 191-197.

Merrington CL, Bailey MJ & Possee RD (1997) Manipulation of baculovirus vectors. Mol Biotechnol 8: 283-297.

Miles CA, Knott L, Sumner IG & Bailey AJ (1998) Differences between the thermal stabilities of the three triple-helical domains of type IX collagen. J Mol Biol 277: 135-144.

van Mourik JBA, Buma P & Wilcox WR (1998b) Electron microscopical study in multiple epiphyseal dysplasia type II. Ultrastruct Pathol 22: 249-251.

van Mourik JBA, Hamel BCJ & Mariman ECM (1998a) A large family with multiple epiphyseal dysplasia linked to COL9A2 gene. Am J Med Genet 77: 234-240.

Müller-Glauser W, Humbel B, Glatt M, Sträuli P, Winterhalter KH & Bruckner P (1986) On the role of type IX collagen in the extracellular matrix of cartilage: type IX collagen is localized to intersections of collagen fibrils. J Cell Biol 102: 1931-1939.

Muir H (1995) The chondrocyte, architect of cartilage. Biomechanics, structure, function and molecular biology of cartilage matrix macromolecules. Bioessays 17: 1039-1048.

Mundlos S & Olsen BR (1997a) Heritable diseases of the skeleton. Part I: molecular insights into skeletal development-transcription factors and signaling pathways. FASEB J 11: 125-132.

Mundlos S & Olsen BR (1997b) Heritable diseases of the skeleton. Part II: molecular insights into skeletal development-matrix components and their homeostasis. FASEB J 11: 227-233.

Muragaki Y, Kimura T, Ninomiya Y & Olsen BR (1990a) The complete primary structure of two distinct forms of human α1(IX) collagen chains. Eur J Biochem 192: 703-708.

Muragaki Y, Mariman ECM, van Beersum SEC, Perälä M, van Mourik JBA, Warman ML, Olsen BR & Hamel BCJ (1996) A mutation in the gene encoding the α2 chain of the fibril-associated collagen IX, COL9A2, causes multiple epiphyseal dysplasia (EDM2). Nat Genet 12: 103-105.

Muragaki Y, Nishimura I, Henney A, Ninomiya Y & Olsen BR (1990b) The α1(IX) collagen gene gives rise to two different transcripts in both mouse embryonic and human fetal RNA. Proc Natl Acad Sci USA 87: 2400-2404.

Myers JC, Dion AS, Abraham V & Amenta PS (1996) Type XV collagen exhibits a widespread distribution in human tissues but a distinct localization in basement membrane zones. Cell Tissue Res 286: 493-505.

Myers JC, Li D, Bageris A, Abraham V, Dion AS & Amenta PS (1997) Biochemical and immunohistochemical characterization of human type XIX defines a novel class of basement membrane zone collagens. Am J Pathol 151: 1729-1740.

Myers JC, Li D, Rubinstein NA & Clark CC (1999) Up-regulation of type XIX collagen in rhabdomyosarcoma cells accompanies myogenic differentiation. Exp Cell Res 253: 587-598.

Myers JC, Sun MJ, D'Ippolito JA, Jabs EW, Neilson EG & Dion AS (1993) Human cDNA clones transcribed from an unusually high-molecular-weight RNA encode a new collagen chain. Gene 123: 211-217.

Myers JC, Yang H, D'Ippolito JA, Presente A, Miller MK & Dion AS (1994) The triple-helical region of human type XIX collagen consists of multiple collagenous subdomains and exhibits limited sequence homology to α1(XVI). J Biol Chem 269: 18549-18557.

Myllyharju J, Lamberg A, Notbohm H, Fietzek PP, Pihlajaniemi T & Kivirikko KI (1997) Expression of wild-type and modified proα chains of human type I procollagen in insect cells leads to the formation of stable $[\alpha1(I)]_2\alpha2(I)$ collagen heterotrimers and $[\alpha1(I)]_3$ homotrimers but not $[\alpha2(I)]_3$ homotrimers. J Biol Chem 272: 21824-21830.

Nagata K (1996) Hsp47: a collagen-specific molecular chaperone. Trends Biochem Sci 21: 22-26.

