IDENTIFICATION AND CHARACTERIZATION OF NOVEL COLLAGEN CHAINS

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Sudheer Kumar Gara

aus Visakhapatnam, Indien

Köln, 2010

REFEREES/ BERICHTERSTATTER

1. Berichterstatter: Prof. Dr. med. Mats Paulsson
2. Berichterstatter: Prof. Dr. med. Jens Brüning

Vorsitzender:
Prof. Dr. rer. nat. Siegfried Roth

Tag der mündlichen Prüfung: 20.04.2010

DEDICATED TO

MY BELOVED PARENTS

INDEX

Abbreviations

Amp	Ampicillin
AD	Atopic dermatitis
AP	Alkaline phosphatase
BM	Bethlem myopathy
bp	Basepair
cDNA	Complementary DNA
DEPC	Diethylpyrocarbonate
DIG	Digoxigenin
DISH	Diffuse idiopathic skeletal hyperostosis
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DPX	Distyrene Plasticizer Xylene
DTT	Dithiothreitol
DVWA	Dual von Willebrand factor A
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
ES	Embryonic stem cells
FCS	Fetal calf serum
G. gallus	*Gallus gallus*
h	hour
H. sapiens	*Homo sapiens*
HRP	Horse radish peroxidase
kDa	Kilodalton
LB	Luria Bertani
M. musculus	*Mus musculus*
MIDAS	Metal ion-dependent adhesion site
min	Minute
mRNA	messenger Ribonucleic acid

OA	Osteoarthritis
OPLL	Ossification of the posterior longitudinal ligament
P. troglodytes	*Pan troglodytes*
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
RNA	Ribonucleic acid
RT	Room temperature
SDS	Sodium dodecyl sulfate
sec	Second
SaOS-2	Sarcoma osteogenic cells (human)
TBS	Tris buffered saline
Tris	Tris (hydroxymethyl) aminomethane
rpm	Revolutions per minute
U	Units
UCMD	Ullrich congenital muscular dystrophy
VWA	Von Willebrand factor A

1 INTRODUCTION

All living organisms are made up of cells. Organisms that exist as single cells are called unicellular while those made of a group of cells are referred to as multicellular. In multicellular organisms, most cells are surrounded by a complex network of non-cellular material 'ExtraCellularMatrix' (ECM). In humans, the ECM accounts for more than 80% of the total body mass whereas the cell content constitutes only 20%. The ECM is composed of three major classes of biomolecules. The first and predominant class is comprised of structural proteins such as collagen and elastin. Proteoglycans and specialized non-collagenous proteins like fibrillin, fibronectin and laminin form the second and third category of ECM molecules. The focus of this study is on the members of the first class of proteins, the collagens.

1.1 Collagen

The word collagen was derived from the Greek word '*kolla*' which means 'glue producer.' Collagens are the most abundant proteins in the human body constituting about 30% of its protein mass. They provide the major organic component with 90% of dry mass in bone and 60% in cartilage. Hence, these proteins offer the structural framework of bone and cartilage and are responsible for shape and biomechanical properties such as resistance to pressure, torsion and tension. The defining feature of a collagen molecule is the sequence motif XaaYaaGly. In 1954, Ramachandran and Kartha first proposed the three-stranded helical model for collagens (Ramachandran & Kartha, 1954) that was later improved to give the generally accepted supercoiled triple helical structure (Rich and Vrick, 1961, Ramachandran *et al.*, 1967 and Fraser *et al.*, 1979). It is now known that the collagen structure is made up of three parallel polypeptide strands in a left-handed, polyproline II-type (PPII) helical conformation, which then coil around each other with a one-residue stagger to form a right-handed triple helix (Figure 1.1). The tight packing of PPII helices within the triple helix requires that every third residue is a Gly, resulting in a XaaYaaGly repeat, where Xaa and Yaa can be any amino acid (Shoulders & Raines, 2009). In 1994, Berman and colleagues (Bella *et al.*, 1994) reported the first high-resolution crystal structure of a triple-helical collagen-related peptide (CRP)

(Figure 1.1a). This structure confirmed the existence of inter-strand N-$H_{(Gly)}$…O=$C_{(Xaa)}$ hydrogen bonds (Figure 1.1c,d) and provided additional insights into the C^{α}-$H_{(Gly/Yaa)}$…O=$C_{(Xaa/Gly)}$ hydrogen bonds that could likewise stabilize the triple helix (Bella & Berman, 1996). Owing to this structural framework, all collagens, or rather collagenous domains, possess the unique XaaYaaGly repeat.

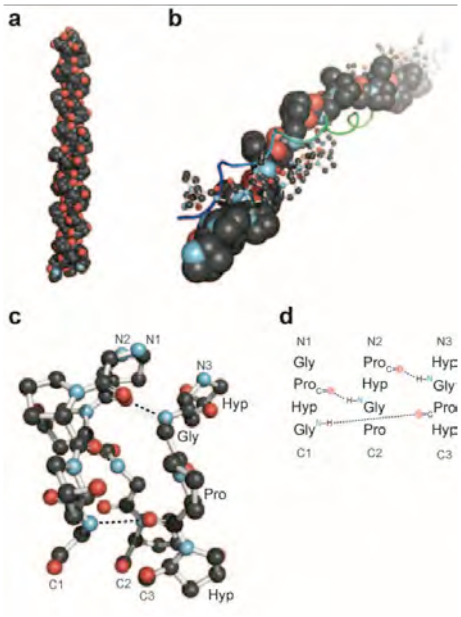


Figure 1.1 **Overview of the collagen triple helix**. (a) First high-resolution crystal structure of a collagen triple helix, formed by $(ProHypGly)_4$-$(ProHypAla)_5$ [PDB-1cag] (Bella *et al.*, 1994). (b) View down the axis of a $(ProProGly)_{10}$ triple helix [PDB-1k6f] with the three strands depicted in space-filling, ball-and-stick, and ribbon representation (Berisio *et al.*, 2002). (c) Ball-and-stick image of a segment of collagen triple helix [PDB-1cag], highlighting the ladder of inter-strand hydrogen bonds. (d) Stagger of the three strands in the segment in panel c. *Adapted from* (Shoulders & Raines, 2009)

1.2 Classification of collagens

The vertebrate family of collagens is encoded by more than 43 genes, which yield at least 28 distinct trimeric collagens (Shoulders & Raines, 2009; Veit *et al.*, 2006). Most collagens form supramolecular assemblies and depending on the structure of these assemblies, the collagen superfamily can be divided into several subfamilies such as fibril forming, network forming, anchoring fibril forming, FACITs (Fibril-associated collagens with interrupted triple helices), MACITs (Membrane associated collagens with interrupted triple helices) and MULTIPLEXINs (Multiple triple helix domains and interruptions) (Figure 1.2 & Table 1.1). In addition, there is also a group of highly heterogenous proteins that contain collagenous domains but are not considered as collagens as they do not primarily fulfil a structural function. The subcomponent C1q of complement, a C1q-like factor, adiponectin, at least eight collectins and three ficolins,

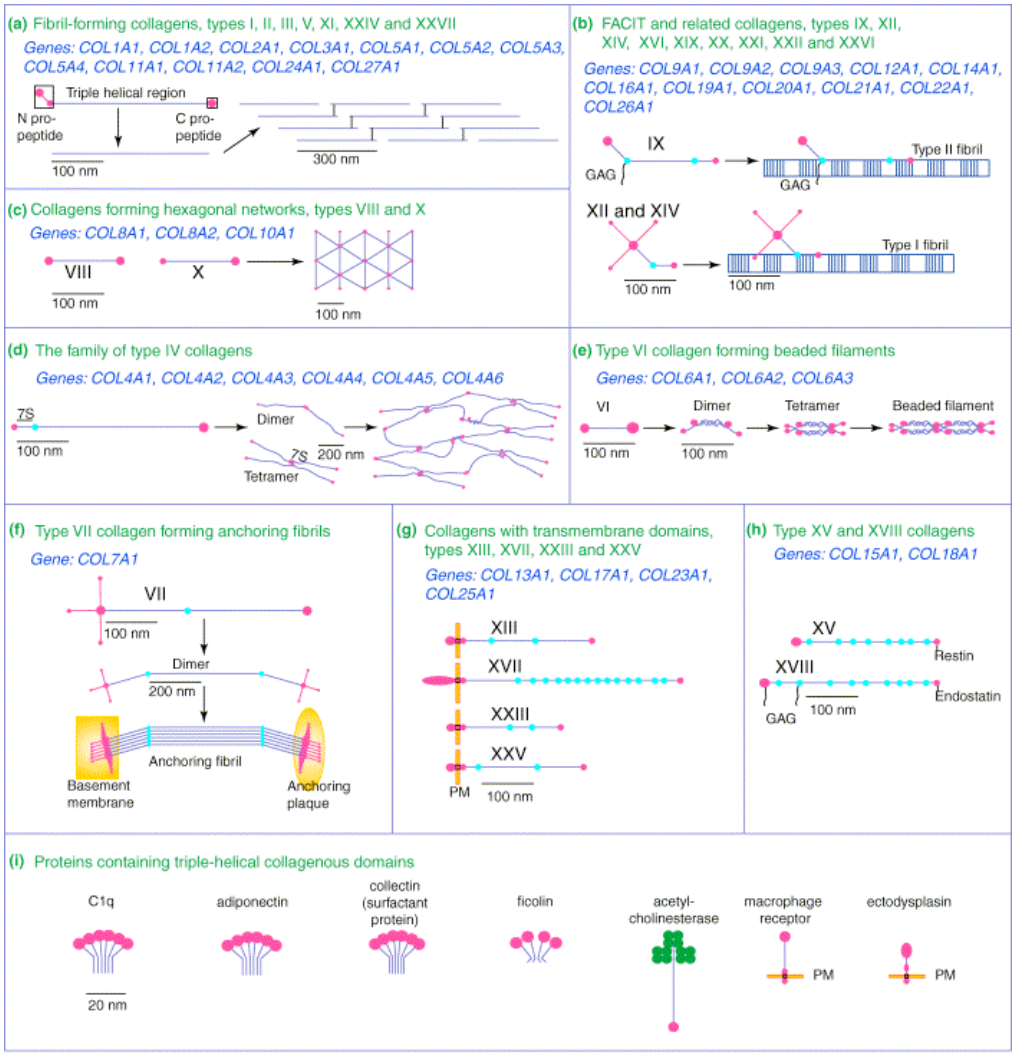


Figure 1.2 **Members of the collagen superfamily and their known supramolecular assemblies**. The collagen superfamily can be divided into nine families on the basis of the supramolecular assemblies and other features of its members: **(a)** fibril-forming collagens; **(b)** fibril-associated collagens with interrupted triple helices (FACITs) located on the surface of fibrils, and structurally related collagens; **(c)** collagens forming hexagonal networks; **(d)** the family of type IV collagens located in basement membranes; **(e)** type VI collagen, which forms beaded filaments; **(f)** type VII collagen, which forms anchoring fibrils for basement membranes; **(g)** collagens with transmembrane domains; and **(h)** the family of type XV and XVIII collagens. The supramolecular assemblies of families **(g)** and **(h)** are unknown and are therefore not shown in this figure. The polypeptide chains found in the 28 collagen types are encoded by 43 genes in total (shown in blue), each molecule consisting of three polypeptide chains that can be either identical or different. An additional highly heterogenous group **(i)** within the superfamily comprises proteins that possess collagenous domains but have not been defined as collagens. Some of these **(i)** proteins could also be defined as collagens, although they do not have a structural function. The collagen domains are shown in purple, the N and C-terminal non-collagenous domains are in dark pink, and the non-collagenous domains interrupting the triple helix in light blue, short interruptions of a few amino acids are not shown. For acetylcholinesterase, the catalytic domain (shown in green) and the tail structure are products of separate genes. Adapted from: (Myllyharju & Kivirikko, 2004) Abbreviation: PM, plasma membrane.

Type	Class	Composition	Distribution	Pathology
I	Fibrillar	$\alpha1[I]_2\alpha2[I]$	Abundant and widespread: dermis, bone, tendon, ligament	OI, Ehlers–Danlos syndrome, osteoporosis
II	Fibrillar	$\alpha1[II]_3$	Cartilage, vitreous	Osteoarthrosis, chondrodysplasias
III	Fibrillar	$\alpha1[III]_3$	Skin, blood vessels, intestine	Ehlers-Danlos syndrome, arterial aneurysms
IV	Network	$\alpha1[IV]_2\alpha2[IV]$ $\alpha3[IV]\alpha4[IV]\alpha5[IV]$ $\alpha5[IV]_2\alpha6[IV]$	Basement membranes	Alport syndrome
V	Fibrillar	$\alpha1[V]_3$ $\alpha1[V]_2\alpha2[V]$ $\alpha1[V]\alpha2[V]\alpha3[V]$	Widespread: bone, dermis, cornea, placenta	Ehlers-Danlos syndrome
VI	Network	$\alpha1[VI]\alpha2[VI]$ $\alpha3[VI]$	Widespread: bone, cartilage, cornea, dermis	Bethlem myopathy and Ullrich congenital muscular dystrophy
VII	Anchoring fibrils	$\alpha1[VII]_2\alpha2[VII]$	Dermis, bladder	Epidermolysis bullosa acquisita
VIII	Network	$\alpha1[VIII]_3$ $\alpha2[VIII]_3$ $\alpha1[VIII]_2\alpha2[VIII]$	Widespread: dermis, brain, heart, kidney	Fuchs endothelial corneal dystrophy
IX	FACIT	$\alpha1[IX]\alpha2[IX]\alpha3[IX]$	Cartilage, cornea, vitreous	Osteoarthrosis, multiple epiphyseal dysplasia
X	Network	$\alpha1[X]_3$	Cartilage	Chondrodysplasia
XI	Fibrillar	$\alpha1[XI]\alpha2[XI]\alpha3[XI]$	Cartilage, intervertebral disc	Chondrodysplasia, osteoarthrosis
XII	FACIT	$\alpha1[XII]_3$	Dermis, tendon	—
XIII	MACIT	—	Endothelial cells, dermis, eye, heart	—
XIV	FACIT	$\alpha1[XIV]_3$	Widespread: bone, dermis, cartilage	—
XV	MULTIPLEXIN	—	Capillaries, testis, kidney, heart	—
XVI	FACIT	—	Dermis, kidney	—
XVII	MACIT	$\alpha1[XVII]_3$	Hemidesmosomes in epithelia	Generalized atrophic epidermolysis bullosa
XVIII	MULTIPLEXIN	—	Basement membrane, liver	Knobloch syndrome
XIX	FACIT	—	Basement membrane	—
XX	FACIT	—	Cornea (chick)	—
XXI	FACIT	—	Stomach, kidney	—
XXII	FACIT	—	Tissue junctions	—
XXIII	MACIT	—	Heart, retina	—
XXIV	Fibrillar	—	Bone, cornea	—
XXV	MACIT	—	Brain, heart, testis	Amyloid formation?
XXVI	FACIT	—	Testis, ovary	—
XXVII	Fibrillar	—	Cartilage	—
XXVIII	—	—	Dermis, sciatic nerve	Neurodegenerative disease?

Table 1.1 **Distribution, composition and pathology of vertebrate collagens.**
Modified from (Shoulders & Raines, 2009)

acetylcholinesterase, three macrophage receptors, ectodysplasin, two EMILINs (elastic fiber-associated glycoproteins and a src- homologous-collagen protein) (Figure 1.2i) (Lu *et al.*, 2002; Franzke *et al.*, 2003; Myllyharju & Kivirikko, 2001) fall in this category. However, the largest group of collagens is formed by the classical fibril forming and FACIT collagens. Some collagens have a very restricted tissue distribution while others are ubiquitously present in almost every tissue. A summary of different collagen types, their distribution and pathology is given in Table 1.1 (Shoulders & Raines, 2009). Although the three polypeptide chains in the triple helix of each collagen type can be identical, heterotrimeric helices are more prevalent. Moreover, collagen fibrils often consist of more than one collagen type. For example, collagen I fibrils contain small amounts of collagen III, V and XII, whereas the collagen II fibrils contain collagen IX and XI. Besides this, collagen V can also form hybrid molecules with collagen XI, *i.e.* the α1(XI) and α2(V) chains can be present in the same heterotrimeric collagen molecule (Myllyharju & Kivirikko, 2004). Some mature collagens also possess non-collagenous domains in addition to the actual triple helical collagen domains.

1.3 Von Willebrand factor A like domain (VWA)

The von Willebrand factor is a large multimeric glycoprotein that is found in blood plasma, platelet α-granules and sub-endothelial connective tissue. In von Willebrand factor, the type A domain (VWA) is the prototype for a protein superfamily. The VWA

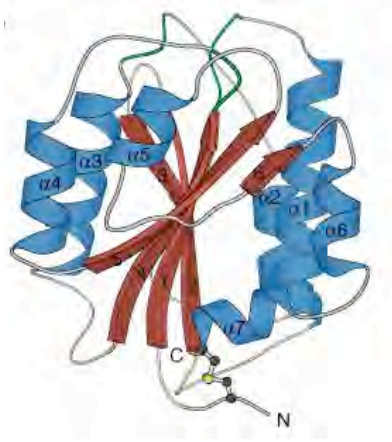


Figure 1.3 **Structure of the vWF A3 domain**. Ribbon drawing with the cysteine bridge that links the N- and C-terminal parts of the molecule shown in ball-and-stick representation. Loop segments that contribute residues to the potential ion-binding site are coloured green, *α* helices are blue and *β* strands red. (Huizinga *et al.*, 1997)

domain is found in various plasma proteins, *e.g.* complement factors B, C2, CR3 and CR4, integrins (I-domains), matrilins, collagen VI, VII, XII, XIV, XXI and XXII and

other extracellular proteins (Whittaker & Hynes, 2002). Some intracellular proteins, like the copines, were also identified as members of the family (Tomsig and Creutz, 2002). Proteins that contain a VWA domain are involved in numerous biological events such as cell adhesion, migration, homing, pattern formation and signal transduction by interacting with a large array of ligands (Colombatti *et al.*, 1993). The typical domain is made up of about 200 amino acid residues and adopts a classic α/β Rossmann fold that is stabilized by cysteine derived disulfide bonds (Figure 1.3). The domain often contains a highly conserved $DxSxSx_nTx_nD$ sequence motif, also referred to as metal ion dependent adhesion site (MIDAS) at its surface (Bienkowska *et al.*, 1997; Huizinga *et al.*, 1997; Lee *et al.*, 1995; Sadler, 1998).

1.4 Collagen VI

Collagen VI is an extracellular matrix protein that is present in almost all connective tissues and is known to form a structurally unique microfibrillar network in close association with basement membranes. It was previously called 'short-chain collagen', 'high molecular weight aggregate' or 'intima collagen' (Chung *et al.*, 1976; Furthmayr *et al.*, 1983). Chung et al., (1976) first isolated collagen VI from human aortic intima after limited pepsin digestion. Thereafter several groups isolated collagen VI in slightly different ways from human and bovine placenta (Jander *et al.*, 1981; Furuto & Miller, 1981; Furuto & Miller, 1980; Odermatt *et al.*, 1983; Jander *et al.*, 1983). Collagen VI was

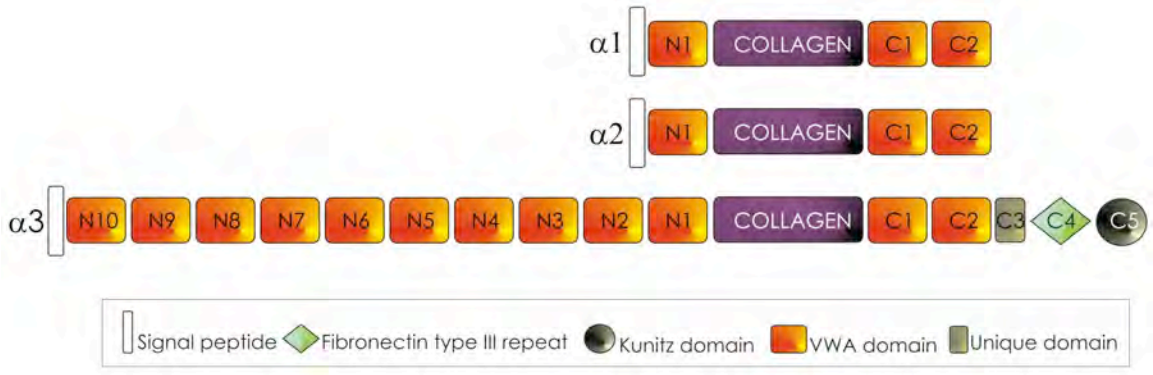


Figure 1.4 **Domain structures of the collagen VI chains.** The domain composition of the collagen VI α1, α2 and α3 chains. The N and C terminal globular domains to the collagenous domain are numbered as in (Chu *et al.*, 1988)

long thought to be composed of only three different polypeptide chains α1(VI), α2(VI) and α3(VI). In humans, the α1(VI) and α2(VI) chains are encoded by two genes *COL6A1* and *COL6A2* that are arranged in a head to tail fashion on chromosome 21q22.3 (NT_011515) and are separated by 150 kb of genomic DNA (Heiskanen *et al.*, 1995). The α3(VI) chain is encoded by C*OL6A3*, mapped to chromosome 2q37 (NT_005120) (Weil *et al.*, 1988). Similarly, the α1(VI) and α2(VI) chains in mouse are encoded by two genes *Col6a1* and *Col6a2* that are located on chromosome 10, 41.1cM, while the *Col6a3* is situated on chromosome 1, 53.9cM. The collagen VI genes are known to be highly polymorphic and more than 25 non-synonymous polymorphic amino acid changes have been described to date (Pan *et al.*, 2003; Lampe *et al.*, 2005). The collagen VI α1 and α2 chain each have a molecular mass of about 120 kDa while the collagen VI α3 chain has a mass of 260 kDa. All three collagen VI chains contain two C-terminal VWA domains, whereas at the N-terminus the α1 and α2 chains carry only one but the α3 chain has ten VWA domains (Chu *et al.*, 1989; Chu *et al.*, 1990) (Figure 1.4). In addition, the collagen VI α3 chain contains a unique domain with similarities to salivary gland proteins, a short proline rich repeat, a fibronectin type III repeat and a bovine pancreatic trypsin inhibitor/Kunitz family of serine protease inhibitor domain (Kunitz domain) at the C-terminus (Chu *et al.*, 1989). Unlike the classical collagen molecules, the triple helical collagenous domain of collagen VI is short and about 335-336 amino acid residues long. Another noticeable feature in the collagenous domain of collagen VI is the presence of RGD motifs. A total of 13 RGD sites could be detected in all the three collagen VI α1, α2 and α3 chains (Chu *et al.*, 1988; Saitta *et al.*, 1991). Collagen VI is known to anchor large interstitial structures such as nerves, blood vessels and collagen fibrils into the surrounding connective tissue. It is also involved in cell migration and differentiation and may play a role in bridging the cells with extracellular matrix.

1.4.1 Alternative splicing of collagen VI

COL6A1 consists of 37 exons (35 coding) with a single promoter and produces a single transcript encoding a protein of 1021 amino acids while *COL6A2* is described to possess 30 exons (29 coding) and two different promoters. It has also been shown that *COL6A2* produces multiple alternatively spliced mRNAs which differ in the 5'- and 3'-

untranslated regions as well as in the 3'- coding part producing at least three collagen VI α2 variants with distinct carboxy termini but retaining the same domain architecture. The *COL6A3* gene comprises of 44 exons (43 coding) and is known to undergo extensive alternative splicing primarily at the N- terminal globular domains in a tissue specific manner, particularly in humans and mouse but not in chicken. However, the three most abundant transcripts lack the N10 and N9 domains but contain N8/N7/N6, N8/N6 or N6 domains. In mouse, transcripts containing the N10 domain are the least abundant in all tissues studied. There are also reports showing the presence of transcripts lacking a large part of N3 domain and also transcripts lacking the entire N5 domain (Dziadek *et al.*, 2002).

1.4.2 Macromolecular assembly of collagen VI

The assembly of collagen VI is a complex multistep process. Several hypotheses have been postulated, however the most widely accepted model is deduced from

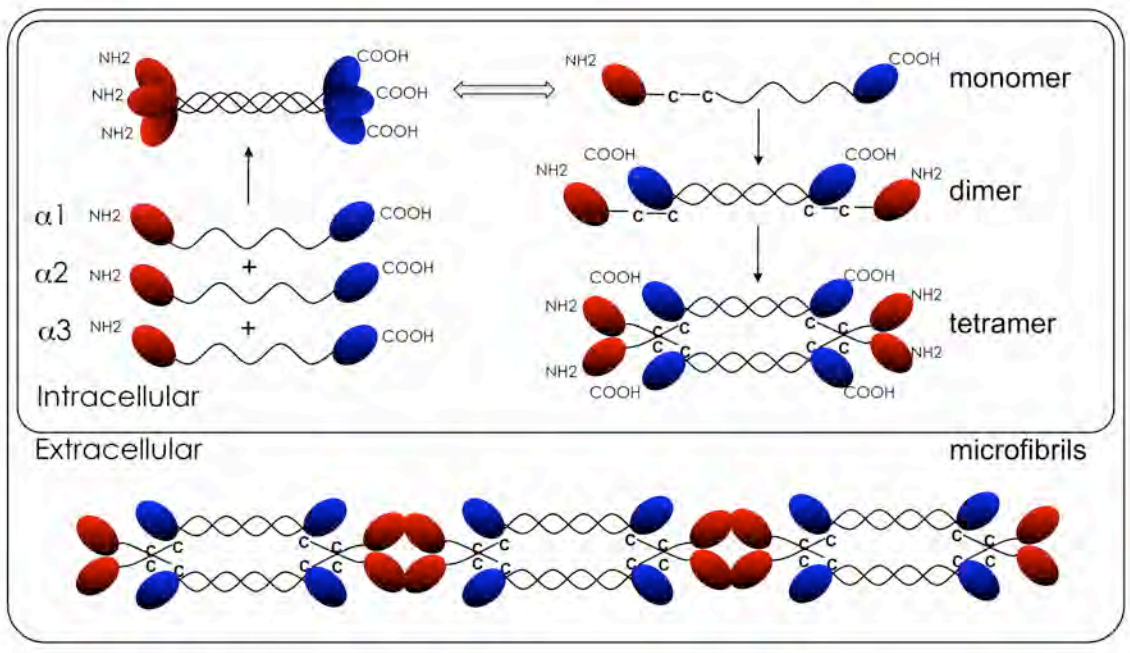


Figure 1.5 **Schematic model of the collagen VI assembly.**
NH2 (red) and COOH (blue) are the amino and carboxy terminal end of the N- and C-terminal globular domains respectively. The letter 'C' represents the cysteine bridges between the collagenous domain and globular domains. Based on (Furthmayr *et al.*, 1983) and (Zhang *et al.*, 2002).

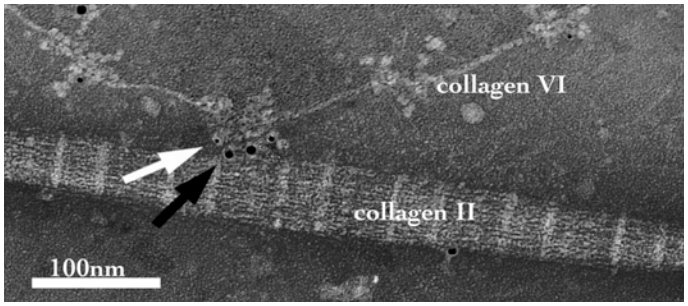


Figure 1.6 **Electron micrograph of collagen VI microfibrils.** Double staining with gold particles of different sizes located complexes of biglycan (white arrow, small gold) and matrilin-1 (black arrow, large gold) between collagen VI microfibrils and striated collagen II fibrils. (Wiberg *et al.*, 2003)

extensive biochemical and rotary shadowing electron microscopy experiments. The three genetically distinct collagen VI subunits α1, α2 and α3 form a triple helical monomer followed by a staggered assembly into disulfide bonded antiparallel dimers. Electron microscopic studies on isolated dimers suggested that they pack with a molecular overlap of ~75 nm leaving non-overlapping single molecular ends of ~30 nm (Furthmayr *et al.*, 1983). Subsequently, the dimers align to form tetramers that are also stabilized by disulfide bonds (Figure 1.5). The tetramers are then secreted outside the cell to form long molecular chains known as microfibrils that have a beaded repeat of 105nm (Figure 1.6) (Furthmayr *et al.*, 1983; Chu *et al.*, 1988). Several mutational studies suggest that the collagen VI triple helix folds from the C to the N-terminus, with folding being nucleated by C-terminal Gly-Xaa-Yaa triplets with a high proportion of prolines in Y position which have the potential of being hydroxylated by prolyl-4-hydroxylase (Lamande *et al.*, 2002). The single cysteine residue located in the triple helical domain of the collagen VI α1 or α2 chain is known to interact with a cysteine residue in the C-terminal globular domain and is thought to be responsible for the assembly/stability of dimers (Figure 1.5) (Colombatti *et al.*, 1995; Furthmayr *et al.*, 1983; Chu *et al.*, 1988). In addition, the cysteine residue in the triple helical domain of the collagen VI α3 chain appears to be involved in tetramer formation and stability. Negative staining electron microscopic studies together with the identification of positions of hydrophobic patches along and around the triple helix of the monomers suggest that the dimers are further stabilized by supercoiling of the overlapping triple helices (Furthmayr *et al.*, 1983). The role of the N-terminally extended region containing N10-N6 domains of the collagen VI α3 chain that is highly prone to the alternative splicing in the assembly of collagen VI microfibrils is not clear. Transfection experiments suggested that these domains are not required for

either dimer or tetramer formation in the assembly of collagen VI (Lamande *et al.*, 1998b; Fitzgerald *et al.*, 2001).

1.4.3 Interaction partners of collagen VI

It has been shown that collagen VI interacts with several other extracellular matrix components, including collagen I (Bonaldo *et al.*, 1990), II (Bidanset *et al.*, 1992) and XIV (Brown *et al.*, 1994), perlecan (Tillet *et al.*, 1994), tenascin-X (Minamitani *et al.*, 2004b; Minamitani *et al.*, 2004a), and the microfibril associated glycoprotein MAGP1 (Finnis & Gibson, 1997) as well as at the cell surface with $\alpha 1\beta 1$ integrin (Loeser, 1997). The N-terminal globular domains of the collagen VI molecule bind to small leucine-rich repeat proteoglycans like decorin and biglycan, which in turn interact with matrilins, mediating contacts to further binding partners (Wiberg *et al.*, 2003; Wiberg *et al.*, 2001). Collagen VI also interacts with other matrix constituents such as hyaluronan, heparan sulfate and NG-2 proteoglycans via its N-terminal domains (Burg *et al.*, 1996; Specks *et al.*, 1992).

1.4.4 Human disorders associated with collagen VI

As collagen VI is one of the important extracellular matrix components maintaining the structural integrity of skeletal muscle, mutations in the genes encoding any of the three collagen VI $\alpha 1$, $\alpha 2$ and $\alpha 3$ chains have been shown to cause serious muscular disorders. The two major diseases associated with collagen VI in humans are Bethlem myopathy (MIM 158810) and Ullrich congenital muscular dystrophy (MIM 254090).

1.4.4.1 Bethlem Myopathy (BM)

In 1976, Bethlem and van Wijngaarden discovered an autosomal dominantly inherited muscular disorder in 28 individuals of three Dutch pedigrees (Bethlem & Wijngaarden, 1976). It is characterized by the combination of proximal muscle weakness and variable contractures affecting most frequently the long finger flexors, elbows and ankles. The onset of BM may be prenatal (characterized by decreased fetal movements), neonatal (hypotonia or torticollis), in early childhood (delayed motor milestones, muscle weakness and contractures), or in adulthood (proximal weakness or Achilles tendon or long finger

flexor contractures). Respiratory involvement is rare and seems to be related to more severe muscle weakness in later life. In most cases, the condition is slowly progressive and more than two thirds of the patients over 50 years of age require aids for ambulation, especially outdoors. In some BM patients, unusual skin features including follicular hyperkeratosis and keloid formation or "cigarette paper" scarring may be present (Pepe *et al.*, 2002b). BM may resemble Emery-Dreifuss muscular dystrophy (EMD), particularly when a rigid spine is present (Pepe *et al.*, 2002a; Pepe *et al.*, 2002b). However, it can be well differentiated from EMD and other similar pathologies such as sarcoglycanopathy, calpainopathy and dysferlinopathy by immunohistochemical (western blotting and immunohistochemistry) testing on muscle biopsies. The histopathological features of a typical BM muscle biopsy are non-specifically myopathic but usually consist of marked variation in muscle fiber diameter with possible increase of fatty tissue and occasional necrotic or dystrophic fibers (Bethlem & Wijngaarden, 1976; Merlini *et al.*, 1994).

Various mutations disrupting the Gly-Xaa-Yaa motif of the triple helical domain in *COL6A1*, *COL6A2* or *COL6A3* have been identified in BM patients (Pepe *et al.*, 1999a; Scacheri *et al.*, 2002; Lampe *et al.*, 2005). In general, the mutations towards the N-terminal region of the triple helix induce kinks in the tetramers, reducing their ability to form microfibrils and also exerting a dominant negative effect (Lamande *et al.*, 2002). The second most frequent group of mutations in BM are the splice site mutations that result in the skipping of *COL6A1* exon 14 leading to the in-frame deletion of 18 amino acids from the triple helical domain (Pepe *et al.*, 1999b; Pan *et al.*, 2003; Lampe *et al.*, 2005; Pepe *et al.*, 1999a; Jobsis *et al.*, 1996; Lamande *et al.*, 1999). As a consequence, the short collagen VI α1 chains may form intracellular monomers but are unable to assemble into dimers and tetramers due to the lack of a unique cysteine residue resulting in half of the normal amount of collagen VI being deposited. Studies on some splice site mutations such as in the C1 domain of *COL6A2* suggested that the pathogenicity is due to nonsense mediated mRNA decay (Lamande *et al.*, 1998a; Vanegas *et al.*, 2002). Considering the highly polymorphic nature of the collagen VI genes, the pathogenicity of missense mutations within the triple helical domain, without confirmatory evidence, has to be interpreted with caution (Pan *et al.*, 2003; Lampe *et al.*, 2005).

1.4.4.2 Ullrich Congenital Muscular Dystrophy (UCMD)

The disease was first described by Ullrich in 1930 and displays an autosomal recessive inheritance and a recognizable pattern of symptoms. The hallmarks of UCMD are muscle weakness of early onset with proximal joint contractures and striking hyperelasticity of the distal joints. Posteriorly protruding calcanei are commonly seen. Due to the profound weakness, most of the children either never achieve the ability to walk independently or walk independently for short periods only (Mercuri *et al.*, 2002; Voit, 1998). With progression of the disease a development of spinal rigidity and scoliosis as well as variable proximal contractures can be seen. On the other hand, the distal hyperlaxity can give way to marked long finger flexion contractures and tight Achilles tendons. Respiratory failure in the first or second decade is a common cause of death unless treated with nocturnal respiratory support. However, cardiac involvement has not been reported to date (Pepe *et al.*, 2002a; Mercuri *et al.*, 2002) and intelligence is normal in these individuals. Other distinctive features observed in UCMD patients are congenital hip dislocations (Furukawa & Toyokura, 1977; Mercuri *et al.*, 2002) and a transient kyphotic deformity at birth as well as follicular hyperkeratosis over the extensor surfaces of upper and lower limbs, soft velvety skin on the palms and soles and the tendency to keloid or 'cigarette paper' scar formation (Mercuri *et al.*, 2004).

