FROM RARE SYNDROMES TO A COMMON DISEASE

Mutations in minor cartilage collagen genes cause Marshall and Stickler syndromes and intervertebral disc disease

SUSANNA ANNUNEN
Department of Medical Biochemistry

OULU 1999
Mutations in minor cartilage collagen genes cause Marshall and Stickler syndromes and intervertebral disc disease

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 December 3rd, 1999, at 12 noon.
Annunen, Susanna, From rare syndromes to a common disease. Mutations in minor cartilage collagen genes cause Marshall and Stickler syndromes and intervertebral disc disease.

Collagen Research Unit, Biocenter Oulu and Department of Medical Biochemistry, University of Oulu, FIN-90220 Oulu, Finland 1999

Oulu, Finland
(Received 1 October 1999)

Abstract

Collagens IX and XI are quantitatively minor components of the collagen fibrils in cartilage. The spectrum of the phenotypes caused by mutations in the COL2A1 gene coding for collagen II, the main cartilage collagen, is relatively well defined, but there is little data on the phenotypes caused by collagen IX and XI mutations.

The structure of the human COL11A1 gene coding for the α1 chain of collagen XI was characterized here. It was found to consist of 68 exons and span 160 kb, excluding introns 1 and 4. Over 50 kb of new intronic sequences were defined. The exon-intron organization coding for the major triple helical domain was found to be identical to that of the human COL11A2 gene, which codes for the α2 chain of collagen XI.

The sensitivity of conformation sensitive gel electrophoresis (CSGE) for mutation detection was improved and tested with a large number of sequence variations in collagen genes. The sensitivity with the revised conditions was found to be close to 100%. In addition, CSGE was found to be a simple and practical method for analyzing large numbers of samples.

Fifteen mutations in the COL11A1 gene and eight in the COL2A1 gene were found by CSGE in patients with Marshall or Stickler syndrome. The genotypic-phenotypic comparison indicated that mutations leading to a premature translation termination codon in the COL2A1 gene resulted in Stickler syndrome and splicing mutations of 54 bp exons in the C terminal half of the COL11A1 gene resulted in Marshall syndrome. The other COL11A1 mutations caused phenotypes overlapping both syndromes.

In an analysis of the COL9A2 gene in 157 patients with intervertebral disc disease, six were found to have a tryptophan for glutamine substitution in the central collagenous domain of the collagen IX molecule. None of 174 control individuals had this substitution. The substitution cosegregated with the phenotype in the families studied, and linkage and linkage disequilibrium analyses supported the association of the locus and the disease with a joint lod score of over 11.

Keywords: mutation detection, sciatica
Acknowledgements

The research presented in this thesis was performed at the Department of Medical Biochemistry, University of Oulu, and Biocenter Oulu, Finland, and at the Center for Gene Therapy, the late Allegheny University of the Health Sciences, Philadelphia, USA, during the last years of the second millennium Anno Domini.

I wish to express my sincere thanks to my supervisor, Docent Leena Ala-Kokko, for always having time to advise me, and sometimes even to help in the practical lab work, for her everlasting optimism, for arranging an opportunity for me to work at the Center for Gene Therapy in Philadelphia for two summers, and most importantly, for guiding me into the field of science. I would also like to thank Professors Kari I. Kivirikko and Taina Pihlajaniemi, the leaders of the Collagen Research Unit, for providing excellent research facilities and a high-standard scientific environment. Moreover, I wish to acknowledge Professor Ilmo Hassinen, whose enormous knowledge and enthusiastic attitude towards science inspired me as a first-year medical student to want to learn and create a piece of the fascinating world of science. I also wish to express my gratitude to Professor Darwin J. Prockop, Director of the Center for Gene Therapy, for giving me the honour of working in his laboratory.

I am grateful to Professor Markku Savolainen and Docent Jaakko Ignatius for their valuable comments on the manuscript of this thesis. I also wish to thank Malcolm Hicks, M.A., for reviewing the language of the manuscripts of this thesis and the original articles II and III.

I am indebted to all my collaborators. The clinicians who examined the patients and took samples made an invaluable contribution to this research. I also learned a lot about medicine from them. Harald H. H. Göring did much invaluable work in calculating the lod scores and trying to make me understand even something of the statistical dimensions.

My friends and colleagues at the Department of Medical Biochemistry deserve my warmest thanks. The fruitful and relaxing scientific and non-scientific conversations and the time spent outside the lab with Jarmo Körkkö, Petteri Paassilta, Tero Pihlajamaa, Jussi and Mirka Vuoristo, Miia Melkonieni, Merja Välkiälä, Merja Periä, Jaana Lohiniva, Veli-Antero Kaakinen and Joni Mäki have been of great value for me. I also wish to thank Tero for teaching me the basics of lab work in the very first steps of this research and Miia for helping me to keep pace with my medical studies. Special thanks
are due to Jarmo and Jaana Körkkö for their invaluable help in numerous things and all the moments of leisure spent with them during my stays in Philadelphia. My sister and colleague Pia Marttila deserves the greatest thanks for much wise advice concerning practical work and the more tortuous sides of scientific life.

I wish to thank Aira Harju, Helena Lindqvist, Jaana Körkkö and Robert Hnatuk for their skilful technical assistance, Pertti Vuokila for ordering all the reagents and Seppo Lähdesmäki for helping with technical problems. I am grateful to Marja-Leena Kivelä and Auli Kinnunen for their always friendly secretarial help. I also wish to acknowledge Ari-Pekka Kvist and Juha Närpänkangas for maintaining the computer systems at the department and helping with the problems concerning computers.

I wish to express my gratitude to my parents for their loving care and support and for giving me an excellent basis for my life. I am also grateful to my sisters, brother, their families and my friends for always being there for me even if I seemed to be constantly in the lab. Above all, I owe my deepest gratitude to my fiancé Markku for all his love, support and understanding and for brightening the dark moments and making the bright moments even brighter.

This work was supported financially by the Emil Aaltonen Foundation and the Finnish Medical Foundation.

Oulu, September 1999
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4Hyp</td>
<td>4-hydroxyproline</td>
</tr>
<tr>
<td>ACGII-HCG</td>
<td>achondrogenesis type II-hypochondrogenesis</td>
</tr>
<tr>
<td>BAP</td>
<td>1,4-bis(acryloyl)piperazine</td>
</tr>
<tr>
<td>bp</td>
<td>basepair</td>
</tr>
<tr>
<td>C</td>
<td>carboxy</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>COL</td>
<td>collagenous domain</td>
</tr>
<tr>
<td>COL(\alpha)</td>
<td>human gene for the (\alpha(x)) chain of collagen (y)</td>
</tr>
<tr>
<td>Col(\alpha)</td>
<td>murine gene for the (\alpha(x)) chain of collagen (y)</td>
</tr>
<tr>
<td>COMP</td>
<td>cartilage oligomeric matrix protein</td>
</tr>
<tr>
<td>CSGE</td>
<td>conformation sensitive gel electrophoresis</td>
</tr>
<tr>
<td>CT</td>
<td>computerized tomography</td>
</tr>
<tr>
<td>DGGE</td>
<td>denaturing gradient gel electrophoresis</td>
</tr>
<tr>
<td>ECM</td>
<td>extracellular matrix</td>
</tr>
<tr>
<td>FACIT</td>
<td>fibril-associated collagen with interrupted triple helices</td>
</tr>
<tr>
<td>GAG</td>
<td>glycosaminoglycan</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>MACIT</td>
<td>membrane-associated collagen with interrupted triple helices</td>
</tr>
<tr>
<td>MED</td>
<td>multiple epiphyseal dysplasia</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MULTIPLEXIN</td>
<td>collagen with multiple triple helix domains and interruptions</td>
</tr>
<tr>
<td>N</td>
<td>amino</td>
</tr>
<tr>
<td>NC</td>
<td>non-collagenous domain</td>
</tr>
<tr>
<td>OA</td>
<td>osteoarthritis or osteoarthrosis</td>
</tr>
<tr>
<td>OMIM</td>
<td>Online Mendelian Inheritance in Man</td>
</tr>
<tr>
<td>OSMED</td>
<td>otospondylomegaepiphyseal dysplasia</td>
</tr>
<tr>
<td>PSACH</td>
<td>pseudoachondroplasia</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase PCR</td>
</tr>
<tr>
<td>SED</td>
<td>spondyloepiphyseal dysplasia</td>
</tr>
<tr>
<td>SSCP</td>
<td>single-strand conformation polymorphism</td>
</tr>
<tr>
<td>WZS</td>
<td>Weissenbacher-Zweymüller syndrome</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>X</td>
<td>any amino acid</td>
</tr>
<tr>
<td>Y</td>
<td>any amino acid</td>
</tr>
</tbody>
</table>
List of original articles

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


In addition, some unpublished data are presented.
Contents

Abstract
Acknowledgements
Abbreviations
List of original articles
Contents
1. Introduction..................................................................................................................13
2. Review of the literature..............................................................................................15
  2.1. Collagens..............................................................................................................15
    2.1.1. Fibrillar collagens.........................................................................................15
    2.1.2. Non-fibrillar collagens.................................................................................16
    2.1.3. Collagen biosynthesis..................................................................................18
  2.2. Cartilage................................................................................................................18
    2.2.1. Collagenous components of cartilage.......................................................19
    2.2.2. Non-collagenous components of cartilage ...............................................22
  2.3. Cartilage collagens in non-cartilaginous tissues ...................................................23
    2.3.1. Inner ear ......................................................................................................23
    2.3.2. Ocular vitreous body ...................................................................................23
    2.3.3. Intervertebral disc.......................................................................................23
  2.4. Disorders affecting cartilage collagens .................................................................24
    2.4.1. Achondrogenesis type II-hypochondrogenesis (ACGII-HCG) ...............25
    2.4.2. Spondyloepiphyseal dysplasia (SED)..........................................................25
    2.4.3. Kniest dysplasia...........................................................................................26
    2.4.4. Stickler and Marshall syndromes .................................................................26
    2.4.5. Otospondylomegaepiphyseal dysplasia (OSMED), Weissenbacher-
          Zwemüller syndrome (WZS) and non-ocular Stickler-like syndrome ......28
    2.4.6. Multiple epiphyseal dysplasia (MED)..........................................................29
    2.4.7. Osteoarthritis (OA).....................................................................................30
  2.5. Animal models for disorders of the cartilage collagens .......................................30
    2.5.1. Collagen II...................................................................................................31
    2.5.2. Collagen IX..................................................................................................31
    2.5.3. Collagen XI..................................................................................................32
  2.6. Intervertebral disc disease.....................................................................................32
1. Introduction

Cartilage is an important form of supportive tissue. It is rich in extracellular matrix and contains only a sparse population of chondrocytes. The most abundant type of cartilage in the body is hyaline cartilage, which forms articular cartilage, growth plates and the cartilaginous anlage for endochondral ossification. The extracellular matrix of hyaline cartilage is highly hydrated and comprises cartilage-specific collagen fibrils and proteoglycans.

Collagen II, the main component of the collagen fibrils in cartilage, is a homotrimer of three identical α chains, α1(II)3, which are encoded by the COL2A1 gene. While collagen II accounts for the majority of the collagens in cartilage, other significant types are collagens IX and XI. Collagen IX is a heterotrimer of three dissimilar α chains, α1(IX), α2(IX) and α3(IX), which are encoded by the COL9A1, COL9A2 and COL9A3 genes, respectively. Collagen IX is known to be located on the surface of the cartilage collagen fibrils, but its precise function is not currently known, although it has been suggested that it may mediate the interaction between collagenous and non-collagenous cartilage components and maintain the homeostasis of cartilage. Collagen XI is likewise a heterotrimer of three distinct α chains, the α1(XI), α2(XI) and α3(XI) chains being encoded by the COL11A1, COL11A2 and COL2A1 genes, respectively. Like collagen II, collagen XI forms fibrils, and they are thought to be buried in collagen II fibrils. It has been suggested that the function of collagen XI may be to regulate collagen fibril diameter.

Collagens II, IX and XI are also expressed in the intervertebral disc, which consists of an outer ring-like structure called the annulus fibrosus and an inner gel-like matrix called the nucleus pulposus. Collagen II is the major collagenous component of the nucleus pulposus, while collagen IX is found in minor amounts throughout the disc and collagen XI accounts for a few percent of the collagens in the nucleus pulposus. In addition to hyaline cartilage and the intervertebral disc, collagens II, IX and XI are also expressed in the ocular vitreous body and the tectorial membrane of the inner ear.

Collagen II is the most intensively studied collagen type in cartilage, and a large number of mutations in the COL2A1 gene have already been characterized. Depending on the type of mutation, they cause a spectrum of phenotypes ranging from lethal achondrogenesis type II-hypochondrogenesis to early onset osteoarthritis with mild chondrodysplasia. Mutations characterized in the genes coding for collagen IX lead to
multiple epiphyseal dysplasia, and mutations in the genes coding for collagen XI lead to Stickler and Marshall syndromes and heterozygous and homozygous otospondylo-megaepiphyseal dysplasia. The number of mutations in these genes known at present is small, however, and the spectrum of phenotypes caused by gene mutations in collagens IX and XI is largely unknown.

Characterization of the genetic defects lying behind such phenotypes is important for obtaining an understanding of the function of the particular molecule and of the whole tissue. Furthermore, knowing the molecular defect that causes the disorder in question helps in the development of a treatment for the disease and the prevention of symptoms, and on the other hand, enables defining the exact genetic defect involved and genetic prenatal diagnosis.

The goal of this thesis is to increase our knowledge of the role of collagens IX and XI in disorders affecting cartilaginous tissues. It involves characterization of the genomic structure of the COL11A1 gene, the description of 23 mutations in the COL11A1 and COL2A1 genes causing Marshall and Stickler syndromes, definitions of the genotypic and phenotypic relations of these syndromes, and a report of an association between an allele of the COL9A2 gene bearing a putative mutation and intervertebral disc disease. In addition, an improvement to the conformation sensitive gel electrophoresis (CSGE) method for mutation detection is described.
2. Review of the literature

2.1. Collagens

Collagens are a significant component of the extracellular matrix (ECM). They are a group of related proteins that share common structural characteristics. Collagens are trimeric molecules consisting of three similar or dissimilar $\alpha$ chains, and they form supramolecular structures in the ECM. All collagens have at least one triple helical region in the molecule. The formation of triple helix requires the presence of glycine as every third amino acid, because any bulkier amino acid would cause steric hindrance to helix formation. So far, 19 collagen types encoded by over 30 different genes are known. These can be divided into several groups based on their structural and other properties (For reviews, see Vuorio & de Crombrugghe 1990, Kielty et al. 1993, Mayne & Brewton 1993, Pihlajaniemi & Rehn 1995, Prockop & Kivirikko 1995.)

2.1.1. Fibrillar collagens

Collagens I, II, III, V and XI aggregate into rod-like fibrils in a quarter-staggered array that gives a characteristic banding pattern. They contain a large collagenous domain of about 1000 amino acids. (See Kielty et al. 1993, Pihlajaniemi & Rehn 1995, Prockop & Kivirikko 1995.) Collagens I, II and III are known as the major fibrillar collagens, since they are quantitatively more significant than collagens V and XI, which are often called minor fibrillar collagens.