Nakata K, Ono K, Miyazaki J-I, Olsen BR, Muragaki Y, Adachi E, Yamamura K-I & Kimura T (1993) Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing α1(IX) collagen chains with a central deletion. Proc Natl Acad Sci USA 90: 2870-2874.

Neame PJ, Tapp H & Azizan A (1999) Noncollagenous, nonproteoglycan macromolecules of cartilage. Cell Mol Life Sci 55: 1327-1340.

Newall JF & Ayad S (1995) Collagen IX isoforms in the intervertebral disc. Biochem Soc Trans 23: 517S.

Ninomiya Y, Castagnola P, Gerecke D, Gordon MK, Jacenko O, LuValle P, McCarthy M, Muragaki Y, Nishimura I, Oh S, Rosenblum M, Sato N, Sugrue S, Taylor R, Vasios G, Yamaguchi N & Olsen BR (1990) The molecular biology of collagens with short triple-helical domains. In: Sandell LJ & Boyd CD (eds) Extracellular Matrix Genes. Academic Press, San Diego, p 79-114.

Ninomiya Y & Olsen BR (1984) Synthesis and characterization of cDNA encoding a cartilage-specific short collagen. Proc Natl Acad Sci USA 81: 3014-3018.

Ninomiya Y, van der Rest M, Mayne R, Lozano G & Olsen BR (1985) Construction and characterization of cDNA encoding the α2 chain of chicken type IX collagen. Biochemistry 24: 4223-4229.

Nishimura I, Muragaki Y & Olsen BR (1989) Tissue-specific forms of type IX collagen-proteoglycan arise from the use of two widely separated promoters. J Biol Chem 264: 20033-20041.

Nishiyama T, McDonough AM, Bruns RR & Burgeson RE (1994) Type XII and XIV collagens mediate interactions between banded collagen fibers in vitro and may modulate extracellular matrix deformability. J Biol Chem 269: 28193-28199.

Niyibizi C, Visconti CS, Kavalkovich K & Woo SL (1995) Collagens in an adult bovine medial collateral ligament: immunofluorescence localization by confocal microscopy reveals that type XIV collagen predominates at the ligament-bone junction. Matrix Biol 14: 743-751.

Nokelainen M, Helaakoski T, Myllyharju J, Notbohm H, Pihlajaniemi T, Fietzek PP & Kivirikko KI (1998) Expression and characterization of recombinant human type II collagens with low and high contents of hydroxylysine and its glycosylated forms. Matrix Biol 16: 329-338.

Noro A, Kimata K, Oike Y, Shinomura T, Maeda N, Yano S, Takahashi N & Suzuki S (1983) Isolation and characterization of a third proteoglycan (PG-Lt) from chick embryo cartilage which contains disulfide-bonded collagenous polypeptide. J Biol Chem 258: 9323-9331.

Notbohm H, Nokelainen M, Myllyharju J, Fietzek PP, Müller PK & Kivirikko KI (1999) Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis *in vitro*. J Biol Chem 274: 8988-8992.

Oh SP, Griffith CM, Hay ED & Olsen BR (1993) Tissue-specific expression of type XII collagen during mouse embryonic development. Dev Dyn 196: 37-46.

Oh SP, Kamagata Y, Muragaki Y, Timmons S, Ooshima A & Olsen BR (1994) Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins. Proc Natl Acad Sci USA 91: 4229-4233.

Oh SP, Taylor RW, Gerecke DR, Rochelle JM, Seldin MF & Olsen BR (1992) The mouse α1(XII) and human α1(XII)-like collagen genes are localized on mouse chromosome 9 and human chromosome 6. Genomics 14: 225-231.

Olsen BR, Hoffmann HP & Prockop DJ (1976) Interchain disulfide bonds at the COOH-terminal end of procollagen synthesized by matrix-free cells from chick embryo tendon and cartilage. Arch Biochem Biophys 175: 341-350.

O'Reilly MS, Boehm T, Shing Y, Fukai N, Vasios G, Lane WS, Flynn E, Birkhead JR, Olsen BR & Folkman J (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88: 277-285.

Paassilta P, Lohiniva J, Annunen S, Bonaventure J, Le Merrer M, Pai L & Ala-Kokko L (1999) COL9A3: A third locus for multiple epiphyseal dysplasia. Am J Hum Genet 64: 1036-1044.