In UCMD patients, most of the mutations reported appear to result in premature termination codons and follow the well-known nonsense mediated mRNA decay thereby lacking the mutated chain. The premature termination codons occur either by direct introduction of a stop codon at the genomic level (Lampe *et al.*, 2005; Demir *et al.*, 2002) or through frameshift inducing genomic deletions (Lampe *et al.*, 2005; Higuchi *et al.*, 2001), insertions (Camacho Vanegas *et al.*, 2001), duplications (Lampe *et al.*, 2005) and splice changes (Camacho Vanegas *et al.*, 2001; Ishikawa *et al.*, 2002). Similar to BM patients, missense mutations substituting glycine in the triple helical Gly-Xaa-Yaa motif were observed in UCMD patients (Lampe *et al.*, 2005).

It has to be noted that although BM and UCMD were originally described as separate entities, they share similarities in both the mutation prone regions and pathogenesis of the disease. Today, it becomes more and more evident that the two myopathies are different forms of the same disease that varies in their severity.

1.4.4.3 Ossification of the Posterior Longitudinal Ligament (OPLL)

A recent genome-wide linkage and linkage disequilibrium analyses identified *COL6A1*, as the locus for Ossification of the Posterior Longitudinal Ligament (OPLL) of the spine (Tanaka *et al.*, 2003). OPLL is a subset of bone-forming diseases that is characterized by an ectopic ossification in the spinal ligaments (Tanaka *et al.*, 2003). However, except for the genetic linkage, there is neither biochemical nor immunological evidence suggesting that collagen VI is the cause of the disease.

1.4.4.4 Diffuse Idiopathic Skeletal Hyperostosis (DISH)

DISH is also a skeletal hyperostotic disease and is related to OPLL. It is characterized by ligamentous ossification of the anterolateral side of the spine (Tsukahara *et al.*, 2005). The *COL6A1* was identified as a susceptible gene in DISH patients from Japan but not in the DISH patients from the Czech Republic. As OPLL and DISH share commonality in the disease, some of the DISH patients exhibit OPLL. However, there was a strong allelic association with the same SNP in DISH patients without OPLL suggesting that DISH is also clearly associated with *COL6A1* (Tsukahara *et al.*, 2005).

1.4.4.5 Osteoarthritis (OA)

It has been shown that collagen VI is a minor component of normal human cartilage and the amount of type VI collagen epitopes increases significantly during early stages of osteoarthritis (Swoboda *et al.*, 1998). Collagen VI was shown to be expressed in a zone specific pattern in knee osteoarthritic cartilage (Pullig *et al.*, 1999). However, due to the high interindividual variability, collagen VI is not very precise in the diagnosis of early osteoarthritic lesions when used as the only marker (Swoboda *et al.*, 1999).

1.4.5 Animal models of collagen VI myopathies

Bonaldo and colleagues generated a mouse model where the *Col6a1* gene is specifically inactivated by targeted gene disruption (Bonaldo *et al.*, 1998). Due to the lack of mRNA, *Col6a1* -/- mice do not synthesize a collagen VI α1 polypeptide. As a consequence, no triple helical collagen VI molecules are deposited in the extracellular matrix (Bonaldo *et al.*, 1998). Although there are no obvious phenotypic consequences, histological signs of

myopathy were detected in skeletal muscle of both homo- and heterozygous mutant animals. A major difference between the human and the mouse pathologies was the preferred location of lesions: extensor muscles of limbs in Bethlem myopathy and diaphragm and auxiliary respiratory muscles in the knockout mice. However, the distinct pattern of dystrophic changes in the two species has been attributed to the different obligatory respiratory work rate per unit mass (~5 times lower in human), a condition which would make the respiratory muscles of the mouse more prone to mechanical damage than the limb ones (Stedman *et al.*, 1991). A recent analysis of this model showed a mitochondrial dysfunction as the origin of the pathogenesis of the myopathic phenotype. Interestingly, the ultrastructural muscle defects are partially reversible with the drug cyclosporin A that influences the mitochondrial pore complex (Irwin *et al.*, 2003).

1.4.6 Cell attachment activity of collagen VI

Several members of the collagen family have been identified as strong cell attachment substrates and collagen VI is no exception. Several cell types, including fibroblasts and tumors cells, were able to attach and spread on substrates of pepsin-solubilized or intact collagen VI (Aumailley *et al.*, 1989). The interactions of the unfolded chains were inhibited by low concentrations of synthetic RGD peptides while the binding of cells to pepsin-solubilized collagen VI was more than 20-fold less sensitive to these peptides suggesting that the attachment is dependent on the RGD motif (Aumailley *et al.*, 1989; Pfaff *et al.*, 1993). Antibody inhibition, affinity chromatography and/or ligand binding studies suggested that the binding of the triple helical substrate is mediated by $\alpha 1\beta 1$ and $\alpha 2\beta 1$ integrins whereas binding to denatured substrates is mediated by $\beta 1$ and $\beta 3$ integrins. It was also shown that the cell adhesion is strictly dependent on the divalent cations Mg^{2+} and Mn^{2+} (Gailit & Ruoslahti, 1988; Pfaff *et al.*, 1993). An independent study suggested that $\beta 1$ integrin dependent cell attachment and spreading is mediated by NG2 proteoglycan (Tillet *et al.*, 2002). Collagen VI was shown to be a strong adhesive substrate for various hematopoietic cell lines and light-density bone marrow mononuclear cells (Klein *et al.*, 1995). However, the adhesive site seems to be restricted to the triple helical domain as individual collagen VI chains were not active in the attachment assays

(Klein *et al.*, 1995). The binding could be completely abolished by heparin, indicating that membrane-bound heparan sulfate might be involved in the binding process (Klein *et al.*, 1995).

1.5. Collagen XXVIII

The most recent member of the collagen superfamily was identified by screening the EST database and subsequently characterized as an authentic tissue component (Veit *et al.*, 2006). The human and mouse collagen XXVIII genes map to syntenic regions on chromosomes 7 (7p21.3) and 6(6A1) respectively. The exon/intron organization of the two genes is very similar and the genes contain 34 exons coding for the translated part of the mRNA (Figure 1.7). The mature secreted protein in human and mouse has a calculated Mr of 113,917 and 116,414 or 116,095, respectively. The collagenous domain is about 528 amino acids long and contains 12 GXG and four GXXXXG imperfections that are uniformly distributed along the sequence. It has one VWA domain on either side of the triple helical collagenous domain. Besides this, it has one unique domain and a domain related to the bovine pancreatic trypsin inhibitor/Kunitz family of serine protease inhibitors at the C-terminus. Collagen XXVIII has a quite restricted tissue distribution and is present at dorsal root ganglia with strong expression in the peripheral nerve fibers, which is unusual for a collagen. In addition, a staining was seen in parts of the sciatic nerve of an adult mouse.

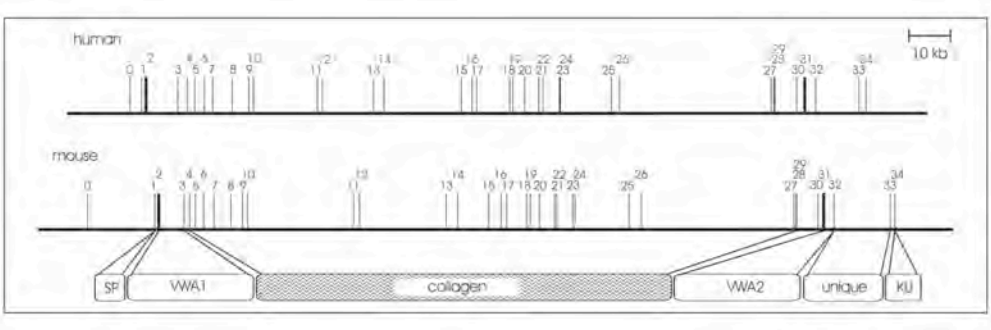


Figure 1.7 **Structure of collagen XXVIII gene in human and mouse.** The exons are numbered and marked with straight lines. The domain boundaries and their coding exons are shown below.

Collagen XXVIII is not only a new collagen but also forms a separate subgroup, as it cannot clearly be assigned to any known collagen subgroup. Although it has some

similarities to collagen VI, it is very unlikely that heterotrimeric molecules can be formed together with collagen VI chains due to the long collagenous domain of collagen XXVIII. Although there are so far not any disease-causing mutations associated with collagen XXVIII, based on the available expression data it could be a potential target for neurodegenerative diseases.

1.6 Aims of the thesis

Despite of the fact that many UCMD and BM patients display no mutations in the known collagen VI chains, there were no reports pointing to the existence of additional collagen VI chains. The identification of the three novel *COL6* genes could provide insight into the pathogenesis of collagen VI associated disorders. At the same time this discovery raises the complexity in the macromolecular assembly of collagen VI. Therefore, it was important to provide in depth knowledge on the characteristic features of these new chains.

The primary aims of my dissertation were:

1. To analyze the characteristic features of gene and protein sequences of new collagen VI chains in mouse and human
2. To examine the gene and protein expression as well as the localization of the new chains in mouse and human
3. To study the fate of the new chains in the collagen VI α1 deficient mouse model and in human UCMD and BM patients.

Collagen XXVIII, another recently identified novel collagen forms a separate subgroup within the collagen superfamily. On the basis of its predominant expression in the nerve tissue, it was proposed that collagen XXVIII may play a role in neurodegenerative diseases but there has been no mouse model available in which the hypothesis could be tested and the role of collagen XXVIII studied.

An additional aim of my dissertation was therefore:

4. To generate a mouse model that allows the study of collagen XXVIII function.

2 RESULTS

2.1 NOVEL COLLAGEN VI CHAINS

In a screen of the mouse and human genomic databases with collagen and matrilin sequences as queries, three genes that code for new VWA domain-containing collagens were identified. Because of their high homology to the α3 chain of collagen VI and their location in the genome, these were designated as the α4, α5 and α6 chains of collagen VI. The occurrence and nature of these new collagen VI chains in mouse and human were here studied in detail.

2.1.1 Cloning, gene and protein structure of new collagen VI chains

The new collagen VI genes in mouse and human were cloned and the sequences deposited in the databases. All the three new chains share sequence similarity with the collagen VI α3 chain. The gene and domain structures of the new chains are conserved and resemble the collagen VI α3 chain.

2.1.1.1 Cloning of cDNAs coding for new mouse collagen VI chains

The mouse cDNAs for the collagen VI α4, α5 and α6 chains were cloned as overlapping partial clones by RT-PCR, using primers deduced from the genomic sequences, and sequenced.

The cloned mouse collagen VI α4 cDNA of 7084 bp (accession numbers AM231151-AM231153) contains an open reading frame of 6927 bp encoding a protein consisting of 2309 amino acid residues including a signal peptide of 22 residues which is predicted by a method using neural networks and/or hidden Markov models (Emanuelsson *et al.*, 2007). The mature secreted protein has a calculated M_r of 248,389 (Figure 2.1). At least nine EST clones exist that extend 207 bp in the 3'-ends. In addition, a partial RIKEN cDNA clone (AK159050) extends 1219 bp and also contains an ATTAAA polyadenylation signal at its 3'-end, indicating the presence of different 3'-UTRs.

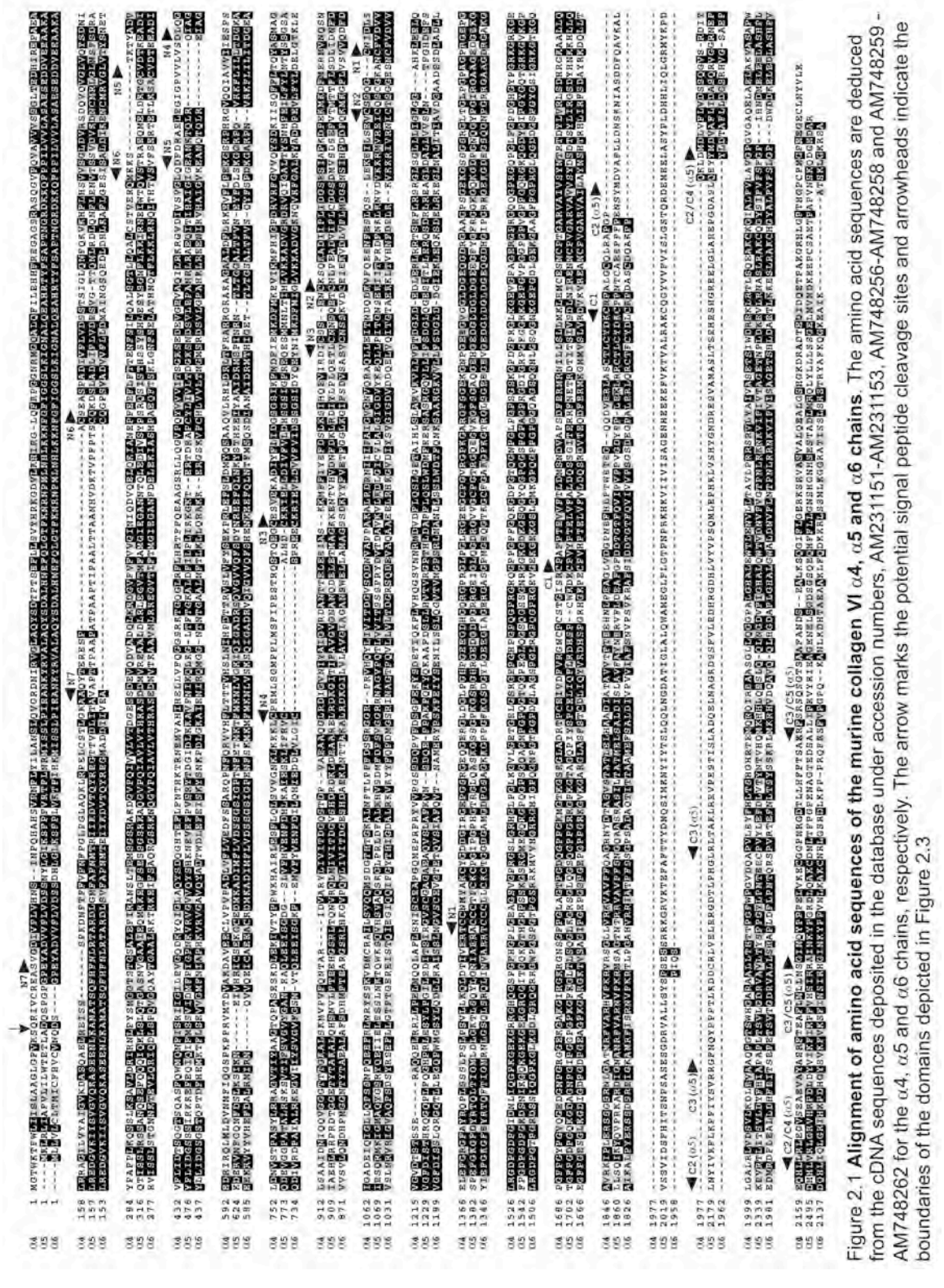


Figure 2.1 **Alignment of amino acid sequences of the murine collagen VI α4, α5 and α6 chains**. The amino acid sequences are deduced from the cDNA sequences deposited in the database under accession numbers, AM231151-AM231153, AM748256-AM748258 and AM748259 - AM748262 for the α4, α5 and α6 chains, respectively. The arrow marks the potential signal peptide cleavage sites and arrowheads indicate the boundaries of the domains depicted in Figure 2.3

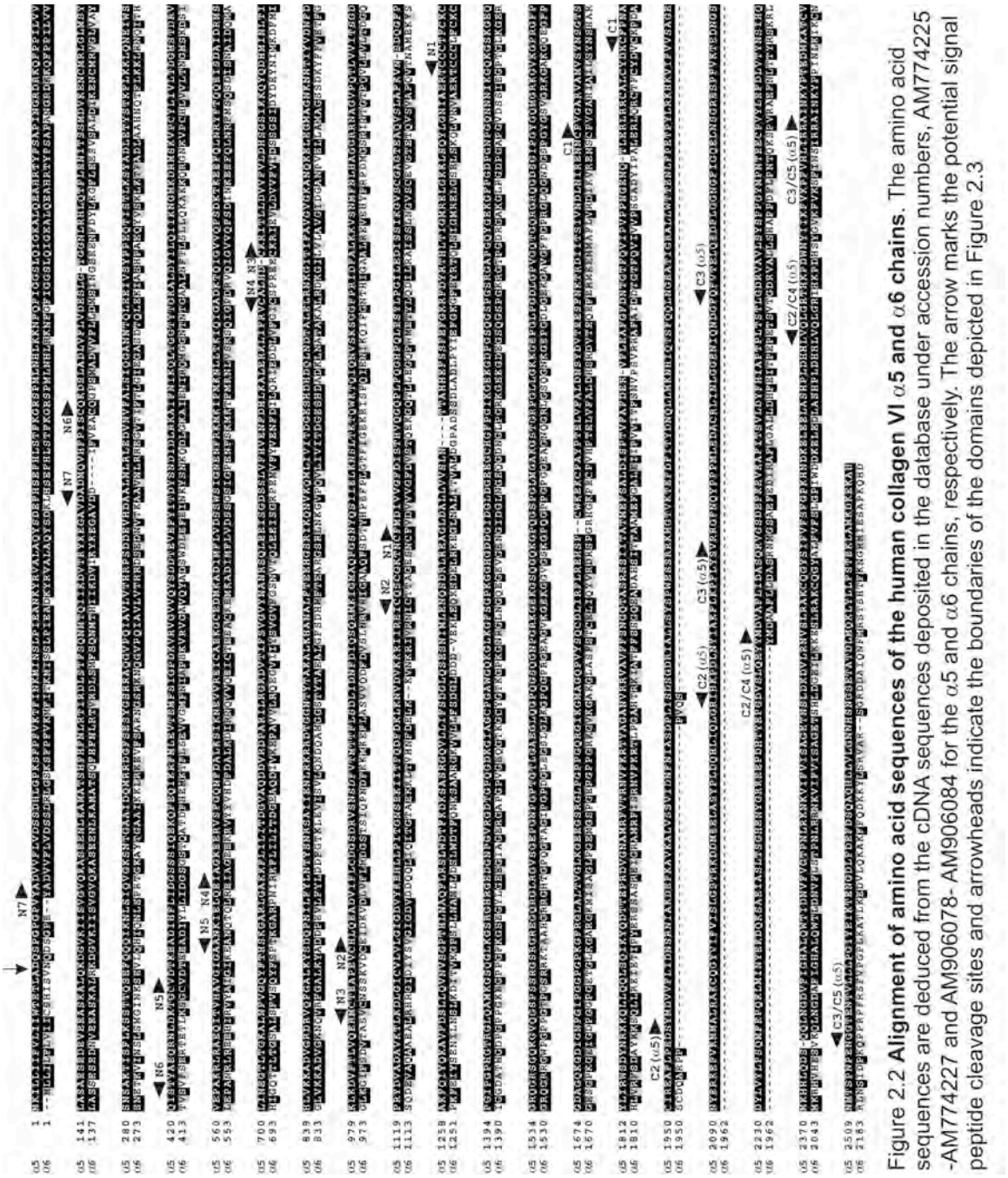


Figure 2.2 **Alignment of amino acid sequences of the human collagen VI α5 and α6 chains**. The amino acid sequences are deduced from the cDNA sequences deposited in the database under accession numbers, AM774225 -AM774227 and AM906078- AM906084 for the α5 and α6 chains, respectively. The arrow marks the potential signal peptide cleavage sites and arrowheads indicate the boundaries of the domains depicted in Figure 2.3

The cloned mouse collagen VI α5 chain cDNA of 8298 bp (accession numbers AM748256-AM748258) contains an open reading frame of 7920 bp, encoding a protein consisting of 2640 amino acid residues including a signal peptide of 18 residues (Emanuelsson *et al.*, 2007). The mature secreted protein has a calculated M_r of 287,502 (Figure 2.1). A partial RIKEN clone (AK134435) extends 751 bp at the 3'-end but does not contain a polyadenylation signal.

The cloned mouse collagen VI α6 chain cDNA of 7097 bp (accession numbers AM748259-AM748262) contains an open reading frame of 6795 bp, encoding a protein consisting of 2265 amino acid residues including a signal peptide of 18 residues (Emanuelsson *et al.*, 2007). The mature secreted protein has a calculated M_r of 244,260 (Figure 2.1).

2.1.1.2 Cloning of cDNAs coding for new human collagen VI chains

The human cDNAs for the collagen VI α5 and α6 chains were cloned as overlapping partial clones by RT-PCR, using primers deduced from the genomic sequences, and sequenced (accession numbers AM774225-AM774227 and AM906078- AM906084).

The cloned human collagen VI α5 chain cDNA contains an open reading frame of 7771 bp encoding a protein consisting of 2590 amino acid residues including a signal peptide of 18 residues (Emanuelsson *et al.*, 2007). The mature secreted protein has a calculated M_r of 287,165 (Figure 2.2).

The cloned human collagen VI α6 chain cDNA contains an open reading frame of 6789 bp encoding a protein consisting of 2263 amino acid residues including a signal peptide of 19 residues (Emanuelsson *et al.*, 2007). The mature secreted protein has a calculated M_r of 247,187 (Figure 2.2).

It should be noted that a functional gene coding for *COL6A4* does not exist in humans (see Figure 2.6 and section 2.1.16.1).

2.1.1.3 Domain architecture of the new chains

The domain structure of the new chains is very similar to that of the collagen VI α3 chain (Figure 2.3). Therefore, the existing nomenclature for the already known collagen VI chains (Chu *et al.*, 1990) was extended to the new chains. The domains N-terminal to the

collagenous domain are designated with N and the domains C-terminal of the collagenous domain with C. Numbering starts at the collagenous domain. At the N-terminus, all three mature proteins contain seven VWA domains (N7-N1) followed by a 336 amino acid residues long collagen triple helical domain (Figure 2.3). Towards the C-terminus they have two VWA domains (C1-C2) that are followed by a unique sequence, C3, that in the new α6 chain also represents the C-terminus. In mouse, the α4 chain carries a short stretch of 17 amino acid residues at the very C-terminus (C4) that resembles a truncated bovine pancreatic trypsin inhibitor/Kunitz family of serine protease inhibitor domain (Figure 2.4A). Interestingly, when searching the genomic databases exons coding for a

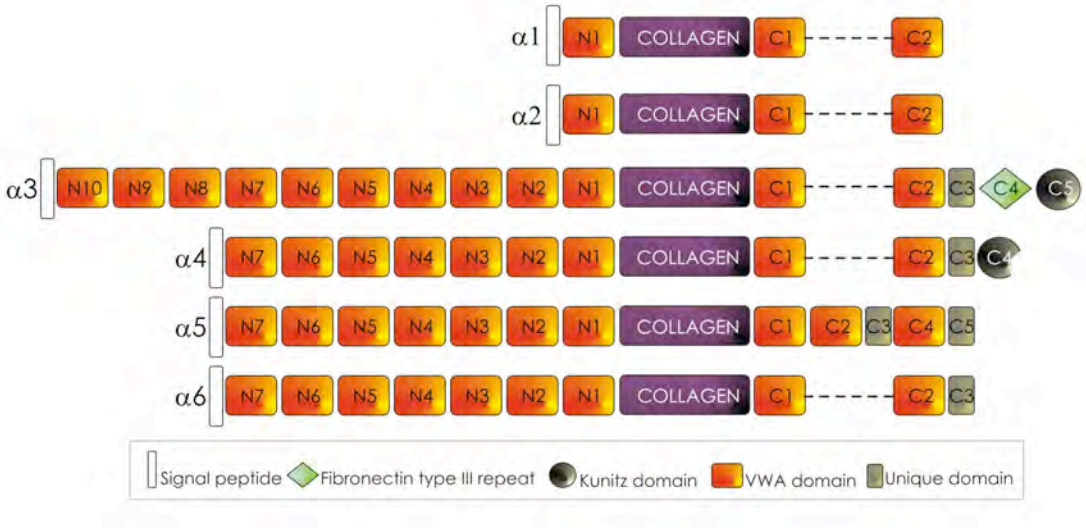


Figure 2.3 **Domain structures of the new collagen VI chains compared with the collagen VI α1, α2 and α3 chains**. VWA domain, von Willebrand factor A domain, Kunitz domain, Kunitz family of serine protease inhibitors domain. The numbering of the domains was according to Chu et al., 1976, and only represents the order of the domains in each chain. The dashed line indicates the absence of the duplication of the second VWA domain and the unique domain in all the collagen VI chains except the collagen VI α5 chain.

complete Kunitz domain could be identified at this position for ortholog genes of several species. Only in mouse and rat the sequences contain a premature stop codon, indicating that except for in mouse and rat, a full Kunitz domain is present at the C-terminus of the collagen VI α4 chain. In the α5 chain the C-terminus is formed by a third VWA domain (C4) followed by another unique domain (C5). According to the sequence similarity the C4 domain of collagen VI α5 groups up with the C2 domain of the other chains.

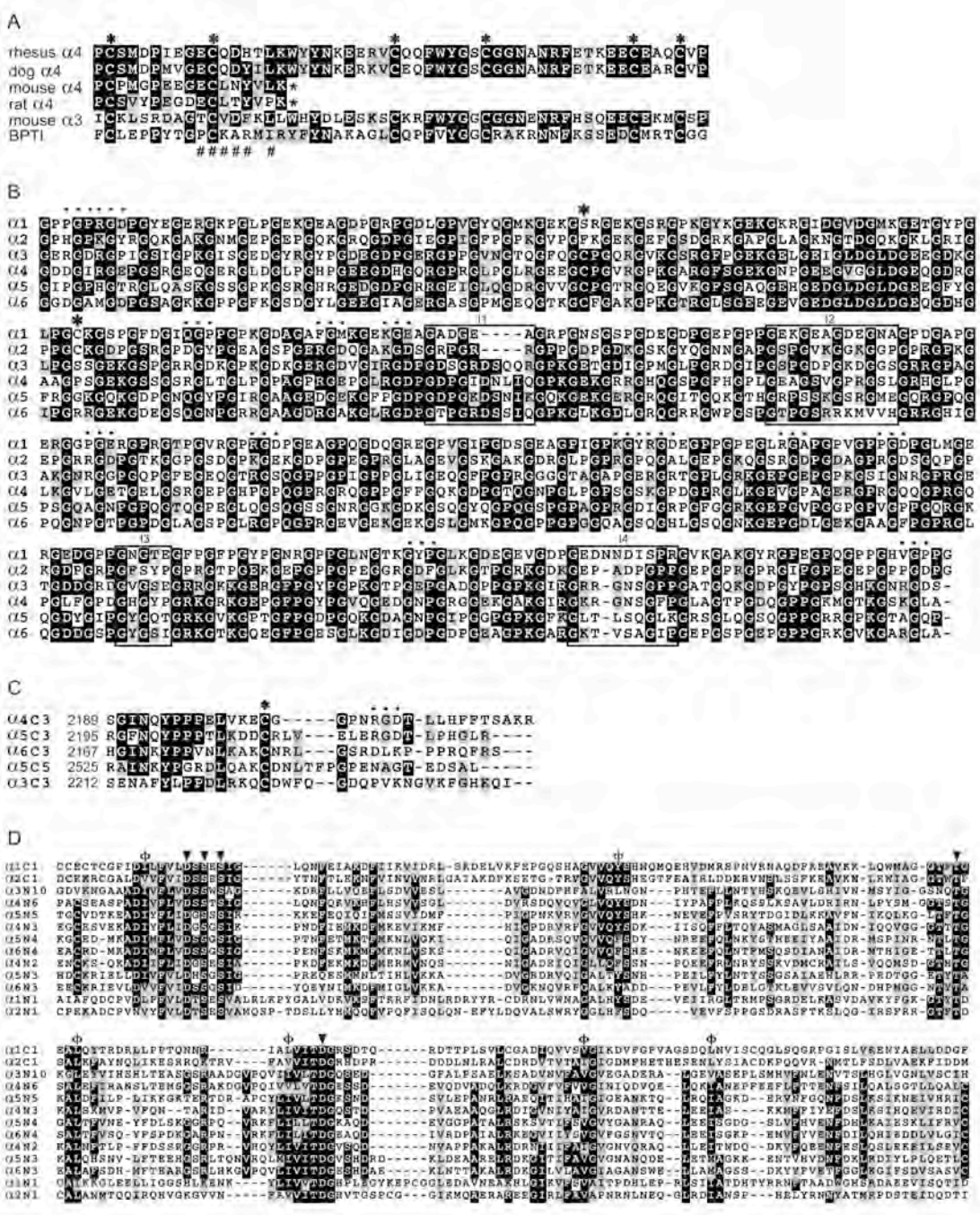


Figure 2.4 **Amino acid alignment of collagen VI Kunitz domains in different species (A), collagenous domains from the different collagen VI chains in mouse (B), the C-terminal part of the unique domains (C) and VWA domains containing the metal ion-dependent adhesion site motif in mouse (D)**. The sequences for the Kunitz domains of rhesus monkey, dog and rat were deduced from genomic sequences. The sequences were aligned by CLUSTAL X using the default parameters. The residues forming the trypsin interaction site in the original bovine pancreatic trypsin inhibitor (BPTI) (Perona *et al.*, 1993) are marked with (#), the cysteine residues with asterisks and the RGD sequences with dots. Imperfections in the collagenous domains are boxed and numbered I1-I4. The conserved metal ion-dependent adhesion site (Lee *et al.*, 1995) and the conserved hydrophobic moieties (Perkins *et al.*, 1994) are denoted with (▼) and (*Φ*), respectively.

Similarly, the C5 domain of collagen VI α5 pairs up with the C3 domain of α3, α4 and α6 chains leaving the C2 and C3 domain as a distinct branch (Figure 2.1). A major difference between the new chains and the collagen VI α3 chain is the presence of three

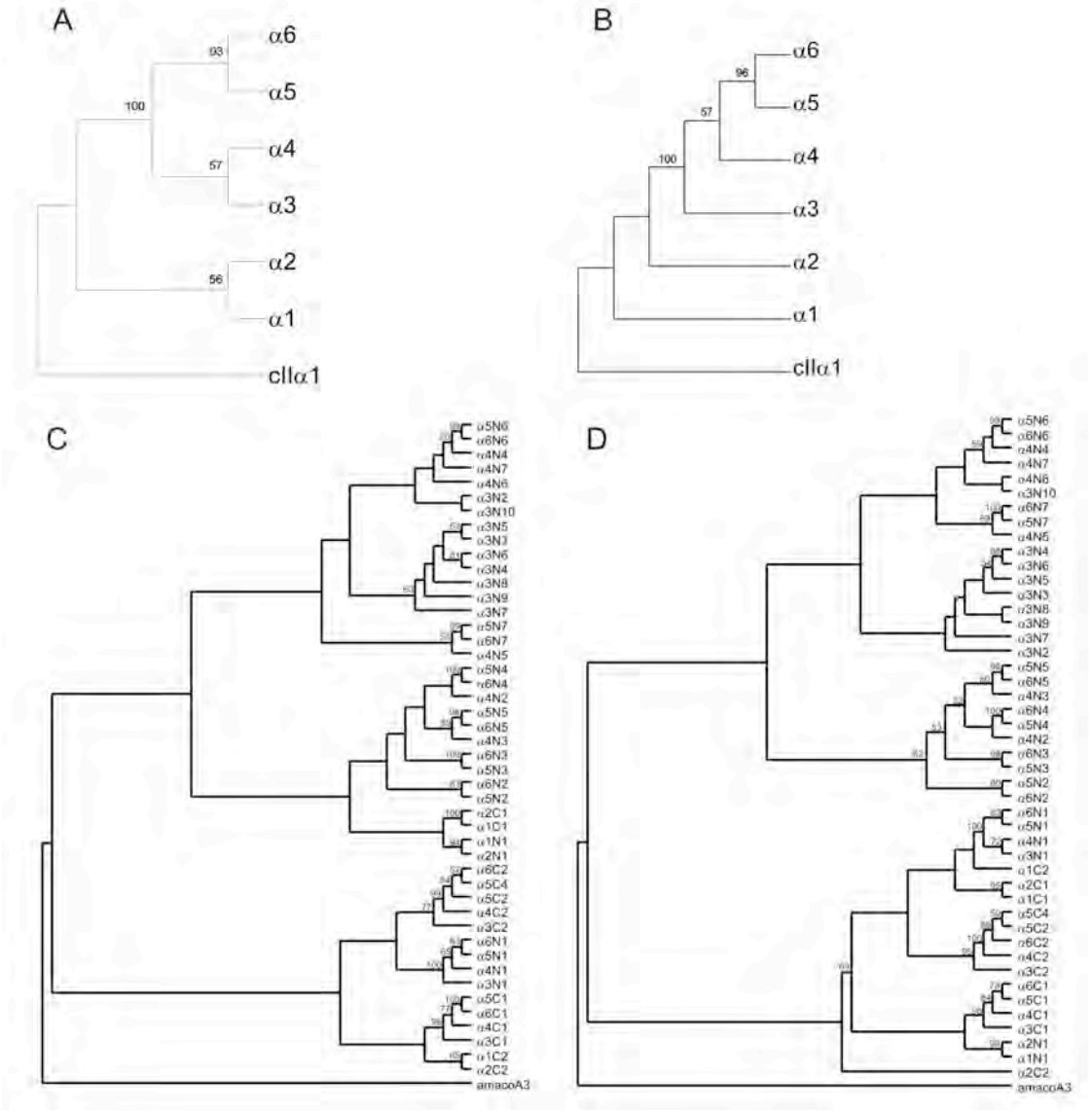


Figure 2.5 **Phylogenetic trees of the collagenous domains (A and B) and VWA domains (C and D) of collagen VI**. The sequences of the collagenous domains of all six collagen VI chains in mouse were aligned using CLUSTAL X. The trees were constructed using the programs PROTEIN DISTANCE, NEAREST NEIGHBOUR and CONSENSE (A and C) and PROTEIN PARSIMONY and CONSENSE (B and D) of the PHYLIP package version 3.66. Bootstrap analyses using 100 replicates were performed to show the significance. The numbers indicate the statistical weight of the individual branches. The collagenous domain of the collagen II α1 chain (cIIα1**)** (A and B) and VWA3 domain of AMACO (C and D) were used as outgroups.

additional VWA domains at the N-terminal end of the α3 chain (Figure 2.3). Interestingly, a splice variant of the collagen VI α3 chain lacks the first, the second and the fourth VWA domain (AAC23667) and thereby contains seven N-terminal VWA domains similar to the new chains. The overall identity at the amino acid level is highest between the α5 and α6 chains (44.7%) and lowest between the α4 and α5 chains (28.0%). The overall identity of the three new chains and the α3 chain varies between 25.9% and 26.7%.

2.1.1.4 Analysis of the collagenous domains

The 336-amino-acid residue long collagenous domains have exactly the same size as that of the collagen VI α3 chain (Figure 2.4B). The identity between the collagenous domain of the α3 chain and those in the α4, α5 and α6 chains is 53.3%, 49.1% and 51.8%, respectively. A cysteine residue that is also present in the collagenous domain of the collagen VI α3 chain and appears to be involved in tetramer formation and stability (Lampe & Bushby, 2005) is conserved in all new chains. The locations of the two imperfections in the Gly-X-Y repeat found in the collagen VI α3 chain are conserved in all new chains, while the α5 and α6 chains have additional imperfections. In both these chains a glycine residue in a Gly-X-Y repeat close to the C-terminus of the collagenous domain is replaced by a leucine or a valine residue, respectively, introducing another imperfection. Interestingly, the position coincides with an imperfection found in the α1 and α2 chains. In addition, in the collagen VI α5 and α6 chains an imperfection is present at the center of the collagenous domains where one or two glycine residues of Gly-X-Y repeats are lacking, respectively.