Collagen I, II and III fibrils create the structural frameworks for many tissues, such as bone, skin, cartilage, tendon and blood vessels. Collagen I is a heterotrimer of two $\alpha1(I)$ chains and one $\alpha2(I)$ chain that are encoded by the COL1A1 and COL1A2 genes, respectively. Homotrimers of $\alpha1(I)$ have also been observed (Uitto 1979). Collagen I is expressed in various tissues, but it is the main structural component of bone, skin, tendon, ligament, teeth and fasciae. Collagen II is a homotrimer of $\alpha1(II)$ chains that are encoded by the COL2A1 gene. In adult tissues it is expressed in cartilage, vitreous humour, the intervertebral disc and the inner ear (for more details on collagen II, see section 2.2.1.1.).
Collagen III is a homotrimer as well, consisting of three α1(III) chains that are encoded by the COL3A1 gene and being expressed in various tissues, although it is most abundant in elastic tissues such as the skin, blood vessels, gut and lung. The major fibrillar collagens share highly similar characteristics. They are synthesized as large precursors with C and N propeptides, which are later cleaved by specific C and N proteinases. Also, all the genes coding for the α chains of major fibrillar collagens are similar in their genomic structure. The numbers and the sizes of the exons are the same, with only a few exceptions. (See Vuorio & Crombrugghe 1990, Chu & Prockop 1993, Prockop & Kivirikko 1995.)

Collagens V and XI share many properties with the major fibrillar collagens, although they also have some features that are different. The exon sizes and numbers of exons in the genes for the minor fibrillar collagens characterized so far, the COL5A1 and COL11A2 genes, are different from those of the genes that code for the major fibrillar collagens (Takahara et al. 1995, Vuoristo et al. 1995). While the entire propeptides are cleaved during post-translational modification in the major fibrillar collagens, the N propeptides of collagens V and XI seem to be partially retained in the mature molecule. Collagens V and XI form heterotypic fibrils with collagens I and II, respectively, and they are thought to be located inside collagen I/II fibrils (Birk et al. 1988, Mendler et al. 1989). It has been suggested that they regulate fibril diameter and thus have a significant role in fibril formation. Collagen V is codistributed with collagen I, and is composed of three different α chains, α1(V), α2(V) and α3(V), encoded by the COL5A1, COL5A2 and COL5A3 genes. The most common composition for the collagen V molecule is [α1(V)]₂α2(V), but others also exist. Collagen XI usually consists of α1(XI), α2(XI) and α3(XI) chains, coded by the COL11A1, COL11A2 and COL2A1 genes (for more details on collagen XI, see section 2.2.1.3.). Collagen V and XI α chains can also appear in the same heterotypic molecule, e.g. the α2(V) chain substitutes for α2(XI) in the vitreous body of the eye. It has therefore been suggested that collagens V and XI could be considered a single type, collagen V/XI. (See Fichard et al. 1994.)

### 2.1.2. Non-fibrillar collagens

Of the currently known collagen types, collagens IV, VI-X and XII-XIX are the ones that do not form fibrils. According to their structural properties, they can be further classified as network-forming collagens, beaded filament-forming collagen, collagen of anchoring fibrils, fibril-associated collagens with interrupted triple helices (FACITs), membrane-associated collagens with interrupted triple helices (MACITs) and collagens with multiple triple helix domains and interruptions (MULTIPLEXINs) (see Pihlajaniemi & Rehn 1995, Prockop & Kivirikko 1995).

Collagens IV, VIII and X form network-like structures. Collagen IV has a large triple helical domain that is interrupted by numerous short non-collagenous sequences, which confer flexibility on the individual molecule. This allows aggregation of the monomers to form the network-like structures that are found in basement membranes. The structures of collagens VIII and X are very similar to each other but differ from that of collagen IV. The triple helical domains of collagens VIII and X are only about 450 amino acids in
length and these types assemble into hexagonal lattices. The genes that code for collagens VIII and X have an atypical structure, since they consist of only three exons, the third of which codes for almost the entire protein. Collagen VIII is expressed primarily in Descemet’s membrane in the eye, and collagen X is a highly specialized collagen of hypertrophic chondrocytes in the deep-calcifying zone of cartilage. (See Kielty et al. 1993, Pihlajaniemi & Rehn 1995, Prockop & Kivirikko 1995.)

Collagen VI molecules form a unique structure of beaded filaments, which are found in most connective tissues. It is located in the close vicinity of cells, large collagen fibrils and some basement membranes, and may thus function as an anchorage between these structures and the surrounding ECM. (See Chu et al. 1990.)

Collagen VII is the major component of the anchoring fibrils that attach the basement membrane to the stroma of the tissues, subject to external frictional forces such as skin, oral mucosa and cervix (see Burgeson et al. 1990, Kielty et al. 1993). The FACITs (IX, XII, XIV, XVI, XIX) do not form fibrils themselves, but are found to be associated with fibrillar collagens. They are not secreted as procollagens like the fibrillar collagens are, and thus do not undergo proteolytic processing after secretion. (See Shaw & Olsen 1991, Mayne & Brewton 1993, Prockop & Kivirikko 1995.)

Collagen IX, the best studied member of the FACIT group, is expressed in the same tissues as collagen II and is covalently crosslinked to the surface of collagen II fibrils (Eyre et al. 1987, Brewton & Mayne 1994, Diab et al. 1996; for more details on collagen IX, see chapter 2.2.1.2.). Collagens XII and XIV bear structural similarities to collagen IX, but their functions are unclear, although it has been suggested that they are associated with collagen I fibrils. There is also evidence of collagen XII and XIV expression in tissues from which collagen I is absent, however. Collagens XVI and XIX have also been classified as FACITs in view of their homology with the other FACITs. (See Mayne & Brewton 1993.) Even though the exact function of the FACITs is unclear, it has been suggested that they may serve as molecular bridges and thus have a role in ECM organization and integrity (see Shaw & Olsen 1991).

Collagens XIII and XVII are not homologous in structure but have a transmembrane domain, and are therefore called MACITs. The mRNA for collagen XIII can be found in a variety of tissues, but the function of the protein has not yet been defined. Even though alternative splicing is relatively common in collagens, that found in collagen XIII is of an unusual complexity, affecting the sequences for ten exons. Collagen XVII is expressed in hemidesmosomes of the skin. Autoantibodies against it have been identified in cases of bullous pemphigoid, an autoimmune disease that causes blistering, suggesting an anchoring function for this collagen. (See Pihlajaniemi & Rehn 1995.)

Collagens XV and XVIII consist of multiple triple helical domains and interruptions and are thus called MULTIPLEXINs. They are homologous proteins that are expressed in a variety of tissues, with overlapping but not similar distributions (Kivirikko et al. 1995, Saarela et al. 1998). A fragment of the C terminus of collagen XVIII is proteolytically cleaved to form endostatin, which is known to prevent angiogenesis (O’Reilly et al. 1997).
2.1.3. Collagen biosynthesis

After the transcription and translation of collagen mRNA, the newly synthesized polypeptide undergoes post-translational modifications both intracellularly, in the endoplasmic reticulum, and extracellularly. The biosynthesis of collagen is a complex multistage process that requires numerous enzymes. Following cleavage of the signal peptide, prolyl residues in X-Pro-Gly sequences are hydroxylated to hydroxyprolines, lysyl residues in X-Lys-Gly sequences to hydroxylysines, and prolyl residues in Pro-4Hyp-Gly sequences to hydroxyprolines by the enzymes prolyl 4-hydroxylase, lysyl hydroxylase and prolyl 3-hydroxylase, respectively. Prolyl 4-hydroxylase plays a central role in collagen biosynthesis, since 4-hydroxyproline is essential for triple helix stability at body temperature. Some of the hydroxylysyl residues are further glycosylated with galactose or galactose and glucose by hydroxylysyl galactosyl transferase or galactosyl hydroxyllysyl glucosyl transferase, respectively. The formation of the triple helix proceeds from the C terminus to the N terminus. The initial interaction between the non-collagenous domains of the \( \alpha \) chains leads to nucleation of the C termini of the collagenous domains and then to the propagation of triple helix formation in a zipper-like manner. Correct chain selection is directed by molecular recognition sequences in the C terminal non-collagenous domains (Lees et al. 1997). In addition, interchain disulphide bonds in the C terminal non-collagenous domains, catalyzed by protein disulphide isomerase (PDI), stabilize the chain assembly, but they are not essential for triple helix formation (Bulleid et al. 1996, see also Bulleid 1996). The correctly folded procollagen molecules are secreted through the Golgi complex. (See Kielty et al. 1993, Kivirikko 1995, Prockop & Kivirikko 1995.)

The C and N propeptides are cleaved from the procollagen molecules by specific C and N proteinases in the extracellular space, removal of the propeptides being a requirement for normal fibril formation. The cleavage of the propeptides converts the soluble procollagen molecules to insoluble collagen molecules that are able to self-assemble and form fibrils spontaneously. The fibril-forming collagens undergo this cleavage through the action of C and N proteinases, but many of the non-fibrillar collagens retain the C and N terminal non-collagenous domains in the mature molecule. After aggregation of the collagen molecules into fibrils, covalent cross-links are formed to provide the supramolecular structures with tensile strength and mechanical stability. The cross-links are formed either by disulphide bonds between cysteine residues or by lysyl and hydroxylysyl aldehydes that are formed in the reaction catalyzed by lysyl oxidase. (See Kielty et al. 1993, Kivirikko 1995.)

2.2. Cartilage

Cartilage is hyperhydrated supportive tissue that does not contain blood vessels, nerves or lymphatics. It consists of about 80% water by weight, and the remaining 20% is mainly collagen II and proteoglycans. Cartilage is rich in ECM, which is produced by chondrocytes of mesenchymal origin. There are three types of cartilage. Hyaline cartilage is the prototypic manifestation that forms articular cartilage, the cartilaginous anlage of
developing bones, the epiphyseal plates, the costochondral cartilages and the cartilages of
the respiratory tract. Fibrocartilage resembles hyaline cartilage but contains collagen I
fibres for tensile and structural strength. It is found in the annulus fibrosus of the
intervertebral disc, tendinous and ligamentous insertions, menisci, the symphysis pubis
and the insertions of joint capsules. Elastic cartilage constitutes the pinna of the ears, the
epiglottis and the arytenoid cartilages of the larynx. (See Schiller 1994, Muir 1995.)

2.2.1. Collagenous components of cartilage

Collagens provide the structural framework for hyaline cartilage. Collagen II, which
constitutes 80-85% of the total collagen content of cartilage, forms fibrils that contain
small amounts of collagens IX and XI. The quantities of these minor cartilage collagens
range from 1% to 10%, depending on the cartilage source and age. Small amounts of
collagens VI, XII and XIV have also been identified in cartilage. Expression of collagen
X is restricted to hypertrophic chondrocytes during endochondral ossification. (See

2.2.1.1. Collagen II

The COL2A1 gene that codes for the α chains of homotrimeric collagen II consists of 54
exons and is about 31 kb in size (Ala-Kokko & Prockop 1990, Ala-Kokko et al. 1995).
Its chromosomal location is 12q13.11-q13.12 (Takahashi et al. 1990). Besides its
expression in cartilage, the ocular vitreous body, intervertebral discs and the inner ear,
collagen II can be found in many non-chondrogenic tissues during development (see
Brewton & Mayne 1994). The sequences coding for the N terminal propeptide contain an
alternatively spliced exon that codes for a 69-amino acid cysteine-rich domain (Ryan &
Sandell 1990). Collagen II molecules including (IIA) or excluding this domain (IIB) have
distinct distributions in the stages of chondrogenesis, type IIA predominating in the pre-
chondrogenic mesenchyme and differentiating chondrocytes, and type IIB in

The collagen II molecules in the fibril overlap with each other by a distance of about a
quarter of their length, thus forming a banded fibril. The molecules are covalently cross-
linked between the triple helical domain and the N or C terminal telopeptides of adjacent
collagen II molecules, i.e. the short non-collagenous regions that are retained in
propeptide cleavage. The particular molecular properties that serve to adapt the collagen
for use in cartilage are unclear, but it has been suggested that the high hydroxylysine
content of the molecule results in a highly hydrated fibril suitable for cartilage. (See Eyre
Collagen IX is a heterotrimeric molecule composed of three separate gene products, \( \alpha_1(IX), \alpha_2(IX) \) and \( \alpha_3(IX) \) chains. Like collagen II, it is expressed in cartilage, the ocular vitreous body, intervertebral discs and the inner ear (Ayad et al. 1982, Slepecky et al. 1992b, see also Brewton & Mayne 1994). Collagen IX belongs to the group of FACITs and consists of three collagenous (COL1-COL3) and four non-collagenous (NC1-NC4) domains numbered from the C terminus of the molecule. The length of the molecule is about two-thirds of that of the collagen II molecule. (See Kiely et al. 1993.) The molecule is covalently crosslinked to the surface of collagen II fibrils in an antiparallel manner; the N terminal regions of the COL2 domains of all three chains being crosslinked to the N telopeptide of \( \alpha_1(II) \) and the central region of the COL2 domain of \( \alpha_3(IX) \) chain to the C telopeptide of \( \alpha_1(II) \) (Eyre et al. 1987, Vaughan et al. 1988, Wu et al. 1992, Diab et al. 1996). In addition, collagen IX molecules have been shown to be crosslinked to one another between the N terminal COL2 domain of the \( \alpha_1(IX) \) or \( \alpha_3(IX) \) chain and the C terminal NC1 domain of the \( \alpha_3(IX) \) chain (Wu et al. 1992, Diab et al. 1996). The length of the NC3 domain is different in all three \( \alpha \) chains, thus forming a kink in the molecule (McCormick et al. 1987). When the COL2 domain is attached covalently to the surface of collagen II fibrils, the kink allows the COL3 and NC4 domains to project away from the fibril surface (Bruckner et al. 1988, Vaughan et al. 1988).

Collagen IX can also be classified as proteoglycan, because the molecule contains a glycosaminoglycan (GAG) side chain (Bruckner et al. 1985, Bruckner et al. 1988). The attachment site for the chondroitin sulphate side chain is located in the NC3 domain of the \( \alpha_2(IX) \) chain (Huber et al. 1986, McCormick et al. 1987). The proportion of molecules with glycosaminoglycan side chain substitution varies depending on the tissue and its source, and the function of the glycosaminoglycan side chain is not known (Yada et al. 1990, Diab et al. 1996, see also Brewton & Mayne 1994).

The \( \alpha \) chains of collagen IX are encoded by the COL9A1, COL9A2 and COL9A3 genes. The COL9A2 and COL9A3 genes consist of 32 exons, while the COL9A1 gene consists of 38 exons. The six additional exons of the COL9A1 gene code for the 243-amino-acid globular NC4 domain that is formed only from the \( \alpha_1(XI) \) chain. There is an alternative promoter and second transcription start site in the COL9A1 gene, giving rise to a short form of the \( \alpha_1(XI) \) chain and the whole collagen IX molecule. In the short form the alternative exon 1 in intron 6 is used and exon 7 is skipped, thus splicing the alternative exon 1 straight to exon 8 of the COL9A1 gene (Nishimura et al. 1989). The short form exists primarily in the eye. (See Brewton & Mayne 1994.) The exon-intron structures of the corresponding regions of these genes are highly similar, but the genes are very different in size. The COL9A1 gene spans about 90 kb, while the COL9A2 and COL9A3 genes are only about 15 and 23 kb, respectively. (Pihlajamaa et al. 1998a, Paasnilta et al. 1999b.) The chromosomal locations of the genes are 6q12-q13 (COL9A1), 1p33-p32.2 (COL9A2) and 20q13.3 (COL9A3) (Warman et al. 1993b, Warman et al. 1994, Tiller et al. 1998).