Peltonen S, Hentula M, Hägg P, Ylä-Outinen H, Tuukkanen J, Lakkakorpi J, Rehn M, Pihlajaniemi T & Peltonen J (1999) A novel component of epidermal cell-matrix and cell-cell contacts: transmembrane protein type XIII collagen. J Invest Dermatol 113: 635-642.

Perälä M, Elima K, Metsäranta M, Rosati R, de Crombrugghe B & Vuorio E (1994) The exon structure of the mouse α2(IX) collagen gene shows unexpected divergence from the chick gene. J Biol Chem 269: 5064-5071.

Perälä M, Hänninen M, Hästbacka J, Elima K & Vuorio E (1993) Molecular cloning of the human α2(IX) collagen cDNA and assignment of the human COL9A2 gene to chromosome 1. FEBS Lett 319: 177-180.

Perälä M, Savontaus M, Metsäranta M & Vuorio E (1997) Developmental regulation of mRNA species for types II, IX and XI collagens during mouse embryogenesis. Biochem J 324: 209-216.

Pierce JC & Sternberg NL (1992) Using bacteriophage P1 system to clone high molecular weight genomic DNA. Methods Enzymol 216: 549-574.

Pihlajaniemi T & Rehn M (1995) Two new collagen subgroups: membrane-associated collagens and types XV and XVII. Prog Nucleic Acid Res Mol Biol 50: 225-262.

Prockop DJ & Kivirikko KI (1995) Collagens: molecular biology, diseases and potentials for therapy. Annu Rev Biochem 64: 403-434.

Prockop DJ (1998) What holds us together? Why do some of us fall apart? What can we do about it? Matrix Biol 16: 519-528.

Pulkkinen L & Uitto J (1999) Mutation analysis and molecular genetics of epidermolysis bullosa. Matrix Biol 18: 29-42.

Ramchandran R, Dhanabal M, Volk R, Waterman MJ, Segal M, Lu H, Knebelmann B & Sukhatme VP (1999) Antiangiogenic activity of restin, NC10 domain of human collagen XV: comparison to endostatin. Biochem Biophys Res Commun 255: 735-739.

Reese CA & Mayne R (1981) Minor collagens of chicken hyaline cartilage. Biochemistry 20: 5443-5448.

Reese CA, Wiedemann H, Kühn K & Mayne R (1982) Characterization of a highly soluble collagenous molecule isolated from chicken hyaline cartilage. Biochemistry 21: 826-830.

Ren ZX, Brewton RG & Mayne R (1991) An analysis by rotary shadowing of the structure of the mammalian vitreous humor and zonular apparatus. J Struct Biol 106: 57-63.

van der Rest M & Garrone R (1991) Collagen family of proteins. FASEB J 5: 2814-2823.

van der Rest M, Mayne R, Ninomiya Y, Seidah NG, Chretien M & Olsen BR (1985) The structure of type IX collagen. J Biol Chem 260: 220-225.

van der Rest M & Mayne R (1987) Type IX collagen. In: Mayne R & Burgeson RE (eds) Structure and Function of Collagen Types. Academic Press, Orlando, p 195-221.

van der Rest M & Mayne R (1988) Type IX collagen proteoglycan from cartilage is covalently cross-linked to type II collagen. J Biol Chem 263: 1615-1618.

Ricard-Blum S, Hartmann DJ, Herbage D, Payen-Meyran C & Ville G (1982) Biochemical properties and immunolocalization of minor collagens in foetal calf cartilage. FEBS Lett 146: 343-347.

Rokos I, Muragaki Y, Warman M & Olsen BR (1994) Assembly and sequencing of a cDNA covering the entire mouse α1(IX) collagen chain. Matrix Biol 14: 1-8.

Rosenberg K, Olsson H, Mörgelin M & Heinegård D (1998) Cartilage oligomeric matrix protein shows high affinity zinc-dependent interaction with triple helical collagen. J Biol Chem 273: 20397-20403.

Roughley PJ & Lee ER (1994) Cartilage proteoglycans: structure and potential functions. Microsc Res Tech 28: 385-397.