In contrast to the collagenous domain of the collagen VI α3 chain, which contains five potentially integrin-binding RGD sequences, in each of the new chains only one RGD motif is present. In the collagen VI α4 and α6 chains the motif is found at exactly the same position where an RGD is present also in the collagen VI α3 chain (Figure 2.4B). The content of proline or hydroxyproline in the X and Y positions is lower (17.4-20.5%) than in the fibril forming collagen I α1 or collagen II α1 chains (Persikov *et al.*, 2005). N- and C-terminal of the collagenous domains several cysteine residues are

present, which might form intermolecular disulfide bridges that enhance the stability of the trimeric collagens. In phylogenetic analyses using protein distance and protein parsimony, the collagenous domains of the α3, α4, α5 and α6 chains group in one clade (Figure 2.5A&B).

2.1.1.5 Analysis of the VWA domains

Out of the 28 VWA domains present in the new collagen VI chains, the MIDAS (metal ion-dependent adhesion site, D-x-S-x-S-x-nT-x-nD) motif is fully conserved only in eight of them (Figure 2.4D). Sequence alignment of the VWA domains of the new chains with their counterparts present in the collagen VI α1-α3 chains highlights the homology (Figure 2.4D and Appendix Fig. 1). The highest sequence identity between two VWA domains of the new chains is 92.1% for α5N7 and α6N7. High identity values were also obtained for the α5N4 and α6N4 (64.5%), α5N5 and α6N5 (51.9%), α5C2 and α6C3 (52.9%), α5C1 and α6C1 (50.5%) and α5N1 and α6N1 (50.3%). Among the various VWA domains found in the collagen VI α1-α3 chains, the N10 domain of the collagen VI α3 chain shows the highest identity value to the N7 domain of the α4 chain (39.5%). Similar identity values were obtained for α3N9 and α4N7 (34.7%) and α3C1 and α4C1 (34.5%). Identity values between the α3 chain VWA domains and α5 and α6 chain VWA domains are not higher than 28.4 and 28.9%, respectively. The identity between the VWA domains of the new chains and those of the collagen VI α1 and α2 chains is always lower than 24.0%. In phylogenetic analyses using protein distance and protein parsimony, all the VWA domains of the α5 and α6 chains pair up together (Figure 2.5C&D). The C-terminal VWA domains of the α3, α4, α5 and α6 chains group to a distinct branch in which the C1 domains are in one subbranch and the C2 domains and the C4 domain of the α5 chain are in another. Similarly, the N1 domains of the α3, α4, α5 and α6 chains all cluster together (Figure 2.5C&D).

2.1.1.6 Analysis of the unique domains

The unique sequences at the C-terminal end follow directly after the second C-terminal VWA domains (C2). In the collagen VI α5 chain a second unique domain is present C-

terminal of the C4 domain. The unique domains are 99-111 amino acid residues long. The unique domain of the α4 chain and the first unique domain of the α5 chain as well as the second unique domain of the α5 chain and the unique domain of the α6 chain share some pairwise similarity, 31.6% and 26.1%, respectively. However, a stretch of 15 amino acid residues at the beginning of each domain is highly identical in all four unique domains and has a cysteine residue at the end (Figure 2.4C). Interestingly, the unique sequence of the collagen VI α3 chain, C-terminal to the C2 domain, also shares some homology to the unique domains of the new chains, most clearly in the C-terminal portions, and particularly the cysteine residue is conserved (Figure 2.4C). Interestingly, shortly after the highly homologous stretch an RGD motif is present in both the α4 chain and the first unique domain of the α5 chain while it is missing in the α6 chain, in the second unique domain of the α5 chain and in the α3 chain (Figure 2.4C). In addition to the single RGD motifs present in each of the collagenous domains, these two RGD motifs are the only ones found in the new collagen VI chains. An RGD motif is lacking in the unique domain of the collagen VI α3 chain. BLAST searches with the unique sequences revealed some weak homologies to intracellular proteins like the REST corepressor 1 (α4 35/83 (42%)), ubiquitin D (α5C3 22/32 (68%)), protein tyrosine phosphatase (α5C5, 34/71 (47%)) and dynein cytoplasmic 2 heavy chain 1 (α6 26/60 (43%)).

2.1.1.7 Structure of the new collagen VI genes

In mouse, the new collagen VI chains map to chromosome 9 (9F1) (Figure 2.7). The genomic sequences are completely contained in the public databases (NT_039477 and NW_001030918). The three new genes lay head to tail in tandem orientation on the minus strand. The *Pik3r4* gene and the *Mirn135a1* gene are located downstream and upstream of the new collagen genes respectively. The exons were identified by flanking consensus splice signals and by comparison with the respective cDNAs. The exon/intron organization of the three genes is very similar (Figure 2.7 and Appendix Tab. 1,2,3) regarding size, exon and intron length and codon phase. The *Col6a4* and *Col6a5* genes are 112 kb and the *Col6a6* gene is 104 kb long. They consist of 38, 44, and 37 exons, respectively, that code for the translated part of the mRNA (Figure 2.8). The first exon in each gene completely encodes the 5' untranslated region. All second exons code for the

signal peptide sequence followed by six exons coding for the first six VWA-domains (N7-N2), whereas the VWA domain N1 is encoded by three exons. The collagenous domains are encoded by exons 12 - 30. Interestingly, intron 24 of the α4 gene is a GC-AG type intron. Exons 31 and 32 code for short spacer regions. The VWA domains C1 and C2 are encoded by exons 33/34 and 35, respectively. The structures of the three genes differ at the 3'-end. In *Col6a4* the unique sequence is encoded by two exons followed by a last exon coding for the truncated Kunitz domain and the 3'-UTR. In *Col6a6* the last two exons code for the unique domain and 3'-UTR. In case of *Col6a5,*

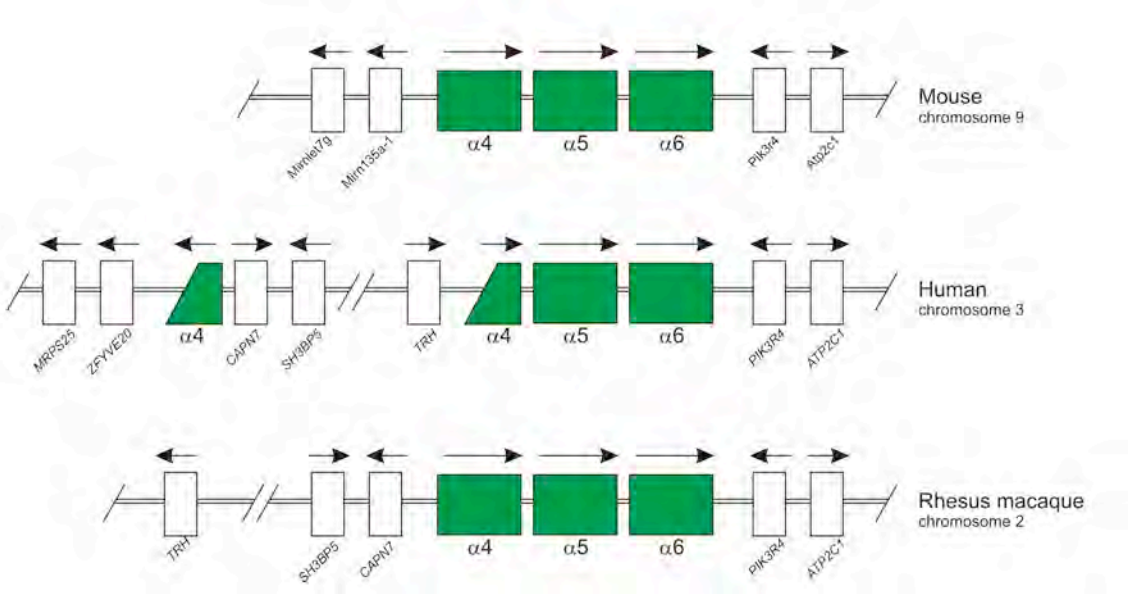


Figure 2.6 **Loci for *COL6A4, COL6A5 and COL6A6* genes in the genomes of man, rhesus monkey and mouse**. The orientation of the genes is indicated by *arrows*. The neighboring genes of the new chains are indicated for comparison.

the unique domain between the VWA domains C2 and C4 is encoded by two exons followed by the exon coding for the additional VWA domain C4. As in *Col6a6* the last two exons of *Col6a5* code for the unique domain and 3'UTR. Although there is only partial homology between the unique domains of the new collagen chains, each unique domain is encoded by two exons where the first exon is always about 95 bp and the second is about 200 bp long, pointing to the likelihood of a common ancestor. The orthologs of the new mouse genes map to human chromosome 3q21 (Figure 2.6). The tandem orientation of the genes is conserved, but the gene coding for the α4 chain is

broken into two pieces and the 5' region of the gene is located at 3p24.3. Only the region downstream of the new collagen VI genes, coding for *PIK3R4*, is in synteny in man and mouse. The breakpoint resembles the large scale pericentric inversion that occurred in the common ancestor of the African apes and is present in modern human chromosome 3 as well as in the chimpanzee and gorilla orthologs, but not in orangutan or old world monkeys (Muzny *et al.*, 2006). In contrast to Rhesus macaque, the *COL6A4* in humans is

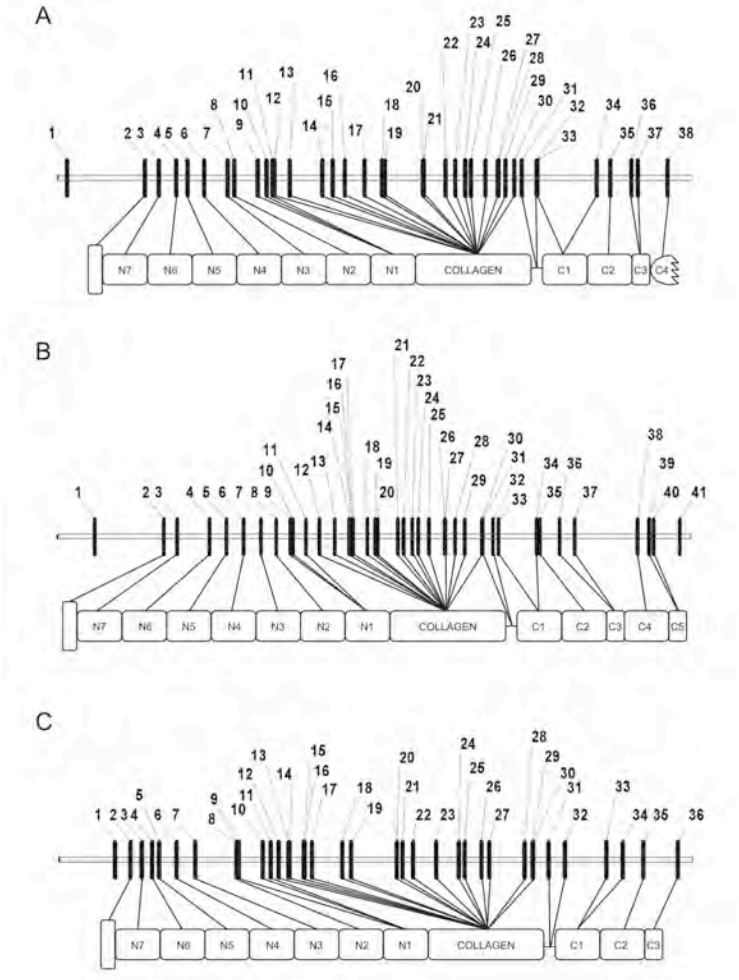


Figure 2.7 **Organization of the murine *COL6A4, COL6A5 and COL6A6* genes**. The structure of *Col6a4* (A), *Col6a5* (B) and *Col6a6* (C) genes in mouse with respect to their coding regions. The numbers indicate the positions of exons.

interrupted after the first exon coding for the collagenous domain. For both parts of the gene, EST clones can be found in the databases. However, due to the presence of stop codons that are distributed over the sequence, both parts of the human *COL6A4* are likely to be transcribed non-processed pseudogenes. In contrast to the mouse collagen VI α5

gene, *COL6A5* in man contains an additional intron in the 3'UTR that by alternative splicing leads to three different C-termini. Full-length proteins of 2526, 2614 or 2615 amino acid residues would result (Appendix Fig. 2).

2.1.1.8 Human collagen VI α5 and α6 chains

The human collagen VI α5 chain has an identity of 73.1% at the amino acid level to the mouse ortholog (Appendix Fig. 2). The variations are not evenly distributed over the sequence. A 32 amino acid long proline-rich stretch at the C-terminus of N7 is missing in man and the unique domains are highly divergent. In addition, at two positions in the C-terminal part an amino acid residue is deleted and at three positions an amino acid residue is inserted into the human α5 chain. Most of the cysteine residues are conserved, but there is an additional cysteine present in the collagenous domain of the human α5 chain. However, the cysteine codon resembles an SNP (rs1497312) leading to a non-synonymous exchange to a serine codon. The positions and sizes of the imperfections in the collagenous domain are identical to those in mouse, whereas the RGD motif in the collagenous domain of the α5 chain is lost. Instead there is a new RGD motif at the N-terminus of the collagenous domain. The two RGD motifs present in the unique domain of mouse are also missing in man. Another SNP (rs11355796) that resembles the deletion of a thymidine at the C-terminus forms a premature stop codon, leading to a full-length protein of 2590 residues (Appendix Fig. 2). No information is available on the population frequency, but both variants are found in the TRACES-WGS database. Interestingly, in the alternative Celera assembly of the human genome the deletion is present, whereas the thymidine is found in the reference assembly, leading to a longer protein. The human collagen VI α6 chain has an identity of 83.4% at the amino acid level to the mouse ortholog and only the last 30 amino acids at the C-terminus show some differences (Appendix Fig. 3). The positions of the signal peptide cleavage site, the RGD motif and all cysteine residues in the mature protein are completely conserved.

2.1.1.9 Alternative splicing

In mouse, two different splice variants of the collagen VI α4 mRNA leading to premature stop codons can be deduced from EST clones. First, the ESTs AU023415 and BG068629

contain a stop codon in an alternative exon following the exon coding for the N4 domain. If translated, it would lead to a protein containing only the first four VWA domains. A second splice variant was detected in the three EST clones BX520360, AI427280 and W48310. Here, an alternative splice donor site in exon 35 coding for the C2 domain and

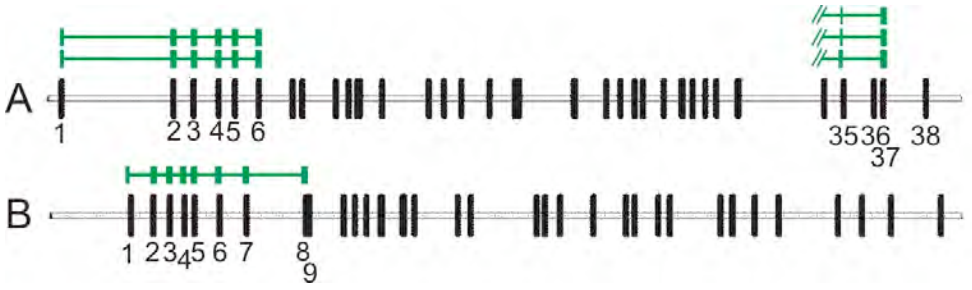


Figure 2.8 **Alternative splicing in the new collagen VI genes of mouse**. The position of the ESTs derived from the *Col6a4* (A) and *Col6a6* (B) genes are given in green. "//" indicates that the N-terminus is not shown or unknown.

an alternative splice acceptor site in exon 37 coding for the unique domain is used. Due to a shift in codon phase, the new exon codes for a different frame and contains a stop codon 101 bp downstream of the alternate splice site. If translated, it would lead to a protein that lacks nearly half of the C2 domain and the unique domain (Figure 2.8). Interestingly, the alternative splice site contains a non-canonical GC-AG motif.

A RIKEN cDNA clone coding for the collagen VI α6 chain (AK054356) shows alternative splicing in the 5'UTR, indicating the presence of two different promoters. Interestingly, due to additional alternative splicing of exon 6 a much shorter open reading frame occurs that would generate a protein that contains only the first six VWA domains and lacks the seventh VWA domain, the collagenous domain and the C-terminal non-collagenous domains (Figure 2.8).

2.1.2 Expression analysis of new collagen VI chains

The gene expression of new collagen VI chains was examined in a panel of newborn and adult mouse tissues. Specific polyclonal antibodies against human and mouse collagen VI chains were generated and their protein expression was analyzed. The new chains exhibit restricted and differential expression patterns in selected tissues of both man and mouse.

2.1.2.1 Analysis of mRNA of the new collagen VI genes in mouse

To determine the length of the new collagen VI chain mRNAs northern hybridization was performed with total RNA or mRNA. The mRNA coding for the α6 chain could be easily detected as a 9.6 kb band in total RNA derived from lung of newborn mice (Figure 2.9A). The messages coding for the α5 and α4 chains could be detected in purified mRNA

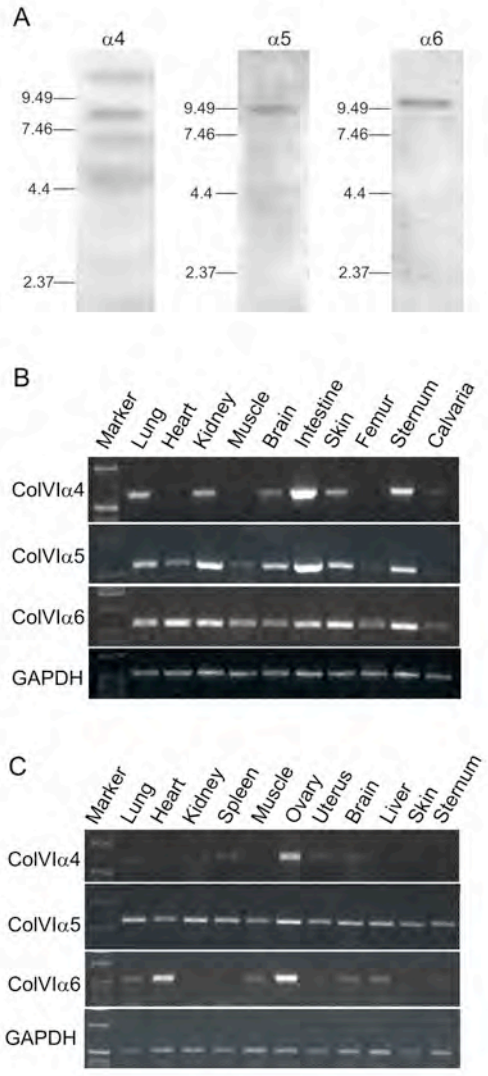


Figure 2.9 **Northern blot (A) and RT-PCR (B and C) analysis of the new mouse collagen VI chain mRNA species.** Northern hybridization was performed for the collagen VI α4 and α6 chains with 5 µg of total RNA from lung of newborn mice and for the α5 chain with 1 µg of poly(A)$^+$ RNA from sternum of 4-week-old mice. Probes were generated using primers α4m8 and α4m9 for the α4 chain, α5m2 and α5m7 for the α5 chain and α6m6 and α6m10 for the α6 chain (*ref*. Materials and methods). Position of the size markers are indicated on the left (A). RT-PCR analysis was performed using the primer pair α4m6 and α4m7 for the α4 chain, α5m4 and α5m5 for the α5 chain and α6m2 and α6m9 for the α6 chain (*ref*. Table 4.1, in *Materials and Methods*). Template RNA was isolated from newborn (upper panel) and adult mice (lower panel). The 1kb ladder was used as a reference.

derived from sternum of 4-week-old mice and from lung of newborn mice, respectively. The length of the α4 chain mRNA is 8.4 kb and that of the α5 chain mRNA is 9.4 kb. RT-PCR was performed to screen the tissue distribution of the new collagen VI chains (Figure 2.9) and mRNA for the α5 and α6 chains could be detected in lung, heart, kidney, muscle, brain, intestine, skin, femur and sternum of newborn mice. In addition,

α6 chain mRNA could be detected in calvaria of newborn mice. The α4 chain mRNA shows a more restricted tissue distribution and could be detected in lung, kidney, brain, intestine, skin, sternum and weakly in calvaria of newborn mice (Figure 2.9B). In adult mice, most of the α4 chain expression is lost and RT-PCR showed only a signal in ovary and very weakly in spleen. In contrast, the α5 chain is widely expressed also in adult mice and mRNA could be detected in lung, heart, kidney, spleen, muscle, ovary, uterus, brain, skin, liver and sternum, whereas the α6 chain was more restricted and could be detected in lung, heart, muscle, ovary, brain, liver and sternum of adult mice (Figure 2.9C).

2.1.2.2 Preparation and evaluation of collagen VI specific antibodies

In order to study the protein expression of the new collagen VI chains by immunohistochemistry and western blot, a panel of collagen VI specific antibodies was generated using recombinant mouse and human protein fragments as antigens in rabbit and guinea pig (*ref.* Table 4.3 in *Material and Methods*). In addition to antibodies raised against the novel collagen VI chains, specific antibodies against the already known collagen VI α3 chain were generated, as no such antibody was commercially available. The recombinant protein fragments representing mouse collagen VI α3 (N10-N4), α4 (N6-N3), α5 (α5a (N3) & α5b (N6-N2)) and α6 (N6-N2) domains were used as antigens for antibody production. Specific antibodies against the recombinant protein fragments for human collagen VI α3 (N4-N1), α5 (N6-N2) and α6 (N6-N2) chains were also generated. The antisera from both rabbit and guinea pig were further affinity purified against the respective protein fragments used as immunogen. The specificity and titer of each antiserum was tested using Enzyme Linked Immunosorbent Assay (ELISA). All antisera showed a specific reactivity against their corresponding mouse and human recombinant protein fragment antigens (Figure 2.10 & 2.11). The two mouse collagen VI α4 and α6 chain antisera that were generated in rabbit and guinea pig showed similar titers whereas the preimmune sera from the same animals did not react with their respective antigens (Figure 2.10B&E). For the collagen VI α5 chain, two distinct

antigens, one consisting of a single VWA domain (α5a) and the other of five VWA domains (α5b) were used for immunization (*ref.* Table 4.3 in *Material and Methods*).

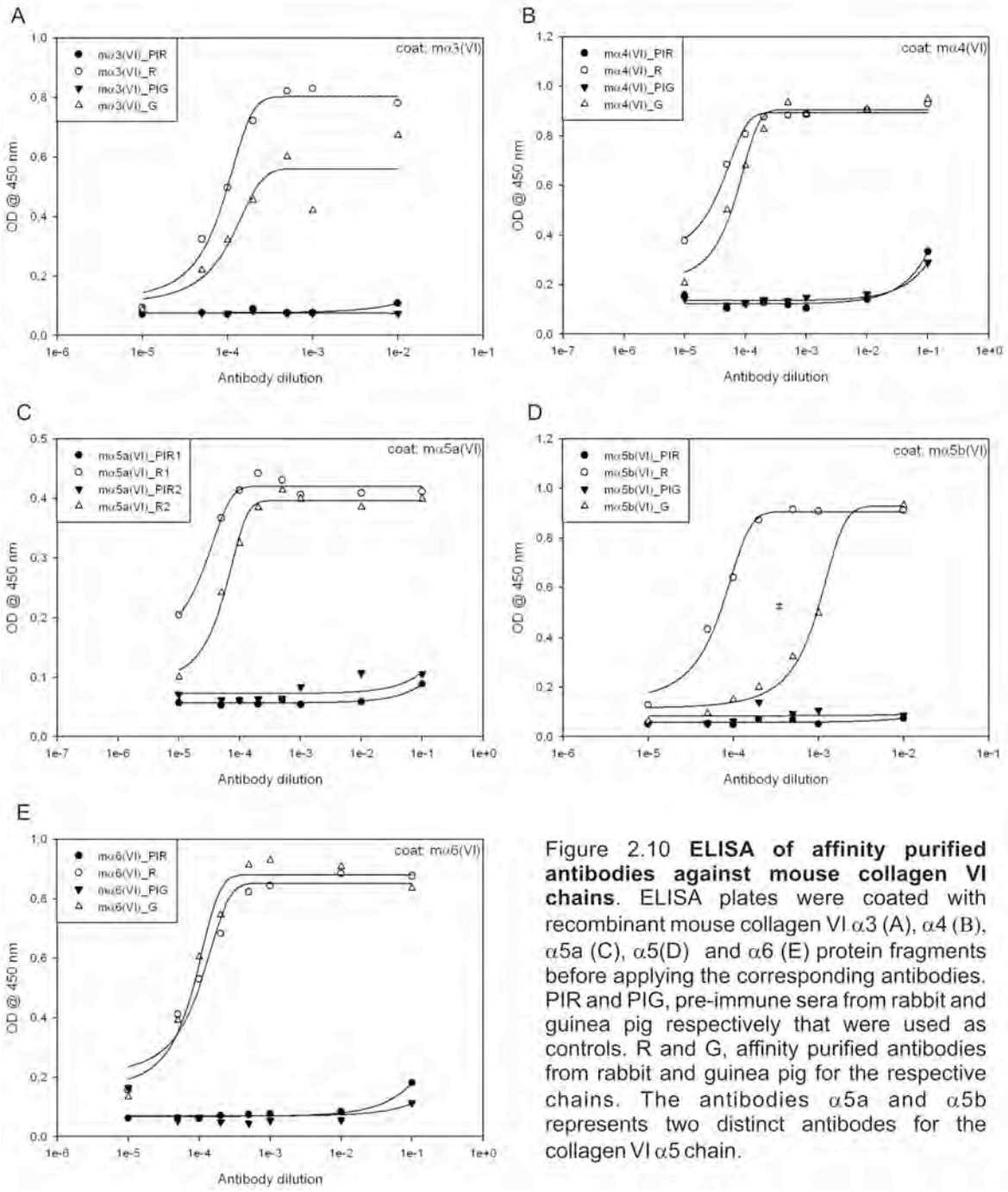


Figure 2.10 **ELISA of affinity purified antibodies against mouse collagen VI chains**. ELISA plates were coated with recombinant mouse collagen VI α3 (A), α4 (B), α5a (C), α5(D) and α6 (E) protein fragments before applying the corresponding antibodies. PIR and PIG, pre-immune sera from rabbit and guinea pig respectively that were used as controls. R and G, affinity purified antibodies from rabbit and guinea pig for the respective chains. The antibodies α5a and α5b represents two distinct antibodes for the collagen VI α5 chain.

The antigen α5a was used to immunize two independent rabbits R1 and R2. Interestingly, both antisera had almost the same titer (Figure 2.10C). The second collagen VI α5 chain antigen (α5b) yielded a reasonably good antibody titer from rabbit but not from guinea

pig (Figure 2.10D). Therefore, the collagen VI α5a antiserum was used for the following experiments unless otherwise mentioned.

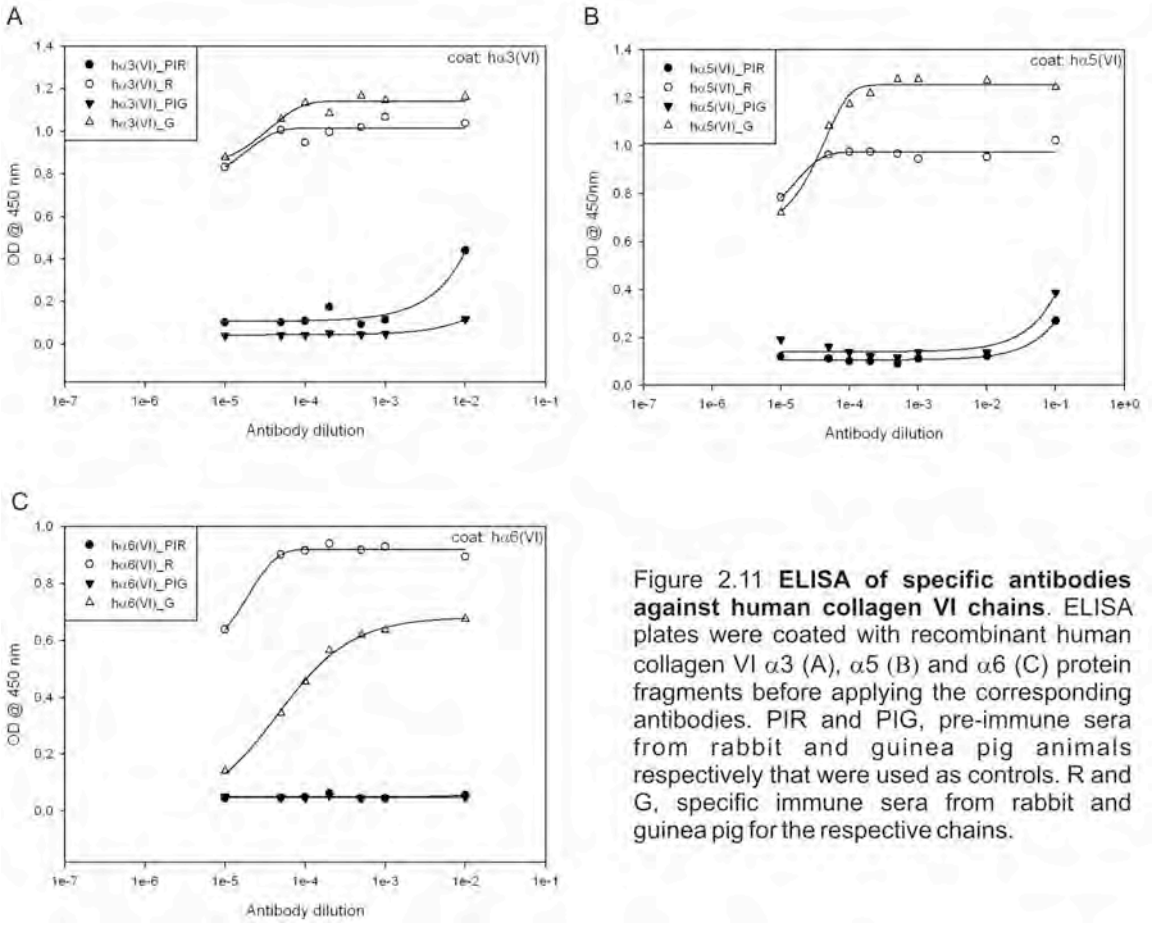


Figure 2.11 **ELISA of specific antibodies against human collagen VI chains**. ELISA plates were coated with recombinant human collagen VI α3 (A), α5 (B) and α6 (C) protein fragments before applying the corresponding antibodies. PIR and PIG, pre-immune sera from rabbit and guinea pig animals respectively that were used as controls. R and G, specific immune sera from rabbit and guinea pig for the respective chains.

Similarly, antisera against human collagen VI chains were also generated and tested for their specificity by ELISA. The human collagen VI α3 and α5 chain antisera, raised in both rabbit and guinea pig, had a high titer (Figure 2.11). The antibody against the human collagen VI α6 chain raised in guinea pig had a lower titer than the other antisera but was sufficiently reactive for use in immunohistochemical and biochemical assays.

2.1.2.3 Analysis of antibody cross-reactivity between the collagen VI chains

As the new collagen VI chains share sequence similarities with the collagen VI α3 chain and with each other, it was important to analyze the cross-reactivity of the affinity purified antisera with the various collagen VI chains. Therefore, the recombi nant antigen proteins were coated onto ELISA plates and the panel of collagen VI antisera was tested.

The antisera raised against the murine recombinant proteins were highly specific and did not show significant cross-reactivity at dilutions beyond 1/500 (Figure 2.12).

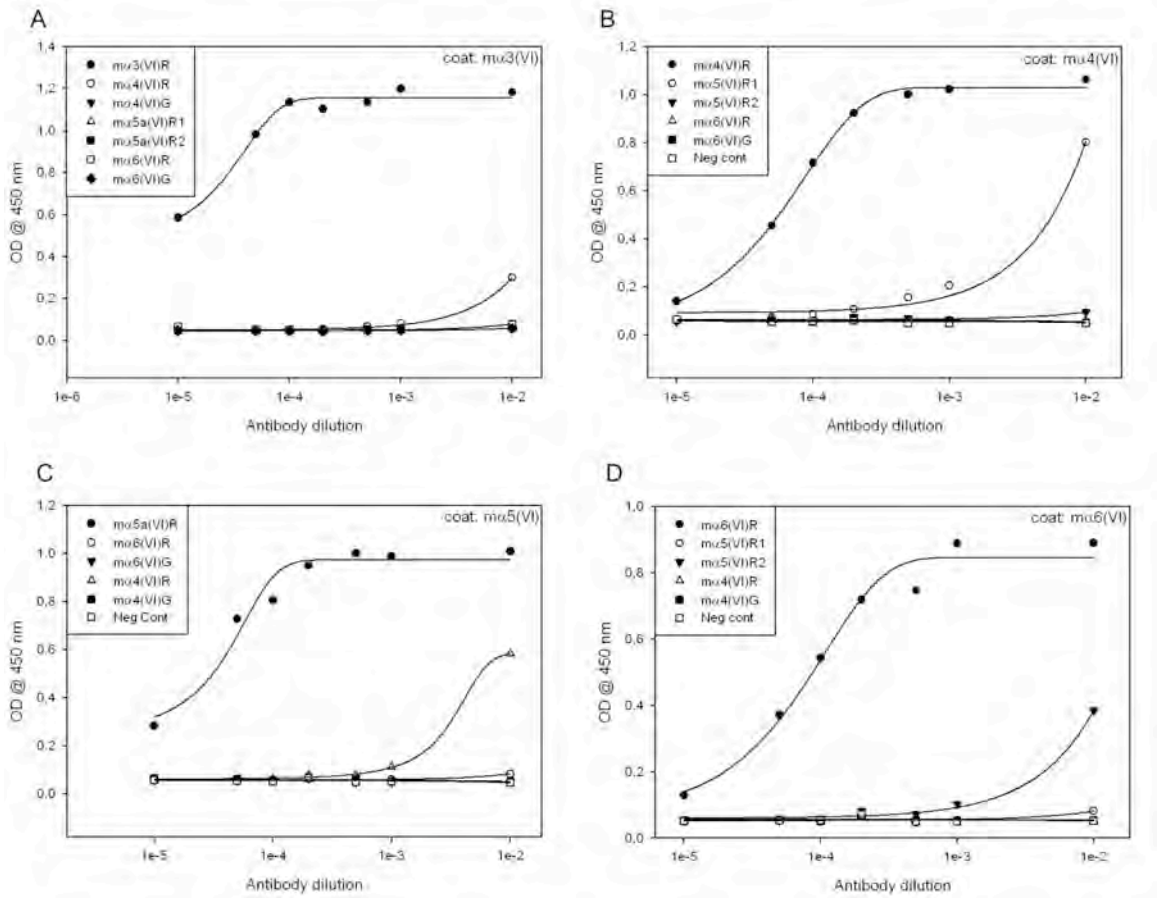


Figure 2.12 **ELISA showing the cross-reactivity of antisera raised against different mouse collagen VI chains**. ELISA plates were coated with recombinant mouse collagen VI α3 (A), α4 (B), α5 (C) and α6 (C) protein fragments before applying the corresponding antisera. R and G, affinity purified antibodies from rabbit and guinea pig for the respective chains. All antisera showed specificity and lacked cross-reactivity with other chains. m, mouse, Neg cont, negative control.