The function of collagen IX is not currently known. The location of collagen IX molecules on the surface of collagen II and the projection of the COL3 and NC4 domains out from the surface of the fibril suggest a role in the intermolecular organization of
cartilage. The pI of the NC4 domain is about 10, thus making the domain a candidate for interaction with the highly anionic glycosaminoglycans of cartilage (Vasios et al. 1988). There is also evidence of a role of collagen IX in the maintenance of cartilage integrity (Fässler et al. 1994, Hagg et al. 1998).

### 2.2.1.3. Collagen XI

Collagen XI, originally called $\alpha_2\alpha_3\alpha$ collagen, is similarly a heterotrimer, and is quantitatively a minor member of the group of fibrillar collagens (Morris & Bächinger 1987). Like collagens II and IX, collagen XI is expressed in cartilage, the ocular vitreous body and the nucleus pulposus (see Eyre & Wu 1987, Brewton & Mayne 1994), but the question of its occurrence in the inner ear is controversial (Slepecky et al. 1992b, Thalmann 1993). Furthermore, the mRNAs for all the $\alpha$(XI) chains can be found in equal or unequal amounts in a variety of tissues (Lui et al. 1995). The molecule consists of a major triple helical region of about 1000 amino acids (COL1) that is flanked by a non-collagenous C terminal region (NC1) and an N terminal region that consists of a minor triple helical region (COL2) flanked by non-collagenous domains (NC2 and NC3). Collagens V and XI seem to resemble each other closely in their structural and biological properties. In addition, their $\alpha$ chains can substitute for each other, so that a fraction of $\alpha_1$(XI) in cartilage has been replaced by $\alpha_1$(V) (see Eyre & Wu 1987) and a fraction $\alpha_1$(V) in bone has been replaced by $\alpha_1$(XI) (Niyibizi & Eyre 1989). Furthermore, the ocular vitreous body contains the $\alpha_1$(XI) and $\alpha_2$(V) chains but no $\alpha_2$(XI) chain (Mayne et al. 1993). These findings have led to the suggestion that collagens V and XI should be classified as a single collagen type, collagen V/XI. (See Fichard et al. 1994.)

The $\alpha_1$(XI) and $\alpha_2$(XI) chains are products of the COL11A1 and COL11A2 genes, respectively, while the $\alpha_3$(XI) chain is an overglycosylated form of the IIB splicing variant product of the COL2A1 gene that also codes for the $\alpha$ chains of collagen II (Eyre & Wu 1987, Wu & Eyre 1995). The COL11A1 gene is located in the p21 region of chromosome 1 (Henry et al. 1988). The cDNA for it has been cloned, but only a minor fraction of the exon-intron organization has been elucidated (Bernard et al. 1988, Zhidkova et al. 1995). On the other hand, the genomic organization of the COL11A2 gene has been thoroughly defined. It spans 28 kb and consists of 66 exons. Even though it shares some properties of the major fibrillar collagens, there are marked differences in the number of exons, exon sizes and codon usage, indicating that the gene has not evolved with those for the major fibrillar collagens. (Vuoristo et al. 1995.) The COL11A2 gene is located at 6p21.2 (Kimura et al. 1989) or 6p21.3 (Hanson et al. 1989).

Both the $\alpha_1$(XI) and $\alpha_2$(XI) chains undergo alternative splicing in the N terminal propeptide. The transcripts for the $\alpha_1$(XI) chain contain either exon IIA, coding for 39 amino acids, IIB coding for 51 amino acids, or neither of them, but not both. In addition, exon IV, coding for 85 amino acids, may or may not be present in the mRNA. It is notable that exons IIA, III and IV are highly acidic, while exon IIB is highly basic. The alternative splicing of the $\alpha_2$(XI) chain involves exons 6, coding for 26 amino acids, and 7, coding for 21 amino acids. Furthermore, Lui et al. (1996) suggest additional splicing of sequences in intron 6 of the transcript to the mRNA, yielding a truncated $\alpha_2$(XI) chain.
because of the translation termination codons in the sequences. (Zhidkova et al. 1995.)

The N terminal domain of the α2(XI) chain has also been shown to give rise to proline/arginine-rich protein (PARP) in cartilage (Neame et al. 1990, Zhidkova et al. 1993).

Collagen XI molecules are covalently cross-linked primarily to each other in head to tail organization. The cross-links occur between the N telopeptides and the C terminal ends of the helical regions. In addition, the N terminal end of the α1(XI) chain is attached to the C telopeptide sequence of the α3(XI) or α1(II) chain, presumably to α1(II). These data indicate that collagen XI may form primarily homopolymers, which are also capable of cofibrillar assembly with collagen II. (Wu & Eyre 1995.) The coexistence of collagens II, IX and XI in the cartilage collagen fibrils and the finding that collagen XI is immunologically masked has led to the suggestion that collagen XI is buried inside the fibril (Mendler et al. 1989). Its hypothesized function is regulation of fibril diameter. This is supported by the finding that collagen XI is restricted to thin fibrils and that mouse and human defective collagen XI occurs in abnormal thick, disorganized fibrils (Seegmiller et al. 1971, Keene et al. 1995, Li Y et al. 1995, van Steensel et al. 1997). Partial retention of the N propeptide has been proposed as the molecular basis for this function. (See Fichard et al. 1994.)

2.2.2. Non-collagenous components of cartilage

While collagens form the structural network for cartilage, proteoglycans (also called mucopolysaccharides) provide hydration and swelling pressure, enabling the cartilage to withstand compressional forces. Proteoglycans are proteins that have one or more GAG side chain, which may be chondroitin sulphate, dermanan sulphate, keratan sulphate, heparan sulphate or heparin. (See Hardingham & Fosang 1992, Roughley & Lee 1994.) The largest and most abundant proteoglycan in cartilage is aggrecan, which consists of a multidomain core protein of over 200 kDa bearing over a hundred chondroitin and keratan sulphate side chains in a bottle brush manner. The core protein of an aggrecan molecule attaches to hyaluronan by non-covalent binding, which is stabilized by link protein, thus forming huge multimolecular aggregates. (See Neame & Barry 1993, Hardingham & Fosang 1995.) Many other proteoglycans, such as decorin, biglycan, fibromodulin, lumican and PRELP (proline/arginine-rich end leucine-rich repeat protein), have also been characterized in cartilage. These have relatively small core proteins and only a few GAGs. While aggrecan is mainly responsible for the hydration of cartilage, other proteoglycans in cartilage have distinct biological functions apart from their hydrodynamic functions, e.g. regulation of collagen fibrillogenesis or cell adhesion and migration. (Grover et al. 1995, Bengtsson et al. 1995, Kinsella et al. 1997, see also Stanescu 1990, Roughley & Lee 1994.)

In addition to link protein, there are other non-collagenous proteins in cartilage that do not belong to the proteoglycans, e.g. cartilage oligomeric matrix protein (COMP), anchorin and chondrocalcin in cartilage (see Heinegård & Oldberg 1989).
2.3. Cartilage collagens in non-cartilaginous tissues

Besides cartilage, cartilage collagens are significant structural components of the tectorial membrane of the inner ear, the ocular vitreous body and intervertebral discs.

2.3.1. Inner ear

Collagen II, which is the most studied cartilage collagen in the inner ear, has been found in various tissues there, such as the tectorial, basilar and vestibular membranes and the cartilaginous rest of the endochondral layer (Thalmann 1986, Slepecky et al. 1992b, Thalmann 1993). Thalmann et al. (1987) have shown that collagen accounts for approximately 40% of the total protein in the tectorial membrane of the guinea pig. This membrane lies over the organ of Corti, and the stereocilia of the outer hair cells are embedded in its gel-like matrix (see Moore 1992a). The co-localization of collagens II and IX in the thick, unbranched radial fibres of the tectorial membrane is well demonstrated, but there is some inconsistency in reports with regard to the expression of collagens XI and V in the tectorial membrane and in the inner ear generally (Slepecky et al. 1992a, Slepecky et al. 1992b, Thalmann 1993).

2.3.2. Ocular vitreous body

The ocular vitreous body is a gel-like structure consisting primarily of hyaluronan and collagen fibrils (Ren et al. 1991, Brewton & Mayne 1992). The collagenous component comprises collagen II, collagen IX and collagen resembling types V and XI (Swann & Sotman 1980, Ayad et al. 1982, Ayad & Weiss 1984, Seery & Davison 1991). Further studies by Mayne et al. (1993) demonstrated that the chains of the collagen V/XI fraction of bovine vitreous body are α1(XI) and α2(V). Since neither of these is known to exist in homotrimmers (see Fichard et al. 1994), it is highly probable that they form heterotypic trimers either with or without the α1(II) chain. Collagen IX in the vitreous lacks the NC4 domain, suggesting utilization of the alternative downstream promoter of the COL9A1 gene (Ren et al. 1991, Warman et al. 1993a). (See Brewton & Mayne 1994, Bishop 1996.)

2.3.3. Intervertebral disc

An intervertebral disc stabilizes the spine by anchoring adjacent vertebral bodies to each other and simultaneously absorbs the compressive forces directed towards the spine. While stabilization of the spine requires firmness, resilience is essential for shock absorption. To fulfil both of these tasks, the intervertebral disc comprises concentrically arranged compartments. The outermost structure is a ring-like outer annulus fibrosus,
which consists of highly oriented, densely packed collagen fibril lamellae, inside which is the fibrocartilaginous inner annulus fibrosus. There is a transition zone separating the central nucleus pulposus, a gel-like structure, from the layers of the annulus fibrosus, and there are no clear boundaries between the structures of the intervertebral disc. The cartilaginous endplates between the vertebral bodies and the intervertebral disc are often considered part of the intervertebral disc. (See Humzah & Soames 1988, Moore 1992b, Buckwalter 1995.)

Collagens I and II are the predominant collagens in the intervertebral disc. Collagen II comprises almost all of the total collagen in the nucleus pulposus, the amount of which decreases towards the outer structures so that it is absent or almost absent in the outer annulus fibrosus, while collagen I comprises almost all of the total collagen composition of the outer annulus fibrosus and decreases to absent or almost absent in the nucleus pulposus (Eyre & Muir 1977, Buckwalter 1995, Nerlich et al. 1998). Many reports indicate the presence of collagen IX in the intervertebral disc, even though Roberts et al. (1991) failed to immunolocalize it in the normal human disc (Ayad et al. 1982, Newall & Ayad 1995, Nerlich et al. 1998). Collagen IX has been found in all the compartments of the intervertebral disc, although it comprises about 2% of the total collagen or even less. Both short and long forms of collagen IX can be found, the short form predominating in the nucleus pulposus (Newall & Ayad 1995). By contrast, collagen XI has been found only in the nucleus pulposus, where it comprises about 3% of total collagen.

Proteoglycans, primarily aggrecan, are a significant component of intervertebral discs accounting for a few percent of the dry weight of the annulus fibrosus, but about 50 percent of the nucleus pulposus. (See Buckwalter 1995.)

### 2.4. Disorders affecting cartilage collagens

Numerous conditions that affect the components of cartilage have been characterized. The development and maintenance of cartilage can be affected at several stages in the life-span: the defects can already occur during embryogenesis, in the early steps of mesenchymal differentiation, or only after middle age. The mechanisms of the faulty processes are various. Transcription factors, growth factors, growth factor receptors, structural proteins or enzymes involved in their modification can be defective, or else the process can be due to inflammatory autoimmune disease, trauma or simply wear and tear. (See Horton & Hecht 1993, Mundlos & Olsen 1997a, Mundlos & Olsen 1997b, Kuivaniemi et al. 1997, Cremer et al. 1998.) Only disorders affecting cartilage collagens are discussed in the following.

In the case of a cartilage collagen defect, the symptoms may appear in all or some of the tissues in which the collagens are expressed. Probably the most common symptom is destruction of the hyaline cartilage of joints. The development of bony structures such as the spine, long bones and facial bones is often defective due to the disruption of endochondral ossification. Interestingly, the eye and ear can be severely affected: high myopia, retinal detachments and hearing deficits are usual manifestations of cartilage collagen defects. (See Spranger et al. 1994, Mundlos & Olsen 1997b.)
2.4.1. Achondrogenesis type II-hypochondrogenesis (ACGII-HCG)

Initially, achondrogenesis type II (Langer-Saldino type) and hypochondrogenesis were considered two separate disorders (Langer et al. 1969, Saldino 1971, Maroteaux et al. 1983), but after individuals with an intermediate phenotype had been characterized they often came to be regarded as different degrees of severity of the same disorder (Borochowitz et al. 1986). Infants with ACGII-HCG (OMIM #200610) die perinatally or within the first weeks of life. The disorder is characterized by a short, barrel-shaped trunk, very short extremities, a large head, a soft cranium, a flat face and hydropic appearance. Radiographically, the affected infants show varying degrees of underossification of the axial skeleton. Histological and electron microscopic examinations reveal hypercellular epiphyseal cartilage with poorly organized or absent growth plate and diminished extracellular matrix that contains thick, irregular collagen fibrils together with large chondrocytes with a dilated rough endoplasmic reticulum. Biochemical studies of hyaline cartilage from ACGII-HCG infants have indicated abnormal and diminished collagen II, even its absence, and the presence of collagen I. (Eyre et al. 1986, Godfrey et al. 1988, Godfrey & Hollister 1988, see also Horton & Hecht 1993, Spranger et al. 1994.)

As suggested by earlier biochemical data, the genetic defect causing ACGII-HCG was found in the COL2A1 gene (Vissing et al. 1989). Over ten mutations causing ACGII-HCG have been reported, and all of them involve the replacement of glycine by a bulkier amino acid in the triple helical region of the $\alpha_1$(II) chain. The disorder is caused by heterozygous mutations, and naturally, because of the severity of the symptoms, the cases are usually sporadic. (See Horton & Hecht 1993, Kuivaniemi et al. 1997.)

2.4.2. Spondyloepiphyseal dysplasia (SED)

Spondyloepiphyseal dysplasia (SED, OMIM #183900) is characterized by a short trunk, short extremities, a barrel-shaped chest, kyphosis and lordosis, severe myopia, and often retinal detachments, cleft palate and clubfoot. Affected infants usually survive. Radiographs show defects in ossification of the spine and primarily proximal extremities. (See Horton & Hecht 1993.) Histological studies show defects resembling those in ACGII-HCG, and biochemical examination indicates abnormalities in collagen II (Yang et al. 1980, Murray et al. 1989). SED refers to a heterogeneous group of disorders that affect the spine and epiphyses, and as the symptoms resemble those of ACGII-HCG, SED can be considered to represent the mild end of the spectrum comprising ACGII-HCG and SED phenotypes (see Spranger et al. 1994).

Most cases of SED show autosomal dominant inheritance, but an autosomal recessive form of inheritance has also been suggested in some cases (Harrod et al. 1984, see also Horton & Hecht 1993). Lee et al. (1989) identified a deletion of an entire exon in the COL2A1 gene, leading to a lack of 36 amino acids in the triple helical region of the $\alpha_1$(II) chain, and subsequently numerous mutations, including insertions, mutations causing aberrant RNA splicing, glycine mutations and a cystein for arginine substitution, have been characterized (see Vikkula et al. 1994, Kuivaniemi et al. 1997). However,
some investigations have excluded the COL2A1 gene as a locus for SED in certain families, indicating that there must also be some other locus or loci for the disorder (Wordsworth et al. 1988, Anderson et al. 1990).