Saarela J, Rehn M, Oikarinen A, Autio-Harmainen H & Pihlajaniemi T (1998) The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. Am J Pathol 153: 611-626.

Sado Y, Kagawa M, Naito I, Ueki Y, Seki T, Momota R, Oohashi T & Ninomiya Y (1998) Organization and expression of basement membrane collagen IV genes and their roles in human disorders. J Biochem (Tokyo) 123: 767-776.

Savontaus M, Ihanamäki T, Perälä M, Metsäranta M, Sandberg-Lall M & Vuorio E (1998) Expression of type II and IX collagen isoforms during normal and pathological cartilage and eye development. Histochem Cell Biol 110: 149-159.

Schnittger S, Herbst H, Schuppan D, Dannenberg C, Bauer M & Fonatsch C (1995) Localization of the undulin gene (UND) to human chromosome band 8q23. Cytogenet Cell Genet 68: 233-234.

Schuppan D, Cantaluppi MC, Becker J, Veit A, Bunte T, Troyer D, Schuppan F, Schmid M, Ackermann R & Hahn EG (1990) Undulin, an extracellular matrix glycoprotein associated with collagen fibrils. J Biol Chem 265: 8823-8832.

Shaw LM & Olsen BR (1991) FACIT collagens: diverse molecular bridges in extracellular matrices. Trends Biochem Sci 16: 191-194.

Shimokomaki M, Duance VC & Bailey AJ (1980) Identification of a new disulphide bonded collagen from cartilage. FEBS Lett 121: 51-54.

Shimokomaki M, Duance VC & Bailey AJ (1981) Identification of two further collagenous fractions from articular cartilage. Biosci Rep 1: 561-570.

Shuttleworth CA (1997) Type VIII collagen. Int J Biochem Cell Biol 29: 1145-1148.

Smith GN Jr. & Brandt KD (1992) Hypothesis: can type IX collagen “glue” together intersecting type II fibers in articular cartilage matrix? A proposed mechanism. J Rheumatol 19: 14-17.

Snellman A, Keränen M-R, Hägg PO, Lamberg A, Hiltunen JK, Kivirikko KI & Pihlajaniemi T (2000) Type XIII collagen forms homotrimers with three triple helical collagenous domains and its association into disulfide-bonded trimers is enhanced by prolyl 4-hydroxylase. J Biol Chem 275: 8936-8944.

Slepecky NB, Cefaratti LK & Yoo TJ (1992) Type II and type IX collagen form heterotypic fibers in the tectorial membrane of the inner ear. Matrix 12: 80-86.

Spector TD, Cicuttini F, Baker J, Loughlin J & Hart D (1996) Genetic influences on osteoarthritis in women: a twin study. Br Med J 312: 940-944.

Spranger J (1976) The epiphyseal dysplasias. Clin Orthop 114: 46-59.

Sugrue SP, Gordon MK, Seyer J, Dublet B, van der Rest M & Olsen BR (1989) Immunoidentification of type XII collagen in embryonic tissues. J Cell Biol 109: 939-945.

Sumiyoshi H, Inoguchi K, Khaleduzzaman M, Ninomiya Y & Yoshioka H (1997) Ubiquitous expression of the α1(XIX) collagen gene (Col19a1) during mouse embryogenesis becomes restricted to a few tissues in the adult organism. J Biol Chem 272: 17104-17111.

Superti-Furga A, Neumann L, Riebel T, Eich G, Steinmann B, Spranger J & Kunze J (1999) Recessively inherited multiple epiphyseal dysplasia with normal stature, club foot, and double layered patella caused by a DTDST mutation. J Med Genet 36: 621-624.

Sutmuller M, Bruijn JA & de Heer E (1997) Collagen type VIII and X, two non-fibrillar, short-chain collagens. Structure homologies, functions and involvement in pathology. Histol Histopathol 12: 557-566.

Svoboda KK, Nishimura I, Sugrue SP, Ninomiya Y & Olsen BR (1988) Embryonic chicken cornea and cartilage synthesize type IX collagen molecules with different amino-terminal domains. Proc Natl Acad Sci USA 85: 7496-7500.