Surprisingly, all collagen VI antisera raised against the human chains showed more pronounced cross-reactivity. For instance, when a recombinant human collagen VI α3 fragment was coated onto an ELISA plate and detected with α3, α5 and α6 chain antisera, there was a significant reaction with the α5 and α6 chain antisera at a dilution of 1/100 (Figure 2.13A). Since the human recombinant proteins carry a double Strep-tag, in contrast to the mouse recombinant proteins that carry a single Strep-tag, the observed cross reactions are likely to be due to antigenic determinants present in the double Strep-tag. Therefore, a dot blot was performed where the recombinant human collagen VI

α3,α5 and α6 chain proteins were dotted together with a Strep-tagged protein marker and independently detected with the human collagen VI and Strep tag antisera (Figure 2.13B). Each antibody detected not only its antigen protein, but also the two other chains and the Strep-tag marker protein. However, the Strep-tag antibody was not able to detect the recombinant α6 chain fragment suggesting that its tag was either cleaved or not accessible. Therefore, also the human α3 and α5 chain antisera failed to detect the

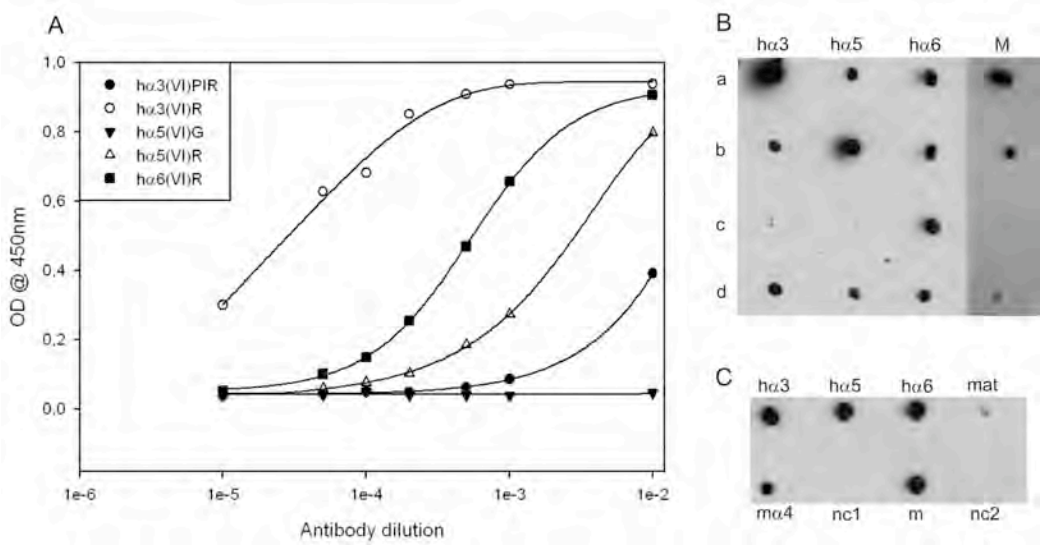


Figure 2.13 **Cross-reactivity of antisera raised against different human collagen VI chains**. A) ELISA assay. ELISA plates were coated with recombinant human collagen VI α3 protein fragment and detected with the panel of human collagen VI antisera. PIR, preimmune serum from rabbit, R and G, affinity purified rabbit and guinea pig antibodies. B&C) Dot blot assay. B) Each row was dotted with the recombinant human collagen VI α3 (a), α5 (b), α6 (c) and Strep-tagged protein marker (d). Each column was detected with the human collagen VI specific antisera hα3, hα5, hα6 and a Strep antiserum as indicated in the top of each column. C) The recombinant protein fragments were spotted and detected with a collagen VI α6 antiserum. mat, matrilin 3 VWA domain; M, Strep-tagged protein marker; nc1, sonic hedgehog (negative control 1) and nc2, water (negative control 2).

recombinant α6 chain fragments. In order to assess the reactivity of the human collagen VI α6 chain antisera against Strep-tag containing proteins, the collagen VI α3 and α5chains and control proteins that either contain a Strep-tag or a different tag were tested (Figure 2.13C). The mouse collagen VI α4 chain fragment and the matrilin 3 VWA domain that carry a single Strep-tag as well as the human collagen VI α3 and α5 chain fragments were detected by the human collagen VI α6 chain antiserum, whereas an N-terminal fragment of sonic hedgehog lacking a Strep-tag did not give a signal with the α6 chain antiserum. (Figure 2.13C). This indicates that the observed cross-reactivity is only

against the artificial Strep tag. Therefore, these antisera could be used for immunohistochemical and western blot analyses of biopsies or tissue extracts.

2.1.2.4 Collagen VI preparations from newborn mice contain the new chains

If the new collagen VI chains assemble with known collagen VI chains, they should be present in conventional collagen VI preparations. Thus native collagen VI was isolated

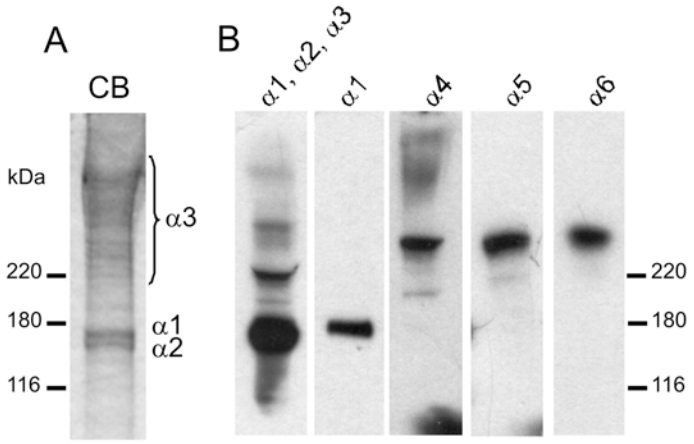


Figure 2.14 **Analysis of collagen VI purified from newborn mice**. Carcasses of newborn mice were natively extracted and collagen VI was isolated by molecular sieve column chromatography as previously described (Colombatti *et al.*, 1989). The proteins were submitted to SDS-PAGE with prior reduction on 4-12% polyacrylamide gradient gels. (A) Coomassie brilliant blue stained gel. (B) Immunoblot of the same preparation with specific antibodies against collagen VI (α1,α2,α3), the collagen VI α1 chain and the new collagen VI α4, α5 and α6 chains. (*in collaboration with Paolo Bonaldo, Padova*)

from newborn mouse carcasses (Colombatti *et al.*, 1989). A Coomassie brilliant blue staining of the isolated collagen VI preparation on a reduced polyacrylamide gel showed the α1 and α2 chains running below 180 kDa and a smear running above 220 kDa, probably representing variants of α3 chains and new chains (Figure 2.14A). The nature of the bands was confirmed by immunoblot (Figure 2.14B). All three new chains were detected under reducing conditions as major bands running above the 220 kDa marker (Figure 2.14B), consistent with their calculated molecular masses. For the collagen VI α4 and α5 chains additional lower migrating bands were detected (Figure 2.14B), indicating alternative splicing or proteolytic processing. The weak smear with lower mobility seen for the α4 chain could indicate the presence of non-reducible cross-linked molecules.

2.1.2.5 Immunohistochemical analysis of the expression of the new chains in mouse

The mRNA expression profile of the new collagen VI chains obtained by RT-PCR suggested that the α5 chain is widely distributed in both newborn and adult mouse tissues. On the other hand, α4 and α6 chains showed a restricted expression in adult but not in newborn mouse tissues (Figure 2.9). In order to determine the exact localization of the new collagen VI chains, extensive immunohistochemical studies with the panel of specific antisera raised against mouse collagen VI chains were performed on both newborn and adult mouse tissue cryosections.

2.1.2.5.1 Collagen VI chain expression in skeletal and cardiac muscle

Immunostaining on longitudinal sections of adult mouse skeletal muscle showed that the

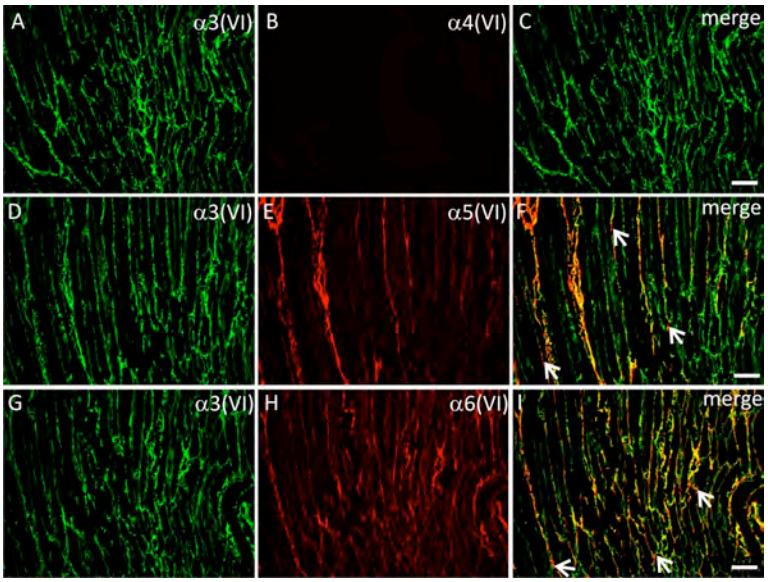


Figure 2.15 **Immunohistochemical analysis of skeletal muscle from adult mouse**. Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. Collagen VI α3 and α6 were strongly expressed in both endomysial and perimysial connective tissue. α5 was strongly present in the perimysial regions and weakly expressed in endomysium. Both α5 and α6 show partial co-localization with α3. Arrows point to selected regions that are exclusively positive for either the collagen VI α5 or α6 chain. The collagen VI α4 chain was completely absent in skeletal muscle. Scale bar, 100 µm.

collagen VI α3 chain is widely expressed in the extracellular matrix surrounding the skeletal muscle cells or fibers (Figure 2.15). It is present in both endomysium and

perimysium of muscle cells. The collagen VI α6 chain is present at similar locations and partially co-localizes with the collagen VI α3 chain. The collagen VI α5 chain showed a strong expression in the perimysium but was sparse in the endomysium of skeletal muscle fibers. Similarly, the collagen VI α6 chain exhibits a partial co-localization with the α3 chain. Interestingly, the collagen VI α5 and α6 chains are either exclusively or less expressed in some distinct regions (Figure 2.15). However, the expression of the collagen VI α6 chain is stronger than that of the collagen VI α5 chain. In contrast, the collagen VI α4 chain is completely absent in the skeletal muscle of adult mouse.

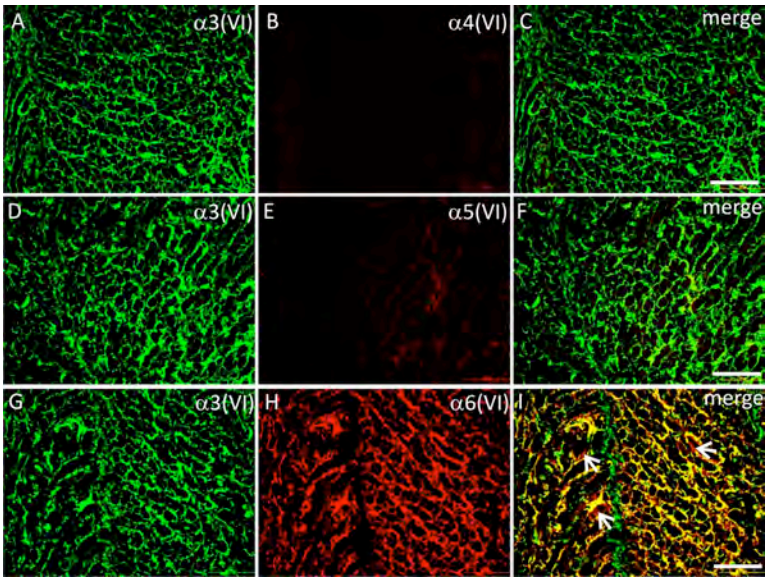


Figure 2.16 **Immunohistochemical analysis of skeletal muscle from newborn mouse**. Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. The antiserum against α6 chain stained endomysial structures but not the perimysium and mostly co-localized with α3. α5 was weakly present and α4 was absent in the newborn mouse skeletal muscle. Arrows point to selected regions that are exclusively positive for the collagen VI α6 chain. Scale bar, 100 µm.

The expression of the new chains in the skeletal muscle of newborn mice differs slightly from that in adult mice (Figure 2.16). The collagen VI α5 chain shows a weaker expression and the collagen VI α6 chain is abundant in the endomysium but not in the perimysium and mostly coincides with where the collagen VI α3 chain is expressed (Figure 2.16). The collagen VI α4 chain is not expressed also in newborn mice.

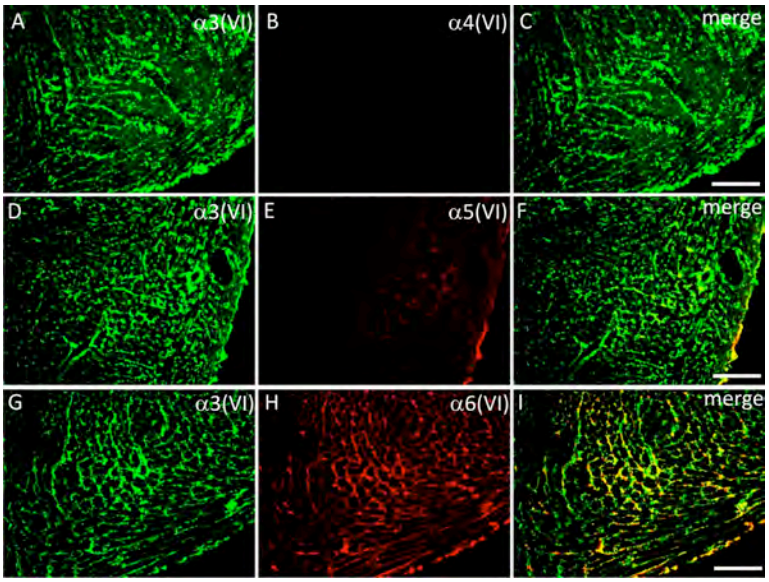


Figure 2.17 **Immunohistochemical analysis of cardiac muscle from adult mouse**. Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. α3 and α6 chains are widely present in the connective tissue of muscle fibers and show partial co-localization. The collagen VI α5 chain was weakly expressed and α4 was absent in the cardiac muscle of adult mice. Scale bar, 100 µm.

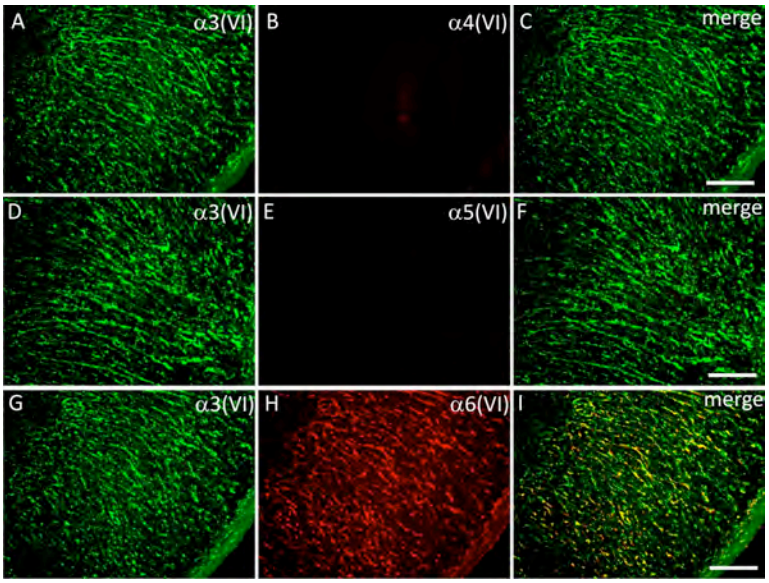


Figure 2.18 **Immunohistochemical analysis of cardiac muscle from newborn mouse**. Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison**.** α3 and α6 chains are widely present in the connective tissue of muscle fibers and show partial co-localization. The collagen VI α5 and α4 chains are absent in the cardiac muscle of newborn mice. 100 µm.

On sections from both newborn and adult mouse hearts, the collagen VI α3 and α6 chains are strongly expressed in the muscle fibers. In contrast, the α5 chain is weakly expressed in adult and absent in newborn mouse hearts. Similar to in skeletal muscle, there is no expression of the collagen VI α4 chain (Figure 2.17 & 2.18).

2.1.2.5.2 Collagen VI chain expression in smooth muscle

Sections of newborn and adult intestine, a tissue that contains layers of smooth muscle,

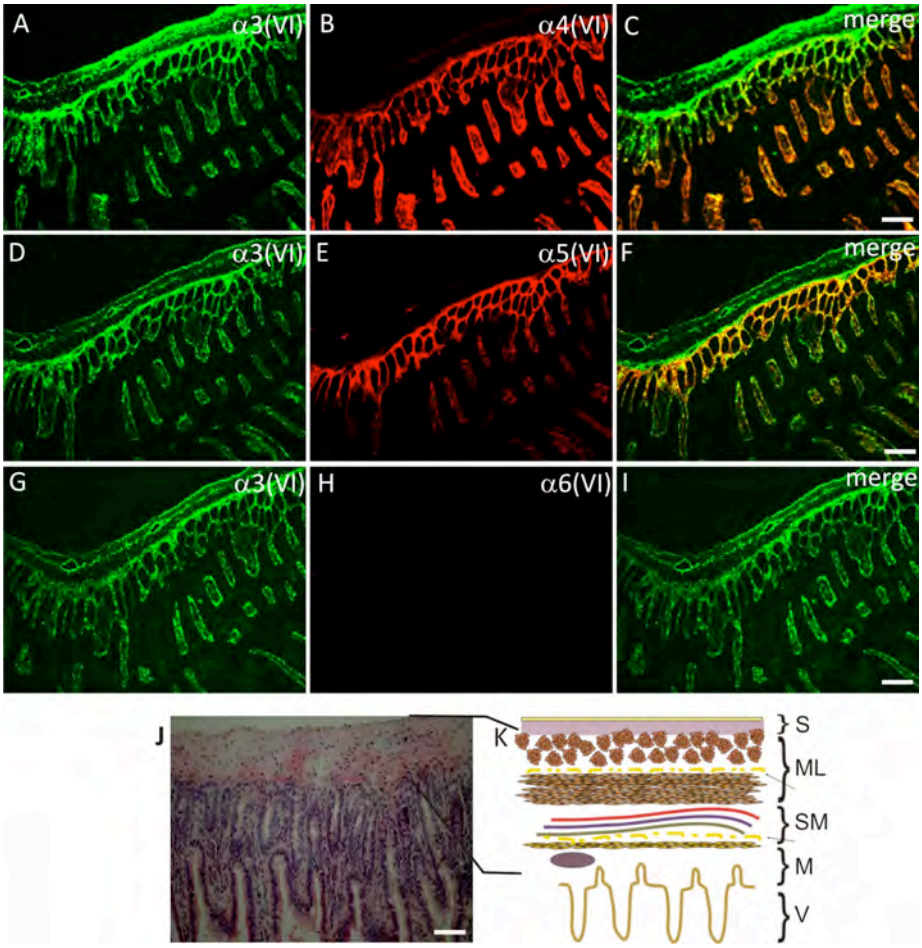


Figure 2.19 **Immunohistochemical analysis of intestine from adult mouse.** Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. For orientation, an intestinal section was stained with H&E (J). A schematic view (K) depicts the different layers of intestine**.** α3 is widely distributed in all layers of intestine. The collagen VI α4 chain is expressed in the submucosal, mucosal and villi regions. The α5 chain is strongly expressed in the mucosal layer and weakly present in villi. The collagen VI α6 chain is not expressed in the intestine. S, sclerosa, ML, muscular layer, SM, sub-mucosal layer, M, mucosal layer, V, villi regions. Scale bar, 100 µm.

were stained with the panel of collagen VI specific antibodies. The collagen VI α3 chain is widely expressed in the inner mucosal layer especially in the villi or papillary mucosal projections towards the lumen, the sub-mucosal connective tissue and the muscularis layer of outer longitudinal and inner circular smooth muscle of adult intestine (Figure 2.19). On the contrary, the collagen VI α4 chain is completely absent in the muscularis layer but abundant in the sub-mucosal and mucosal layers. It shows strong expression in the lining of villi regions, which occupies the major part of the inner mucosal layer and finely co-localizes with the collagen VI α3 chain. The collagen VI α5 chain showed prominent expression in the mucosal layer and distinctly co-localizes with collagen VI α3. However, it is not as strongly expressed as the collagen VI α4 chain. Moreover, it exhibits a gradient of decreasing expression towards the lumen. Interestingly, the collagen VI α6 chain that is present in both cardiac and skeletal muscle is absent in the intestinal smooth muscle tissue of adult mice.

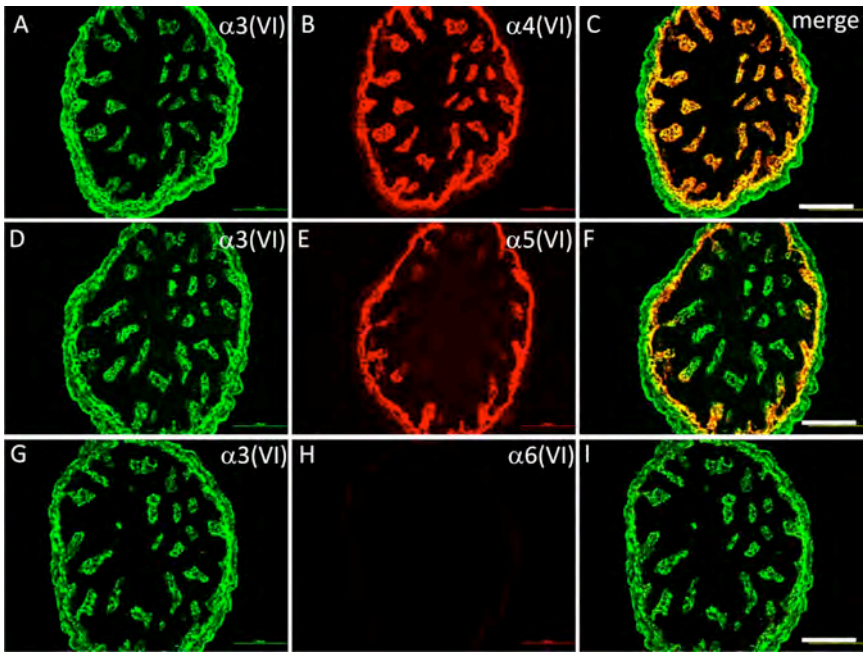


Figure 2.20 **Immunohistochemical analysis of intestine from newborn mouse.** Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. The α3 chain is widely distributed in all layers of intestine. The collagen VI α4 chain is expressed in the submucosal, mucosal and villi regions. The α5 chain is strongly expressed in the mucosal layer and weakly present in villi. The collagen VI α6 chain is absent in the newborn intestine. Scale bar, 100 µm.

The expression and distribution of the new chains in the intestinal smooth muscle of newborn mouse is similar to that in adult mice (Figure 2.20).

2.1.2.5.3 Collagen VI chain expression in kidney

Particularly selective and restricted expression pattern of the new collagen VI chains is seen in kidney. In sections of adult mouse kidney, the collagen VI α3 chain is widely distributed in the cortical (Figure 2.21) and medullar (not shown) regions. The capillary

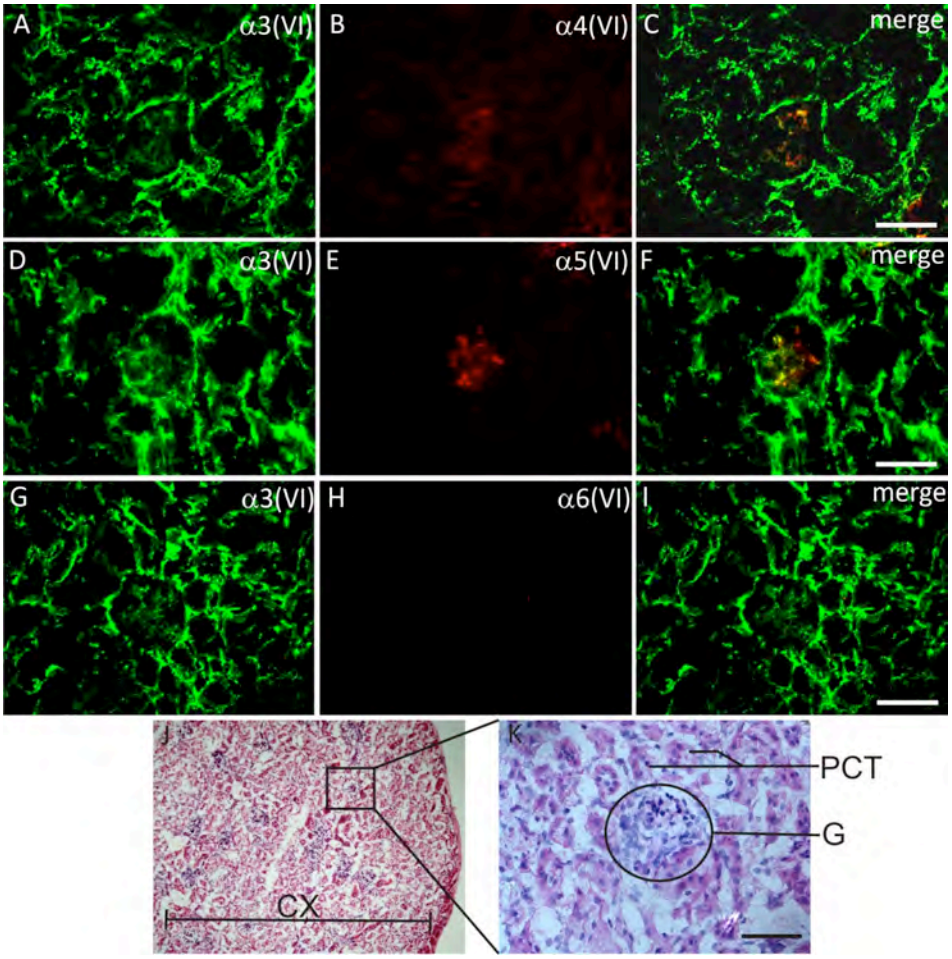


Figure 2.21 **Immunohistochemical analysis of kidney glomeruli from adult mouse.** Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of the α4(C), α5(F) and α6(I) chains with α3 is shown for comparison. The α5 chain is restricted to glomeruli and partially co-localizes with the α3 chain. The collagen VI α4 chain stains weakly in glomeruli while the α6 chain shows no expression. For orientation, a kidney section was stained with H&E (J and K). G, glomerulus, PCT, proximal convoluted tubules, and CX, cortex. Scale bar, 50 µm.

tuft/glomerulus and proximal convoluted tubules were labeled by the collagen VI α3 chain antiserum. Interestingly, the collagen VI α4 and α5 chains are highly restricted to the glomeruli where they are associated with basement membranes and do not completely co-localize with the collagen VI α3 chain. The expression of the collagen VI α5 chain is stronger and more specific when compared to the collagen VI α4 chain. The collagen VI α6 chain is completely absent in the cortical region and glomeruli (Figure 2.21). Collagen VI α3 is also strongly present in the kidney capsule. Here, the collagen VI α5 chain partially co-localizes with the α3 chain (not shown). In addition, the collagen VI α4 and α6 chains are weakly expressed in adult kidney capsule (not shown).

2.1.2.5.4 Collagen VI expression in mouse reproductive organs

Another striking differential and selective expression of the new collagen VI chains can be observed in the reproductive tissues of adult mice: ovary and testis. In sections of adult mouse ovary, the collagen VI α3 chain is present almost throughout the ovary. It is expressed in parts of the cortical and medulla regions of the ovary. The collagen VI α4 chain is primarily seen in the cortical layer of the ovary, particularly in the capsule-like structures or in the connective tissue surrounding the follicles, referred to as theca, which also contribute to the wall of the follicle. In addition, it is expressed in parts of the vascular medulla region of the ovary. Interestingly, the staining for the collagen VI α4 chain in ovary mostly co-localizes with that for the collagen VI α3 chain (Figure 2.22). The collagen VI α3 and α4 chains are also co-expressed in the outer connective tissue layer of the ovary, the tunica albuginea. On the other hand, the collagen VI α5 chain is only present in the stroma between follicles in the cortical region of the ovary and co-localizes with the collagen VI α3 chain. It is completely absent from the follicular granulosa or theca cells. Taken together, the expression patterns of the collagen VI α4 and α5 chains in ovary are complementary (Figure 2.22). In adult mouse testis, the collagen VI α3 and α4 chains are widely expressed and completely co-localize. Both chains are expressed in the peripheral surrounding layer of testis, the tunica albuginea, which is a fibrous connective tissue primarily made up of collagens. In addition, both

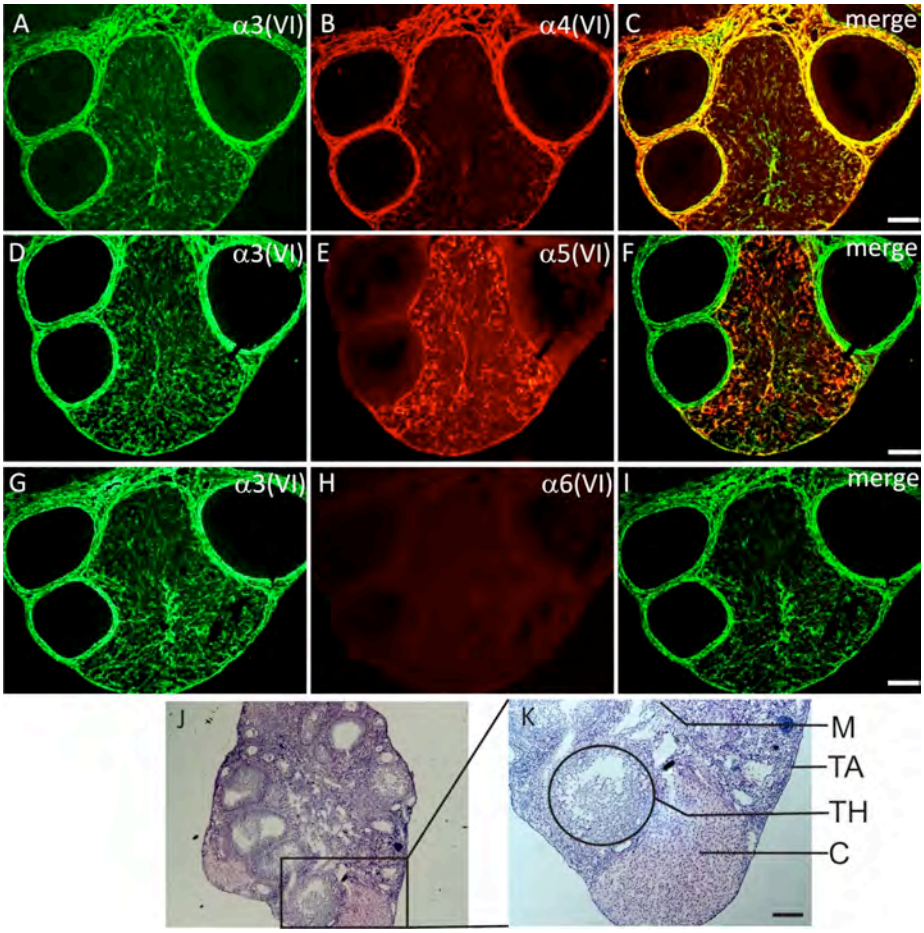


Figure 2.22 **Immunohistochemical analysis of adult mouse ovary.** Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of α4(C), α5(F) and α6(I) with the α3 chain is shown for comparison. For orientation, a section of the ovary was stained with H&E (J and K). The α3 chain is widely distributed in tunica albuginea, cortical stroma and follicular granulosa or theca cells. The collagen VI α4 chain shows strong expression in the follicular granulosa and a weak expression in cortical stroma. The α5 chain antiserum stains predominantly in the cortical stroma and not in the follicular granulosa. The α6 chain is absent in ovary. M, medulla, TA, Tunica albuginea, TH, theca cells, C, cortical stroma. Scale bar, 100 µm

chains are significantly present in the thin loose fibrous connective tissue surrounding the seminiferous tubules (Figure 2.23). In contrast, the collagen VI α5 chain is present in the tunica albuginea where it clearly co-localizes with the collagen VI α3 chain but is absent in the connective tissue of seminiferous tubules. The collagen VI α6 chain was only weakly present in the tunica albuginea but is clearly not expressed in the tissue surrounding seminiferous tubules (Figure 2.23).

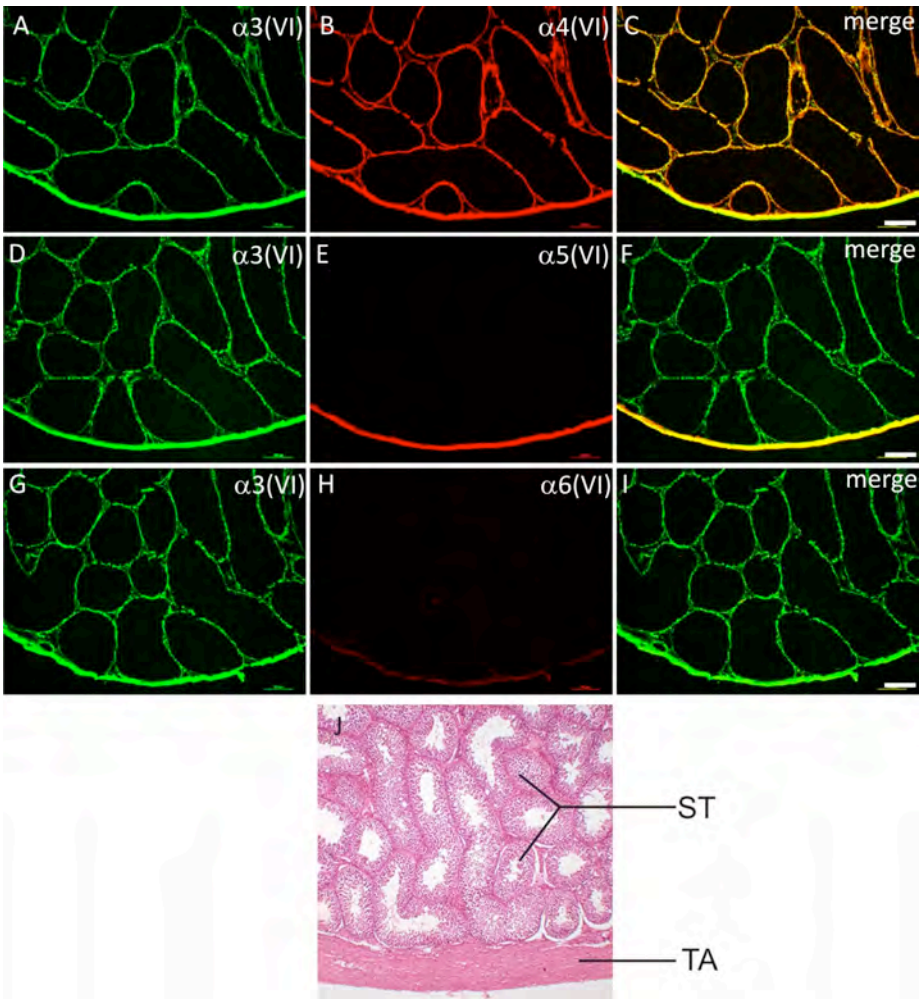


Figure 2.23 **Immunohistochemical analysis of adult mouse testis**. Frozen testis sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of α4(C), α5(F) and α6(I) with the α3 chain is shown for comparison. For orientation, a section of the testis was stained with H&E (J). The collagen VI α3 and α4 chains show strong expression and co-localize in tunica albuginea and in the loose connective tissue surrounding the seminiferous tubules. The α5 chain is restricted to tunica albuginea and co-localizes with the α3 chain. A very weak staining for the collagen VI α6 chain is observed at the tunica albuginea. TA, tunica albuginea, ST, seminiferous tubules. Scale bar, 100 µm.