### 2.4.3. Kniest dysplasia

Patients with Kniest dysplasia (OMIM #156550) have a round face, midfacial hypoplasia, prominent eyes and high myopia. Cleft palate, hearing loss and retinal detachment are also frequent manifestations of the disorder. The trunk and extremities are shortened and the joints are enlarged and exhibit restricted mobility. Numerous abnormal characteristics are seen in radiographic examinations. There is generally flattening of the vertebral bodies, the long bones are shortened and of a dumbbell shape due to expanded metaphyses, while the epiphyses are irregular and fragmented, the joint spaces are narrow, and there is generalized osteopenia in their ends. Kniest dysplasia is usually more severe than SED and is lethal in some cases. (Chen et al. 1980, see also Horton & Hecht 1993, Spranger et al. 1994.) Histologically, the cartilage is soft and vacuolar, and electron microscopy shows abnormally thin and irregular collagen fibrils in the matrix and dilated rough endoplasmic reticulum (Horton & Rimoin 1979, Poole et al. 1988, Gilbert-Barnes et al. 1996.)

Kniest dysplasia is an autosomal dominant disorder. More than ten heterozygous mutations in the COL2A1 gene have been reported to date, including that in the patient originally described by Dr. Kniest in 1952. All except two of them lead to whole or partial exon deletions, and all but one are clustered in the N terminal half of the molecule, thus indicating that the cause of Kniest dysplasia lies in short in-frame deletions in this particular region of the collagen II molecule. (Spranger et al. 1997, Fernandes et al. 1998, see also Kuivaniemi et al. 1997.) Recent studies have shown that a significant fraction of the defective collagen II molecules are secreted and incorporated in cartilage collagen fibrils. Furthermore, it has been hypothesized that the normal $\alpha_1$(II) chains would form a loop and thus allow in-register triple helix formation in both directions from the deletion site. Moreover, due to the susceptibility of the loop to protease cleavage, this could also account for the degenerate vacuolar appearance of the cartilage matrix in patients with Kniest dysplasia. (Weis et al. 1998, Fernandes et al. 1998.)

### 2.4.4. Stickler and Marshall syndromes

Stickler syndrome (OMIM #108300), originally called hereditary progressive arthro-ophthalmopathy, is an autosomal dominantly inherited disorder. Patients have high myopia, vitreoretinal degeneration and frequently retinal detachments and cataracts. Midfacial hypoplasia and micrognathia are often accompanied by cleft palate or a lesser degree of clefting, and sensorineural hearing defect is common. The skeletal manifestations include juvenile progressive arthropathy, irregularity of the vertebral bodies and hypermobile joints. Skeletal growth is usually normal. Mild epiphyseal
dysplasia and overtubulation of long bones are seen in radiological examination. (Stickler et al. 1965, Stickler & Pugh 1967, see also Herrmann et al. 1975, Temple 1989, Snead & Yates 1999.) In addition, an increased prevalence of mitral valve prolapse has been reported (Liberfarb & Goldblatt 1986), but this was not supported in the echocardiographic studies of over 100 patients (see Snead & Yates 1999).

Marshall syndrome (OMIM #154780) is also an autosomal dominantly inherited disorder, and closely resembles Stickler syndrome. The ocular, facial and auditory features are much alike, but patients with Marshall syndrome are usually short in stature and have intracranial ossifications, abnormal frontal sinuses and thickened calvaria. In addition, Dr. Marshall originally reported some kind of hypohidrotic ectodermal dysplasia associated with the other symptoms, but it is thought nowadays that ectodermal dysplasia is not involved in the syndrome. (Marshall 1958, O’Donnell et al. 1976, Aymé & Preus 1984, Stratton et al. 1991, Shanske et al. 1997, Warman et al. 1998.)

Numerous new cases of Stickler or Marshall syndrome have been reported since the original kindreds, and these have broadened the range of characteristics, so that it is now clear that there is clinical heterogeneity within the disease phenotypes, especially in Stickler syndrome (Herrmann et al. 1975, Zlotogora et al. 1992). The similar features of Stickler and Marshall syndromes have led to a clinical debate as to whether the syndromes are two distinct entities or just different manifestations of a single entity (Baraitser 1982, Winter et al. 1983, Aymé & Preus 1984, Stratton et al. 1991, Shanske et al. 1997). Furthermore, the diagnosis of Marshall syndrome has aroused some debate (Shanske et al. 1998, Warman et al. 1998).

Another autosomal dominantly inherited syndrome with ocular manifestations is Wagner syndrome, or hyaloideoretinal degeneration of Wagner (OMIM #143200). This was originally characterized by myopia, cataract, vitreoretinal degeneration and often retinal detachment, but no manifestations in non-ocular tissues. There are some reports of Stickler-like non-ocular symptoms in Wagner syndrome, however, leading to a suggestion that the two could represent the same disorder. (See Liberfarb et al. 1979, Horton & Hecht 1993.)

Linkage studies have shown that the COL2A1 gene is a locus for Stickler syndrome (Francomano et al. 1987, Knowlton et al. 1989). Altogether nine mutations have been reported to date, all of them leading to premature termination of translation (Ahmad et al. 1991, Brown et al. 1992, Ahmad et al. 1993, Ritvaniemi et al. 1993, Brown et al. 1995a, Ahmad et al. 1995, Williams et al. 1996). The same gene was also found to be a locus for Wagner syndrome in one family (Körkkö et al. 1993), even though Wagner syndrome, together with erosive vitreoretinopathy, is normally linked to 5q13-q14 (Brown et al. 1995b). The COL2A1 gene has also been excluded in several families with Stickler syndrome (Knowlton et al. 1989, Vintiner et al. 1991, Bonaventure et al. 1992), and Snead et al. (1994) have suggested that Stickler families with congenital vitreous anomaly (Stickler syndrome type 1) have linkage to the COL2A1 gene, while in those with congenitally defective vitreous gel architecture but no congenital vitreous anomaly (Stickler syndrome type 2) linkage is excluded. Subsequently, Richards et al. (1996) demonstrated linkage to the COL11A1 gene in a family with Stickler syndrome type 2, and traced it to a missense mutation converting glycine to valine in the N terminal part of the triple helix of the α1(XI) chain. In addition, Wilkin et al. (1998) have suggested still another locus for Stickler syndrome apart from the genes coding for collagens II and XI.
It was also hypothesized that Marshall syndrome was caused by mutations in the COL11A1 gene (van Steensel et al. 1997). A splicing mutation that led to a deletion of 18 amino acids in the C terminal half of the α1(XI) triple helical region was found to be associated with the phenotype (Griffith et al. 1998).

2.4.5. Otospondyloepiphyseal dysplasia (OSMED), Weissenbacher-Zweymüller syndrome (WZS) and non-ocular Stickler-like syndrome

Otospondyloepiphyseal dysplasia (OSMED, OMIM #215150) is an autosomal recessive chondrodysplasia characterized by sensorineural deafness, spondyloepiphyseal dysplasia with large epiphyses, disproportionate shortness, midfacial hypoplasia and often micrognathia and cleft palate, but no ocular manifestations (Giedion et al. 1982, Kääriäinen et al. 1993, van Steensel et al. 1997). Electron microscopic examination of cartilage shows disorganized matrix with extremely thick collagen fibres that consist of aggregated thinner fibrils (van Steensel et al. 1997). Weissenbacher-Zweymüller syndrome (WZS, OMIM #277610) is highly similar to OSMED and is also thought to be an autosomal recessive disorder. One characteristic of WZS is its favourable progression during childhood. (Cortina et al. 1977, Chemke et al. 1992.)

Vikkula et al. (1995) demonstrated linkage to the COL11A2 gene in a family with OSMED, and also identified a homozygous mutation in the gene. The mutation led to a substitution of arginine for glycine in the N terminal half of the triple helical region of the collagen XI molecule. The original patient with WZS described by Drs. Weissenbacher and Zweymüller in 1964 was also found to have a mutation in the COL11A2 gene. The mutation converted a glycine to glutamate in the C terminal end of the triple helix, and, surprisingly, was heterozygous. (Pihlajamaa et al. 1998b.) Furthermore, a kindred with an autosomal dominant disorder resembling Stickler syndrome but without eye involvement was linked to chromosome 6 near to the COL11A2 gene (Brunner et al. 1994), and subsequently Vikkula et al. (1995) reported of a heterozygous mutation in the kindred causing exon skipping in the COL11A2 mRNA and thus deletion of 18 amino acids near the C terminal end of the triple helical region in the α2(XI) chain. A mutation causing deletion of 9 amino acids in the middle of the triple helical region was described in another family with a Stickler-like syndrome without eye involvement (Sirko-Osadsa et al. 1998).

The nomenclature for the chondrodysplasias caused by mutations in the COL11A2 gene was established before the common molecular background for the disorders had been discovered, and thus the terminology is confusing, with various names applying to the same disorders. Recent findings have led to a suggestion for a new nomenclature for these disorders. The term heterozygous OSMED was introduced to include the former WZS and non-ocular Stickler-like syndrome, while homozygous OSMED is recommended for recessive OSMED. (Pihlajamaa et al. 1998b, Spranger 1998.)
2.4.6. Multiple epiphyseal dysplasia (MED)

The typical finding in patients with multiple epiphyseal dysplasia (MED; OMIM #132400), an autosomal dominantly inherited disorder, is prominent and frequently painful joints with restricted mobility. This usually appears in early childhood. The disorder affects multiple joints rather than being generalized, and it leads to precocious osteoarthritis. The patients have a waddling gait and may have short and stubby hands. Stature is normal or moderately short with normal proportions. Radiographic examination shows typically small, irregular epiphyses. (Barrie et al. 1958, Hoefnagel et al. 1967, Spranger et al. 1974, van Mourik et al. 1998a.) The severity of symptoms varies considerably, and some patients with radiological evidence of MED do not have any complaints (van Mourik et al. 1998a). MED is often divided into two subgroups, the mild Ribbing type and the severe Fairbank type (see International Working Group on Constitutional Diseases of Bone 1998). Because of the wide phenotypic spectrum of MED, not all patients with the disorder can be classified into either the Ribbing or the Fairbank type, and it has been suggested that these terms should be abandoned (van Mourik et al. 1998a, Haga et al. 1998). The Ribbing type has also been reported to be recessive, but molecular genetic studies on such families suggest germline mosaicism as an explanation (Deere et al. 1995, Loughlin et al. 1998).

Pseudoachondroplasia (PSACH, OMIM #177170) is an autosomal dominant disorder that resembles MED but is more severe, causing disproportionate dwarfism and ligamentous laxity. Radiographs show abnormalities in the epiphyses, metaphyses and spine. (See Horton & Hecht 1993.)

Genetic linkage of both MED and PSACH to the pericentromeric region of chromosome 19 has been demonstrated (Hecht et al. 1993, Briggs et al. 1993, Oehlmann et al. 1994). This is where the gene for cartilage oligomeric matrix protein (COMP), a pentamerical calcium binding glycoprotein of the extracellular matrix, is located (Newton et al. 1994). A large number of mutations causing either MED or PSACH have been reported in the COMP gene (Briggs et al. 1995, Hecht et al. 1995, Ikegawa et al. 1998), but it has also been excluded in a MED family, where linkage was demonstrated to a region of chromosome 1 containing the COL9A2 gene (Briggs et al. 1994). Muragaki et al. (1996) identified a splicing mutation of exon 3 in the COL9A2 gene, leading to a lack of 12 amino acids in the COL3 domain of the α2(IX) chain. Recently, another mutation in collagen IX was reported in a family with MED, causing a splicing defect of exon 3 with a similar consequence to that reported earlier but applying to the COL9A3 gene. (Paassilta et al. 1999a.)

Similar dilatations of the rough endoplasmic reticulum in chondrocytes can be seen in MED and PSACH (Stanesca et al. 1982, Stanescu et al. 1993). These were studied further in a PSACH patient who was shown to have a mutation in the COMP gene. The dilatations contained aggregates consisting of at least COMP and collagen IX, suggesting a common pathogenic mechanism for MED and PSACH. (Maddox et al. 1997b.) No dilatation of the rough endoplasmic reticulum was seen in patients with MED caused by the COL9A2 mutation, however (van Mourik et al. 1998b).
2.4.7. Osteoarthritis (OA)

Osteoarthritis, or osteoarthrosis (OA), is the most common articular disorder and is characterized by joint pain and tenderness, stiffness, crepitus and limitation of motion. Radiographs typically show joint-space narrowing, osteophytes, subchondral bone sclerosis and subchondral cyst formation. In severe cases deformity of the bone ends is also seen. OA is classified into primary (idiopathic) or secondary subsets; in a primary case there is no known predisposing factor, while in a secondary case such a factor is known, e.g. trauma, infection, a mechanical factor, some other joint disorder or a metabolic disease. Further classification divides OA into localized, in which only certain joints are affected, and generalized subsets. Even though it is often considered only an inevitable consequence of ageing, female sex, obesity, occupational load and heavy sports exertion are risk factors for it. (See Moskowitz 1989, Altman 1995, Creamer & Hochberg 1997.)

The suggestion of genetic factors underlying OA in certain families was made several decades ago (Stecher 1941, Stecher et al. 1953, Kellgren et al. 1963). In 1989, Palotie et al. demonstrated a linkage to the COL2A1 gene in a family with primary OA. Knowlton et al. (1990) made a similar finding in a family with precocious OA and associated mild chondrodysplasia, and Ala-Kokko et al. (1990) subsequently demonstrated a mutation in this family. The mutation converted arginine to cysteine, an amino acid not normally found in the triple helical domains of collagens, in the middle of the triple helix of the collagen II molecule. Altogether five families with this same mutation have been reported to date, and three of them have been shown to have a common ancestor (see Bleasel et al. 1995, Kuivaniemi et al. 1997, Bleasel et al. 1998). In addition, a mutation converting arginine to cysteine at position 75 near the N terminus of the triple helix has been reported in three families with OA associated with mild SED (see Bleasel et al. 1995, Bleasel et al. 1996).

2.5. Animal models for disorders of the cartilage collagens

In order to fully understand the mechanisms of a cartilage disorder, histological and molecular studies are essential. Human cartilage specimens for such studies are difficult to obtain, because that would require a surgical operation, and perhaps more importantly, spare cartilage is rarely available, at least in large amounts. Furthermore, it is well known that chondrocytes dedifferentiate in cell culture (Grundmann et al. 1980). Animal models, often mice, are thus used for cartilage disorders. The phenotype is either a natural variant of the species or is generated by the scientist by introducing mutagenic agents into the animal, or more elegantly, by manipulating the genome by modern techniques that allow targeted mutation, inactivation or specific expression of virtually any gene. Large amounts of cartilage specimens can be obtained from animal models, and disorders can also be studied in vivo.
2.5.1. Collagen II

Several approaches have been used for studying collagen II defects in mice. Two transgenic mouse lines with deletions in the triple helical domain have been generated, and two with glycine substitutions. The phenotypes caused by these mutations are usually perinatally lethal chondrodysplasias characterized by dwarfism, short snout, cleft palate, disorganized growth plates and reduced ossification. The collagen fibrils are disorganized and reduced in amount. (Vandenberg et al. 1991, Garofalo et al. 1991, Metsäranta et al. 1992, Maddox et al. 1997a.) A phenotype resembling this has also been produced in a mouse line with disproportionate micromelia (Dmm) generated by irradiation (Brown et al. 1981). These mice have a deletion in the region encoding the C propeptide of procollagen II, which leads to a substitution of one amino acid for two (Pace et al. 1997).