Swiderski RE & Solursh M (1992) Differential co-expression of long and short form type IX collagen transcripts during avian limb chondrogenesis *in ovo*. Development 115: 169-179.

Takahara K, Hoffman GG & Greenspan DS (1995) Complete structural organization of the human α1(V) collagen gene (COL5A1): divergence from the conserved organization of other characterized fibrillar collagen genes. Genomics 29: 588-597.

Tiller GE, Warman ML, Gong Y, Knoll JHM, Mayne R & Brewton RG (1998) Physical and linkage mapping of the gene for the α3 chain of type IX collagen, COL9A3, to human chromosome 20q13.3. Cytogenet Cell Genet 81: 205-207.

Tillet E, Mann K, Nischt R, Pan TC, Chu M-L & Timpl R (1995) Recombinant analysis of human α1(XVI) collagen. Evidence for processing of the N-terminal globular domain. Eur J Biochem 228: 160-168.

Timpl R & Brown JC (1996) Supramolecular assembly of basement membranes. Bioessays 18: 123-132.

Ting K, Petropulos LA, Iwatsuki M & Nishimura I (1993) Altered cartilage phenotype expressed during intramembranous bone formation. J Bone Miner Res 8: 1377-1387.

Tomita M, Kitajima T & Yoshizato K (1997) Formation of recombinant human procollagen I heterotrimers in a baculovirus expression system. J Biochem 121: 1061-1069.

Tomita M, Ohkura N, Ito M, Kato T, Royce PM & Kitajima T (1995) Biosynthesis of recombinant human pro-α1(III) chains in a baculovirus expression system: production of disulphide-bonded and non-disulphide-bonded species containing full-length triple helices. Biochem J 312: 847-853.

Tomita M, Yoshizato K, Nagata K & Kitajima T (1999) Enhancement of secretion of human procollagen I in mouse HSP47-expressing insect cells. J Biochem (Tokyo) 126: 1118-1126.

Trueb J & Trueb B (1992a) The two splice variants of collagen XII share a common 5' end. Biochim Biophys Acta 1171: 97-98.

Trueb J & Trueb B (1992b) Type XIV collagen is a variant of undulin. Eur J Biochem 207: 549-557.

Uitto J & Pulkkinen L (1996) Molecular complexity of the cutaneous basement membrane zone. Mol Biol Rep 23: 35-46.

Vasios G, Nishimura I, Konomi H, van der Rest M, Ninomiya Y & Olsen BR (1988) Cartilage type IX collagen-proteoglycan contains a large amino-terminal globular domain encoded by multiple exons. J Biol Chem 263: 2324-2329.

Vaughan L, Mendler M, Huber S, Bruckner P, Winterhalter KH, Irwin MI & Mayne R (1988) D-periodic distribution of collagen type IX along cartilage fibrils. J Cell Biol 106: 991-997.

Vaughan L, Winterhalter KH & Bruckner P (1985) Proteoglycan Lt from chicken embryo sternum identified as type IX collagen. J Biol Chem 260: 4758-4763.

Vikkula M, Metsäranta M, Syvänen AC, Ala-Kokko L, Vuorio E & Peltonen L (1992) Structural analysis of the regulatory elements of the type-II procollagen gene. Conservation of promoter and first intron sequences between human and mouse. Biochem J 285: 287-294.

Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T & Kivirikko KI (1997) Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. EMBO J 16: 6702-6712.

Vuorela A, Myllyharju J, Pihlajaniemi T & Kivirikko KI (1999) Coexpression with collagen markedly increases the half-life of the recombinant human prolyl 4-hydroxylase tetramer in the yeast *Pichia pastoris*. Matrix Biol 18: 519-522.

Vuoristo MM, Pihlajamaa T, Vandenberg P, Prockop DJ & Ala-Kokko L (1995) The human COL11A2 gene structure indicates that the gene has not evolved with the genes for the major fibrillar collagens. J Biol Chem 270: 22873-22881.

Wälchli C, Koch M, Chiquet M, Odermatt BF & Trueb B (1994) Tissue-specific expression of the fibril-associated collagens XII and XIV. J Cell Sci 107 Pt 2: 669-681.