2.1.2.5.5 Collagen VI chain expression in skin

In skin, the new collagen VI chains show differences in the expression patterns between newborn and adult stages. The collagen VI α3 chain is widely distributed. It is present at the dermal-epidermal junction, throughout the dermis and also in the subcutaneous

muscular layer below the dermis. It is also expressed surrounding the blood vessels and hair follicles, but it is absent in the epidermis. Interestingly, the collagen VI α5 chain shows a restricted and specific staining surrounding the small blood vessels of the

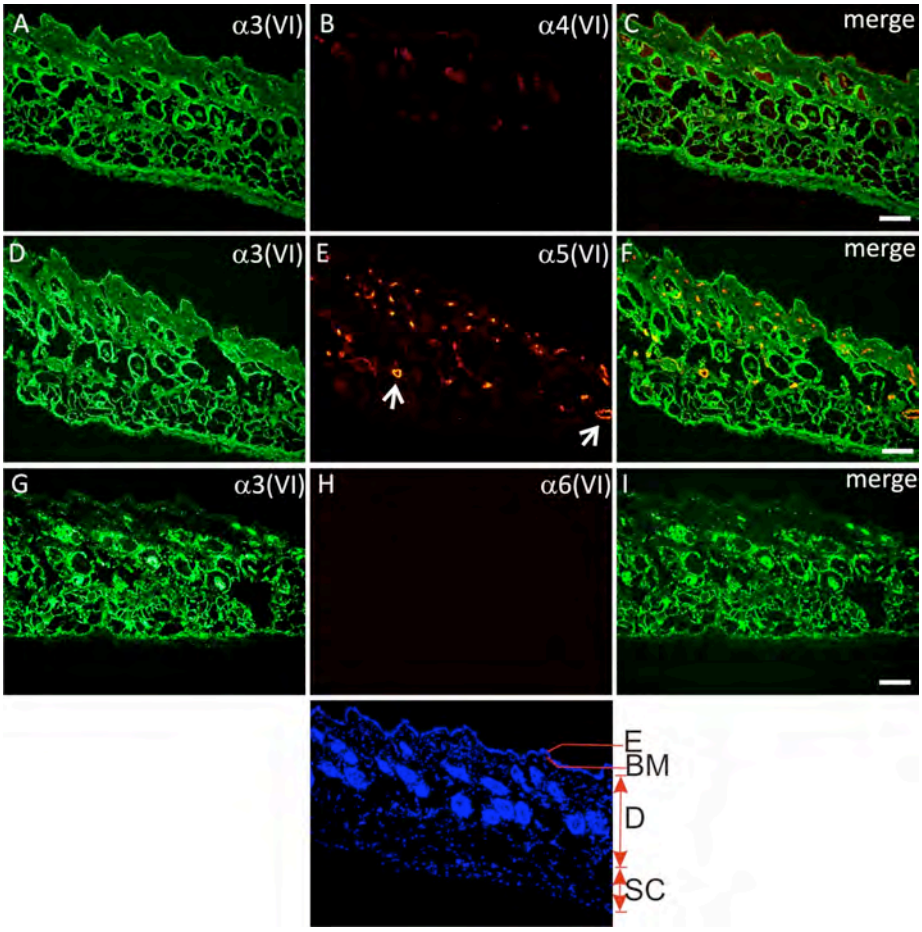


Figure 2.24 **Immunohistochemical analysis of adult mouse skin.** Frozen skin sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of α4(C), α5(F) and α6(I) with the α3 chain is shown for comparison. For orientation, a consecutive section was stained with DAPI (J). The α3 chain is widely distributed in the dermis but absent in the epidermis. The collagen VI α4 shows a weak, patchy staining in the dermis. The weak staining of the epidermis is not specific. The α5 chain shows restricted expression around the blood vessels (arrows) and a patchy expression, probably at nerves or macrophages, in the papillary dermis. The collagen VI α6 chain is absent in adult skin. E, epidermis, BM, basement membrane, D, dermis, SC, subcutaneous muscle layer. Scale bar, 100 µm

papillary dermis (Figure 2.24). Besides this, the α5 chain displays a patchy expression most probably representing nerves or macrophages in the dermis of adult skin. In all

regions it partially co-localizes with the widely expressed collagen VI α3 chain (Figure 2.24). The collagen VI α4 chain is very weakly present in the dermal region of adult mouse skin. None of the new chains are found in the basement membranes of the skin. The collagen VI α6 chain is completely absent in adult skin. However, the situation is

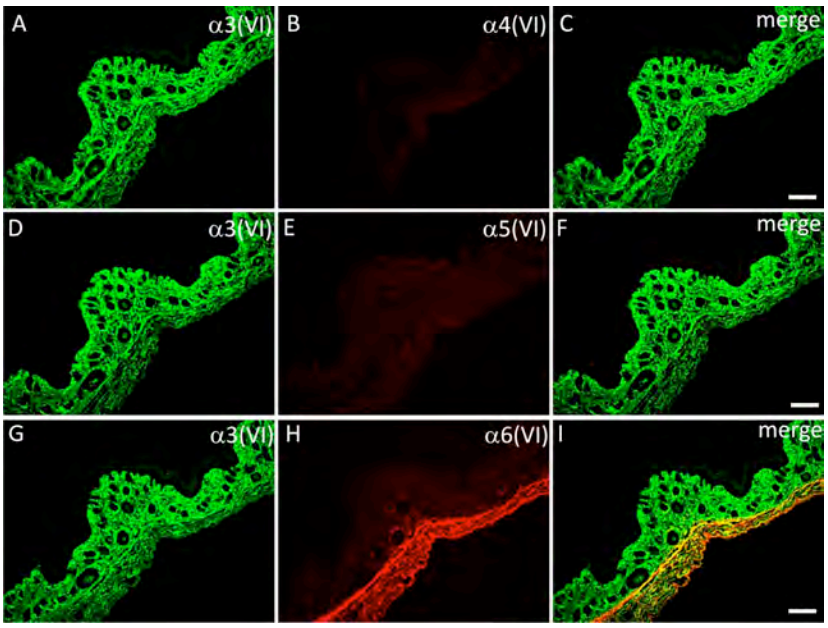


Figure 2.25 **Immunohistochemical analysis of newborn mouse skin**. Frozen skin sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of α4(C), α5(F) and α6(I) with the α3 chain is shown for comparison. The collagen VI α6 chain is present in the muscular layer below the newborn skin and co-localizes with the α3 chain. No expression of the new collagen VI chains was observed in the epidermis, the dermal epidermal basement membrane and dermis. Scale bar, 100 µm

completely different in newborn mice. The skin of a newborn mouse is absolutely devoid of the new collagen VI chains whereas the collagen VI α3 chain is present throughout the dermis and in the basement membranes of skin. Nevertheless, the collagen VI α6 chain is predominantly present in the muscle layer below the dermis and partially co-localizes with the collagen VI α3 chain (Figure 2.25).

2.1.2.5.6 The new collagen VI chains are absent in cartilage

The collagen VI α3 chain shows a very distinct and prominent expression in various layers of the growth plate of femur (thigh bone). It is weakly expressed in the resting

zone and stronger in the proliferating and hypertrophic cartilage. However, none of the new collagen chains is present in the growth plate. Interestingly, the α5 and α6 chains together with the α3 chain show a specific expression, probably in a ligament (Figure 2.26). In sagittal sections of a newborn mouse, the collagen VI α3 chain is broadly

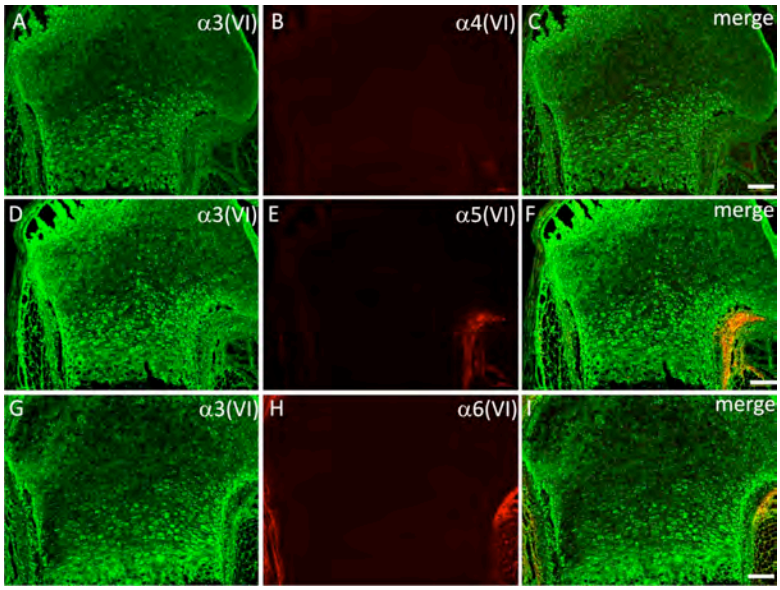


Figure 2.26 **Immunohistochemical analysis of newborn mouse knee cartilage** Frozen sections were incubated with affinity-purified collagen VI α3 (A, D and G), α4 (B), α5 (E) and α6 (H) antisera followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The overlay of α4(C), α5(F) and α6(I) with the α3 chain is shown for comparison. The collagen VI α4, α5 and α6 chains are completely absent in the cartilage while the collagen VI α3 chain is weakly expressed in resting zone and strongly present in proliferating and hypertrophic zones. α5 and α6 chains show specific staining in the ligament that connects the muscle and bone. Scale bar, 100µm

expressed in the intervertebral discs of fibrocartilage. Similar to in the knee cartilage, the new collagen VI chains are not present in the intervertebral fibrocartilage (not shown).

2.1.3 Fate of the new collagen VI chains in the Col6a1 null mice

The immunohistochemical analysis of the expression of the new collagen VI chains showed that they can occur together with α3 chain at specific sites. On the other hand the new chains are also present at sites where the α3 chain is not expressed. Based on structural information, it is likely that the new chains replace the α3 chain in collagen VI assemblies. The assembly of collagen VI containing α1, α2 and α3 chains has been

studied in detail. Triple helical molecules are not formed in *Col6a1* knockout mice leading to a complete absence of collagen VI in these mice (Bonaldo *et al.*, 1998). Therefore the expression of the new chains was studied in the knockout mice.

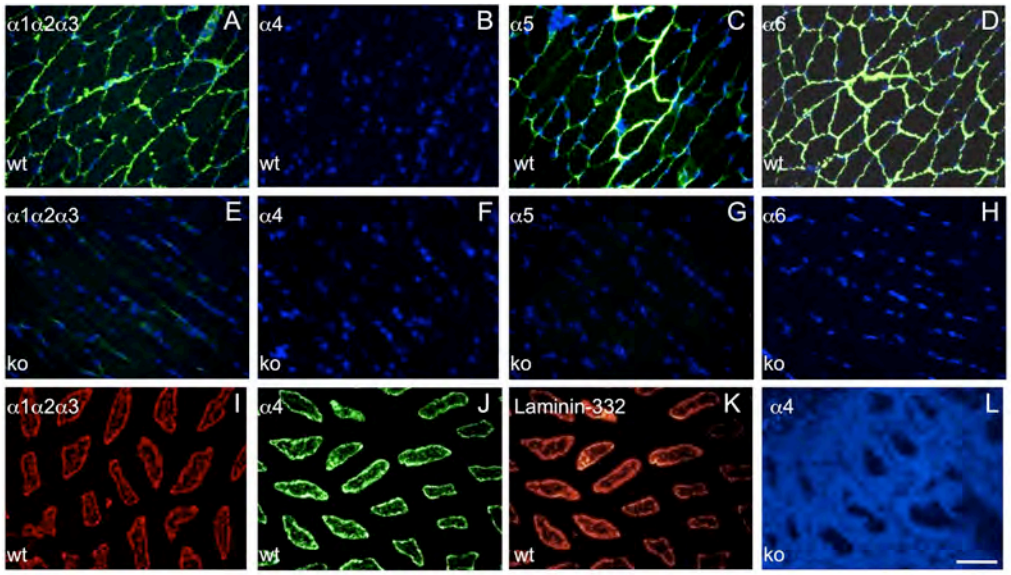


Figure 2.27 **Immunohistochemical analysis of wild type and *Col6a1* knockout mouse**. Immunohistochemistry was performed on frozen sections from mouse quadriceps femoris muscle (*A–H*) and small intestine (*I–L*) from wild type (*wt*) (*A–D* and *I–K*) or *Col6a1* knockout mice (*ko*) (*E–H* and *L*). Sections were incubated with the affinity-purified antisera against the collagen VI α4 (*B*, *F*, *J*, and *L*), α5 (*C* and *G*), and α6 (*D* and *H*) chains, human collagen VI from placenta (detecting the classical collagen VI chains α1, α2 and α3; *A*, *E*, and *I*) and laminin 332 (*K*), followed by an AlexaFluor (*green*, *A–H*), AlexaFluor 546 (*red*, *I* and *K*), or AlexaFluor 488 (*green*, *J* and *L*) conjugated secondary antibodies. Antibodies against the classical collagen VI chains (α1α2α3) and the α5 and α6 chains, but not such against the α4 chain, strongly stained the extracellular matrix surrounding the muscle fibers of wild type (*A–D*) mice. In small intestine, antibodies against classical collagen VI (α1α2α3) (*I*) and antibodies against the collagen VI α4 chain show co-staining with those against the basement membrane marker laminin-332 (*K*). In collagen VI α1 chain-deficient mice, staining for the new collagen VI chains is absent (*L*). Nuclei were counterstained with DAPI (*blue*, *A–H* and *L*). Scale bar is 100 µm.

Immunohistochemical staining on adult quadriceps femoris skeletal muscle of *Col6a1* deficient mice revealed that the α5 and α6 chains as well as the α3 chain are completely absent indicating a participation of the α1 chain in the assembly of collagen VI molecules containing the α5 and α6 chains (Figure 2.27). As there is no expression of the collagen VI α4 chain in skeletal muscle (See Figure 2.1), intestinal sections from the wild type and *Col6a1* deficient mice were tested. In the wild type intestine, the collagen VI α4 chain co-localizes with laminin 332 which is a marker for basement membrane structures.

Similar to the collagen VI α5 and α6 chains, the α4 chain is completely absent in the collagen VI α1 knockout mice. This confirms that all the new chains require *Col6a1* to form a triple-helical monomer. Furthermore, the absence of collagen VI α5 and α6 in *Col6a1* knockout mice was confirmed by immunoblot analysis of diaphragm extracts

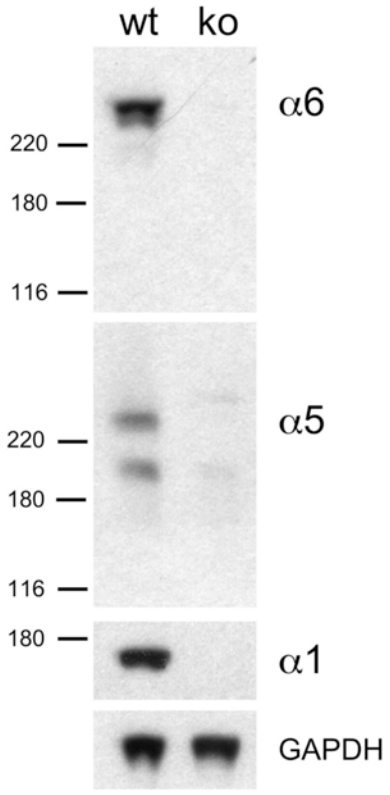


Figure 2.28 **Immunoblot analysis of collagen VI α5 and α6 chains in wild type and collagen VI α1 chain deficient mouse diaphragm extracts.** Diaphragm muscle of adult wild type (*wt*) and collagen VI α1 chain deficient mice (*ko*) were extracted with lysis buffer. The proteins were submitted to SDS-PAGE with prior reduction on a 4-12% polyacrylamide gradient gel, transferred to a membrane and immunostained with specific antibodies against the collagen VI α1, α5 and α6 chains or glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (*In collaboration with Paolo Bonaldo, Padova*).

(Figure 2.27). For wild type mice, incubation with antibodies specific for either the collagen VI α5 or α6 chain resulted in clearly identifiable bands above 220 kDa. When the same method was applied to diaphragm from collagen VI α1 chain deficient mice, no bands were detected, supporting the results from the immunohistochemical analysis. A GAPDH specific antibody was also used on the same extracts as a control to compare the total protein content of both wild type and knockout tissue extracts (Figure 2.28).

2.1.4 Expression of new collagen VI chains in man

As described earlier, only the genes coding for the collagen VI α5 and α6 chains have remained intact in humans (*see 2.1.7*) while the gene coding for the collagen VI α4 chain

was transformed to a pseudogene as a consequence of a large pericentric inversion in chromosome 3 (*see 2.1.7*). Therefore only the expression of the collagen VI α5 and α6 chains was studied. Affinity purified polyclonal antibodies that are specific to human collagen VI chains were used in sections of skeletal muscle and skin biopsies (*in collaboration with Patrizia Sabatelli, Bologna*). Although collagen VI is ubiquitously expressed, the main reason behind selecting these tissues was the clinical phenotypes associated with mutations in collagen VI genes in humans.

2.1.4.1 Expression in human skeletal muscle

Immunostaining on cryosections from healthy human skeletal muscle showed a prominent expression of the collagen VI α5 chain in the myotendinous junction (Figure 2.29A) that co-localizes well with the laminin α2 chain (Figure 2.29B). In contrast, the

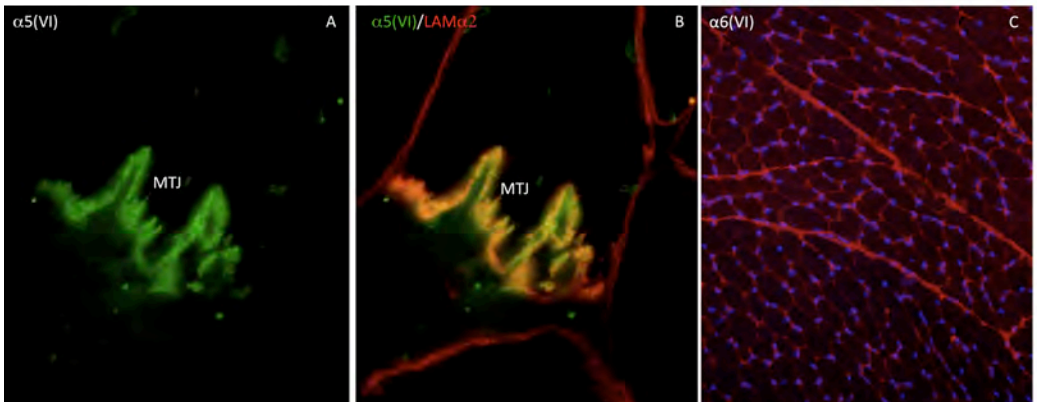


Figure 2.29 **Immunohistochemical analysis of human skeletal muscle**. Frozen sections were incubated with affinity purified collagen VI α5 (A&B) and α6 antisera (C) followed by AlexaFluor 488 (green) and 546 (red) secondary antibodies. The collagen VI α5 chain is strongly expressed in the myotendinous junctions (A) and co-localizes with the laminin α2 chain (B). The collagen VI α6 chain shows a broad distribution in both perimysium and endomysium of human skeletal muscle (C). Nuclei are stained with DAPI (blue).

collagen VI α6 chain is abundant and widely present in both perimysium and endomysium of the human skeletal muscle fibers (Figure 2.29C).

2.1.4.2 Expression in human skin

Immunofluorescence microscopy of normal skin sections labeled with monoclonal antibodies against the collagen VI α3 chain showed expression throughout the dermis

(Figure 2.30A,D), while labeling with polyclonal antibodies against human collagen VI α5 and α6 chains revealed a more restricted expression pattern. Collagen VI α5 is mainly expressed in the narrow zone of the papillary dermis (Figure 2.30B,E) just below the dermal-epidermal junction. Here the expression of collagen VI α6 is much weaker and discontinuous (Figure 2.30C,F). In contrast, the α5 and α6 chains were both strongly

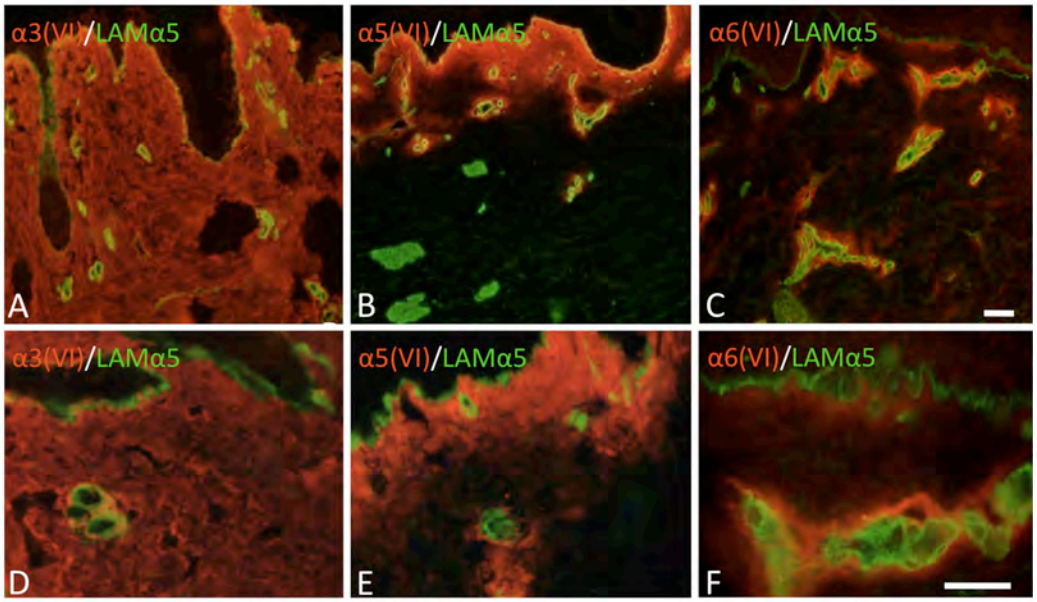


Figure 2.30 **Immunohistochemical analysis of human skin.** Frozen skin sections from a healthy donor with a monoclonal antibody against the collagen VI α3 chain (A,D) or polyclonal antibodies against the collagen VI α5 (B,E) or α6 chain (C,F) (red). Sections were double labeled with an antibody against laminin α5 (green), as a marker of basement membranes. Collagen VI α3 shows a broad distribution in the papillary and reticular dermis, including vasculature and nerves, and hypodermis. In contrast, the collagen VI α5 chain is localized in the papillary dermis, close to dermal-epidermal junction, and around some vessels of the reticular dermis. The α6 chain appears around the vessels of the papillary and reticular dermis, with a weaker and discontinuous labeling below the dermal-epidermal junction. A, B, C and D, E, F are at the same magnification. Bar, 100 µm

expressed around blood vessels at the interface between the papillary and reticular dermis, but not around the annexes in the deeper layers of the skin (Figure 2.30B,C). This pattern was confirmed by double labeling with antibodies against laminin α5, used as a marker for basement membranes (Figure 2.30).

2.1.5 The role of new collagen VI chains in human diseases.

A variety of pathological conditions have been linked to collagen VI and the disease association of the genes coding for the α1, α2 and α3 chains has been thoroughly studied

(Lampe & Bushby, 2005). In addition recent studies showed the association of the *COL6A5* and *COL6A4* genes to atopic dermatitis (Söderhall *et al.*, 2007) and knee osteoarthritis (Miyamoto *et al.*, 2008), respectively. As the lack of deposition of the new chains in the *Col6a1* knockout mouse show a close association with the expression of collagen VI α1 chain, the new chains may play a role also in the etiology of *COL6A1*, *COL6A2* and *COL6A3* linked diseases. Collagen VI is known to be primarily associated with muscular dystrophies such as Ullrich congenital muscular dystrophy and Bethlem myopathy. These patients have been shown to carry mutations in either of the genes coding for the collagen VI α1, α2 and α3 chains. Nevertheless, there are significant numbers of UCMD and BM patients in whom no mutation was detected in the old collagen VI genes.

2.1.5.1 Association to muscular dystrophies

A preliminary immunohistochemical analysis of human skeletal muscle biopsies from collagen VI deficient patients with UCMD phenotype who do not carry mutations in any of the three collagen VI α1, α2 and α3 chains revealed that the collagen VI α6 chain expression is remarkably reduced in endomysium (Figure 2.31) (*in collaboration with Patrizia Sabatelli, Bologna*). Nevertheless, the perimysium staining of collagen VI α6 could still be seen in both the patients N49 and N50. While in depth expression analysis of collagen VI α5 and α6 in UCMD and BM patients is underway, an initial attempt to

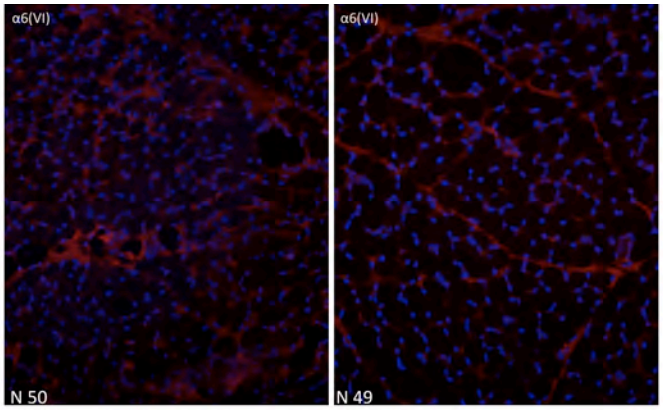


Figure 2.31 **Immunohisto-chemical analysis of the collagen VI α6 chain in UCMD patients**. Frozen sections of human skeletal muscle from two patients, N49 and N50 with UCMD phenotype but no mutation in *COL6A1*, *COL6A2* and *COL6A3* genes was stained with collagen VI α6. The expression of α6 was reduced in the perimysium of muscle fibers compared to Figure 2.29C. The nuclei are stained with DAPI (blue).

identify mutations in both the collagen VI α5 and α6 genes in patients suffering from UCMD and BM did not yield positive results (*in collaboration with Volker Straub,*

Newcastle and Patrizia Sabatelli, Bologna). However, this analysis needs to be extended also to patients carrying mutations in the old collagen VI genes, because additional mutations in the new genes modulate the severity of the disease.

2.1.5.2 Analysis of skin of patients with COL6A1, -A2 and -A3 mutations

BM and UCMD patients display distinct skin phenotypes in addition to the muscular defects. As the new chains are also expressed in skin, biopsies from four UCMD patients carrying mutations in the three genes (one in *COL6A1*, two in *COL6A2* and one on *COL6A3*) and one BM mutation with a recessive mutation in *COL6A3* were analyzed to assess whether the mutations in *COL6A1*, *COL6A2*, and *COL6A3* may effect the secretion and localization of the collagen VI α5 and α6 chains and thereby contribute to the observed phenotypes such as keloid scarring, cigarette paper skin or keratosis pilaris (*in collaboration with Patrizia Sabatelli, Bologna*). Immunofluorescence analysis with a

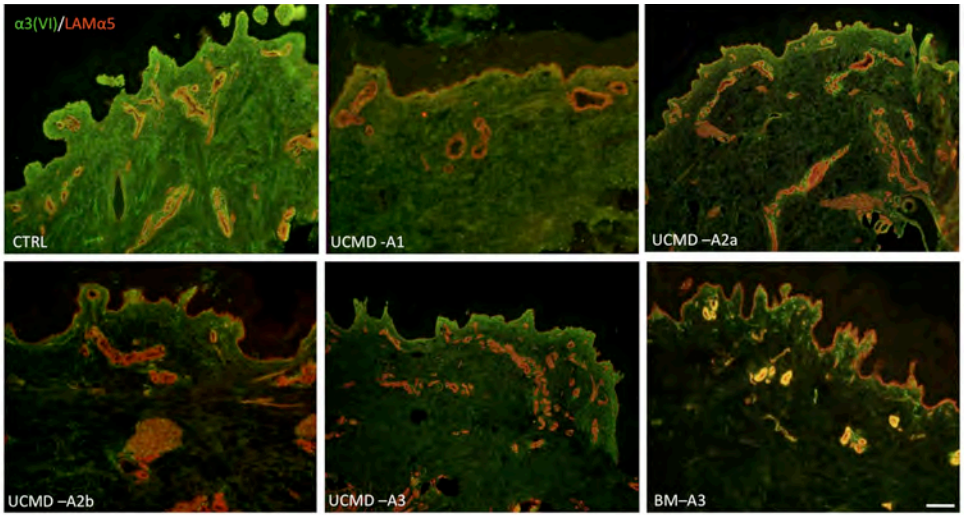


Figure 2.32 **Immunostaining for collagen VI α3 in skin of normal controls and of UCMD and BM patients**. Analysis of collagen VI in skin sections of four UCMD patients carrying mutations in *COL6A1* (UCMD-A1), *COL6A2* (UCMD-A2a and UCMD-A2b) and *COL6A3* (UCMD-A3) and one BM patient with a homozygous mutation in the *COL6A3* gene (BM-A3), compared with a healthy donor (CTRL). Staining was performed with antibodies against collagen VI α3 (green) and laminin α5 (red). The merged images show a reduced expression of collagen VI α3, ranging from mild (UCMD-A2a) to moderate (UCMD-A1 and UCMD-A3) or severe (UCMD-A2b and BM-A3). Bar, 100 µm.

monoclonal antibody against collagen VI α3 showed a collagen VI deficiency in the skin of all patients, ranging from moderate to severe (Figure 2.32). Immunostaining with the

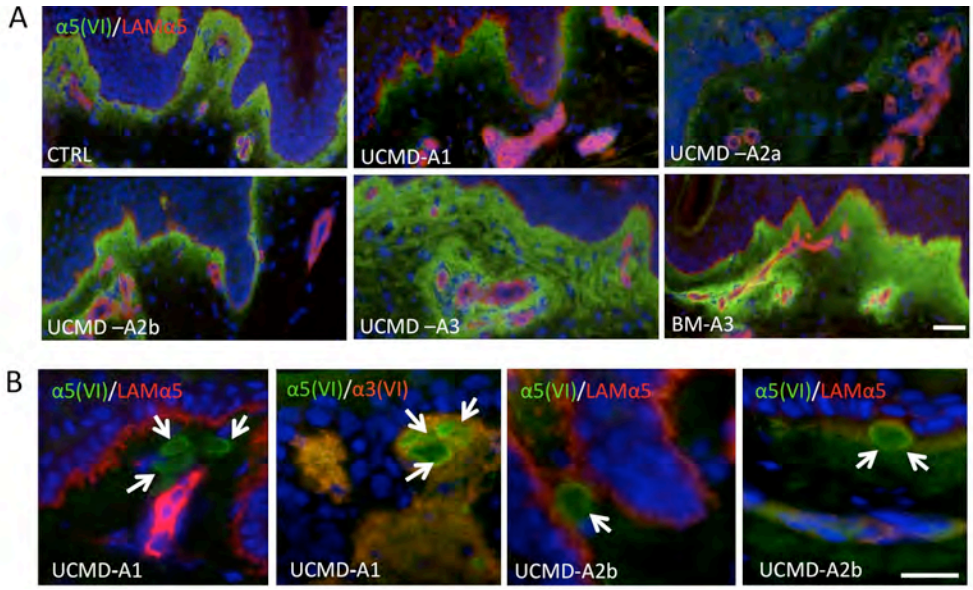


Figure 2.33 **Immunostaining for collagen VI α5 in skin of normal controls and of UCMD and BM patients.** Staining was performed with antibodies against collagen VI α5 (green) and laminin α5 (red). The nuclei are stained with DAPI (blue). The collagen VI α5 chain was reduced in UCMD-A1, UCMD-A2a and UCMD-A2b, but normally expressed in UCMD-A3 and BM-A3 (A). Staining with collagen VI α5 (green) and laminin α5 (red) or collagen VI α3 (red), in UCMD-A1 and UCMD-A2b (B). Anomalous roundish deposits (arrows) that stain for collagen VI α5 are seen at the epidermal-dermal interface, just below the basement membrane while collagen VI α3 was absent within these deposits (arrows). Bars, 100 µm.

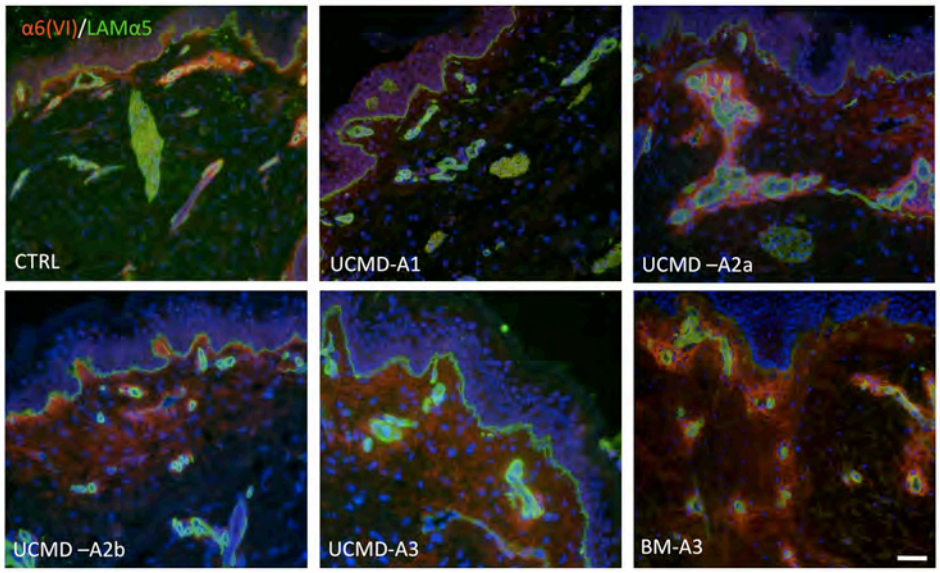


Figure 2.34 **Immunostaining for collagen VI α6 in skin of healthy controls and of UCMD and BM patients.** Staining was performed with antibodies against collagen VI α6 (red) and laminin α5 (green). The nuclei are stained with DAPI (blue). The collagen VI α6 chain was reduced in UCMD-A1 and normally expressed in UCMD-A2a. A wide distribution of α6 was observed in the papillary dermis of UCMD-A2b, UCMD-A3 and BM-A3 patients, Bars, 100 µm.

collagen VI α5 chain antibody showed altered expression in the papillary dermis of the patients with mutations in *COL6A1* or *COL6A2*. In the patient UCMD-A1 the α5 chain was markedly reduced at the papillary dermis and around blood vessels (Figure 2.33). Interestingly, the impact on the expression of the α5 chain varied in patients carrying different compound heterozygous mutations in *COL6A2*. While in the patient UCMD-A2a the signal for the α5 chain was evenly reduced, in patient UCMD-A2b the intensity of labeling was decreased in some areas of the papillary dermis both at the dermal-epidermal junction and around the vessels (Figure 2.33). In contrast, patients UCMD-A3 and BM-A3 that carry mutations in *COL6A3* showed a collagen VI α5 chain staining of normal intensity, however it appeared more widely distributed in both patients and particularly in UCMD-A3 (Figure 2.33). Moreover, in patients UCMD-A1 and UCMD-A2b double labelling for the collagen VI α3 and α5 chains revealed the presence of roundish deposits containing only the α5 chain (Figure 2.33B). These deposits, usually detected close to the basement membrane of the dermal-epidermal junction, showed an irregular shape and the diameter was variable, ranging from 0.5 to 20 μm. The α6 chain immunolabeling was markedly reduced in patient UCMD-A1, but largely unaffected in patients carrying mutations in *COL6A2* and *COL6A3* (Figure 2.34). However, it appeared more widely distributed in patients carrying mutations in *COL6A3*. In contrast, the roundish deposits were not labeled with the antibody for the α6 chain and were absent from normal skin (not shown).