Garofalo et al. (1993) overexpressed collagen II. The mice died at birth, but they did not exhibit cleft palate, abnormal craniofacial features, or other skeletal deformities, except for a slight reduction in the length of the long bones. Electron microscopy of the cartilage showed abnormally thick, extensively branched fibrils, which led to the suggestion that the imbalance in cartilage components caused by the overexpression of collagen II disrupted the mechanism that controls the diameter of the fibrils. On the other hand, total inactivation of the Col2a1 gene generated by S.-W. Li et al. (1995) in mice led to severe perinatally lethal chondrodysplasia. This mouse model shed some light on the role of cartilage in skeletal development by showing that collagen II is essential for the development of endochondral bone and epiphyseal cartilage but not that of many other skeletal structures such as membranous and periosteal bone.

2.5.2. Collagen IX

Three mouse lines with defective collagen IX have been generated. Mice expressing a truncated α1(IX) chain, which lack part of the COL2 domain, the entire NC3 domain and part of the COL3 domain, were shown to develop osteoarthritis, which was mostly obvious in the knee joints. Ones with a higher level of transgene expression also developed mild chondrodysplasia with mild proportionate dwarfism and a slight ossification defect in the spine. Abnormally thin collagen fibrils were seen in electron microscopy. (Nakata et al. 1993.) Furthermore, long-term follow-up of the mice revealed accelerated intervertebral disc degeneration, which occasionally led to herniation and osteophyte formation (Kimura et al. 1996).

Fässler et al. (1994) generated a mouse line lacking the α1(IX) chain. The mice showed no detectable abnormalities at birth, but later developed a severe progressive degenerative joint disease resembling human OA, which was most prominent in the knee joints. The fibril morphology of the cartilage was nevertheless visually identical in the homozygous and control mice. Absence of the α1(IX) chain was shown to lead to absence of the α2(IX) and α3(IX) chains as well, even though they were transcribed. Thus this mouse line exhibited functional inactivation of the entire collagen IX. These findings lead to a suggestion that collagen IX is essential for the integrity of the cartilage matrix, but not for skeletal development. (Hagg et al. 1997.)
Transgenic mice expressing multiple copies of a minigene coding for the NC4 domain of collagen IX have also been created. This overexpression of the NC4 domain induced osteoarthritis analogous to human OA. (Haimes et al. 1995.)

2.5.3. Collagen XI

One mouse model for collagen XI disorders has been reported. Mice with autosomal recessive chondrodysplasia (cho) have disproportionately short limbs with wide, flared metaphyses and abnormal facial features consisting of a short snout, short mandible, protruding tongue and cleft palate. They have soft tracheal cartilage rings, which is probably the cause of their perinatal death. Microscopy showed disorganized growth plates and abnormally thick collagen fibrils. (Seegmiller et al. 1971.) Linkage studies mapped the cho locus to the region containing the Col11a1 gene, and a frame shift mutation causing premature translation termination was identified. The mutation thus leads to absence of the α1(XI) chain. The finding indicates the importance of this chain for skeletal morphogenesis and, more specifically, for the formation of proper collagen fibrils in cartilage. (Li Y et al. 1995.)

2.6. Intervertebral disc disease

Degenerative changes in the intervertebral discs begin early in childhood. The degeneration occurs in all parts of the discs, but the nucleus pulposus undergoes the most extensive changes. The annulus fibrosus generally expands at the expense of the nucleus pulposus, the number of viable cells in the disc decreases and the water concentration also decreases. The amounts of proteoglycans decrease, and fragmentation of the aggregating proteoglycans can be detected. The pathomechanism of these changes is not known, but one potential cause could be a decline in nutrition due to the increasing volume of the disc during growth and a simultaneous age-related decrease in the arteries supplying nutrients and removing waste. Even if disc degeneration is an inevitable consequence of ageing, it can vary greatly in rate and extent within the same person and between individuals. (See Buckwalter 1995.)

The degeneration impairs the structural integrity of the disc tissue and increases the probability of functional deficiency and disc herniation. In intervertebral disc herniation part of the nucleus pulposus extrudes through the annulus fibrosus. The extruded material may cause mechanical compression or chemical irritation to the adjacent nerve root and lead to sciatica, i.e. radiating pain in the dermatome of the root. (See Hasue 1993, Alaranta & Poussa 1994.) Herniation of the lumbar discs and sciatica, also called lumbar disc syndrome, is a common disorder affecting the Finnish population with a prevalence of about 5% (Heliovaara et al. 1987). Many environmental risk factors for intervertebral disc disease have been proposed, including hard physical work, driving motor vehicles and smoking (see Heliovaara 1989), but genetic predisposition to advanced disc degeneration and intervertebral disc disease has also been suggested, and it has even been
claimed that this may be the primary factor (Varlotta et al. 1991, Scapinelli 1993, Battié et al. 1995, Sambrook et al. 1999). In addition, polymorphisms in the vitamin D receptor gene have been associated with intervertebral disc degeneration (Videman et al. 1998).

2.6.1. Cmd mice

Mice that are homozygous for cartilage matrix deficiency (cmd) have short limbs, trunk, tail and snout, cleft palate and hearing impairment, and they die at birth, probably of respiratory failure (Rittenhouse et al. 1978, Yoo et al. 1991). The mice were found to lack aggrecan (Kimata et al. 1981), and a deletion causing premature translation termination in the gene coding for aggrecan has been identified (Watanabe et al. 1994). Despite the original understanding that the heterozygous mice are normal, Watanabe et al. (1997) demonstrated that they show dwarfism and spinal misalignment at the age of one year. They exhibited deformation of the vertebral bodies, advanced disc degeneration and disc herniations, but no changes in other cartilaginous tissues. This finding supports results obtained with mice expressing a truncated α1(IX) chain, indicating that mutations in the genes coding for cartilage matrix protein can cause intervertebral disc disease.

Mice with hereditary kyphoscoliosis (ky) also develop advanced intervertebral disc degeneration, but these changes are secondary to a neuromuscular disorder that leads to atrophy of the muscles supporting the spine (Mason & Palfrey 1984, Bridges et al. 1992).

2.7. Mutation detection methods

Knowing the genetic defect behind a disease provides a basis for understanding the disease and treating it. The detection of new, previously unknown mutations in sequences of a candidate gene requires an appropriate method. Sequencing is a thorough method, but if the sequences to be searched span thousands or tens of thousands of base pairs, the procedure turns out to be expensive and excessively laborious. A large number of methods for mutation detection, most of them based on PCR, have been introduced to make it more efficient with regard to costs, time and work. The methods usually take advantage either of the change in electrophoretic mobility produced by the mutation or of recognition and cleavage of the mismatched site in a heteroduplex comprising a mutant and a wild strand. (See Cotton 1989, Cotton 1993, Eng & Vijg 1997, Nollau & Wagener 1997.)

2.7.1. Commonly used methods based on electrophoresis

Single-strand conformation polymorphism (SSCP) is based on the fact that single-stranded DNA has a defined secondary structure under certain conditions. This secondary structure changes if the sequence is altered, and this is detected in the form of mobility
shifts on a polyacrylamide gel. (Orita et al. 1989a, Orita et al. 1989b.) Band visualization has to be performed with radiolabelling or silver staining, as ethidium bromide, which is widely used to stain double-stranded DNA, does not stain single-stranded DNA efficiently. The advantage of SSCP is the simplicity of the procedure, but its sensitivity, which is best with fragments less than 200 bp, is only 60-95%. (See Cotton 1993, Eng & Vijg 1997.)

The melting properties of double-stranded DNA depend on the sequence of the DNA fragment. This phenomenon is utilized in denaturing gradient gel electrophoresis (DGGE) by electrophoresing DNA fragments on a polyacrylamide gel under increasing denaturing conditions. Thus sequence variations alter the melting and mobility of the mutant fragment as compared with a control fragment. (Fischer & Lerman 1983, see also Myers et al. 1987.) The original method was able to detect sequence variations only in the low melting domains of the fragment, i.e. sequences with a low melting temperature, giving a sensitivity of about 50%, but the addition of an artificial high melting GC-rich sequence (GC-clamp) to one end of the fragment using PCR primers converts the whole fragment to a low melting region and greatly improves the sensitivity (Sheffield et al. 1989). Fragments of up to 1 kb can be analyzed, but ones of 150-500 bp are usually most appropriate. The advantage of DGGE is its high sensitivity of almost 100%, but optimization of the melting properties of the fragment and electrophoresis conditions is laborious and time-consuming. (See Cotton 1989, Cotton 1993, Ganguly & Prockop 1995.)

Heteroduplex analysis (HA or HET) is based on the different mobilities of homoduplexes and heteroduplexes on a non-denaturing polyacrylamide gel. Heteroduplexes, i.e. double-stranded DNA which contains one strand of the mutant type and one of the wild type, are formed by heat denaturation and subsequent reannealing of the two forms of DNA. (See Glavac & Dean 1995.) HA is a simple method that does not require optimizing, but its sensitivity is estimated to be about 80% (Cotton 1993). Many modifications of HA have been used to improve the sensitivity (see Glavac & Dean 1995).

In conformation sensitive gel electrophoresis (CSGE), mildly denaturing agents are used to enhance the conformational changes in the heteroduplexes. Sensitivity is also improved by using 1,4-bis(acryloyl)piperazine (BAP) instead of the commonly used Bis as a crosslinker in the polyacrylamide gels. Even though CSGE is a simple method, it has been reported to detect most mutations in fragments of 200-800 bp. (Ganguly et al. 1993, Ganguly & Prockop 1995.)

### 2.7.2. Commonly used methods based on cleavage of a mismatch

A mismatch in a heteroduplex creates a short region of single-stranded DNA, which can be detected by an enzyme or a chemical that cleaves single-stranded DNA. In enzyme mismatch cleavage (EMC), bacteriophage T4 endonuclease VII and/or T7 endonuclease I is used to cleave the site of the mismatch. EMC can be used for fragments of up to 1.5 kb. The method is relatively sensitive, but some unspecific background can also be observed. (Youil et al. 1995, Mashal et al. 1995.) Chemical mismatch cleavage (CMC) is a two-
35

phase procedure in which the DNA is first treated with hydroxylamine and osmium
tetroxide, which react with mismatched cytosines and thymines, respectively, and
piperidine is then added to cleave the modified mismatched base. Fragments up to 2 kb
can be screened with a sensitivity of 100%, but the toxic reagents are a major drawback.
(Cotton et al. 1988, see also Cotton 1993, Nollau & Wagener 1997.)

2.7.3. Other methods

Carbodiimide (CDI) is a bulky reagent that reacts with thymine and guanine residues.
The CDI modification procedure for mutation detection is based on the observation that
CDI reacts more rapidly with unpaired bases than with paired ones. Electrophoresis,
nuclease cleavage, immuno electron microscopy and primer extension have been used to
detect the modified bases (Novack et al 1986, Thomas et al. 1988, Ganguly et al. 1989,
Ganguly & Prockop 1990). In the primer extension method, which is probably the most
practical of these, CDI blocks primer extension and shorter products can be identified by
electrophoresis. DNA fragments of up to 1 kb can be used, and the location of the
mutation can be estimated to an accuracy of about 15 bp. The advantage of the immuno
electron microscopy method is that a mutation can be detected in a heteroduplex of
several kilobases. (See Cotton 1993.)

Cleavase fragment length polymorphism (CFLP) is based on the secondary structure
changes that a sequence variation generates. Cleavase I endonuclease cleaves the single-
stranded DNA at the bifurcated end of a double-stranded region, and thus generates DNA
fragments of different lengths depending on the sequences. (Lyamichev et al 1993, see
also Nollau & Wagener 1997.)

A mismatch binding protein MutS, which is a component of the Eschericia coli DNA
repair system, can also be used to detect mutations in heteroduplexes. Various methods,
including mobility shift assays, biotinylated MutS, solid-phase assay with immobilized
MutS and an optical biosensor have been used to detect MutS binding (Lishanski et al.
been taken of the protective feature of the MutS protein against endonucleases for
detection purposes (Ellis et al. 1994).

The protein truncation test (PTT) is specific to mutations causing premature
translation termination. A T7 promoter and a eucaryotic translation initiation sequence
are added to the fragment of interest by PCR, after which the fragment is transcribed and
translated. A premature translation termination codon results in shortening of the
translation product, which can be detected in electrophoresis. (Roest et al. 1993, see also
Nollau & Wagener 1997.)

The most recent methods for mutation detection require special equipment and
computerized analysis. Denaturing high-performance liquid chromatography (DHPLC) is
based on the high resolution separation of heteroduplexes from homoduplexes under
controlled partially denaturing conditions. The method is highly sensitive and the
commercially available procedure (Transgenomic WAVETM System) is fully automated.
microarray (DNA chip) technology will be a promising method for mutation detection in
the future. It is based on hybridization of the sample to a large number of immobilized oligonucleotide microarrays fabricated by advanced manufacturing processes. The method has an enormous capacity, and many applications have been demonstrated for it. (See Marshall & Hodgson 1998, Ramsay 1998, Hacia 1999.)
3. Outlines of the present research

Collagen II is the main collagen in cartilage, and mutations in the COL2A1 gene coding for the collagen II homotrimer are known to cause a spectrum of chondrodysplasias ranging from lethal ACGII-HCG to early-onset OA associated with mild chondrodysplasia. In general, a certain phenotype predicts a certain type of mutation in the COL2A1 gene. Kniest dysplasia, for example, is usually caused by a mutation leading to a short in-frame deletion in the N terminal half of the triple helix, while a glycine substitution at the C terminal end of the triple helix often leads to ACGII-HCG.

Collagens IX and XI are minor cartilage collagens which are also expressed in the intervertebral disc, ocular vitreous body and tectorial membrane of the inner ear. At the time when this research project was initiated, only one mutation had been reported in collagen IX. This was in the COL9A2 gene and led to MED. Four mutations in collagen XI had been reported, two of which were in the COL11A1 gene, one leading to Stickler syndrome and the other to Marshall syndrome, while two of them were in the COL11A2 gene, one leading to OSMED and the other to Stickler-like syndrome without ocular involvement.

Based on the collagen II studies it was hypothesized that mutations in collagens IX and XI would also cause a spectrum of phenotypes. The object of this work was therefore to increase our knowledge of the role of collagens IX and XI in disorders affecting cartilaginous tissues. The specific aims were:

1. to improve the sensitivity of conformation sensitive gel electrophoresis (CSGE),
2. to characterize the genomic structure of the human COL11A1 gene and to set up the CSGE method for the gene,
3. to clarify the relation between the Marshall and Stickler syndromes by analyzing the COL11A1 and COL2A1 genes in patients with Marshall and Stickler syndromes for mutations by CSGE and comparing the two syndromes on the basis of the genetic and clinical data, and
4. to analyze the COL9A2 gene in patients with intervertebral disc disease for mutations by CSGE.
4. Materials and methods

The materials and methods are described in detail in the original papers I-III.

4.1. Subjects and clinical examination methods (I-III)

Altogether 30 patients with a suspected diagnosis of Marshall or Stickler syndrome were referred to us from several departments of genetics, ophthalmology and surgery in Finland, Israel, the Netherlands, Norway, Poland and the USA. The clinical diagnosis was established by a clinician familiar with this morphology and experienced with these disorders. Informed consent was obtained from the patients.