Wälchli C, Trueb J, Kessler B, Winterhalter KH & Trueb B (1993) Complete primary structure of chicken collagen XIV. Eur J Biochem 212: 483-490.

Wallis GA, Kadler KE, Starman BJ & Byers PH (1992) A tripeptide deletion in the triple-helical domain of the proα1(I) chain of type I procollagen in a patient with lethal osteogenesis imperfecta does not alter cleavage of the molecule by N-proteinase. J Biol Chem 267: 25529-25534.

Walmsley AR, Batten MR, Lad U & Bulleid NJ (1999) Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase. J Biol Chem 274: 14884-14892.

Warman ML, McCarthy MT, Perälä M, Vuorio E, Knoll JHM, McDaniels CN, Mayne R, Beier DR & Olsen BR (1994) The genes encoding α2(IX) collagen (COL9A2) map to human chromosome 1p32.3-p33 and mouse chromosome 4. Genomics 23: 158-162.

Warman M, Kimura T, Muragaki Y, Castagnola P, Tamei H, Iwata K & Olsen BR (1993a) Monoclonal antibodies against two epitopes in the human α1(IX) collagen chain. Matrix 13: 149-156.

Warman ML, Tiller GE, Polumbo PA, Seldin MF, Rochelle JM, Knoll JHM, Cheng SD & Olsen BR (1993b) Physical and linkage mapping of the human and murine genes for the α1 chain of type IX collagen (COL9A1). Genomics 17: 694-698.

Watt SL, Lunstrum GP, McDonough AM, Keene DR, Burgeson RE & Morris NP (1992) Characterization of collagen types XII and XIV from fetal bovine cartilage. J Biol Chem 267: 20093-20099.

Wieringa B, Hofer E & Weissmann C (1984) A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. Cell 37: 915-925.

Wilson R, Lees JF & Bulleid NJ (1998) Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. J Biol Chem 273: 9637-9643.

Wotton SF, Duance VC & Fryer PR (1988) Type IX collagen: a possible function in articular cartilage. FEBS Lett 234: 79-82.

Wright DW & Mayne R (1988) Vitreous humor of chicken contains two fibrillar systems: an analysis of their structure. J Ultrastruct Mol Struct Res 100: 224-234.

Wu J-J & Eyre DR (1989) Covalent interactions of type IX collagen in cartilage. Connect Tissue Res 20: 241-246.

Wu J-J & Eyre DR (1995) Structural analysis of cross-linking domains in cartilage type XI collagen. Insights on polymeric assembly. J Biol Chem 270: 18865-18870.

Wu J-J, Lark MW, Chun LE & Eyre DR (1991) Sites of stromelysin cleavage in collagen types II, IX, X, and XI of cartilage. J Biol Chem 266: 5625-5628.

Wu J-J, Woods PE & Eyre DR (1992) Identification of cross-linking sites in bovine cartilage type IX collagen reveals an antiparallel type II-type IX molecular relationship and type IX to type IX bonding. J Biol Chem 267: 23007-23014.

Yada T, Arai M, Suzuki S & Kimata K (1992) Occurrence of collagen and proteoglycan forms of type IX collagen in chick embryo cartilage. Production and characterization of a collagen form-specific antibody. J Biol Chem 267: 9391-9397.

Yada T, Suzuki S, Kobayashi K, Kobayashi M, Hoshino T, Horie K & Kimata K (1990) Occurrence in chick embryo vitreous humor of a type IX collagen proteoglycan with an extraordinarily large chondroitin sulfate chain and short α1 polypeptide. J Biol Chem 265: 6992-6999.

Yamagata M, Yamada KM, Yamada SS, Shinomura T, Tanaka H, Nishida Y, Obara M & Kimata K (1991) The complete primary structure of type XII collagen shows a chimeric molecule with reiterated fibronectin type III motifs, von Willebrand factor A motifs, a domain homologous to a noncollagenous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site. J Cell Biol 115: 209-221.

Yoshioka H, Zhang H, Ramirez F, Mattei M-G, Moradi-Ameli M, van der Rest M & Gordon MK (1992) Synteny between the loci for a novel FACIT-like collagen locus (D6S228E) and α1(IX) collagen (COL9A1) on 6q12-q14 in humans. Genomics 13: 884-886.