2.1.5.3 Association to knee osteoarthritis

While the characterization of the new collagen VI chains was underway, Miyamoto and colleagues identified the 5' part of the split *COL6A4* gene (see *section 2.1.2)* on chromosome 3p24.3 to be associated with a susceptiblity for knee osteoarthritis. They named the gene *DVWA* (Dual von Willebrand factor A domain) as it contains exons coding for the N-terminal VWA domains of *COL6A4*. However, they were unable to recognize the real nature of the locus although the bioinformatic analysis yielded an ambiguous result. A FASTA similarity search of the DVWA/COL6A4 sequence with the murine collagen VI α4 chain sequence revealed a 72% identity between the two sequences at the amino acid level (Appendix Fig. 3). However, only 49% of the amino

acid sequence forming the first (N7) and 75% of that forming the second VWA domain (N6) are present in the published DVWA sequence (Figure 2.35). Analysis of the exon-

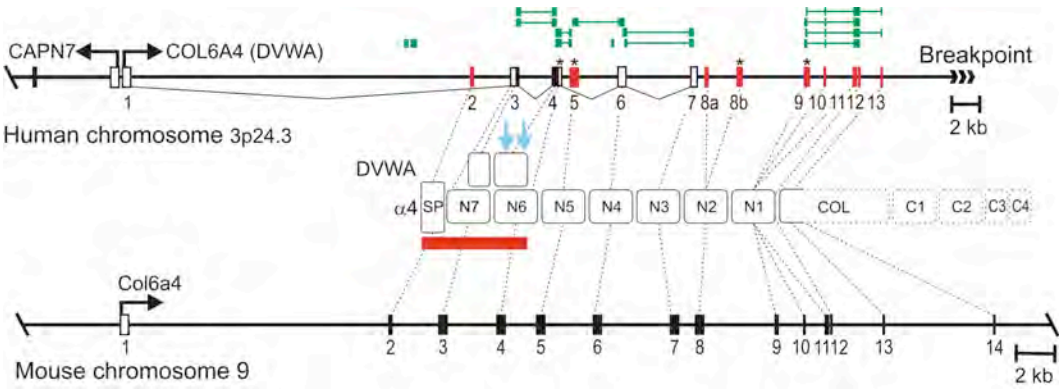


Figure 2.35 **Comparison of the 5' regions of the murine and the human *COL6A4/DVWA* genes.** Exons in the proposed human DVWA gene are depicted in black and their splicing is indicated by lines. Translated exons are shown as filled, untranslated exons as open boxes. Additional exons of the human COL6A4 gene are depicted in red. Asterisks mark exons in *COL6A4/DVWA* that contain at least one stop codon. The position of EST clones derived from the DVWA/COL6A4 locus are given in green. The predicted domain structures of the collagen VI α4 chain and the proposed DVWA protein are given between the gene structures. SP, signal peptide, N7-N1, N-terminal VWA domains, COL, collagenous domain, C1 and C2, C-terminal VWA domains, C3, unique domain and C4, Kunitz domain. Domains that are encoded by downstream exons are dashed. The red bar indicates the open reading frame of the truncated human COL6A4. Blue arrows mark the positions of the polymorphisms linked to knee osteoarthritis.

intron organisation of *COL6A4* (*DVWA*) revealed that the first exon is located just 266 bp 5' of the neighboring CAPN7 gene (Figure 2.35). In the published *DVWA* cDNA sequence the first exon is spliced to an exon that is orthologous to the third exon of the murine *Col6a4* gene. The *DVWA* cDNA lacks the corresponding second exon coding for the signal peptide. However, the reading frame for a complete VWA domain is maintained in the third exon and therefore the proposed reading frame for *DVWA* starts at a conserved methionine residue that corresponds to a methionine residue in the most N-terminal VWA domain (N7) of the collagen VI α4 chain (Appendix Fig. 4). At the DNA level the next exon of *DVWA* is 82% identical to the corresponding fourth exon of murine *Col6a4*, coding for the VWA domain N6. However, due to the insertion of an adenine at position 1280 a frameshift occurs that leads to a premature stop codon at position 1332 (Appendix Fig. 5). Interestingly, the only polymorphism that shows global association

with knee osteoarthritis (Meulenbelt *et al.*, 2009), the cysteine 260 to tyrosine polymorphism, is located within in the frameshifted sequence close to the C-terminus, whereas the tyrosine 169 to asparagine polymorphism is located in the truncated N6

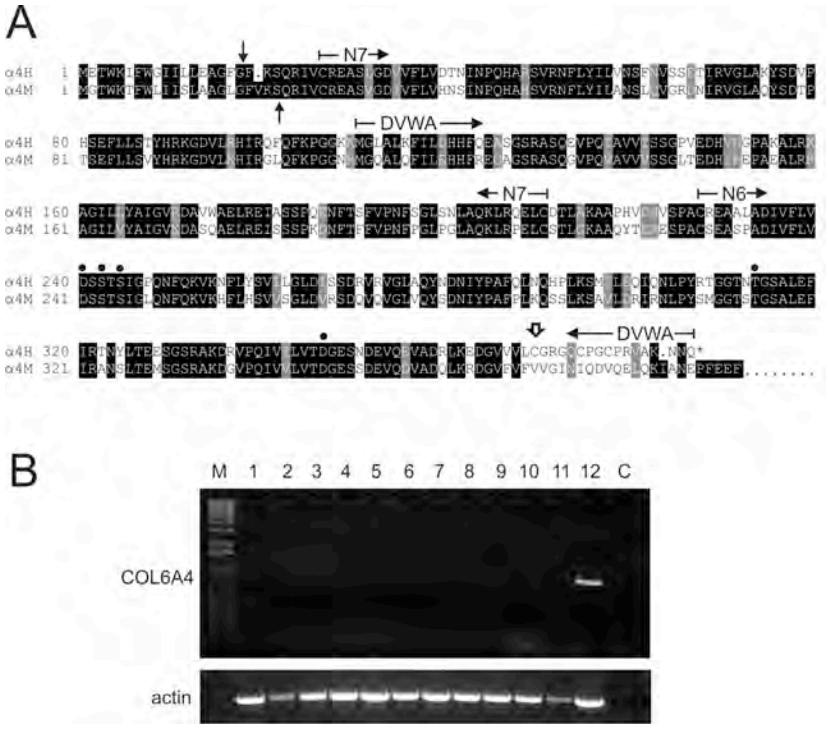


Figure 2.36 Human collagen VI α4 chain. A) Amino acid alignment of human and mouse collagen VI α4 chains. The amino acid sequence of the human collagen VI α4 chain (α4H) was deduced from the cDNA sequence deposited in the database under accession number FN394065, the partial murine sequence (α4M) is from the database, accession number A2AX52. Vertical arrows mark the potential signal peptide cleavage sites. Horizontal arrows indicate the boundaries of the domains N6 and N7 and of DVWA as depicted in Fig. 1B. An open arrow indicates the position of the frameshift. The residues forming the conserved metal ion-dependent adhesion sites are denoted with dots. B) RT-PCR analysis of the COL6A4/DVWA mRNA was performed using primer pairs BK417 and BK425. Template RNA was isolated from primary human chondrocytes (1-7), human cartilage (8-11) and HEK293 cells (12). All cartilage donors were of european origin. Lanes 9 and 10 show RTPCRs performed on different RNA preparations obtained from the same donor. C, negative control. RT-PCR for actin was used to control cDNA quality. The 1kb ladder (M) from GIBCO BRL was used as a reference.

domain. As all collagens need a signal peptide to enter the secretory pathway, the lack of a sequence coding for a signal peptide in the DVWA cDNA is unlikely. Indeed, in a search for an exon downstream of exon 1 in *COL6A4* (*DVWA*), the signal peptide sequence coding exon could be identified. It corresponds to the murine exon 2, which in

addition contains a putative splice site 13 bp upstream of the AUG codon (Appendix Fig. 6). Several human EST clones could be identified that code for the N-terminal VWA domains of the collagen VI α4 chain, however similar to the published *DVWA* cDNA none of these contained the putative signal peptide coding sequence (Figure 2.35). Therefore RT-PCR was performed to clone the predicted signal peptide coding cDNA, using primers deduced from the genomic sequence. The signal-peptide-containing collagen VI α4 chain sequence could be amplified from the human embryonic kidney cell line HEK293 (Figure 2.36). The protein coding sequence comprises exons 2-4 (Figure 2.36) and as in the published *DVWA* cDNA the frameshift in the fourth exon leads to a premature stop codon (Fig. 2.36A). The cloned human α4 cDNA of 1269 bp (Appendix Fig. 7; accession number FN394065) contains an open reading frame of 1155 bp, encoding a protein consisting of 385 amino acid residues preceded by a signal peptide of 18 residues, as predicted by a method using neural networks or hidden Markov models (Bendtsen et al. 2004). The mature secreted protein has a calculated Mr of 40,564 and contains only the complete most N-terminal VWA domain (N7) and, as DVWA, a truncated VWA domain N6. Interestingly, although we were able to amplify cDNA transcribed from the 3' part of the split gene (not shown), we were not able to amplify the 5' *COL6A4/DVWA* cDNA from human cartilage or primary human chondrocytes (Figure 2.36B). cDNA starting from the non-coding exon 1 of DVWA could also not be amplified from human HEK293 cells.

2.2 TARGETING OF THE COLLAGEN XXVIII GENE IN MOUSE

Recently, collagen XXVIII, similar to collagen VI in that it contains VWA domains N and C terminal of the collagenous domain, was identified. However, only its gene structure and tissue distribution is known (Veit *et al.*, 2006). Collagen XXVIII shows a very specific expression close to the basement membranes in dorsal root ganglia and peripheral nerves (Veit *et al.*, 2006). To study the function of collagen XXVIII the generation of a knockout mouse was initiated.

2.2.1 Strategy for the construction of the Col28a1 knockout targeting vector

The mouse *Col28a1* is located on chromosome 6A1 and contains 34 exons that are spread over a length of 195 kb. The exon 0 encodes the 5' untranslated region whereas the exon 1 codes for the signal peptide sequence. The N-terminal VWA domain is encoded by exon 2 and exon 3 (see Figure 1.7). The strategy was to insert a stop codon into exon 2 so that only a short unstable truncated version of collagen XXVIII is produced. The final targeting vector was constructed in two steps (Figure 2.37). In the first step a preliminary targeting vector was made with a selection cassette containing the neomycin resistance gene flanked by FLP recognition sequences and thymidine kinase gene sequences for positive and negative selections respectively. A reporter gene, PALP (Placental ALkaline

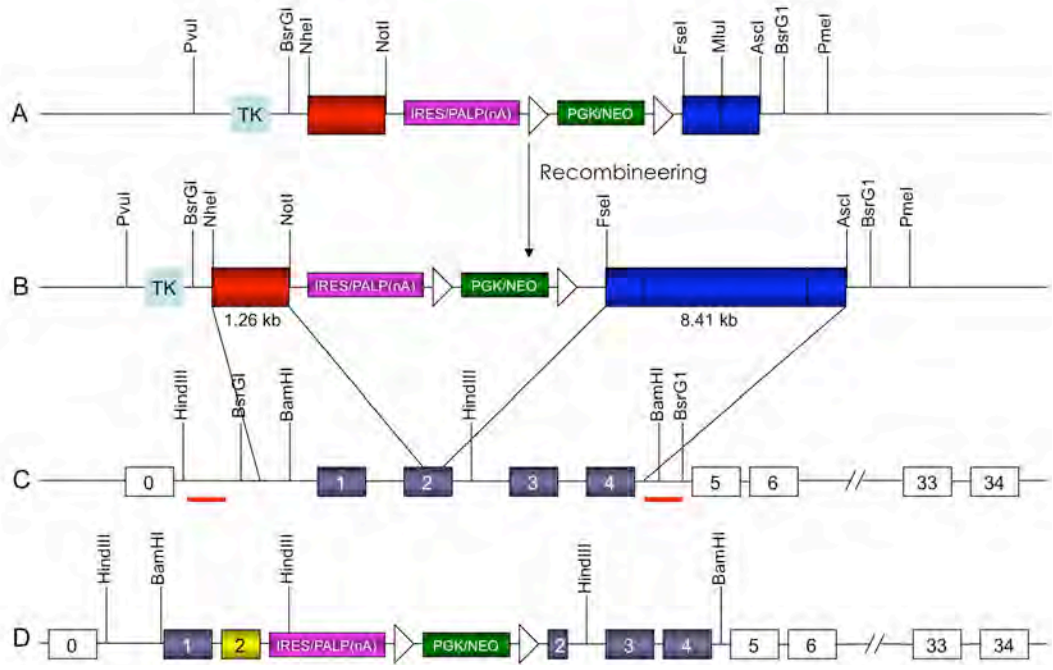


Figure 2.37 **Schematic representation of the *Col28a1* knockout strategy.** The 5' homology arms (red) and two short 3' homology arms (blue) were cloned together with the selection cassette before proceeding to recombineering. After recombineering, the final targeting vector containing the long 3' homology arm of around 8.41 kb was incorporated where the final targeting vector was then transfected to ES cells that contain the wild type collagen XXVIII allele to generate a mutant or recombinant allele. The mutated exon 2 is represented in yellow. TK, thymidine kinase. A, preliminary targeting vector. B, final targeting vector, C, wild type allele, D, targeted allele,

Phosphatase), that later allows tracing the expression of the targeted gene by an internal ribosomal entry site, IRES, is inserted upstream of the neomycin resistance gene. The cloned 5' homology arm is about 1.26 kb and contains the complete first exon and a part

of the second exon. On the other hand, the final 3' homology arm is about 8.4 kb and consists of a remaining part of the second, complete third and fourth exons. However, initially two short fragments (~500 bp) from either side of the intended 8.4 kb long 3' homology arm were joined and inserted into the 3' end of the cassette. The second step was to introduce the preliminary targeting vector into DH10B cells that contain a BAC clone with the collagen XXVIII gene. In this step, as a result of homologous recombination (recombineering), the region (~7.4 kb) between the two short fragments was retrieved from the BAC-DNA into the preliminary targeting vector. In order to confirm the integrity of the final targeting vector, it was subjected to several restriction digestion analyses with enzymes *NheI* and *EcoRI* (Appendix Fig. 8).

2.2.2 ES-cell transfection and screening for positive clones

Homologous recombination was used to modify the mouse chromosome 6 A1 encoded *Col28a1* gene locus. The final targeting vector was linearized by use of an restriction site with *AscI* that is present downstream to the 3' homology arm and then transfected into the embryonic stem (ES) cells V6.5, derived from the 129xC57BL/6 mouse strain. Upon

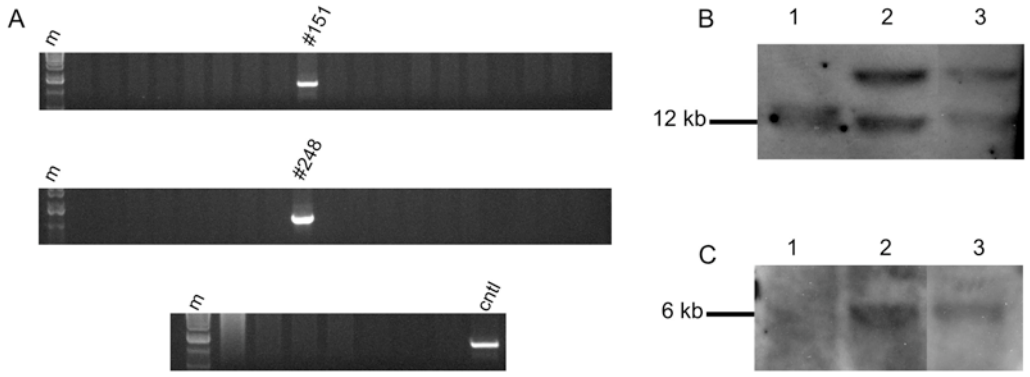


Figure 2.38 **Screening of ES cells**. The isolated embryonic stem cell clones were tested for the 5' and 3' integration of the targeting vector using PCR (A) and Southern hybridization (B) respectively. The presence of selection cassette in both the ES cell clones was confirmed with neo probe (C). In panel A, three independent agarose gels were selected to show the positive clones. In panel B, lane 1, control (wild type ES cell DNA); lane 2, and 3 are ES cell clones #151, and #248 respectively. m, 1 kb ladder. cntl, positive control (targeting vector).

the homologous recombination of the introduced final targeting vector DNA and the C57BL/6 genomic DNA, the ES cells were grown under G418 and gancyclovir selection markers. Around 300 ES cell clones that survived the selection were isolated, cultured

and screened initially for the integration of the 5' homology arm with PCR probes (Figure 2.38A). Only two out of 300 clones (#151 and #248) gave a positive signal at the expected size, which corroborated the insertion of 5' end of the targeting vector (Figure 2.38A). The genomic DNA from the two clones were isolated and subjected to restriction digest with *BsrGI* followed by Southern hybridization with the 3' probe to verify the integration of the 3' end of the targeting vector (Figure 2.38B). The targeted allele should run at a size of 12 kb while the wild type allele should run at 16.7 kb. Both the ES cell clones (#151 and #248) prominently contain both the wild type as well as targeted allele. Similarly, the integrity of the neo-cassette was confirmed by *HindIII* digestion of ES cell DNA followed by a southern hybridization with neo-probe. The expected band for the integration of neo-cassette was around 6 kb. As expected, the wild type ES cell DNA does not show any signal but both the clones show a band at the respective size confirming the presence of the neo-cassette (Figure 2.38C).

2.2.3 Generation of chimeric mice

As both the ES cell clones #151 and #248 were completely positive for the integration of neo-cassette, 5' and 3' homology arms, these ES cell clones were independently injected into CB20 blastocysts and subsequently implanted into the uterus of pseudo-pregnant foster mice strain. However, despite of several attempts, the mice that received #248 ES cell clones did not become pregnant. The mice that received #151 ES cell clones

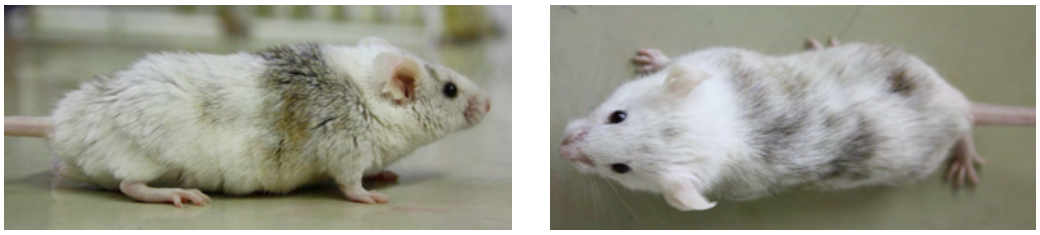


Figure 2.39 **Chimeric mice generated from engineered ES cell DNA**. A representative of the three chimeric mice (65%) is shown. Snap shot from side view (left) and top view (right).

conceived and gave birth to several pups. However, only three viable chimeric male mice (65%, 60% and 25%) resulted (Figure 2.39). These were used for analyzing the germ-line transmission of the targeted allele.

2.2.4 Examining the germ-line transmission of the targeted gene

In order to generate a heterozygous mutant mouse for *Col28a1* gene, the three viable chimeric male mice were backcrossed with C57BL/6 female mice. The tail biopsies of their progeny were analyzed by PCR with neo and 5' probes. A total of seven litters from each male chimeric mouse were tested to observe the germline transmission of the targetted allele. A representative genotyping analysis from one litter is shown (Appendix Fig. 9). However, the targeted allele did not enter the germline.

3. Discussion

3.1 Collagen VI

Collagen VI is unique within the large family of collagenous proteins in that it can form abundant microfibrils in a variety of extracellular matrices. It was first identified as a collagen of probable basement membrane origin occurring in several tissues (Chung *et al.*, 1976). Collagen VI serves as a substrate for cell attachment and anchors the meshwork that connects collagen fibers, nerves and blood vessels to the surrounding matrix (Keene *et al.*, 1988; Bonaldo *et al.*, 1990). For about three decades, collagen VI was considered to contain triple helical monomers made up from three different α1, α2 and α3 polypeptide chains. This study describes the identification and biochemical characterization of yet another three collagen VI chains, α4, α5 and α6, that appear to serve as alternatives to the α3 chain. The detection of new collagen VI chains raised the question why these have not been detected earlier as components of collagen VI assemblies. Indeed, several studies gave results that could be explained by the existence of the new chains. For instance, a heterogeneity in the α3 chains produced by different cell lines was observed by immunoprecipitation using polyclonal antibodies (Colombatti *et al.*, 1995). It was also reported that the ratio between the α1/α2 chains and the α3 chain may differ depending on the proliferative state of cells (Kielty *et al.*, 1990) or the culture system in which cells are grown (Hatamochi *et al.*, 1989). The *in silico* and biochemical characterization of the new chains presented here widens our understanding of collagen VI and the variability of this multifunctional protein.

3.1.1 Gene structure and evolution of new chains

In mouse, the three new genes are arranged in tandem on chromosome 9 and were numbered according to their appearance from 5’ to 3’ on the coding strand. These genes have previously only been incompletely annotated or incorrectly predicted by conceptual translation or gene prediction programs. Moreover, the three new chains possess a similar gene structure. Interestingly, the mouse collagen VI α1 and α2 genes are also arranged in tandem on chromosome 21 (Weil *et al.*, 1988) and have similar gene structures.

However, the new genes structurally resemble the collagen VI α3 gene, which is present on chromosome 2, rather than *Col6a1* and *Col6a2*. The close relationship between the α3 chain and the new chains is reflected in the almost identical exon/intron organization of the portions of the respective genes encoding the collagenous domains. Among the 19 coding exons, the exon lengths and the number of the exons coding for collagenous domain are the same with the exception of the last exon. Although the lengths of the N- and C-terminal globular domains of the new chains vary among themselves and also compared to the α3 chain, their overall structure is still conserved pointing to the importance of the globular domains in collagen VI function. Interestingly, the major splice variants of the collagen VI α3 mRNA code for only seven N-terminal VWA domains, similar to the major variants of the new chains (Dziadek *et al.*, 2002). The sequences and the domain structures of the new chains show that they have evolved as the consequence of a series of gene duplications originating from their common ancestor, the collagen VI α3 gene. Probably the first duplication gave rise to the α4 chain, as this chain still carries a C-terminal Kunitz domain. The tandem orientation of the new collagen VI genes indicates that the α5 and α6 genes resulted from sequential duplication of the α4 gene. Furthermore, the sequence identity of 95% at the DNA level of the N7 domains of the α5 and α6 chains supports the likelihood of their sequential duplication. Moreover, another duplication of exons coding for the VWA and unique domains occurred exclusively in the C-terminal region of the collagen VI α5 gene, which suggests that the α5 chain has been more extensively modified than the other new chains. From sequence analysis, it is quite evident that the collagen VI α1 and α2 chains must have originated from a gene duplication (Francomano *et al.*, 1991). Therefore, the six chains in collagen VI are a result of duplications from two common ancestral genes (*Col6a1*/*Col6a2* precursor and *Col6a3*). Moreover, also these genes possess a certain similarity in sequence and gene structure at the collagenous domain indicating that they may have a common ancestor.

Also collagen IV consists of six genetically distinct chains. In this case, the six genes evolved through three consecutive gene duplications, probably from a common ancestral gene (Khoshnoodi *et al.*, 2008). However, although the six genes may have their own common ancestor both in case of collagen IV and collagen VI, the collagen IV

genes are more closely related than the collagen VI ones. The collagen IV genes occur in three pairs on different chromosomes (Khoshnoodi *et al.*, 2008). The collagen VI genes are present in head to tail orientation, whereas in case of collagen IV a head to head arrangement is found. This allows sharing a common promoter in the middle, directing the bidirectional transcription of the two genes (Heikkila & Soininen, 1996). However, in order to allow a differential expression of *COL4A5* and *COL4A6*, alternative promoters may be used (Oohashi *et al.*, 1994; Sugimoto *et al.*, 1994; Zhou & Reeders, 1996; Zhou *et al.*, 1993). In contrast, the collagen VI genes have uni-directional promoters.

A preliminary analysis of the vertebrate genomic databases shows that only one of the new collagen VI genes is present in *Danio rerio* and *Gallus gallus* while all the three new genes are found in mammals like *Bos taurus, Canis familiaris* or *Equus caballus*. As most mammals possess all three genes, the sequential duplication could be a rather novel event in evolution. In humans the new genes are present on a very interesting part of chromosome 3 where a large pericentric inversion occurred after the division between Homininae and Ponginae (Muzny *et al.*, 2006). The 3' breakpoint of the inversion is located within *COL6A4* and leads to its inactivation. Although both parts of *COL6A4* are still present and can be easily identified by their sequence, both have become transcribed non-processed pseudogenes. Thereby Homininae have become natural *COL6A4* knockouts (Figure 2.6). However, the collagen VI α5 and α6 chains are still intact in humans. This raises the question if one of the remaining genes, *COL6A5* or *COL6A6*, has taken over the function of *COL6A4*. The major structural difference between the collagen VI α4 chain and the α5 and α6 chains is at the C-terminus, where the fibronectin type III domain and the Kunitz domain occur only in the α4 chain. However, comparing human and mouse collagen VI α5 and α6 chains, the α5 chain diverges especially at the C-terminus, which in addition shows alternative splicing and could represent an adaptation to a need to replace the α4 chain in Homininae.

3.1.2 Molecular assembly and interactions of new collagen VI chains

The present model of collagen VI assembly includes only the α1, α2 and α3 chains which assemble intracellularly from heterotrimeric monomers into dimers and tetramers, that are then secreted to form extracellular beaded filaments by interactions of their non-

collagenous domains (Furthmayr *et al.*, 1983). There is good evidence that α3 chain expression is essential for the formation of functional collagen VI molecules, as human SaOS-2 cells that are deficient for the α3 chain do not produce triple helical collagen VI (Lamande *et al.*, 1998b). Although the length of the collagenous domain of the collagen VI α1 chain is identical to those of the new chains and the α2 chain is only one amino acid residue shorter, there are other criteria that clearly show the closest relationship of the new chains with the α3 chain. First, the exact position of the single cysteine residue within the collagenous domain is conserved in the α3, α4, α5 and α6 chains. In the α3 chain this cysteine appears to be involved in tetramer formation and stability (Lampe & Bushby, 2005). The α1 and α2 chains also contain one cysteine each, but these are at a different position and it has been suggested that they are involved in the stabilization of the supercoil that is formed during antiparallel dimer formation (Chu *et al.*, 1988). Second, it has also been proposed that the supercoiled dimer is partially stabilized by ion pairs between different segments along the supercoil (Knupp & Squire, 2001). In the α1, α2, α3 chain heterotrimer the supercoiled part of the α1 chain carries a high negative net charge, while that of the α3 chain has a high positive net charge and that of the α2 is close to neutral. All three new chains carry a positive net charge that is even higher than that of the α3 chain. In addition, the positions of the two imperfections present in the α3 chain, giving the supercoil a clearly segmented character (Knupp & Squire, 2001), are conserved. Third, phylogenetic analyses based on the collagenous domains, using protein distance and protein parsimony methods, clusters the new chains to the collagen VI α3 chain, whereas the α1 and α2 chains form a different branch (Figure 2.5). Together, these data indicate that the new chains may substitute for the α3 chain probably forming α1α2α4, α1α2α5 and α1α2α6 heterotrimers.

Besides the collagenous domains, all three new chains contain N- and C-terminal globular domains that are similar to those of the old chains. However, the domain structure of the new chains most closely resembles that of the α3 chain. Similar to in the collagen VI α3 chain, the major part of the globular domain in the new chains is made up by VWA domains. A detailed bioinformatic analysis of the distribution and evolution of VWA domains showed the existence of ~200 VWA domain coding exons in the human

genome (Whittaker & Hynes, 2002). The collagen VI chains alone contribute 46 VWA domains and collagen VI could be considered the prototype member of the VWA domain containing protein superfamily. The VWA domain is known to participate in protein-protein, cell-cell and cell-matrix interactions. A variety of such interactions involving the VWA domain have been identified for collagen VI. Interestingly, in a phylogenetic analysis of all VWA domains of collagen VI, only the C1, C2 and N1 domains of the new chains cluster together with the respective domains of the α3 chain, suggesting that these domains are crucial and could potentially play a similar role as in the α3 chain. In contrast, the corresponding domains of the α1 and α2 chains cluster together. Similar to for many other collagens, the assembly of collagen VI starts from the C-terminal ends where specific interactions between the VWA domains C1 direct the stoichiometric assembly of chains into a triple helical molecule (Lamande *et al.*, 2002). However, for the initiation of monomer formation the C-terminal domain does not necessarily need to be a VWA domain. For instance, the structurally distinct C-terminal non-collagenous (NC1) domains of both collagen VIII and X have been shown to be critical for chain association (Marks *et al.*, 1999; Illidge *et al.*, 1998). In collagen VI, the C1 domain of the α3 chain is known to be crucial for monomer formation (Ball *et al.*, 2001). All three new chains share a high sequence similarity at the C1 domains (~62%) that possess exactly the same size and are flanked by conserved cysteine residues, which might be important for assembly. Moreover, they form a single cluster together with the α3 chain in the phylogenetic trees (Figure 2.5). In addition, the C1 domain of the collagen VI α1 chain is also known to interact with the C-terminal domain (NC1) of collagen IV, thereby anchoring the endothelial basement membrane around blood vessels and nerves. Interestingly, the highly similar α2 chain C1 domain does not show any interaction (Kuo *et al.*, 1997). In contrast, this domain appears to be essential for higher assemblies such as dimers and tetramers. There is no involvement of the C2 domain of the α3 chain in dimer and tetramer formation. In phylogenetic analyses the C2 domains of the α4, α5 and α6 chains cluster together with that of the α3 chain in a single branch suggesting that they may have similar functions (Figure 2.5).

Expression analysis in SaOS2 cells of human collagen VI α3 cDNAs that do not encode N-terminal domains demonstrated that the N6-N10 domains are neither crucial

for the assembly of collagen VI monomers, dimers or tetramers nor for their extracellular assembly. However, it was shown that the N5 domain of the collagen VI α3 chain is critical for the extracellular assembly into microfibrils (Fitzgerald *et al.*, 2001; Lamande *et al.*, 2006). The exact region or sequence signatures within the N5 domain that are responsible for the microfibril formation are not known. Therefore, based on the sequence information alone, it is not possible to identify which domains in the new chains play a similar role to the N5 domain of the α3 chain.

The VWA domains N9, N7 and N3 of the α3 chain showed binding to heparin that was abolished already by moderate salt concentrations (Specks *et al.*, 1992). However, it was not possible to identify the corresponding domains in the new chains on the basis of sequence information. Also, a stretch of eight N-terminal globular domains (N9-N2) showed strong binding to hyaluronan and required denaturing agents for full dissociation suggesting that the interaction is not simply ionic but conformation dependent (Specks *et al.*, 1992). Although the specific region of interaction is not known, as the new chains possess seven VWA domains at their N-termini, hyaluronan could potentially be an interaction partner. The unique domain of the new chains (C3) is only partially conserved as compared to the α3 chain. However, a conserved cysteine residue that is present in the α3 chain is seen in all the new chains. Although the role of this cysteine residue is not yet clear, based on the positional conservation it could be involved in forming stable disulfide bonds during the assembly. Interestingly, 4-6 residues removed from the cysteine residue, a RGD motif is present in both the α4 and α5 chains that is missing in the α3 chain. Although collagen VI is known to bind to cells, this affinity appears to be restricted to the collagenous domains (Aumailley *et al.*, 1991; Aumailley *et al.*, 1989; Pfaff *et al.*, 1993; Klein *et al.*, 1995). This could mean that the new α4 and α5 chains differ in their cell attachment activity compared to the collagen VI α3 chain. The C-terminal domain C5 resembling a Kunitz domain was found to play a critical role in microfibril assembly (Lamande *et al.*, 2006). Only the α4 chain contains a clearly homologous C5 domain. Interestingly, the Kunitz domain is truncated in mouse and rat, but the N-terminal part, which contains the trypsin interaction site in the original bovine pancreatic trypsin inhibitor (Perona *et al.*, 1993), is still present in the truncated

molecules and could serve as an interaction module. Collagen XXVIII also contains a Kunitz domain and shares similarity with collagen VI in carrying N- and C-terminal VWA domains (Veit *et al.*, 2006). However, the collagenous domain of collagen XXVIII is long and remains unique due to the high number of imperfections as compared to collagen VI. The collagen VI α5 and α6 chains lack a Kunitz domain, which may indicate differences in the assembly of α5- and α6-chain-containing microfibrils. Indeed, it will be interesting to study the process by which the collagen VI fibrils of different chain composition assemble. For instance, do the fibrils contain α1, α2 and only one of the four related α3, α4, α5 and α6 chains or are also mixed assemblies possible? The latter alternative would lead to a very high number of possible permutations.

The participation of the new chains in collagen VI assembly was studied in the *Col6a1* knockout mouse. It was earlier shown that the absence of the α1 chain blocks secretion of the α2 and α3 chains (Bonaldo *et al.*, 1998), indicating that the α1 chain is essential for the assembly of collagen VI molecules. Immunohistochemical and western blot analyses showed that also the new chains are completely absent in *Col6a1* deficient mice (Figure 2.27 and 2.28). This means that the α1 chain is a prerequisite also for their secretion and rules out the possibility that new chains can form homotrimers. Moreover, the fact that they can replace the α3 chain explains why the new collagen VI chains are expressed in the regions where the α3 chain is absent. These observations strongly indicate that the new chains assemble in a similar manner as proposed for the α1, α2 and α3 chain-containing collagen VI. In contrast to the results presented here, Fitzgerald and colleagues found that only the mouse α4 chain, but not the α5 or α6 chains assemble with the human α1 and α2 chains when transfected into SaOS2 cells (Fitzgerald *et al.*, 2008). Interestingly, although a functional collagen VI α4 chain is lacking in humans, the human α1 and α2 chains were able to assemble with the mouse collagen VI α4 chain. In contrast, although both the α5 (N5-C1) and α6 (N7-C2) chain constructs used in their study contained the C1 domain, which is critical for assembly, they did not form trimers. As both constructs lacked domains, this could point a possible role of these in the assembly. From automated electron tomography it was calculated that about ten VWA domains are present in the N-terminal globular domain of the collagen VI heterotrimer

purified from bovine nuchal ligament or aorta, while six domains were found for the C-terminal part (Baldock *et al.*, 2003). Alternative splicing of the α3 chain mRNA (Dziadek *et al.*, 2002) could explain that less than the 12 expected VWA domains are present when the full-length α3 chain is assembled. Nevertheless, nine N-terminal VWA domains and six or seven C-terminal domains would be present in assembly forms containing one of the new collagen chains, which is close to the calculated numbers. While assembly of other heterotrimeric collagen types mostly involves recognition between two or three different α chains, assembly of collagen VI molecules requires selection among six different α chains similar to in the assembly of collagen IV. Theoretically these six chains could form 106 different combinations including six homotrimers and 100 heterotrimers. However, in case of collagen IV only three heterotrimers (α1α1α2, α3α4α5 and α5α5α6) have been identified, pointing to remarkably specific chain interactions (Borza *et al.*, 2001; Boutaud *et al.*, 2000; Hudson *et al.*, 1994). In contrast to in collagen IV, only one heterotrimer (α1α2α3) and no homotrimers have so far been unambiguously identified in collagen VI.