157 unrelated Finnish patients with sciatica, defined as unilateral pain of duration over one month radiating from the back to below the knee, and the family members of four of the patients were referred to us from Department of Physical Medicine and Rehabilitation and the Department of Internal Medicine at Oulu University Hospital. The patients were evaluated clinically and radiologically by experienced clinicians. Altogether 174 Finnish controls consisting of 101 asymptomatic subjects, 54 patients with OA and 19 patients with various chondrodysplasias but no history of sciatica were included in the study. The OA patients were referred to us from the Department of Surgery at Kuopio University Hospital.

The radiological examinations of the patients with sciatica were performed by magnetic resonance imaging (MRI) in 156 cases and by computed tomography (CT) in one case. The MRI scans, obtained with a 1.5-T imaging system (Signa, General Electric, Milwaukee, Wisconsin), consisted of sagittal images with a TR/TE of 4000/95 ms and axial images with a TR/TE of 640/14 ms. The CT images (Hi Speed Advantage, GE Medical Systems, Milwaukee, Wisconsin) consisted of scans through the L2-L3 interspace to the L5-S1 interspace. The radiographs were read by two neuroradiologists blinded to the results of the clinical and genetic examination and the clinical histories of the patients. The following radiological findings were considered indications of intervertebral disc disease: (1) disc extrusion, (2) any herniation at two or more levels, (3) endplate degeneration at one or more levels in patients under 30 years of age, at two or more levels in patients aged 30-50 years or at four or more levels in patients over 50
years, or (4) bulging and/or protrusion at four or more levels (Jensen et al. 1994, Weishaupt et al. 1998).

Genomic DNA was extracted from the venous blood or cultured fibroblasts by standard protocols and used for the analyses.

4.2. Characterization of the COL11A1 gene (II)

To characterize the genomic structure of the COL11A1 gene and obtain intronic sequences of the gene required for the CSGE analysis, screenings for genomic clones were ordered from Genome Systems, Inc. PCR probes were designed based on the published cDNA sequences (Bernard et al. 1988) and the sequences obtained here by sequencing intronic PCR products amplified with exon specific primers. PCR probes for the subsequent screenings were designed based on sequences defined from clones obtained in previous screenings. The probes were from 5’UTR, exon/intron 2, intron 4, intron4/intron5, exons 27/28, intron 28/intron 29, intron 52/intron 53 and 3’UTR. Altogether five clones, one P1, two PACs, and two BACs were found, each containing a part of the gene. The P1 clone was transferred from the *Escherichia coli* NS3529 strain to the NS3516 strain, which gives a higher yield, via transduction according to the instructions provided by Genome Systems. The plasmid DNA preparations were performed using a standard plasmid isolation protocol recommended by the manufacturer.

The genomic organization and intronic sequences were defined, either directly from the clones or from PCR products amplified from the clones, using an automated sequencer (ABI PRISM 377 Sequencer and dRhodamine or BigDye Termination Cycle Sequencing Kit with AmpliTaq DNA Polymerase, FS, Perkin-Elmer; Expand Long Template PCR System, Boehringer Mannheim). For regions that were not included in the clones, the PCR amplifications for sequencing were performed from genomic DNA. The sequencing primers were designed based on the cDNA sequences or sequences defined in this work.

The sizes of the introns up to about 1.5 kb were defined by sequencing, and the sizes of the other introns were defined by PCR amplification (Expand Long Template PCR System, Boehringer Mannheim). Unspecific amplification was ruled out either by sequencing the ends of the PCR products or by using two different sets of PCR primers for each intron.

4.3. CSGE (I-III)

The PCR products for CSGE analysis were amplified from genomic DNA. The primers were designed to amplify the target sequence and about 50 bp of both the 5’ and 3’ flanking sequences. The primers were designed initially to generate PCR products of 200-800 bp, as suggested in the previous reports (Ganguly et al. 1993, Ganguly & Prockop 1995), but during experiments designed to improve the method the size
requirement of the PCR product was changed to 200-450 bp. The PCR amplifications were typically carried out in a reaction volume of 40 μl containing 50-100 ng of genomic DNA, 200μM of each dNTP, 0.25 μM of each primer and 1 unit of Taq polymerase (AmpliTaq Gold, Perkin-Elmer). The cycling conditions consisted of initial denaturation for 10 min at 95°C followed by 30-35 cycles of 95°C for 20-40 sec, 54-60°C for 30-40 sec and 72°C for 20-40 sec, followed by a final extension at 72°C for 4-10 min. Heteroduplexes were generated at the end of the PCR cycling by denaturing the PCR products at 95-98°C for 3-5 min and reannealing at 68°C for 30 min. An aliquot of 50-100 ng was mixed with loading buffer (10 x stock solution of 30% glycerol - 0.25% bromphenol blue - 0.25% xylene cyanol FF) and loaded onto a CSGE gel.

CSGE analysis was performed with a 1 mm gel in a standard DNA sequencing apparatus with 33 x 40 or 37.5 x 45 cm glass plates. The initial gel composition was 10% of a 1:99 ratio of 1,4-bis(acrolyol)piperazine (BAP; Fluka) to acrylamide (Intermountain Scientific), 10% ethylene glycol (Sigma), 15% formamide (Gibco), 0.1% ammonium persulphate (APS; U.S. Biochemicals) and 0.07% N,N,N',N'-tetramethylethylenediamine (TEMED; Sigma) in 0.5 xTTE buffer (44mM Tris - 14.5 mM taurine - 0.1 mM EDTA, pH 9.0), according to the previously published protocol, and the gel was electrophoresed at 400 V for 16 hours (Ganguly et al. 1993, Ganguly & Prockop 1995). In the course of experimentation to improve the method, the gel composition was changed to contain 12-15% BAP-acrylamide, and the electrophoresis was performed at 40 W for 6 to 8.5 hours depending on the size of the glass plates. After the electrophoresis, the gel was stained with ethidiumbromide (1 μg/ml) on a glass plate and a relevant piece was cut out, transferred onto a UV transilluminator and photographed with a high quality charge-coupled device (CCD) camera (Fotodyne or UVP).

4.4. Screening of the COL11A1 and COL2A1 genes of patients with Marshall and Stickler syndromes for mutations by CSGE (II)

PCR primers suitable for CSGE analysis of the COL11A1 gene were designed based on the intronic sequences defined in this work. The entire coding region of the gene, excluding exons 2 and 4, was amplified in 67 fragments. The CSGE analysis for the COL2A1 gene was performed by the procedure set up by Körkkö et al. (submitted), according to the protocol described in section 4.3. After amplification, the PCR products were analyzed on an agarose gel to check the quantity and quality of the products and to reveal any large deletions or insertions. The CSGE analysis was then performed using the improved conditions described in paper I. The samples showing heteroduplexes in CSGE were subjected to automated sequencing (ABI PRISM 377 Sequencer and dRhodamine or BigDye Termination Cycle Sequencing Kit with AmpliTaq DNA Polymerase, FS, Perkin-Elmer) using one of the PCR primers as the sequencing primer. The PCR products had been treated prior to this with exonuclease I and shrimp alkaline phosphatase to remove residual PCR primers and nucleotides (Werle et al. 1994, Hanke & Wink 1994).

The physicians of the patients found to have a mutation were asked to provide complete clinical data to substantiate the diagnosis of either Marshall or Stickler...
syndrome and to perform a genotypic-phenotypic comparison. All available family members of such patients were also analyzed for the presence of the mutation.

4.5. RT-PCR (II)

RT-PCR analysis was performed on a patient, who had a sequence variation at the +3 position in intron 50 of the COL11A1 gene to study the effect of the variation in splicing. Total RNA was extracted from cultured skin fibroblasts (RNasey Midi Kit, Qiagen) and used as a template for first strand synthesis (Superscript Preamplification System, Gibco BRL), which was performed with the oligo(T) primer under the conditions suggested by the manufacturer for transcripts with a high GC content. The subsequent PCR amplifications of the illegitimate transcripts were performed using α1(XI) cDNA specific primers, and the products obtained were sequenced.

4.6. Screening of patients with sciatica for mutations in the COL9A2 gene (III)

The patients with sciatica were screened for mutations in the COL9A2 gene using PCR primers suitable for CSGE analysis designed on the basis of previously defined sequences (Pihlajamaa et al. 1998a). Initially, a subset of patients with sciatica were screened for mutations in the COL9A2 gene by the method described in section 4.3. After a unique sequence variation in exon 19 had been found in one of these patients, exon 19 was amplified in the samples from all the other patients and controls using the PCR primers 5’-TGG ATC TCA GTT TCC CTA CCT G (-92 to -71 in intron 18) and 5’-CAA GAG GTG GTG ATT GAG CAA GAG C (+99 to +75 in intron 19). The PCR products were analyzed for sequence variation by CSGE and sequencing (ABI PRISM 377 Sequencer and dRhodamine or BigDye Termination Cycle Sequencing Kit with AmpliTaq DNA Polymerase, FS, Perkin-Elmer), and some of the PCR products were cloned prior to sequencing (T7 Sequencing Kit, Pharmacia).

The family members of four patients who were found to have the sequence variation in exon 19 were evaluated clinically and radiologically, and genetic analysis for the presence of the mutation was performed to study the co-segregation of the genotype with the phenotype.

4.7. Statistical analysis (III)

The statistical analysis to evaluate the significance of the sequence variation was performed by Harald H. H. Göring at the Department of Genetics and Development of Columbia University, New York, USA, using the ILINK program of the FASTLINK computer package.
According to the clinical and radiological examinations, pedigree members of the four families who tested positive for the disease in either the clinical or the radiological test were coded as affected, and those testing negative in both tests were coded as unaffected. Individuals who were not examined or examined only clinically or radiologically and found to be normal were coded as being of unknown phenotypic status. In addition, artificial pedigrees were created from the case-control data to enable linkage and linkage disequilibrium analyses to be performed jointly on the pedigrees and singletons by coding two case or control individuals as the parents of a hypothetical offspring of unknown phenotype (Kainulainen et al. 1999).
5. Results

5.1. Optimization of CSGE (I)

To test and improve the sensitivity of CSGE for mutation detection, genes coding for collagen \( \alpha \) chains were selected for examination. Collagen genes are challenging targets for mutation detection because they are complex, consisting of large numbers of exons, and the sequences are repetitive and GC-rich due to the numerous glycine and proline residues. Previously known sequence variations in the COL1A1, COL1A2, COL2A1, COL3A1, COL9A1 and COL9A2 genes were used as references in the optimization. In order to improve the sensitivity of CSGE, the lengths of the PCR products used for the analyses, the electrophoresis conditions and the acrylamide concentrations in the gel were altered to test the sensitivity under different conditions.

5.1.1. Improved conditions for mutation detection by CSGE (I)

The original report on CSGE suggested that the procedure could be applied to PCR products of size 200-800 bp (Ganguly et al. 1993). However, sequence variations were detected less frequently in PCR products of 500-800 bp than in shorter products, and a common polymorphism in exon 25 of the COL1A2 gene (Constantinou et al. 1990) was not detected in a PCR product of 755 bp whereas it was clearly detected when the region containing the polymorphism was amplified to generate a PCR product of less than 300 bp (Fig. 1 in paper I). Some other PCR products of over 500 bp were also found to contain polymorphisms using shorter PCR products (data not shown). This finding led to a suggestion that only PCR products of up to 450 bp should be used to ensure the sensitivity of the method.

The electrophoretic conditions were altered to test whether the detection of poorly detectable sequence variations could be improved and to speed up the procedure. The electrophoresis was performed with 40 W for 6 hours instead of a overnight run with 400 V. The separation of heteroduplexes on the gel was improved under the altered electrophoretic conditions in several cases (see Fig. 2A and 2B in paper I), and a specific
test with 48 heteroduplexes detected previously under the original conditions did not indicate any disappearance of these (data not shown). Furthermore, a few previously undetected sequence variations became detectable with the new electrophoretic conditions (data not shown).

In addition, the BAP-acrylamide concentration in the gel was increased from 10% to 15% to see if it would improve the separation of the heteroduplex bands in CSGE. With the higher concentration of the polymer the running time was increased to 8.5 hours. This improved the detection of a single base polymorphism in intron 11 of the COL1A1 gene, which was previously poorly detected (Fig. 2C and 2D in paper I). To test these alterations further, over 200 single base pair changes that had previously been detected using 10% polyacrylamide gels and electrophoretic conditions of 400 V overnight or 40 W for 6 hours were analyzed using 15% polyacrylamide gels with running conditions of 40 W for 8.5 hours. The separation was found to be better under the new conditions.

5.1.2. Testing the sensitivity and specificity of CSGE

Altogether 76 previously identified single base changes detected by nucleotide sequencing of the COL1A1 gene (Körkkö et al. 1998), DGGE analysis of the COL1A1 and COL2A1 genes (Ritvaniemi et al. 1993, Ritvaniemi et al. 1995, Körkkö et al. 1997), or sequencing of the cDNAs for α chains of collagens I and III (see Ganguly et al. 1993) were used to test the sensitivity of CSGE. All the changes except one, which was not consistently revealed, were detected using the improved CSGE procedure (Table 1 in paper I).

The specificity of the method was tested by sequencing 223 PCR products that generated heteroduplexes in CSGE. In addition, over 200 PCR products that did not generate heteroduplexes in CSGE were sequenced as controls. All the products that generated a heteroduplex yielded at least one sequence variation, while no sequence variations were found in the controls (Table 2 in paper I). These results indicate that CSGE does not generate false-positive or false-negative results to any significant extent.

A sequence variation in the PCR product resulted in a unique heteroduplex band that was usually readily distinguishable from other possible sequence variations or their combinations in the product (Fig. 6 in paper I). Only rarely were heteroduplex bands observed that looked similar but were generated by different sequence variations in the PCR product (data not shown).

5.2. Characterization of the human COL11A1 gene (II)

All three genes coding for collagen IX α chains have been characterized (Pihlajamäa et al. 1998a, Paassilta et al. 1999b) as have the COL11A2 and COL2A1 genes, which code for the α2(XI) and α3(XI) chains of collagen IX (Ala-Kokko et al. 1995, Vuoristo et al. 1995). For the COL11A1 gene, which codes for the α1(XI) chain, only the cDNA sequence and a few exon-intron boundaries had previously been known (Bernard et al.
1988, Zhidkova et al. 1995). To complete the information on the genomic structures of the genes coding for collagen XI and to set up a mutation screening method for this gene, the human COL11A1 gene was characterized here.

5.2.1. Characterization of clones for the COL11A1 gene

Altogether five clones for the human COL11A1 gene, one P1, two PAC and two BAC clones, were obtained from the screenings ordered from Genome Systems, Inc. The locations of the introns and the intronic sequences flanking the exons were defined by sequencing with exon-specific primers and primers designed on the basis of the intronic sequences defined here. The exons were numbered in a manner corresponding to the numbering of exons in the COL11A2 gene (Vuoristo et al. 1995). The P1 clone (address DMPC-HFF#1-140(B10)) was found to contain the sequences from intron 4 to intron 29, both of the PAC clones (addresses PAC-84:8A and PAC-154-1M) contained the sequences from intron 4 to intron 54, one of the BAC clones (address 170O21) contained the sequences from intron 4 to intron 63, and the other BAC clone (address BACH-86N5) contained the sequences from an undefined distance from the 5’-end of the gene to intron 1 (Fig. 1 in paper II).