3.1.3 Differential and restricted expression of the new chains

In order to study the expression and localization of new chains in mouse and human tissues, specific antibodies that detect each of the new chains had to be generated. Although the recombinant expression of one VWA domain in bacteria often leads to a correctly folded protein suited for the generation of specific antisera (Dumas *et al.*, 2004), we used an eukaryotic expression system to achieve not only correctly folded but also appropriately post-translationally modified proteins. In addition, we mostly made antigens containing a stretch of four or five VWA domains to enhance specificity of the generated antibodies. The domains were carefully selected on the basis of their sequence identity with other collagen VI chains. VWA domains that possess high sequence identity were excluded to avoid cross-reactivity. For instance, the N7 domain of the collagen VI α5 and α6 chains share a sequence identity of about 92.1% at the protein level, which could lead to the production of cross-reacting antibodies. We initially produced His-tagged collagen VI fragments, but later employed the Strep-tag that binds more tightly to

affinity matrices. The recombinant antigens were used to immunize both rabbits and guinea pigs to allow co-staining. Indeed, the affinity purified antisera against the mouse collagen VI α3, α4, α5 and α6 chains showed no cross-reactivity. Despite of the sequence similarity between mouse and human collagen VI chains, antisera raised against the mouse α5 chain did not cross-react with the human α5 chain (not shown). In contrast, the mouse α6 antiserum reacted with the human α6 chain. The reason could be that the sequence identity between mouse and human is lower in the α5 chains than in the α6 chains. In particular, the domains used for the generation of α5 chain antisera (N6-N2) are only 73% identical in contrast to 86% for the corresponding α6 domains (N6-N2). For these reasons we needed to generate also a separate set of antisera against the human chains. We now used the novel double Strep-tag to affinity purify the human collagen VI fragments. With this system the tagged proteins can be harvested directly from serum-containing medium with high yields. However, when used for immunization of rabbits and guinea pigs, the antisera against human collagen VI chains showed a cross-reactivity due to significant titer for the double Strep-tag. Nevertheless, they could be used for immunohistochemistry and western blots of human tissue sections and extracts as these do not contain any Strep-tag antigens.

All three new chains could be detected in collagen VI preparations from newborn mouse carcasses. A Coomassie stained gel showed distinct α1 and α2 bands and a very heterogenous group of bands above 220 kDa (Figure 2.13A). Most likely, this group of bands is comprised by splice variants of the α3 chain as well as the new collagen VI chains. All the new chains were identified in this mixed preparation by immunoblots using affinity-purified antibodies. This demonstrates that all three new chains are expressed and translated *in vivo*. The presence of bands above the expected size for the full length α4 chain under reducing conditions indicates the presence of non-reducible crosslinks involving the α4 chain. In addition, antibodies against the α4 and α5 chains detected lower bands indicating either alternatively spliced isoforms or proteolytic processing.

The immunohistochemical analysis of mouse tissues revealed that the new chains have a very specific and narrow tissue distribution, which is in contrast to the broadly

distributed collagen VI α3 chain. Immunostaining with the affinity purified antibodies on three different muscle types from both newborn and adult mice revealed that the expression of the new chains does not change significantly over age. However, the new chains show variation in their expression pattern. The collagen VI α4 chain is predominantly associated with basement membrane structures in the smooth muscle of intestine, but absent in both skeletal and cardiac muscle of adult mice. In contrast, the collagen VI α6 chain is abundantly present in both perimysial and epimysial basement membrane structures of skeletal and cardiac muscle, but does not show expression in smooth muscle. Although the α5 chain is present in all three muscle types, the expression in skeletal and cardiac muscle is relatively weak compared to that of the α6 chain in adult mice. In newborn skeletal muscle, the collagen VI α6 chain is restricted mainly to the endomysium. Interestingly, the new chains are only partially co-localized with the collagen VI α3 chain. In the specific regions where the new chains are exclusively present they may replace the α3 chain and assemble together with the collagen VI α1 and α2 chains. Similar to in muscle, there is a partial co-localization of the α4 and α5 chains with the collagen VI α3 chain in kidney, eye and oesophagus (not shown). The collagen VI α4 and α5 chains are present in adult mouse testis and this is the only tissue that exhibits an absolute co-localization with the α3 chain. The collagen VI α5 chain is specifically associated with basement membranes surrounding the nerves and blood vessels in the dermis. However, the α4 and α6 chains do not show specific staining at those sites. This differential and restricted distribution of the new collagen VI chains indicates that they may have tissue specific functions allowing a modulation of collagen VI properties. The new chains appear to be less abundant than the α3 chain. This could partially explain why the new chains were not identified earlier when analysis was mainly performed at the protein level. In contrast to the α3 chain, the new chains are not expressed in articular cartilage and fibrocartilage suggesting that they do not play a role in the pathogenesis of cartilage associated disorders such as osteoarthritis and chondrodysplasias. In contrast to our results from mouse, Fitzgerald and colleagues reported that the α6 chain is present in human articular cartilage and localized to the territorial matrix (Fitzgerald *et al.*, 2008). This could point to the differences in the tissue

distribution between man and mouse. Taken together, the results from immunohistochemistry show that the new chains are closely associated with basement membrane structures in several tissues pointing to a potential role in anchoring epithelia, nerves and blood vessels to the extracellular matrix.

3.1.4 Significance of the new chains in human disease

There is clear evidence for a role of collagen VI in the etiology of Bethlem myopathy (BM) and Ullrich congenital muscular dystrophy (UCMD). A variety of disease causing mutations have been identified in the collagen VI α1, α2 and α3 genes (Lampe & Bushby, 2005). Interestingly, patients have also been described which have phenotypes typical of BM or UCMD, but where mutations in the collagen VI α1, α2 and α3 chains could not be detected (Baker *et al.*, 2005; Lucioli *et al.*, 2005; Lampe *et al.*, 2005). In man, the new collagen VI α5 and α6 chains show specific expression in myotendinous junctions as well as in perimysium and endomysium of skeletal muscle. However, similar to in mouse, the collagen VI α6 chain is more abundantly present in human skeletal muscle than the collagen VI α5 chain. Indeed, a recent analysis of skeletal muscle biopsies of two patients with an UCMD phenotype who do not have mutations in the *COL6A1, COL6A2* and *COL6A3* genes yielded promising results. In both patients, there was a dramatic reduction in expression of the collagen VI α6 chain in the perimysium, which could indicate that also mutations in the new collagen VI chains may cause muscular dystrophy. Genome sequencing to identify these mutations is under progress and it is reasonable to test a larger cohort of patients suffering from BM and UCMD of unclear etiology for mutations in the new genes.

Besides the muscle phenotype, BM and UCMD patients develop skin abnormalities such as keratosis pilaris or keloid scarring. As the new chains are expressed in skin, their potential contribution to the skin phenotypes was studied in selected patients carrying mutations in the α1, α2 and α3 chains. As a byproduct, information on the assembly of collagen VI was obtained. The lack of an obvious defect in secretion or assembly of the α5 and α6 chains in patients with mutations in the α3 chain shows that the deposition of these is not dependent on the presence of α3 and that the new chains

may instead substitute for α3 by the formation of α1α2α5 or α1α2α6 heterotrimers. Interestingly, in the UCMD-A3 and BM-A3 patients the α5 and α6 chain staining was more widely distributed than in normal skin. This indicates that substitution of the α3 chain by the α5 and/or α6 chains may alter the assembly or distribution of collagen VI containing microfibrils. The strongly reduced expression of the α5 and α6 chains in the UCMD-A1 patient carrying a dominant mutation in the *COL6A1* gene matches the complete lack of the new collagen VI chains in *Col6a1* knockout mice. However, as seen in two UCMD patients that carry different compound heterozygous mutations in the *COL6A2* gene, the impact on the expression of the collagen VI α5 and α6 chains is highly dependent on the nature of the actual mutations. Patient UCMD-A2a carries a nonsense mutation in *COL6A2* exon 12, coding for a part of the triple helical domain, and a mutation in *COL6A2* intron 8 leading to a mixture of normally spliced transcripts and transcripts with retention of the entire intron 8, and containing an in-frame premature stop codon. Interestingly, the expression of the α3 and α5 chains is clearly reduced whereas the α6 chain remained unaffected. Due to the strong reduction of the α2 chain in this patient, the α2 chain concentration may become rate limiting in the triple helix formation. Possibly, the α6 chain is preferred to the α3 and α5 chains in this situation. Similarly, in the UCMD-A2b patient, staining for the α3 and α5 chain is reduced and the α6 chain is unaffected. This patient has two different splice site mutations in the *COL6A2* gene, leading to frameshifts either within the triple helical domain or just downstream of the triple helical domain but including cysteine residues that are important for the triple helix formation. It is therefore uncertain if any proper triple helix can be formed. Interestingly, as in the UCMD-A1 patient, roundish deposits are seen, usually close to the basement membrane of the dermal-epidermal junction. These could represent aggregates of a smaller proportion of the collagen VI α5 chain, which are secreted but not properly assembled into microfibrils. However, it is not clear why such structures were not seen for the α6 chain and why they were also not detected in the UCMD-A2a patient. Most likely, the mutations in the *COL6A1* and *COL6A2* genes impair the assembly, so that either collagen VI molecules are not formed at all or the formation of collagen VI microfibrils in the extracellular space fails. On the other hand, the α6 chain is expressed

to a lesser extent at the dermal-epidermal junction, where ring-formed structures are detected.

In a recent linkage study, the human *COL6A5* gene was associated with the very common skin disease atopic dermatitis and designated *COL29A1* (Söderhall *et al.*, 2007). The published protein sequence is, except for a single amino acid exchange, identical to the third potential splice variant presented here. A variety of nonsynonymous coding SNPs were described, but none could explain the association of *COL6A5* with atopic dermatitis on its own. It was therefore proposed that several variants or combinations associated with the most common haplotype of *COL6A5* are involved in the etiology of the disease. In addition, a strongly maternal transmission pattern was found, which could be due either to imprinting or to maternal effects through an interaction of the child's genotype with the maternal environment during prenatal life. Another susceptibility locus for atopic dermatitis was linked to 3p24-22 (Bradley *et al.*, 2002; Bischof *et al.*, 2006), which is exactly the breakpoint area where the 5' part of the *COL6A4* pseudogene is located. It could be that the mechanism leading to atopic dermatitis is more complex and that the expression of the non-processed α4 chain pseudogenes by a yet unknown mechanism influences α5 chain expression. The maternal transmission pattern could point to such a mechanism. A number of pseudogenes have been described where gene conversion between a functional copy of a gene and a neighboring pseudogene causes disease (Bischof *et al.*, 2006). However, in the present case the mechanism is likely to be more complex.

The newly identified locus for atopic dermatitis in *COL6A5* could correlate to another susceptibility locus on chromosome 21p21 found in a Swedish patient cohort, which may contain a susceptibility gene modulating the severity of atopic dermatitis especially in combination with asthma (Bu *et al.*, 2006). Both 21p21 and 3p24 have also been described as asthma susceptibility loci (Ober *et al.*, 1998). Interestingly, *COL6A1* and *COL6A2* are located on 21p21, which could point to a more general role of collagen VI in the development of atopic dermatitis or asthma.

The tissue distribution of the collagen VI α5 chain that is presented here is in clear contrast to Söderhäll et al., where the collagen VI α5 chain (*COL29A1*) was found

to be located in the epidermis (Söderhäll *et al.*, 2007). The reason for this discrepancy is not clear. However, there is no other literature evidence for the presence of collagen VI in the epidermis. Generally, only a few collagens are located in the epidermis, *e.g.* collagen XXIII (Koch *et al.*, 2006). As the lack of the new collagen VI chains in the *Col6a1* knockout mouse clearly indicates that they assemble together with the α1 chain and, probably, the α2 chain to form heterotrimers, an exclusive presence of the collagen VI α5 chain in the epidermis would be surprising. It should be noted that Söderhäll *et al.* detected the collagen VI α5 chain in epidermis with an antiserum that was raised against a short peptide sequence. It is likely that the linear epitope of the peptide does not yield an antibody that detects conformational epitopes, as it is needed for successful immunolocalization on tissue sections, and the epidermal staining by the peptide antibody could therefore be an artifact. In contrast, the antiserum used in this study was raised against a large portion of the N-terminal, non-triple helical domain of the collagen VI α5 chain (N2–N6), expressed in a eukaryotic system to ensure correct folding and purified under native conditions. In addition, the antiserum was affinity-purified against the protein that was used for immunization. The specific staining in a narrow zone just below the basement membrane, which is reminiscent of that of matrilin-2 (Piecha *et al.*, 2002), AMACO (Gebauer *et al.*, 2009) or tenascin C (Latijnhouwers *et al.*, 2000), points to a specialized function of the collagen VI α5 chain in this zone of the papillary dermis, which is important for the resistance to tensile stress. Nevertheless, the α5 chain may play a role in the maintenance of the barrier function of the skin, as indicated by the observation of polymorphisms in the *COL6A5* gene that were found to be linked to atopic dermatitis (Söderhäll *et al.*, 2007) or atopy (Castro-Giner *et al.*, 2009). However, the mechanism by which the α5 chain contributes to the barrier function would be different from the earlier proposed loss of epidermal integrity in the outer epidermis of atopic dermatitis patients (Söderhäll *et al.*, 2007). Perhaps, the presence of collagen VI microfibrils containing the α5 chain close to the basement membrane and around the vessel walls in the papillary dermis is important to prevent the diffusion of antigens or the migration of inflammatory cells. The collagen VI α6 chain, which is mainly expressed around vessels of the papillary and reticular dermis, could also play a role in the proposed barrier function. However, due to the differences in the localization of the two chains,

they most likely have different functions in skin. Interestingly, the collagen VI α5 and α6 chains in adult mouse skin exhibit a different expression pattern. Unlike to in humans, the papillary dermis of mouse skin is not stained by the collagen VI α5 chain specific antibody whereas the staining surrounding the blood vessels is seen both in man and mouse. In contrast to in humans, the collagen VI α6 chain is completely absent in adult mouse skin. This suggests that the new chains may have different roles in skin of man and mouse.

In addition to the linkage of *COL6A5* to atopic dermatitis, recently the chromosomal region 3p24.3 in humans that resembles the 5' part of the split *COL6A4* gene was identified as a susceptiblity locus for knee osteoarthritis in Japanese and Chinese patients (Miyamoto et al 2008). The authors proposed that a gene designated as *DVWA* (Dual von Willebrand factor A domain) located in the associated region codes for a novel protein containing two VWA domains without a signal peptide sequence. Furthermore, they proposed that a recombinantly overexpressed DVWA/COL6A4 binds to tubulin suggesting a mechanism for disease modulation (Miyamoto *et al.*, 2008). The DVWA sequence unequivocally resembles sequences of the collagen VI α4 pseudogene. Although, they could show the presence of *DVWA/COL6A4* transcripts in the human cartilage, it is quite unlikely that the proposed *DVWA/COL6A4* cDNA is translated into a functional protein since truncated VWA domains have not been described to form a stable folded protein. Furthermore, point mutations that affect conserved residues within the β-sheets of the single VWA domain of matrilin-3 cause the protein to misfold and prevent its secretion from the rER, both *in vitro* and *in vivo* (Otten *et al.*, 2005; Cotterill *et al.*, 2005). The truncated VWA1 of DVWA lacks the region where the first three out of five conserved residues in the MIDAS motif are located, only VWA2 contains the full consensus. Nevertheless, it cannot be excluded that if the protein is synthesized in the cytoplasm, a novel stable fold is created. However, the presence of such an intracellular protein would need to be demonstrated *in vivo*, either by specific antibodies directed against DVWA/COL6A4 or by more direct methods like mass spectrometry. The presence of an aberrant mRNA alone does not support the conclusion that the mRNA is translated into a folded protein, despite the fact that it could be recombinantly expressed in eukaryotic and prokaryotic cells. It should be noted that the cDNA constructs used for

recombinant expression and β-tubulin interaction studies lacked the large 5'UTR of the *DVWA/COL6A4* mRNA (Miyamoto *et al.*, 2008). Moreover, it was not possible to amplify the 5' part of *COL6A4/DVWA* cDNA from either human cartilage specimens of patients of European origin or from their cultured chondrocytes. In addition, a large replication study and meta-analyses of *DVWA/COL6A4* locus in a European population found no association to hip osteoarthritis and a significant heterogeneity with regard to knee osteoarthritis (Meulenbelt *et al.*, 2009). Another independent meta analyses study also found significant heterogeneity in the *DVWA/COL6A4* locus for knee osteoarthritis between UK and Asian populations (Valdes *et al.*, 2009). Therefore, the effect of DVWA amino acid changes on tubulin binding is unlikely to influence the risk of OA in caucasians (Valdes *et al.*, 2009). The reason for the differences between the European and Asian populations could be due to the presence of a dominant allelic frequency in the tested sample size that can be minimized by increasing the sample size. In addition, it is also possible that the differences are due to environmental influences or genetic background differences between the two tested groups. The presence of a disease locus in the interrupted *COL6A4* gene at the genome breakpoint in chromosome 3p24.3 could be a consequence of the evolutionary rearrangement and is reminiscent to the linkage of atopic dermatitis close to the other genomic breakpoint of chromosome 3.

However, it could well be that other yet not well characterized diseases are linked to the new collagen VI genes. Immunohistochemical analyses of glomeruli from the patients of diabetic glomerulosclerosis (GS) suggest that the collagen VI deposition is altered at various stages of disease (Nerlich *et al.*, 1994). The uninterrupted staining along the glomerular basement membrane in normal glomeruli was discontinous in diffusely sclerotic glomeruli. In contrast, the collagen VI deposition was increased in nodular GS and evenly distributed throughout the nodular lesion. A comparison of immunohistochemical analysis with clinical parameters of diabetic nephropathy suggested that increased collagen VI deposition may be an indicator of the irreversible remodeling of glomerular matrix to nodular GS which is associated with functional insufficiency (Nerlich *et al.*, 1994). Interestingly, the expression of collagen VI α5 in kidney is quite restricted to the glomeruli structures. In contrast, the collagen VI α6 chain is absent in kidney glomeruli. Based on the distinct and specific localization of the α5

chain in the glomeruli, it is tempting to speculate that the collagen VI α5 chain could play a role in diabetic nephropathy.

Moreover, the strong expression of the new chains in ovary and testis raises the question if mutations could lead to infertility. It would be interesting to monitor the expression of new chains in the genital organs in UCMD and BM patients, even though infertility in such patients has not been described in the literature.

The discovery of three new collagen VI chains will have a great impact in understanding the structure, assembly and function of collagen VI. Further biochemical characterization of the new collagen VI chains and their roles in the development of inherited diseases will contribute to the knowledge of extracellular matrix function and pathology.

3.2 Collagen XXVIII

Recently, another novel VWA domain containing collagen that is capable of forming homotrimeric molecules was identified and designated as collagen XXVIII (Veit *et al.*, 2006). In mouse, the *Col28a1* gene is present on chromosome 6A1. The protein was found to be specifically associated with basement membranes in the peripheral nervous system. However, the functional role of collagen XXVIII is still unknown. Therefore, the inactivation of this gene in mouse embryonic stem cells via homologous recombination was initiated. In order to speed up the process of generating a targeting vector, BAC recombineering was employed (Copeland *et al.*, 2001). As the *Col28a1* gene has only a single start codon in the first exon, the second exon was selected for targeting. In order to eliminate the random insertions of the targeting vector, a negative selection with gancyclovir was performed while selecting the transfected ES cells. The integration of the 3' arm was confirmed by Southern blots. However, Southern blots with two different probes for the 5' integration did not work well. Since the 5' end of the gene has AT rich repeat sequences, there was less possibility to design further probes in this region. Therefore, the 5' integration was tested by PCR with one primer specific for the targeting vector and the other for the wild type allelic region. The positive ES cells clones

when injected into the blastocyst of pseudopregnant mouse gave rise to chimeric mice of variable percentage. However, when the chimeras were further backcrossed with C57BL/6 mice to generate a heterozygote allele for the targeted *Col28a1* gene, there was no germline transmission observed in the F1 generation of chimeras for eight litters. Germline transmission depends on the quality of ES cells used, but the V6.5 cells are known to function well in this regard. However, the cell culture conditions used may have introduced problems. For instance, long intervals between passaging the cells to new feeder layers may result in undesirable changes in ES cells (Doetschman *et al.*, 1985). ES cells that lose their ability to differentiate *in vitro* when cultured without feeder layers may become incapable of differentiating *in vivo* (Doetschman *et al.*, 1985). Therefore the targeting vector will be used to transfect new ES cell clones for further attempts to generate a knockout mouse for *Col28a1* gene.

4. MATERIALS AND METHODS

4.1 MATERIALS

The materials/chemicals used in this study were purchased from the following companies unless otherwise mentioned.

Biozyme (Oldendorf)	FLUKA (Milwaukee, USA)
GERBU (Gaiberg)	GibcoBRL (Paisley, Scotland)
NEB (New England Biolabs, Beverly, USA)	Qiagen (Hilden)
SERVA (Heidelberg)	Sigma (Deisenhofen)

4.1.1 General solutions and buffers

6x DNA loading dye	0.25% (w/v) Bromophenol blue and/or Xylene cyanol 70% (v/v) Glycerol
10x RNA loading dye	50% Glycerol 1 mM EDTA (pH 8.0) 0.4% Bromophenol blue
1x TAE	40 mM Tris 20 mM Acetic acid 1 mM EDTA
1x SDS loading buffer	63 mM Tris HCl 10% (v/v) Glycerol 2% (w/v) SDS 0.0025% (w/v) Bromophenol blue pH 6.8
SDS-PAGE running buffer	200 mM Glycine 0.1% (w/v) SDS 0.04% (w/v) Bromophenol blue 62.5 mM Tris/HCl pH 6.8
20x SSC	3 M NaCl 0.3 M Sodium citrate pH 7.0

1x PBS	140 mM NaCl 10 mM KCl 8 mM Na_2HPO_4 2 mM KH_2PO_4 pH 7.4
1x TBS	150 mM NaCl 50 mM Tris/HCl pH 7.4

4.1.2 Bacterial media and strains

LB-Medium

1.0% (w/v)	Bacto-tryptone
0.5% (w/v)	Bacto-yeast extract
1.0% (w/v)	NaCl

LB-Agar

1.5% (w/v) Agar in LB-medium

Bacterial strains

DH5α (Invitrogen)
Genotype: F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-

DH10B (Invitrogen)
Genotype: F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G

4.1.3 Nucleic acid and protein standards

Protein standard marker:	Broad range marker (NEB) Prestained broad range marker (NEB)
DNA standard marker:	1 kb ladder (Life Technologies)
RNA standard marker:	0.24-9.5 kb (GIBCO-BRL)

4.1.4 Oligonucleotide primers

Table 4.1 Primers used for mouse collagen VI chains

Name	Oligonucleotide sequence	Designation / Purpose
α3m1(f)	5'-AAA**GCTAGC**ACAACAGCATGGAGATGTCAAAA-3'	Recombinant protein
α3m2(r)	5'-TAT**CTCGAG**CTGTGAGGTTAGAGTGGTGATG-3'	Recombinant protein
α4m1(f)	5-TGAGGGATGAGGACAAGGAG-3'	AM231151
α4m2(r)	5-AAGGTAGGAAGCAGGTGTAGAC-3'	AM231151
α4m3(f)	5-TTCCTTCCAGAGAAGGGCAG-3'	AM231152

α4m4(r)	5-ACATTTCCCACCCCGATAG-3'	AM231152
α4m5(f)	5-GCAGCATCAAACCCAACGAC-3'	AM231153
α4m6(r)	5-GCCTACACCATGTCTCACACTC-3'	AM231153, RT-PCR
α4m7(f)	5-TCGCATTATTCGTGCTCGCC-3'	RT-PCR
α4m8(f)	5-CAGAGGATTGTCTGCAGGG-3'	Northern probe
α4m9(r)	5-GCGGAAAGGCTGGGTAAATG-3'	Northern probe
α4m10(f)	5-CAAT**GCTAGC**GCGTGCTCAGAAGCATCCCCGG-3'	Recombinant protein
α4m11(r)	5'-CAAT**AGATCT**CAGGAGCAGATGTCCCGAATG-3'	Recombinant protein
α5m1(f)	5'-CAGGAACCCAGCAGCGAG-3'	AM748256
α5m2(r)	5'-CCCACAGCTCTGGCAAATTTC-3'	AM748256, Northern probe
α5m3(f)	5'-TTCCTTTGTGATGGCTCTGAC-3'	AM748257
α5m4(r)	5'-GTGTGTCTCCCCTCTCTAACTC-3'	AM748257, RT-PCR
α5m5(f)	5'-GCTCAGTGATTGACAGCTTCC-3'	AM748258, RT-PCR
α5m6(r)	5'-TCTTCTTTTGGCAGGCACTCAC-3'	AM748258
α5m7(f)	5'-ATGAACCTCACCATCCACTTG-3'	Northern probe
α5m8(f)	5'-AAT**GCTAGC**TGCAAGAGGATCGAACTCTTA-3'	Recombinant protein
α5m9(r)	5'-AAT**GGATCC**TCACTCTTGTGAATTATTACACAGAG-3'	Recombinant protein
α5m10(f)	5'-CAAT**GCTAGC**TCCCTTTCCCACCTCTTGC-3'	Recombinant protein
α5m11(r)	5'-CAAT**CTCGAG**ATTGGCTTTCCCACAGCTCTG-3'	Recombinant protein
α6m1(f)	5'-CGTGGAGAGCAGCATTCATC-3'	AM748259
α6m2(r)	5'-CCAACAGGAACACCACATCAG-3'	AM748259, RT-PCR
α6m3(f)	5'-CAATGCTAGCTGGCCCCGAGTACGCAG-3'	AM748260
α6m4(r)	5'-CAATCTCGAGATTGCTCTCACCACCTGAGG-3'	AM748260
α6m5(f)	5'-TGTCAGCCAGTGTCTGTAACTC-3'	AM748261
α6m6(r)	5'-AAAGAGGGCATACCCATGAC-3'	AM748261, Northern probe
α6m7(f)	5'-AAGAAAGAGTCCCTACGAGCC-3'	AM748262
α6m8(r)	5'-TGGTGAGCAGCCCAGTTAAG-3'	AM748262
α6m9(f)	5'-TTCCTCAGCCTACAACCTG-3'	RT-PCR
α6m10(f)	5'-ATCCGCTTCTCAGATGCCTAC-3'	Northern probe
α6m11(f)	5'-CAAT**GCTAGC**TGGCCCCGAGTACGCAG-3'	Recombinant protein
α6m12(r)	5'-CAAT**CTCGAG**ATTGCTCTCACCACCTGAGG-3'	Recombinant protein

f, forward; r, reverse; m, mouse; restriction sites are indicated in bold

Table 4.2 Primers used for human collagen VI chains

Name	Oligonucleotide sequence	Designation / Purpose
α3h1(f)	5'-AAT**ACTAGT**AGCAAGCGAGACATTCTG-3'	Recombinant protein
α3h2(r)	5'-AAT**CTCGAG**TCACAAAACTTGCTCGCTCAG-3'	Recombinant protein
α5h1(f)	5'-TGAAACAAAGCCAGCAAACAG-3'	AM906078
α5h2(r)	5'-CAAAGCATCAAAGTTCTGCC-3'	AM906078
α5h3(f)	5'-TATGGCAGCAGAAGAGCAC-3'	AM906079
α5h4(r)	5'-TCCCTGCCATACCCTCAAGTTC-3'	AM906079
α5h5(f)	5'-ACACTGGAGGGAACACCTACAC-3'	AM906080
α5h6(r)	5'-TGTGTCCTCTGCTGCCTTTC-3'	AM906080
α5h7(f)	5'-GAATCATCCGTGAAATCTGCC-3'	AM906081

α5h8(r)	5'-AAACACAGCAACTCTCCTCAC-3'	AM906081
α5h9(f)	5'-AAAACACGGGACATCATCAC-3'	AM906082
α5h10(r)	5'-AAGCATCTATGAGGAAAGCCAC-3'	AM906082
α5h11(f)	5'-CACATCCAGACTTCCTTCCAAC-3'	AM906083
α5h12(r)	5'-CCCTGTCATTTACCACAGTACC-3'	AM906083
α5h13(f)	5'-CCCAAAGAGTAGGAAGTGATGAG-3'	AM906084
α5h14(r)	5'-GAAATGTTAGGGTCACCCCAG-3'	AM906084
α5h15(f)	5'-AAT**GCTAGC**CACTTCCCCATATCCTG-3'	Recombinant protein
α5h16(r)	5'-AAT**GGATCC**TCAATTGGTTTTCCCACAGCTC-3'	Recombinant protein
α6h1(f)	5'-ATTTGAAGTTGAAGATTTTTCAGGTC	AM774225
α6h2(r)	5'-TCATCAGCATACTTCAGAGCC	AM774225
α6h3(f)	5'-TGTTTGGCTCCAATGTCACC	AM774226
α6h4(r)	5'-TGCCATCTCCTCCAATGCAC	AM774226
α6h5(f)	5'-GCCTGAATGCCCTCATAAC	AM774227
α6h6(r)	5'-TGACCATAACAGGAAAGACAGAC	AM774227
α6h7(f)	5'-AAT**GCTAGC**ATGAGGGTTGGCCTTGTGGCC-3'	Recombinant protein
α6h8(r)	5'-AAT**GGATCC**TCAGTTGCTTTCACCCGCTGTGG-3'	Recombinant protein

f, forward; r, reverse; h, human, restriction sites are indicated in bold

Table 4.3 Primers used for mouse collagen XXVIII gene knockout

Name	Oligonucleotide sequence	Purpose
α1m1(f)	5'-GAA**GCGGCCGC**GGATCCCACTACAGATGGTTGTG-3'	5'homology arm
α1m2(r)	5'-AAA**GCTAGC**CAATTAAATTCAGAGACTTGACC-3'	5'homology arm
α1m3(f)	5'-ATAA**GGCCGGCC**TACTATTGAGTGACCCAACACT-3'	3'homology arm (1)
α1m4(r)	5'-AAA**ACGCGT**TGCAAGTCTGAATTTTAAAGTAT-3'	3'homology arm (1)
α1m5(f)	5'-AAA**ACGCGT**GTTGTTTAGCTTATTACATTAAG-3'	3'homology arm (2)
α1m6(r)	5'-TTT**GGCGCGCC**ATAGGAGACCTTGGAGTTTTAAC-3'	3'homology arm (2)
α1m7(f)	5'-AAGCTTCCTGCTGCCTGCTATTTG-3'	Southern probe (5')
α1m8(r)	5'-CTTCATCAGGTTCCACAGTGGAAC-3'	Southern probe (5')
α1m9(f)	5'-TAATAATAACAACTAGTGATAGTG-3'	Southern probe (3')
α1m10(r)	5'-ACAGGACCTCTTTCTCCAGTTTCTCC-3'	Southern probe (3')
α1m11(f)	5'-CCCTTGGCTGCTCTCGGACTCC-3'	Southern probe, Screening* (Neo)
α1m12(r)	5'-CCAGCCTCTGAGCCCAGAAAGC-3'	Southern probe, Screening* (Neo)
α1m13(f)	5'-CATCACATTACAATTCCTTAGAAACC-3'	Screening* (5')
α1m14(r)	5'-GATCAGCTAGATCTTCAGGATCTGC-3'	Screening* (5')

f, forward; r, reverse; m, mouse, restriction sites are indicated in bold, '*' represents the primers used for screening of tail biopsies

4.1.5 Restriction enzymes (NEB)

AscI	*BamHI*	*BglII*
BsrGI	*EcoRI*	*FseI*
HindIII	*MluI*	*NheI*
NotI	*SpeI*	*XhoI*

4.1.6 Antibodies

Table 4.3 Primary antibodies

Origin	Name	Antigen	WB / ELISA	Host	IF
Mouse (*sm*)	α3(VI)	N10-N4	1:2000	Rabbit	1:500
Mouse (*sm*)	α3(VI)	N10-N4	1:2000	Guinea pig	1:500
Mouse (*sm*)	α4(VI)	N6-N3	1:2000	Rabbit	1:500
Mouse (*sm*)	α4(VI)	N6-N3	1:2000	Guinea pig	1:500
Mouse (*sm*)	α5a(VI)	N3	1:2000	Rabbit	1:200
Mouse (*sm*)	α5b(VI)	N6-N2	1:1000	Rabbit	1:50
Mouse (*sm*)	α5b(VI)	N6-N2	1:1000	Guinea pig	1:50
Mouse (*sm*)	α6(VI)	N6-N2	1:2000	Rabbit	1:500
Mouse (*sm*)	α6(VI)	N6-N2	1:2000	Guinea pig	1:500
Human (*sm*)	α3(VI)	N4-N1	1:2000	Rabbit	1:500
Human (*sm*)	α3(VI)	N4-N1	1:2000	Guinea pig	1:500
Human (*sm*)	α5(VI)	N6-N2	1:2000	Rabbit	1:500
Human (*sm*)	α5(VI)	N6-N2	1:2000	Guinea pig	1:500
Human (*sm*)	α6(VI)	N6-N2	1:1000	Rabbit	1:200
Human (*sm*)	α6(VI)	N6-N2	1:1000	Guinea pig	1:200
Human (ca)*	Laminin α5	Laminin	-	Mouse	1:2000
Human (ca)*	Laminin α2	Merosin	-	Mouse	1:2000

sm, self made (affinity purified polyclonal); ca, commercially available (Chemicon); * indicates monoclonal antibody, WB, Western blot; ELISA, Enzyme-linked immunosorbent assay; IF, Immunofluorescence

Table 4.4 Secondary antibodies

Source	Conjugate	Provider	Applied	Dilution
Rabbit IgG	HRP	Dako	WB/ELISA	1:2000
Guinea pig IgG	HRP	Dako	WB/ELISA	1:2000
Rabbit IgG	Alexa 546	Molecular probes	IF	1:2000
Guinea pig IgG	Alexa 488	Molecular probes	IF	1:2000
Mouse IgG	FITC	Dako	IF	1:2000
Mouse IgG	TRITC	Dako	IF	1.2000
Anti-Digoxigenin	AP	Roche	NB	1:15000
			SB	1:20000

IgG, Immunoglobulin G; AP, Alkaline phosphatase; HRP, Horse radish peroxidase; FITC, Fluorescein isothiocyanate isomer 1; TRITC, Tetramethylorhodamine isothiocyanate isomer R; WB, Western blot; ELISA, Enzyme-linked immunosorbent assay; IF, Immunofluorescence; SB, Southern blot; NB, Northern blot

4.2 MOLECULAR BIOLOGY METHODS

4.2.1 Culture and storage of bacteria

A single colony from a LB-Agar plate was inoculated in 8 ml of sterile LB medium (with or without antibiotics) and incubated under 200-220 rpm shaking at 37 ^{o}C in an incubator (Heraeus, Type B5050 E, Heraeus Instruments) overnight. When required, the bacterial cells were preserved in 25% (v/v) glycerol and stored at -80 ^{o}C.

4.2.2 Preparation of competent cells

The overnight culture (1 ml) was inoculated in 100 ml of LB media and allowed to grow at 37 ^{o}C incubator until the A_{600} reached 0.5-0.6. The culture was cooled in an ice/water slurry for 20 min and centrifuged at 4 ^{o}C for 10 min at 3500 rpm. The cells were resuspended in 1/10 of original volume of TSS buffer and snap-frozen in liquid nitrogen in 100 µl aliquots. The aliquots were stored at -80 ^{o}C.

TSS- Buffer

85% (v/v)	LB-Medium
10% (w/v)	PEG 8000
5% (v/v)	DMSO
50mM	$MgCl_2$

4.2.3 Bacterial transformation

The plasmid/ligation reaction mixture (10-100 ng DNA) were added to one aliquot of competent *E. coli* cells and gently mixed. The cells were incubated on ice for 30 min and then heat shocked for 1 min at 42 ^{o}C without shaking. The tubes were immediately transferred to ice and incubated for 2 min. 50-500 µl of LB-Medium were added and the tubes were shaken at 37 ^{o}C for 45 min. The transformation mixture (50-100 µl) was spread on LB-Agar plates containing antibiotics at the appropriate concentration. The plates were incubated for 5-10 min at RT, then inverted and incubated overnight at 37 ^{o}C

4.2.4 Isolation of plasmid DNA

Isolated single colonies from the transformation were inoculated into 8 ml of LB media containing the appropriate antibiotic. After overnight culture at 37 ^{o}C, the cells were centrifuged at 5000 rpm at 4 ^{o}C for 10 min. The plasmid DNA from the bacterial cultures was purified using Qiagen Plasmid Mini Prep Kit.