5.2.2. Genomic structure of the human COL11A1 gene

Although PCR probes for screening of the P1, PAC and BAC libraries were used from exon/intron 2, intron 4, and 3’-UTR, no clones containing the sequences from intron 1 to intron 4 or sequences in the 3’-direction of intron 63 were found. Most of the structure and sequences of these regions were defined by sequencing PCR products amplified from genomic DNA, but several attempts to amplify introns 1 and 4 failed and thus their size remained undefined.

The gene was found to consist of 68 exons (Fig. 1 in paper II, Table 1 in paper II), which were numbered from 1 to 67, using the numbers 6A and 6B for the sixth and seventh exons (previously called IIA and IIB), because they are alternatively spliced and do not exist in the same mRNA (Zhidkova et al. 1995). The exon numbers 9-15 previously used by Bernard et al. (1988) correspond to exons 16-22 in this numbering. The locations of the exon-intron boundaries in the major triple helical region of the COL11A1 gene were exactly the same as those in the COL11A2 gene (Table 1 in paper II, Vuoristo et al 1995).

The sizes of the introns were defined by sequencing or PCR amplification. The size of the gene, excluding introns 1 and 4, was about 160 kb (Table 1 in paper II). The intronic sequences flanking the exons were sequenced and over 50 kb of new sequences were defined for the gene. The sequences were submitted to GenBank (accession numbers AF101079-AF101112).
5.2.3. Resequencing of the coding region

Resequencing of the coding region indicated some differences relative to the published cDNA sequence (Bernard et al. 1988). No cysteinyl residue was found in the triple helical region, and the amino acid sequence at positions 413-416, calculated from the first glycine of the main triple helical domain, was Lys-Asp-Gly-Leu instead of the previously reported Arg-Met-Gly-Cys. In addition, the amino acid at position 690 was methionine instead of tryptophan, an amino acid rarely found in collagen triple helices. These sequences were defined from 150 alleles.

5.3. Identification of mutations in the COL11A1 and COL2A1 genes in patients with Marshall or Stickler syndrome (II)

The CSGE method was set up for the COL11A1 gene based on the sequences defined here and a corresponding analysis for the COL2A1 gene was performed by the procedure set up by Körkkö et al. (submitted). A group of 30 patients who had been referred to us with a suspected diagnosis of Marshall or Stickler syndrome were screened for mutations in these two genes. Altogether 23 mutations were identified, fifteen in the COL11A1 gene and eight in the COL2A1 gene.

5.3.1. Identification of mutations in the COL11A1 and COL2A1 genes by CSGE

All exons and exon boundaries except those of exons 2 and 4 of the COL11A1 gene were amplified by PCR and heteroduplexes were generated. The PCR products were analyzed on an agarose gel to check their quantity and quality and reveal any large insertions or deletions. The agarose gel analysis of the products for exon 53 suggested a heterozygous deletion of about 150 bp in one patient. The sequencing indicated a 162 bp deletion that contained 85 bp of intron 52, exon 53, and 23 bp of intron 53. CSGE analysis was performed on the rest of the patients, and several unique heteroduplexes were observed. Sequencing identified ten mutations that altered the splicing consensus sequences, two small in-frame deletions in the coding sequences, and two glycine substitutions (Table 2 in paper II).

A similar procedure was performed to analyze the COL2A1 gene in the patients for whom no mutations were found in the COL11A1 gene, and altogether eight novel mutations were found by CSGE analysis and sequencing. Six of these caused a frameshift leading to premature translation termination codons, and two of them altered the splicing consensus sequences (Table 2 in paper II). In the familial cases the mutation was found to cosegregate with the phenotype (data not shown).

All except two of the mutations altering the splicing consensus sequences affected the conventional AG-dinucleotide at the acceptor splice site or the GT-dinucleotide at the donor splice site. One of the remaining two was an A to C change and the other was an
insertion of T at position +3 in intron 50 of the COL11A1 gene. To study the effect of the A to C change on splicing, RT-PCR analysis was performed for illegitimate mRNA transcripts extracted from the patient’s cultured skin fibroblasts. Two PCR products were obtained, one corresponding to the wild type cDNA sequence and one lacking the sequences for exon 50. No RNA was available from the patient with the insertion, but DNA was obtained from the parents, who were unaffected, and analyzed for the presence of the insertion. Neither parent had the insertion (data not shown).

In the familial cases those family members who were available were examined for the presence of the mutation and for clinical symptoms. The mutation was found to cosegregate with the phenotype (data not shown).

5.3.2. Genotypic-phenotypic comparison

All the COL11A1 mutations were found in the region coding for the major triple-helical domain. Ten out of fifteen mutations altered the splicing consensus sequences for 54 bp exons, and one genomic deletion led to the loss of a 54 bp exon. Eight of these mutations affected exons 48-54, four of them exon 50. The analysis also identified two in-frame deletions of 18 bp and 9 bp in exons 36 and 52, respectively. In addition, two glycine substitution mutations were found (Fig. 1, table 2 in paper II).

Fig. 1. Schematic drawing of the COL11A1 transcript and the locations of the mutations. The major triple helix is left blank and all the other regions are shaded. The areas corresponding to the exons are separated by the horizontal lanes and are drawn to scale. The locations of all the mutations found in the COL11A1 gene are marked with the identification number of the corresponding patient (Table 2 in paper II). The mutations marked above the drawing are splicing mutations in the C terminal half of the molecule (1 to 8 and 10) and a genomic deletion including a 54 bp exon (9). Those marked below the drawing are glycine substitution mutations (12 and 15), small in-frame deletions (13 and 14), and a splicing mutation in the N terminal end of the triple helix (11).
The patients who were found to have a mutation, were evaluated clinically, with particular emphasis on the major findings characteristic of the Stickler and Marshall syndromes, in order to evaluate the phenotypic consequences of the COL11A1 and COL2A1 mutations. Particular attention was paid to the characteristics reported to differ between these syndromes. It has been suggested that the syndromes differ in that patients with Marshall syndrome are more often of short stature, deaf and have abnormalities in cranial ossification and more pronounced dysmorphic features, including a retracted midface with flat nasal bridge, short nose, anteverted nostrils and a long philtrum. Furthermore, it has been suggested that retinal detachment occurs less frequently in patients with Marshall syndrome. (O’Donnell et al. 1976, Aymé & Preuss 1984, Stratton et al. 1997.)

In general, the phenotypes in the COL11A1 and COL2A1 genes resembled each other (Table 1, table 3 in paper II). The clinical findings in the patients who had a mutation in either of the genes typically consisted of high myopia, midfacial hypoplasia, palatal defects and retro/micrognathia.

There were some major differences, however. The vast majority of patients with COL11A1 mutations had moderate to severe hearing impairment that was congenital or detected in early childhood, whereas the patients with COL2A1 mutations had normal hearing or minor hearing loss that usually developed later in life. The ocular findings were generally more severe in the patients with COL2A1 mutations than in those with COL11A1 mutations, in that almost all of the former had vitreoretinal degeneration and retinal detachment. Furthermore, cataracts were more common in patients with COL2A1 mutations than in the patients with COL11A1 mutations. Slight differences were also seen in stature and facial characteristics (Table 3 in paper II). Midfacial hypoplasia, a short nose and a flat nasal bridge were more clearly pronounced in Marshall syndrome, and did not disappear when the child grew older (Fig. 2 in paper II). Radiographic skeletal evaluations could be performed only on a few patients, as radiographs were available only for those with skeletal complaints. Osteoarthritic changes seemed to be more common in the cases with COL2A1 mutations, but many of the patients were too young to be evaluated for such changes. Cranial radiographs were available only for six cases with COL11A1 mutations and two with COL2A1 mutations. None of the latter had cranial abnormalities, but four of the former group had abnormalities such as abnormal frontal sinuses, intracranial ossifications and/or a thickened calvarium.

The patients with COL11A1 and COL2A1 mutations could be divided into three groups based on the phenotype and mutational type (Table 1). The first group had a splicing mutation of a 54 bp exon or a genomic deletion of a 54 bp exon in the COL11A1 gene region coding for the C terminal half of the \( \alpha_{1}(XI) \) molecule, and their characteristics resembled those reported in Marshall syndrome (Aymé & Preuss 1984, Shanske et al. 1997, Griffith et al. 1998). The second group had other mutations in the COL11A1 gene and phenotypes overlapping both the Marshall and Stickler syndromes, while the third group had a mutation in the COL2A1 gene and a phenotype that closely resembled the classical Stickler syndrome (Stickler et al. 1965, Stickler & Pugh 1967, Herrmann et al. 1975).
Table 1. Summary of the clinical data.

<table>
<thead>
<tr>
<th>Findings</th>
<th>COL11A1 mutations</th>
<th>COL2A1 mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>patients 1-10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>patients 11-15&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>moderate to severe hearing loss</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>ocular findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>high myopia</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>retinal detachment</td>
<td>1/10</td>
<td>2/5</td>
</tr>
<tr>
<td>vitreoretinal degeneration</td>
<td>3/9</td>
<td>3/5</td>
</tr>
<tr>
<td>cataract</td>
<td>4/10</td>
<td>3/5</td>
</tr>
<tr>
<td>facial features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>short nose</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>antverted nares</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>midface hypoplasia</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>flat nasal bridge</td>
<td>10/10</td>
<td>4/5</td>
</tr>
<tr>
<td>palate defect</td>
<td>8/10</td>
<td>4/5</td>
</tr>
<tr>
<td>micro/retrognathia</td>
<td>8/10</td>
<td>3/5</td>
</tr>
<tr>
<td>long philtrum</td>
<td>7/9</td>
<td>3/5</td>
</tr>
<tr>
<td>epicanthus</td>
<td>7/9</td>
<td>2/5</td>
</tr>
<tr>
<td>hypertelorism</td>
<td>6/10</td>
<td>0/4</td>
</tr>
<tr>
<td>lowered auricles</td>
<td>4/8</td>
<td>0/3</td>
</tr>
<tr>
<td>stature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>short</td>
<td>6/10</td>
<td>3/5</td>
</tr>
<tr>
<td>tall</td>
<td>0/10</td>
<td>1/5</td>
</tr>
<tr>
<td>cranial findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>abnormal frontal sinuses</td>
<td>2/3</td>
<td>1/1</td>
</tr>
<tr>
<td>intracranial ossifications</td>
<td>1/3</td>
<td>1/2</td>
</tr>
<tr>
<td>thick calvarium</td>
<td>2/4</td>
<td>1/2</td>
</tr>
<tr>
<td>dental abnormalities</td>
<td>1/9</td>
<td>0/2</td>
</tr>
</tbody>
</table>

<sup>a</sup> splicing mutation of a 54 bp exon or genomic deletion causing a loss of 54 bp in exons coding for the C terminal half of the α1(XI) molecule

<sup>b</sup> other mutations in the COL11A1 gene

<sup>c</sup> 3 of the cases had a minor hearing deficit
5.4. Identification of an allele of the COL9A2 gene associated with intervertebral disc disease (III)

5.4.1. Identification of a tryptophan for glutamine/arginine substitution in the COL9A2 gene

The CSGE method for the COL9A2 gene was set up based on the published sequences (Pihlajamaa et al. 1998a), and ten patients with intervertebral disc disease were initially analyzed for mutations in the gene. A unique CSGE pattern was found in the PCR product for exon 19 in one of the patients (Fig. 1 in paper III), and sequencing identified a heterozygous substitution of Trp (TGG) for Gln (CAG) at amino acid position 326. Analysis of exon 19 in all 157 patients with intervertebral disc disease indicated that six of them had the Trp substitution. The analysis was also performed on the 174 controls, who consisted of 101 asymptomatic individuals, 54 with OA and 19 with various chondrodysplasias but no history of sciatica. None of the controls had the Trp substitution. Sequencing of exon 19 also indicated that Arg (CGG) was found relatively often at position 326, suggesting that the Trp allele had arisen from the Arg allele by a single base substitution (Fig. 2 in paper III). The frequency of the Arg substitution was about the same in the patient group and all of the control groups, indicating that it was a neutral polymorphism (Table 1 in paper III).

5.4.2. Co-inheritance of the Trp allele and intervertebral disc disease

To study whether the Trp allele is co-inherited with intervertebral disc disease, the families of four patients with this allele were included in the study. The two other families were not available for examination. The families were evaluated clinically and by CT (family A) or MRI (families B, C and D), and analyzed for the presence of the Trp allele. As stated in the Materials and methods section, the individuals were considered affected if they had either clinical or radiological evidence of the disease, or both. All the family members who had inherited the allele had intervertebral disc disease (Fig. 2).

5.4.3. Statistical analysis

Linkage analysis and linkage disequilibrium analysis were performed to evaluate the statistical evidence for a connection between the Trp allele in the COL9A2 gene and intervertebral disc disease. Since the aetiology of the disease is multifactorial (Heliövaara 1989), a high phenocopy rate was specified. In the selected model the disease locus accounts for 10% of cases, in other words, 90% of patients with intervertebral disc disease are affected for reasons unrelated to this particular disease locus. A fully penetrant dominant disease model was chosen for the analysis and the disease prevalence
Fig. 2. The pedigrees of the four families studied. The right half of each symbol indicates the clinical findings, and the left half the radiological findings. The symbol is black when the finding is positive, and a dot indicates that the family member was examined and found to be normal. A blank symbol indicates that the individual was not examined. The proband is marked with an arrow. The family members with the Trp allele are indicated by (+), and the members who did not have the allele are indicated by (-).
was set at 4.8%, which is the estimated frequency of sciatica in Finland (Heliövaara et al. 1987).

Artificial pedigrees were created from the case-control data to enable a linkage and linkage disequilibrium analysis to be carried out jointly on the pedigrees and singletons (Kainulainen et al. 1999). Linkage disequilibrium analysis gave a lod score of 4.5 at a recombination fraction of 0.12. Subsequent linkage disequilibrium given the presence of linkage was analyzed, and an additional lod score of 7.1 was obtained. The joint lod score was therefore 11.6 (Table 2 in paper III). A similar analysis was performed with the recombination fraction fixed at 0.0 (Table 3 in paper III). The lod scores were lower, but that does not necessarily contradict the hypothesis that the Trp allele itself caused the disease. The explanation is rather the inaccuracy of the model used. For common diseases such as intervertebral disc disease it is practically impossible to specify a model with 100% accuracy, and thus reduced lod scores and overestimated recombination fractions are obtained as a result of model inaccuracies. (Clerget-Darpoux et al. 1986, Risch & Giuffra 1992, see also Terwilliger & Ott 1994.)

To examine the extent to which the results were dependent on the assumed disease model, individual parameters of the model were varied keeping the predicted population prevalence unchanged at 4.8%. The linkage and linkage disequilibrium lod scores remained high for a dominant model when the proportion of cases explained by the disease locus was varied and when high but incomplete penetrance was used (Table 2 in paper III).

5.4.4. Scanning the entire COL9A2 gene for mutations in patients with the Trp allele

To determine whether there were other sequence variations in the COL9A2 gene among the patients whose disease was linked to this gene, all the exons and exon boundaries were analyzed by CSGE. The analysis did not identify any other possible disease causing mutation of coding sequences or RNA splice sites, but it did identify two additional neutral polymorphisms, an A to G change at the third nucleotide in the codon for proline at nucleotide 9 in exon 21 and a G to A change at nucleotide +17 in intron 30. These were analyzed in 90 and 286 individuals, respectively, but were not found in any individuals other than those with the Trp allele.
6. Discussion

6.1. Detection of mutations by CSGE

The detection of unknown mutations in one or more candidate genes is a challenging task, especially with large, complex genes. Sequencing would be a thorough method, but it is too laborious and expensive. Numerous PCR-based screening methods, each having typical advantages and disadvantages, are available for mutation detection (see Cotton 1989, Cotton 1993, Eng & Vijg 1997, Nollau & Wagener 1997), but typical drawbacks are the radioactive or toxic chemicals required for the procedure and problems of sensitivity or specificity. Screening a large number of samples in complex genes requires an easy, powerful method that maintains sensitivity in all sequence contexts but does not require separate optimization for different PCR products. In addition to being large in size and containing dozens of exons, collagen genes are problematic for mutation detection due to their repetitive sequences and high content of G and C residues. This is a consequence of the obligatory sequence for collagens, which consists of a glycine residue as every third amino acid, which is coded by a codon that always starts with two G residues, and has a high content of proline residues, which are coded by codons that always start with two C residues.

CSGE is a relatively recent method that has many advantages, including non-radioactivity, a standard procedure for all fragments, large capacity and the use of standard equipment. Furthermore, it has been reported to have relatively high sensitivity, detecting 60 out of 63 single base pair changes in PCR fragments of 200-800 bp. (Ganguly et al. 1993, Ganguly & Prockop 1995.) No reports on its sensitivity have been published, however, since the initial one by Ganguly et al. (1993).

To study CSGE further, modifications of the procedure were developed and tested with large number of known sequence variations, including the three changes that were not detected in the initial report. The procedure was improved by using shorter PCR products, a higher polymer concentration in the gel and more powerful electrophoretic conditions. To test the sensitivity of the improved method, 76 previously identified sequence variations were tested and all the changes except one, which was not consistently identified, were detected. Furthermore, 223 samples with unique CSGE patterns, along with over 200 PCR products with only homoduplexes on the gel, were
sequenced to test the specificity of the procedure. All of the samples with heteroduplexes on CSGE and none of the samples with only homoduplexes had sequence variations. This indicates that the sensitivity and specificity of CSGE under the improved conditions are close to 100%.

In addition to the high sensitivity and specificity, CSGE proved to be a simple and practical method. No special optimization is required, which allows different fragments to be analyzed simultaneously. The generation of heteroduplexes, which can be added to the end of the PCR program, is the only special treatment required for the samples. Standard sequencing apparatus is used for the electrophoresis, and the loading of multiple samples of different sizes into the same lane provides for a high capacity. Unique CSGE patterns are usually seen for the different sequence variations on the gel. This makes it easy to avoid sequencing the same neutral polymorphisms in all the samples carrying it, and thus reduces the cost of the analysis and the work required.

Although CSGE proved to be a highly sensitive and specific method, it has the same drawbacks as the other PCR-based procedures that make use of the formation of heteroduplexes. The alleles with a deletion or some other sequence variation at the PCR primer binding site of the fragment to be analyzed may not be amplified and thus will not be detected by CSGE, while insertions between the PCR primer sites may lengthen the fragment to such an extent that the allele is not amplified. Furthermore, equal amplification of both a normal and mutant allele cannot be ensured, and thus unequal amplification due to a low amount of the mutant allele in the PCR reaction, or its absence, can cause false negatives in the analysis. In addition, as the formation of heteroduplexes is based on the presence of two alleles with different sequences at the site of the mutation, the mutation will remain undetected if it is present in both alleles, as in the case of a homozygous mutation in a recessive disease. This can be overcome by mixing the patient sample with a control sample.

6.2. Genomic structure of the human COL11A1 gene

Characterization of the human COL11A1 gene in P1, PAC and BAC clones and in genomic DNA indicated that the gene consists of 68 exons and spans at least 160 kb. The sizes for introns 1 and 4 remained undefined because no clones containing these sequences were found and attempts to amplify them by PCR failed. The exon-intron organization of the major triple helical region of the COL11A1 gene was identical to that of the COL11A2 gene, indicating a common evolution for these genes differing from that of the genes for the major fibrillar collagens, which have different numbers and sizes of exons (Vuorio & de Crombrugghe 1990, Vuoristo et al. 1995). The COL11A1 and COL11A2 genes are completely different in size, however, the former being over 160 kb and the latter only 28 kb. The COL11A1 gene resembles COL5A1 in size, the latter being about 150 kb excluding the first intron, the size of which has not been defined (Takahara et al. 1995). These two genes are also identical in the exon-intron organization of the major triple helical region, except for one area in which there are two exons of 54 bp in the COL11A1 gene corresponding to one exon of 108 bp in the COL5A1 gene. Furthermore, the exon-intron organization of the C propeptides of these genes is
identical. These findings support the previous assumption of a close relation between the
\( \alpha \) chains of collagens V and XI. The size of intron 1 of the COL5A1 gene has been
thought to be at least 600 kb, and that of intron 4 is 26 kb (Takahara et al. 1995). This
suggests that the failure to amplify the sequences of introns 1 and 4 of the COL11A1
gene by PCR was caused by their large size.

6.3. COL11A1 and COL2A1 gene mutations in patients with Marshall
or Stickler syndrome

Two mutations in the COL11A1 gene have been identified, one in Marshall syndrome
and the other in Stickler syndrome (Richards et al. 1996, Griffith et al. 1998). The
COL11A1 gene is the only locus for Marshall syndrome identified thus far, but Stickler
syndrome has also been shown previously to be caused by mutations leading to a
premature translation termination in the COL2A1 gene (see Kuivaniemi et al. 1997). A
clinical debate on the identity and diagnosis of Marshall and Stickler syndromes has been
1998). To study the issue by a genetic approach, the COL11A1 and COL2A1 genes in
patients with Marshall or Stickler syndrome were screened for mutations by CSGE.

Altogether 23 mutations were identified, fifteen in the COL11A1 gene and eight in the
COL2A1 gene. The majority of those in the COL11A1 gene altered the splicing
consensus sequences, all of them affecting the 54 bp exons, as in the previously reported
case (Griffith et al. 1998). Furthermore, one patient had a genomic deletion resulting in the
loss of a 54 bp exon. Ten out of these eleven mutations were in a region spanning
exons 38 to 54 of the gene. Although more than one third of the exons in this region are
90 or 108 bp in size, no splicing mutations or deletions affecting them were found. The
only splicing mutation in the N terminal part of the \( \alpha_1(\text{XI}) \) triple helical region caused a
very different phenotype from the other splicing mutations, one in which only severe
ocular symptoms and shortness occurred, with no hearing deficit or facial features other
than the high arched palate typical of the Stickler and Marshall syndromes.

Six of the COL2A1 mutations resulted in a premature translation termination codon, and
two altered the splicing consensus sequences. Fairly good correlation has been
observed between the mutation type in the COL2A1 gene and the phenotype it causes
(see Kuivaniemi et al. 1997). The two patients who had a splicing mutation had features
typical of Stickler syndrome, with no signs of more severe Kniest dysplasia or SED,
which are known to be consequences of splicing mutations leading to in-frame deletions
in the COL2A1 gene. It is therefore likely that the mutations in the splicing consensus
sequences of these patients lead to cryptic splice sites, and thus to premature translation
termination codons, as was also reported in the original kindred described by Dr. Stickler
(Williams et al. 1996).

The finding that mutations in the COL11A1 and COL2A1 genes lead to similar
phenotypes is not surprising. Types II and XI are both fibril-forming collagens and form
components of the same fibril in cartilage. Furthermore, mutations in the COL11A2 gene,
which codes for the \( \alpha_2(\text{XI}) \) chain, have been reported to cause homozygous and
heterozygous OSMED, which has features similar to the Stickler and Marshall syndromes (see Spranger 1998). The phenotypes caused by COL11A2 mutations nevertheless lack ocular involvement, because the \( \alpha_2(V) \) chain rather than the \( \alpha_2(XI) \) chain is expressed in the vitreous body (Mayne et al. 1993). Also, the correlation of the mutational types with the severity of the phenotypes they cause fits well with the amounts of these collagen types in the fibrils. Mutations leading to defective collagen molecules are known to cause more severe phenotypes than mutations causing haploinsufficiency (Kuivaniemi et al. 1997). In the cases described here, the mutations leading to a premature translation termination codon in a quantitatively more significant molecule in the fibrils generated a similar phenotype to the mutations causing expression of defective molecules of a quantitatively less significant fibril component.

The present results indicate that patients with a COL11A1 mutation, especially those who had a splicing mutation or genomic deletion of a 54 bp exon in the C terminal half of the \( \alpha_1(XI) \) molecule, had findings related to Marshall syndrome, while the patients with a COL2A1 mutation had a more classical Stickler phenotype. The genotypic-phenotypic correlation detected here supports the old clinical suspicion of two separate entities. Other mutations in the COL11A1 gene, however, resulted in phenotypes overlapping both the Marshall and Stickler syndromes, and it is these that possibly explain the conflicting reports as to whether Stickler and Marshall syndromes are separate entities or not.

The clinical distinction between these phenotypes provides important information on the prognosis for the patient’s symptoms. Furthermore, if an analysis of the genetic defect causing the phenotype is desired, the clinical distinction helps in choosing the right gene for this analysis. The hearing deficit is probably the most outstanding feature distinguishing between the phenotypes caused by mutations in the COL11A1 and COL2A1 genes. Fourteen out of fifteen patients with a COL11A1 mutation had early-onset hearing loss and required hearing aids, but the patients with a COL2A1 mutation had normal hearing or only a mild hearing defect. Furthermore, ocular involvement was less severe in the patients with COL11A1 mutations, although there were several exceptions. The classification of the ocular symptoms suggested by Snead et al. (1994) could not be performed, because the type of vitreal abnormality was not defined here. Another interesting means of distinction could be the evaluation of facial characteristics. The patients with Marshall syndrome were found to have more clearly pronounced midfacial hypoplasia, a short nose and a flat nasal bridge. These features did not disappear when the child grew older, as can be observed in individuals with Stickler syndrome (Temple 1989, H. Kääriäinen, personal communication).

Cranial abnormalities including abnormal frontal sinuses, intracranial ossifications and/or a thickened calvarium have been described in Marshall syndrome (Marshall 1958, Aymé & Preuss 1984, O’Donnell et al. 1976, Griffith et al. 1998), and although cranial radiographs were available for only a third of the present patients, the findings supported the suggestion that a mutation in the COL11A1 gene may cause abnormalities in cranial ossification. An attractive explanation for this would be the presence of high levels of mRNA for the COL11A1 gene in the calvaria and brain but only low amounts of the mRNAs for the COL11A2 and COL2A1 genes, as observed by Lui et al. (1995).
6.4. The association of a COL9A2 allele with intervertebral disc disease

Six out of 157 patients with intervertebral disc disease carried the Trp for Gln/Arg substitution in the central collagenous domain of the α2(IX) chain, while none of the 174 controls did so. Furthermore, the substitution co-segregated with the phenotype in the families studied. The statistical analyses provided extremely strong evidence that the Trp allele was associated with intervertebral disc disease.

Even though the Trp allele of the COL9A2 gene was tightly linked to the phenotype, the results do not prove a direct causal role for it in the aetiology of the disease, as it is possible that some other gene in close proximity to the COL9A2 gene that is in linkage disequilibrium with the Trp allele is the true disease locus. A role for collagen IX mutations in intervertebral disc disease has nevertheless been suggested in experiments with transgenic mice, where long-term follow-up of the mice expressing a mutant Col9a1 gene showed intervertebral disc degeneration and herniations (Kimura et al. 1996). On the other hand, the only known disease-causing collagen IX mutations in humans lead to MED, which is characterized primarily by articular complaints. However, only two mutations in the genes coding for collagen IX that cause MED have been reported, one in the COL9A2 gene and one in the COL9A3 gene, and both lead to similar deletions of 12 amino acids at exactly same location in the COL3 domain of the molecule, suggesting the importance of this domain for the pathogenesis of MED (Muragaki et al. 1996, Paassilta et al. 1999a). This does not contradict the present results, as it is well known that different mutations in the same collagen gene may lead to very different phenotypes. A mutation in the COL2A1 gene, for example, may lead to a cartilage disorder without ocular involvement, while another mutation in the same gene leads to a disorder with only ocular manifestations (Ala-Kokko et al. 1990, Körkkö et al. 1993). Furthermore, a 9 bp deletion in the region coding for the COL1 domain of collagen IX that represents a neutral variant without any phenotypic consequences has been identified in the COL9A3 gene (Paassilta et al. 1999b). Thus it would not be surprising for a splicing mutation in the COL3 domain and a Trp substitution in the COL2 domain of the collagen IX molecule to lead to different phenotypes.

The Trp substitution was the only putative disease-causing sequence variation in the COL9A2 gene among the patients whose phenotype was linked to the gene. Even though there might be another such mutation in an intron or in the promoter region of the gene, or alternatively, the Trp substitution or the presumably neutral polymorphisms detected in the gene might cause aberrant splicing, there are several arguments to suggest that the Trp substitution itself is the cause of the disease. Trp is rarely found in the triple helices of collagens, and there are no Trp residues in any of the collagenous domains of human or mouse collagen IX (Muragaki et al. 1990, Perälä et al. 1993, Rokos et al. 1994, Perälä et al. 1994, Brewton et al. 1995), which suggests that it fits poorly into the collagen triple helix. Computerized molecular modelling of the sequence with Trp and the normal sequence indicated that the substitution alters the conformation of the helix (A. Fertala, unpublished data). It could also interfere with the covalent cross-link between the collagen II and collagen IX molecules. In the triple helical heterotrimer, the Lys residue in the α3(IX) chain that is involved in the cross-link is located only three amino acids away from the Trp substitution in the α2(XI) chain.
Studies on disc tissue from a patient carrying the Trp allele could perhaps reveal the mechanism by which the Trp causes the disease, but disc tissue from such patients is difficult to obtain for this purpose, at least in sufficient amounts. Preliminary results of *in vitro* studies performed with normal and Trp-carrying recombinant human collagen IX expressed in insect cells and recombinant human collagen II expressed in a transfected tumour cell line (HT1080) nevertheless indicate that collagen IX with the Trp allele has a higher affinity for collagen II than does normal collagen IX (S. Annunen, A. Fertala, D.J. Prockop, L. Ala-Kokko, unpublished data).

The co-existence of the Trp allele and two other sequence variants indicated that the individuals with the Trp allele had inherited the same, relatively rare, ancestral haplotype. Interestingly, the same haplotype was also found in a Chinese immigrant, but he was not included in the series because an isolated Finnish population was thought to be a more accurate target for linkage disequilibrium analysis. This finding does indicate, however, that the haplotype with the Trp is not restricted to this genetically isolated Finnish population.

Even though all the individuals with the Trp allele had intervertebral disc disease, there were some family members who had the disease but not the Trp allele. Intervertebral disc disease is common in the Finnish population, with a prevalence of 4.8%, and the Trp allele was found in about 4% of the patients, which leaves the remaining 96% of the cases to be explained by other genetic and environmental factors. Thus, it is not surprising that there are more than one aetiologies explaining the phenotype, even within one family.

Although it is known that intervertebral disc disease has many environmental and anthropometric risk factors (see Heliövaara 1989), the present results indicate a role for genetic factors as well. This suggestion is supported by reports of familial intervertebral disc disease in the literature (Varlotta *et al.* 1991, Scapinelli 1993, Battié *et al.* 1995, Sambrook *et al.* 1999).
7. References