4.2.5 Determining of DNA concentration

The plasmid DNA concentration was determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer. For reliable DNA quantification, A_{260} readings should lie between 0.1 and 1.0. An absorbance of 1 unit at 260 nm corresponds to 50 µg plasmid DNA per ml ($A_{260} = 1 \Rightarrow 50$ µg/ml). The measurements were made at neutral pH (Tris-HCl, pH 7.0). For a relatively pure DNA, the ratio of A_{260}/A_{280} should be 1.8-2.0.

For purified PCR products or DNA fragments extracted from agarose gels, the DNA concentration was estimated from agarose gels after ethidium bromide staining.

4.2.6 Agarose gel electophoresis

0.7 – 2.0% (w/v) agarose (Seakem) gels in TAE buffer were prepared by heating in a microwave oven for about 2 min. After cooling at RT for 5 min, ethidium bromide to a final concentration of 0.5 ug/ml was added and properly mixed. Then the gel was poured slowly into the tank without any bubbles. The combs were selected depending upon the volume of the sample and inserted into the gel. The samples were mixed with 1xDNA-loading buffer and loaded onto the gel. A standard DNA molecular marker was also loaded onto the gel. The gels were run at 4-8 Volt/cm with 1xTAE buffer and visualized under UV-light.

4.2.7 Gel-elution of DNA fragments

Elution of DNA fragments was performed with the Qiagen Gel-extraction kit according to the manufacturers instructions. The agarose gels were run at low voltage for efficient separation of DNA fragments without smearing. The fragments were excised quickly with a sharp scalpel under low-strength UV-light to limit DNA damage.

4.2.8 Polymerase chain reaction (PCR)

Polymerase chain reactions were performed either on a cDNAs from mouse tissues or on plasmid DNAs. The PCR reaction mixture was prepared using the standard procedure and the PCR buffer in the reaction was varied depending upon the enzyme used. After preparing the mixture, the PCR was performed with a standard cycler program. The annealing and the extension temperatures were varied depending upon

the melting temperature of primers and the length of the expected PCR product. The amplified products were visualized by running the samples on an agarose gel.

Standard reaction mixture

10-100 ng	DNA
1 µM	Forward primer (10 µM)
1 µM	Reverse primer (10 µM)
200 µM	dNTP Mix (2.5 mM)
1x	PCR buffer (10x AmpliTaq buffer (Perkin Almer) 10x Expand HF buffer, 10x Long Expand buffer 3 (Roche))
2.5 U	Polymerase (Amplitaq (Perkin Almer) Expand HF enzyme mix Long expand PCR system (Roche)

and made up to a desired final volume with MilliQ H_20

Cycler program

1.	Initial hold	94 °C	2-5 min	1x
2.	Denaturation	94 °C	30 sec	
3.	Annealing	48-65 °C	30 sec	25-40x (Steps 2-4)
4.	Extension	68-72 °C	60 sec	
5.	Final extension	72 °C	7 min	1x

4.2.9 Restriction digestion of DNA

The plasmid DNA or PCR products were digested using several restriction enzymes provided by New England BioLabs. After preparing the reaction mixture the tubes were briefly spun and incubated at 37 °C for 2 h to overnight. The digested products were purified using PCR-purification kit from Qiagen.

Reaction mixture

1.0 µg	DNA
2.0 µl	10x Restriction enzyme buffer
0.2 µl	100x BSA (NEB) (if necessary)
2-10U	Restriction enzyme

made up to final volume of 20 µl with H_2O

4.2.10 Ligation of DNA

The digested plasmid DNA and the PCR product (insert) with compatible ends were ligated using the Rapid DNA ligation kit from Qiagen. The vector to insert molar ratio was between 1:1 and 1:10. After preparing the reaction mixture the tubes were incubated for 10 min at RT and used for transformation.

Reaction mixture

2 µl	Plasmid DNA
1 µl	Insert
2 µl	5x Rapid ligation buffer
0.5 µl	T4-DNA Ligase (Fermentas)

made up to final volume of 10 µl with H_2O

4.2.11 DNA sequencing and analysis

The DNA sequencing was performed at the core facility of Center for Molecular Medicine, University of Cologne. The PCR product (50-100 ng) or plasmid DNA (200-400 ng) with specific sequencing primers (1 µM) was used. The sequencing result were analyzed using the 4Peaks software.
(http://mekentosj.com/science/4peaks/)

4.2.12 DNA precipitation

The plasmid DNA (targeting vector) was digested with AscI in 300 µl reaction volume at 37 ^{o}C overnight. The volume was raised to 500 µl with Tris-HCl (pH 8.0), 1 ml chloroform was added, the sample vigorously vortexed and centrifuged at 5000 rpm for 5 min. The upper aqueous phase was transferred to a separate tube, sodium acetate (pH 5.2) to a final concentration of 0.3 M and with 2 volumes of ethanol were added and the sample kept at -80 ^{o}C for 30 min. The solution was centrifuged at 13000 rpm for 15 min and washed twice with 70 % ethanol. The ethanol traces were removed, the pellet air dried for 5-10 min, resuspended in TE and quantified with a spectrophotometer.

4.2.13 Isolation of total RNA

Total RNA was isolated from both newborn and adult tissues using Trizol (Invitrogen). The tissue samples were homogenized in 1 ml of TRIZOL reagent per 50 to 100 mg of tissue. The sample volume did not exceed 10% of the volume of TRIZOL reagent used for the homogenization. 0.2 ml chloroform was added per 1 ml of TRIZOL reagent. The tubes were tightly capped and vortexed vigorously for 15 sec and incubated at RT for 2 to 3 min. The tubes were centrifuged at 12,000 x g for 15 min at 2 to 8 ^{o}C. The upper aqueous phase containing RNA was transferred carefully without disturbing the interphase into fresh tube. The RNA was precipitated by

mixing with isopropanol (1:2). The tubes were incubated at 15 to 30 ^{o}C for 10 min and centrifuged at 12,000 x g for 10 min at 2 to 4 ^{o}C. The supernatant was completely removed from the pellet. The RNA pellet was washed once with 1 ml of 75% ethanol. The tubes were mixed by vortexing and centrifuged at 7,500 x g for 5 min at 2 to 8 ^{o}C. The washing procedure was repeated once. The ethanol traces were removed and the RNA pellet air-dried for 5-10 min. The RNA pellet was not completely dried to maintain its solubility. The RNA was dissolved in DEPC-treated water and quantified by spectrophotometer.

4.2.14 Synthesis of cDNA

The cDNA was prepared in a 20 µl reaction volume from total RNA of newborn and adult mouse tissues. The following components were added in a nuclease free microcentrifuge tube.

1 µl	Oligo (dT)12-18 (500 µg/ml)
x µl	1 ng – 5 µg of total RNA
1 µl	dNTP (10 µM)

The final volume was made up to 12 µl with distilled water and the mixture was heated to 65 ^{o}C for 5 min and immediately chilled on ice. After brief centrifugation

4 µl	5x Strand buffer
2 µl	0.1 M DTT
1 µl	RNase OUT (40 U/µl)

were added and gently mixed and incubated at 42 ^{o}C for 2 min. 1 µl (200 U) of SuperScript II RT were added and the mixture was incubated at 42 ^{o}C for 50 min. The reaction was inactivated by heating at 70 ^{o}C for 15 min.

4.2.15. Northern blot

The vectors containing the probes were linearized with either HindIII or BamHI. The digested products were purified using the Qiagen gel extraction kit.

4.2.15.1 DIG labeling of RNA

The purified template DNA (1 µg) was added to a sterile, RNase-free reaction vial and sterile RNase-free DMPC treated water was added to make the total sample volume up to 13 µl. On ice the following reagents were added.

Reagent	Volume
10x dNTP labeling mixture	2µl
10x Transcription buffer	2µl
RNase inhibitor	1µl
RNA polymerase T3 or T7	2µl

The reagents were gently mixed and briefly centrifuged before incubating for 2 h at 37 ^{o}C. 2 µl of RNase-free DNaseI was added and incubated for 15 min at 37 ^{o}C to remove the template DNA. The reaction was stopped with 2 µl of 0.2 M EDTA (pH 8.0). The RNA was incubated with 4M LiCl and absolute ethanol at -20 ^{o}C for 30 min and precipitated for 15 min at 4 ^{o}C. The pellet was washed with 50 µl of 70% ethanol and dissolved in 100 µl of MilliQ H_2O before storing at -20 ^{o}C.

4.2.15.2 Estimation of labeled RNA

The labeling efficiency was estimated by running the samples on a 1% agarose gel. Before loading, 5 µl of RNA loading dye and 0.5 µl of 5xMOPS was added to 5 µl of RNA solution and the samples were heated for 10 min at 100 ^{o}C and immediately cooled on ice. A DNA ladder (1 kb) was used to estimate the RNA concentration.

4.2.15.3 Formaldehyde gels

Agarose (0.9-1.2%) were dissolved in water and the solution was cooled to 65 ^{o}C. Formaldehyde and 10xMOPS buffer were added to a final concentration of 0.8% and 1x respectively. Total RNA (5-10 µg) or mRNA (1-3 µg) was incubated at 65 ^{o}C for 20 min after mixing with 14.5 µl mix (see below). The solution was cooled on ice for 2 min and 2 µl of loading buffer and 0.5 µl of ethidium bromide solution were added. The entire sample volume and also 5 µl of RNA ladder were loaded on the formaldehyde gel. The gels were run at 5V/cm. After electrophoresis, the gels were incubated twice in 20x SSC for 15 min and transferred onto the Roche Nylon membrane with 20x SSC using a vaccum blotter for 2 h at 55 mbar. The blot was fixed under UV and processed for hybridization.

Mix

1 µl	10x MOPS
3.5 µl	Formaldehyde
10 µl	Formamide

4.2.15.4 Blotting

The membrane was pre-hybridized with RocheEasyHyb for 1 h at 68 ^{o}C. The DIG-labeled RNA (20-50 ng) was denatured for 10 min at 100 ^{o}C in 30 µl of H_2O and rapidly cooled on ice-water. The labeled probe was added to pre-heated hybridizing solution (BBR) and incubated overnight at 68 ^{o}C with gentle agitation. The membrane was washed twice with 2x SSC/0.1% SDS for 5 min at RT followed by two more washes with 0.1x SSC/0.1% SDS for 15 min at 68 ^{o}C. The membrane was incubated for 2 min in buffer I containing 0.3 % Tween-20 and blocked with 1x BBR in buffer I for 30 min. The membrane was incubated with anti-DIG-AP antibody in blocking solution for 30 min and washed twice for 15 min with buffer I containing 0.3% Tween-20. The membrane was incubated in detection buffer for 5 min and about 20 drops of CDP-star substrate was added for 5 min. The signals were visualized by exposing a film for 15-30 min.

RNA dilution buffer

50% MilliQ H_2O
30% 20xSSC
20% Formaldehyde

10xMOPS

0.2 M MOPS (N-Morpholino propane sulfonic acid)
80 mM Sodium acetate
10 mM EDTA (pH 8.0)

Buffer I

100 mM Tris (pH 7.5)
150 mM NaCl

Buffer II

1% Boehringer blocking reagent
50 µg ss-DNA in buffer I

Buffer III

100 mM Tris
100 mM NaCl
50mM $MgCl_2$
pH 9.5

Running buffer

1x MOPS
0.8% Formaldehyde

Detection buffer

100 mM Tris
0.1 M NaCl
pH 9.5

BBR

10% Blocking reagent in maleic acid buffer

4.2.16 Southern blot

The evaluation of 3' end and neo cassette insertion in the targeted ES cell clones was performed on digested ES cell DNA using Southern blot.

4.2.16.1 DIG-labeling of DNA

The PCR probes for the 5', 3' and neo selection cassette were labeled using the PCR-DIG probe synthesis kit (Roche).

Reaction mixture

5 µl	PCR buffer
5 µl	DIG-labeling mix
1 µl	Forward primer (10 µM)
1 µl	Reverse primer (10 µM)
1 µl	Template
0.75 µl	Enzyme mix

made up to final volume of 50 µl with H_2O.

After preparing the reaction mixture, a standard PCR program chosen according to the annealing temperature of the primers and the length of the product was used for labeling the probes. The labeled probes were analyzed on an agarose gel. The DIG-labeled probe should be larger than the unlabeled probe due to the tagging

4.2.16.2 Agarose gel electrophoresis and blotting

The targeted ES cell DNA was digested with *BsrGI* (3' probe) and *HindIII* (Neo probe) at 37 ^{o}C overnight and separated on a 1% agarose gel. The gel was depurinated with 0.125 M HCl for 10 min followed by slight rinse in water. It was then denatured twice with 0.5 M NaOH/1.5 M NaCl for 15 min followed by a slight rinse in water each time. The gel was neutralized twice with 0.5 M Tris/1.5 M NaCl (pH 7.5) for 15 min and blotted onto the Roti Nylon Plus membrane (Roth) with a vaccum blotter in 20x SSC and cross-linked to the membrane under UV light. After rinsing with 2x SSC, the membrane was prehybridized in 1x BBR for at least 1 h at 42 ^{o}C with approximate 20 ml/100 cm^2 membrane. The DIG labeled DNA probe (10-20 ng/ml hybridization solution) was denatured by boiling for 10 min and rapid cooling on ice water. At least 3.5 ml of hybridization solution/100 cm^2 membrane were added and the probe hybridized at 42 ^{o}C overnight. The membrane was washed twice under constant shaking for 5 min with 2x SSC/0.1% SDS and twice for 15 min with 0.5x SSC/0.1% SDS. For detection, the membrane was incubated for 2 min with maleic acid buffer/0.3% Tween 20, blocked for 30 min with 1x BBR in maleic acid buffer and the anti-DIG-AP antibody was applied in blocking solution and incubated for 30 min at RT. The membrane was washed again two times for 15 min with maleic acid buffer/0.3% Tween 20, incubated in detection buffer for 5 min followed by adding CDP-star substrate (10-30 drops depending upon the membrane size) for 5 min and a film exposed for 10-15 min.

Maleic acid buffer

100 mM	Maleic acid
150 mM	NaCl
pH 7.5	

10x BBR (Blocking reagent)

10 % (w/v) BBR in Maleic acid buffer

Detection buffer

100 mM Tris
200 mM NaCl
pH 9.5

4.2.17 Isolation of genomic DNA from mouse tails

To analyze the germline transmission of the targeted *col28a1* allele by PCR, the genomic DNA was prepared from the mouse tail biopsies. 600 µl of lysis buffer and 10 µl proteinase K (10 mg/ml) solution were added to each tail biopsy. The tubes were incubated overnight at 55 ^{o}C to allow efficient proteinase K digestion. The debris was removed by centrifuging at 13,000 rpm for 5 min, 700 µl isopropanol were added to the supernatant and the sample centrifuged at 13,000 rpm for 20 min. The pellet was washed in 70% ethanol. After final centrifugation, the pellet was air dried for 10-20 min, resuspended in TE buffer and solubilized overnight at 55 ^{o}C.

Lysis buffer

50 mM	Tris-HCl (pH 8.0)
50 mM	EDTA (pH 8.0)
0.5%	SDS

TE

10 mM	Tris-HCl (pH 8.0)
1 mM	EDTA

4.2.18 Isolation of genomic DNA from ES cells

ES cells grown in 48 well plates were used for preparing genomic DNA. After removing the media, the cells were washed once with PBS. 400 µl lysis buffer was added to each well, the plates sealed with parafilm and the cells incubated at 55 ^{o}C overnight to allow efficient proteinase K digestion. The lysate was transferred to a sterile tube, 800 µl isopropanol added to the supernatant and the sample centrifuged at 13,000 rpm for 20 min. The pellet was washed in 70% ethanol. After final centrifugation, the pellet was air dried for 10-20 min, resuspended in TE buffer and solubilized overnight at 55 ^{o}C. Quantification of DNA was performed with a Nanodrop photometer (Thermo Scientific).

Lysis buffer

50 mM	Tris-HCl (pH 8.5)
5 mM	EDTA
200 mM	NaCl
0.2%	SDS
1 mg/ml	Proteinase K (Qiagen)

4.3 BIOCHEMICAL METHODS

4.3.1 Affinity purification of His-tagged proteins

2 ml of Ni-NTA superflow were equilibrated in a column with 2 CV (column volumes) of wash buffer under gravity flow. Before applying the cell culture supernatants to the column, these were centrifuged at 4 ^{o}C, 10,000 rpm for 30 min followed by filtering through Whatman filter paper to remove any protein aggregates. The cell supernatant was then loaded onto the Ni-NTA column. In order to ensure proper binding the loading was repeated twice. The column was washed with 10 CV and proteins were eluted with varying concentrations of imidazole; 10 mM, 20 mM, 50 mM, 80 mM, 100 mM and 200 mM.

Wash buffer		**Elution buffer**
50 mM	Tris-HCl	1 M Imidazole in wash buffer
150 mM	NaCl	
pH 7.5		

4.3.2 Affinity purification of Strep-tagged proteins

A 2 ml Strep-Tactin Sepharose column was equilibrated with 2 CV of wash buffer under gravity flow. Before applying the cell culture supernatants to the column, these were centrifuged at 4 ^{o}C, 10,000 rpm for 30 min followed by filtering through Whatman filter paper to remove any protein aggregates. The cell supernatant was then loaded onto the Ni-NTA column. In order to ensure proper binding the loading was repeated twice. The column was washed with 20 CV and proteins were eluted with 20 ml elution buffer. The column was regenerated by washing with 3 CV of regeneration buffer. The column activity status was monitored by change in color from yellow to red. The column was washed and stored in wash buffer at 4 ^{o}C.

Wash buffer		**Elution buffer**	**Regeneration buffer**	
100 mM	Tris	2.5 mM Desthiobiotin	100 mM	Tris
150 mM	NaCl	in wash buffer	1 mM	HABA
pH 8.0				

4.3.3 Ligand coupling and affinity purification of antibodies

Before coupling, the recombinant proteins were dialyzed three times against coupling buffer at 4 ^{o}C. For activation, 150 mg of CNBr-Sepharose were incubated with 1 ml 1 mM HCl. The treatment was repeated five times and the activated beads added to the coupling solution containing the ligand and immediately vortexed to avoid cross-linking of beads. The mixture was incubated for 1 h at RT or overnight at 4 ^{o}C under shaking, centrifuged at 1000 rpm at RT and the pellet washed twice with 2 ml of coupling buffer. Remaining active groups were blocked by washing with 0.1 M Tris-HCl buffer, pH 8.0, for 2 h. The coupled CNBr-Sepharose column was washed with 3 CV of 0.1 M Tris-HCl buffer, pH 8.0, and at least with three cycles of 10 CV with regeneration buffer I followed by a 10 CV with regeneration buffer II.

The antiserum (10 ml) was centrifuged at 13,000 rpm for 1 h at 4 ^{o}C and diluted with PBS or 10mM Tris-HCl (pH 7.5) at a ratio 1:1 before loading onto the column. The flowthrough was recycled over the column at least three times or overnight for efficient binding of the antibodies. The column was washed with 5 CV of 10 mM Tris-HCl (pH 7.5) followed by TBS and the antibodies were eluted in 1 ml fractions with either high pH (Elution buffer I) or low pH (Elution buffer II). In both cases, the pH of the eluate was set to neutral by adding Tris-HCl, pH 8.0. The A_{280nm} was measured and the solution was dialyzed against TBS. The antibody titer was analyzed using ELISA. The column was regenerated by washing with three cycles of alternating pH buffers followed by equilibration in 10 mM Tris-HCl (pH 7.5).

Coupling buffer	**Regeneration buffer I**	**Regeneration buffer II**
0.1 M $NaHCO_3$	0.1 M Sodium acetate	0.1 M Tris-HCl
0.5 M NaCl	0.5 M NaCl	0.5 M NaCl
pH 8.3	pH 4.0	pH 8.0

Elution buffer I	**Elution buffer II**
0.1 M Triethylamine	0.1 M Glycine
0.15 M NaCl	pH 2.5
pH 11.5	

4.3.4 SDS-polyacrylamide gel electrophoresis

The recombinant proteins or tissue extracts were separated by SDS polyacrylamide gel electrophoresis using Laemmli discontinuous gels (Laemmli, 1970). The resolving gel (pH 8.8) was casted followed by a stacking gel (pH 6.8) in glass plates of 10 cm x 7.5 cm and spacers of 0.5 cm thickness. The acrylamide concentration was adjusted according to the requirement. The samples were loaded after mixing with the sample buffer. DTT or β-mercaptoethanol was added when the samples were to be reduced.

Resolving gel (10%)

12.3 ml	H_2O
7.5 ml	1.5 M Tris-HCl (pH 8.8)
150 µl	20% SDS
9.9 ml	Acrylamide/Bis-acrylamide (30%/0.8%)
150 µl	10% APS
20 µl	TEMED

Stacking gel (4%)

3 ml	H_2O
1.25 ml	0.5 M Tris-HCl (pH 6.8)
25 µl	20% SDS
670 µl	Acrylamide/Bis-acrylamide (30%/0.8%)
250 µl	10% APS
5 µl	TEMED

Sample buffer (5x)

10% (w/v)	SDS
10 mM	DTT/β-mercaptoethanol
20%	Glycerol
0.2 M	Tris-HCl (pH 6.8)
0.05% (w/v)	Bromophenol blue

Running buffer

25 mM	Tris-HCl
200 mM	Glycine
0.1% (w/v)	SDS

4.3.5 Coomassie staining

The resolved proteins were visualized by staining the gel with Coomassie Brilliant Blue (CBB) solution at RT with gentle agitation for at least 60 min. Then, the staining solution was removed and destaining solution was added. The gel was destained at RT with gentle agitation with at least two changes of destaining solution until protein bands were clearly visible.

Coomassie brilliant blue solution

0.1% w/v Coomassie brilliant blue R250
50% v/v Ethanol
10% v/v Acetic acid
filter the solution before use

Destaining solution

7% v/v Acetic acid
20% v/v Ethanol

4.3.6 Western blotting

SDS-PAGE gels were blotted onto a nitrocellulose membrane (Protran BA45, Schleicher Schüll) using a Tankblot (Hoefer TE 22 Mighty Small Transphor Tank

Transfer Unit) apparatus at 4 °C under 400 mA for 2 h or at 100 mA overnight with the blotting buffer. After the transfer, the membrane was stained for 2 min in Ponceau S (Serva) in order to visualize the efficiency of transfer. Excess stain was removed by rinsing the membrane in water. The membrane was washed with TBS for 2 min to remove the Ponceau-S, blocked with 5% milk powder in TBS for 1 h at RT and the appropriate primary antibody, diluted in 1% milk powder in TBS, was applied. After incubating for 1 h at RT under constant shaking, the membrane was washed three times with TBS-T for 5 min each and the appropriate secondary antibody was applied for 1 h at RT. The membrane was washed three times with TBS-T for 5 min each and once in TBS, incubated in ECL solution (Enhanced chemiluminescence) for 2-3 min and a film exposed for 1 min or more depending on the background to signal ratio.

Blotting buffer

50 mM Boric acid (pH 8.5)
10% (v/v) Methanol

ECL-Solution

333 µl	3 M Tris-HCl (pH 8.5)
22 µl	90 mM *p*-Coumaric acid
50 µl	250 mM Luminol
3 µl	30% (v/v) H_2O_2
10 ml	H_2O

4.3.7 Enzyme Linked Immunosorbent Assay (ELISA)

The recombinant proteins were dialyzed against TBS and coated onto ELISA plates at a concentration of 500 ng/well. A minimum of 50 µl was used to ensure even coating of the recombinant protein. The plates were incubated overnight at 4 °C and washed three times with a minimum volume of 400 µl of TBS-T per well. 55 µl of the blocking solution was applied to each well and the plates incubated at RT for 2 h. After removal of the blocking solution, 50 µl of primary antibody solution was added. The antibodies were diluted 1:10, 1:100, 1:500, 1:1000, 1:2000, 1:5000, 1:10,000, 1:100,000 in 1% milk powder in TBS. The plates were incubated for 1 h at RT under shaking and washed three times with TBS-T. The secondary antibody (conjugated with HRP) was added to each well and incubated at RT for 1 h. After washing four times with TBS-T and twice with distilled water, 50 µl of the developing solution (substrate) was added to each well and the plates kept in the dark 10-90 min. The appearance of blue color was monitored. As soon as the negative control started developing blue colour, the stop solution (10% H_2SO_4) was added and A_{450} was measured.

TBS-T	**Blocking solution**
0.05% (v/v) Tween 20 in 1xTBS	5% (w/v) Milk powder in 1xTBS

Developing solution

1 ml	Sodium acetate/Citric acid (pH 6.0)
62.5 µl	3,3'-5,5'-Tetramethylbenzidine (10 mg/ml in DMSO)

4.4 CELL BIOLOGY METHODS

4.4.1 Culture and expansion of cells

The overexpression of recombinant proteins was performed in HEK-293-EBNA cells (Invitrogen) which are derived from human embryonic kidney cells. These cells contain a neomycin resistance gene, which can be selected with Geneticin (G418), and the EBNA gene for episomal replication of the pCEP-Pu vectors used here. The cells were cultured at 37 ^{o}C and 5% CO_2 in standard culture medium. For selection, the medium was supplemented with 0.5 µg/ml of puromycin. The cells were grown in 10 cm cell culture dishes. Confluent cells were detached by 0.25% trypsin and 2 mM EDTA (T/E) at 37 ^{o}C for 5 min followed by neutralizing with DMEM/F12 medium corresponding to at least 5 times the volume of trypsin solution. The cells were expanded to triple flasks (2:1) for large-scale culture.

Standard cell culture medium

500 ml	DMEM/F12 (1:1) (Invitrogen)
50 ml	FCS (Fetal calf serum) (10%)
5 ml	Glutamine (2 mM) (Gibco BRL)
5 ml	Penicillin/Streptomycin (Gibco BRL)

4.4.2 Cell storage

When the cells attained 80% confluency they were washed, trypsinized and pelleted at 4000xg for 5 min. 1 ml of freezing medium was added per 10 cm plate and the cells resuspended before aliquoting into the cryotubes. The cryotubes were placed in an isopropanol containing cryo-freezing box (Nalgene) for 24 h at -80 ^{o}C and stored in liquid nitrogen for longer periods.

Freezing medium

10% FCS
10% DMSO
in standard culture medium

4.4.3 Stable transfection of cells

The cells were transfected using FuGene6 (Roche). Cells were grown to 50-70% confluency and the medium was changed just before the transfection. 6 µl of FuGene 6 was directly added to 94 µl of serum free medium in a sterile tube and incubated for 5 min at RT. 3 µg of plasmid DNA was added to a second tube, the diluted FuGene 6 mix dropwise added, gently mixed and incubated for 15 min at RT. The mixture was dropwise added to the cells and gently swirled. 24 h after transfection, fresh standard medium containing puromycin was added and the selection continued to attain stable transfection.

4.4.4 Large-scale cell culture

For each construct, the stably transfected cells were expanded to four triple flasks. When the cells were confluent, serum free medium containing puromycin was added. The medium was changed every 2-3 days and the supernatant stored at 4 ^{o}C or -20 ^{o}C after adding 0.02% (w/v) sodium azide to avoid bacterial contamination.

4.4.5 ES cell culture

4.4.5.1 Expansion of ES cells

Prior to ES cell culture, Embryonic Mouse Feeder cells (EMF1) were thawed and cultured on a 10 cm plate. The EMF1 cells were inactivated by using Mitomycin (70 µl/10ml) for 2-3 h. The 24-well or 6-well plate surface was treated with gelatine (0.1 % in PBS) for 20 min and washed twice with PBS. The EMF1 cells were trypsinized for 5 min FCS containing media was added and by pipette up and down a single cell suspension was achieved. The EMF1 cells were centrifuged, resuspended in DMEM/FCS and plated on either 24-well or 6-well plates. The following day, they were washed twice with ES cell medium and finally resuspended in 300 µl medium. The ES cells were thawed and transferred to a 15 ml Falcon tube containing 13 ml pre-warmed ES cell medium. The cells were centrifuged at 700 rpm for 8 min and the supernatant carefully discarded. The pellet was resuspended in 300 µl ES cell

medium, added to the EMF1 cells and finally 400 µl of fresh media was added to evenly distribute the cells.

ES cell culture media (KO media)

500ml	Knockout DMEM
6ml	Glutamine (200mM)
2.5ml	Penicillin/Streptomycin
90ml	Serum replacement
6ml	Sodium pyruvate
6ml	Non-essential amino acids (NEAA)
7µl	β-mercaptoethanol
100µl	LIF

4.4.5.2 Transfection of ES cells via electroporation

The medium of the ES cells was changed four hours prior to transfection to ensure that the cells pass mitosis during the transfection event, since incorporation of DNA via homologous recombination is only possible in the M-phase. For transfection of 1 x 10^7 ES cells, 40 µg of targeting vector DNA was used. The purifed DNA that was linearized with Asc I was resuspended in 400 µl RPMI (transfection buffer). The ES cells were washed twice with 12 ml PBS and trypsinated for 5 min at 37 °C. The trypsin digestion was stopped by adding 5 ml ES medium. The cells were centrifuged for 5 min at 1000 rpm, the supernatant discarded and the cell pellet resuspended in 400 µl RPMI buffer. DNA and ES cells were mixed and transferred into an electroporation cuvette. The transfection was performed in an electroporator at 500 µF and 240 mV. For regeneration, the cells were incubated for 10 min at RT. Afterwards the cells were resuspended in 40 ml ES medium and distributed onto plates of mitomycin-treated EF3 feeder cells.

4.4.5.3. Positive and negative selection of ES cell clones

For the enhancement of homologously recombined cells, the transfected ES cells were subjected to positive and negative selection. For positive selection, the neomycin analog, G418, was used. It binds to the 80S ribosomal subunit and thus inhibits translation in eukaryotic cells. Only those cells, which have the neo resistance gene stably, integrated into their genome survive. The nucleoside analog gancyclovir was used as a negative selection marker. A viral thymidine kinase is capable to transform gancyclovir into a nucleotide monophosphate, following conversion into a nucleotide

triphosphate by cellular kinases. Integration of this nucleotide triphosphate into a DNA strand during replication stops elongation. The targeting vector contains the gene for the Herpes simplex virus-derived thymidine kinase (HSV-tk). Since the HSV-tk gene is located outside the recombination site, the gene is integrated into the genome only in case of random integration and not in case of homologous recombination. For this reason only homologously recombined clones, which do not contain the gene for the thymidine kinase, are able to grow in gancyclovir containing medium. The positive G418 (250 µg/ml) selection was started 48 hours after transfection. The negative selection with a gancyclovir (2 µM) was started on the fifth day. During 9-10 days of selection, the medium was changed daily.

4.4.5.4 Picking of ES cell clones

ES cell clones, which were not differentiated and maintained during selection, were isolated 8-10 days after transfection. The cells were washed on the 10 cm plate and immersed in 10 ml of chilled PBS under a stereo-binocular microscope, single clones were picked and transferred into the wells of a round-bottom 96 well plate filled with 25 µl 1 x trypsin. To pick clones, the pipette was adjusted to 40 µl and the tip placed directly in front of a clone. With a slight forward movement, the clone was picked. Picking of clones from one 96 well plate was performed within 30 minutes to prevent damage of picked clones by trypsin. The entire 96 well plate was then incubated for 5 min at 37°C and the trypsin digestion stopped by adding 100 µl ES medium, each well thus containing a total volume of 165 µl (40 µl 1x PBS, 25 µl trypsin, 100 µl ES medium). After thorough resuspension, 50 µl transferred onto each of three new flatbottom 96 well plates coated with mitomycin-treated EF cells, resulting in triplicates of each clone. The plates were cultivated in an incubator and the medium changed daily.

4.4.5.5 Freezing and passage of ES cell clones

After cultivation of the isolated ES cell clones for 2-3 days, the cells were washed with 150 µl 2x PBS and detached from the plate by adding 150 µl 2x trypsin solution and incubated for 10 min at 37 °C. The trypsin reaction was stopped by adding 25 µl of trypsin inhibitor (5 mg/ml) followed by 700 µl of KO medium. Subsequently, a single cell suspension was generated by repeated up and down pipetting and 500 µl of cell suspension was transferred into a freezing vial. 500 µl of freezing medium was

added and the vials were kept in cryobox at -80 °C for 24 h and eventually stored in liquid nitrogen. For DNA isolation, the remaining ES cells (150 µl) were transferred to 48 well plates and cultured till confluent.

2x Freezing medium

25 ml KO-medium
15 ml Serum replacement
10 ml DMSO

4.4.5.6 Preparation of ES cells for injection into blastocysts

The ES cells of one 10 cm plate were cultivated for three days prior to injection. On the day of the injection, cells were washed once with PBS and detached from the plate with trypsin. The cells were centrifuged for 5 min at 1000 rpm and plated on a 10 cm dish that was coated with gelatine. After 35-40 min the supernatant was transferred into a 15 ml Falcon tube. Since EMF1 cells settle down faster than ES cells, after 35-40 minutes EMF1 cells were attached to the bottom of the plate, while the ES cells were still floating in the supernatant. The ES cells were centrifuged for 5 min at 1000 rpm, resuspended in 0.5 ml chilled injection medium and stored on ice until injection into blastocysts.

4.5 HISTOLOGICAL METHODS

4.5.1 Cryosectioning

Freshly dissected tissues from newborn and adult mice were immersed in a plate containing tissue-embedding medium, Tissue Tek (Sakura), for a couple of minutes. The tissue was carefully placed in the desired orientation in a cryomold containing Tissue Tek without bubbles and the mold finally completely filled with Tissue Tek. The tubes were labeled and the tissue immediately frozen on a metal block that was already submerged in liquid nitrogen. The frozen tissues were stored at -80 ^{o}C. Sections were made with a Leica Cryotome (CM3050) at -20 ^{o}C. The thickness of all sections was 7 µm. After sectioning the sections were stored at -20 ^{o}C.

4.5.2 Immunofluorescent staining

The cryosections were left at RT for about 1 h and then incubated in PBS to remove the Tissue Tek. The edges of the specimen were marked with a wax marker pen. The

sections were fixed with 2% paraformaldehyde in PBS for 10 min and washed three times in PBS or TBS for 5 min each. The sections were incubated with freshly prepared hyaluronidase solution (50 U/ml hyaluronidase in HB) for 30 min at 37 ^{o}C, washed twice with TBS for 5 min each and blocked with 5% BSA in 0.1% Triton in TBS at RT. The primary antibody was applied at the required dilution in 1% BSA/TBS for 1 h at RT. The sections were washed three times for 5 min each in wash buffer and the specific secondary antibody diluted in 1% BSA/TBS was applied and kept in dark for 1 h as the fluorescent antibodies are light sensitive. After washing three times with wash buffer for 5 min each, the slides were kept in water for 5 min. 5-6 drops of Dako fluorescent mounting medium were added, the cover slip slowly mounted without trapping bubbles and the slides kept at 4 ^{o}C overnight.

Hyaluronidase buffer (HB)

0.1 M NaH_2PO_4
0.1 M Sodium acetate
pH 5.0 (adjust with acetic acid)

Wash buffer

0.1% Triton in 1x TBS

4.5.3 H&E staining

The cryosections were left at RT for about 15 min, and then washed in TBS for 5 min. The sections were kept in filtrated haematoxylin for 3 min and rinsed shortly in tap water, quickly immersed in HCl-alcohol and rinsed in tap water for 5-10 min. The sections were shortly rinsed in water followed by a 3-5 min incubation in eosin. The sections were rinsed carefully with water and shortly immersed in 70% ethanol, followed by 80% and 96% ethanol for 3 min each. The sections were rinsed twice in either isopropanol or 100 % ethanol and twice in xylol. The sections were covered with DPX.

Haematoxylin solution (1 lt)

1 g	Haematoxylin (dissolve first)
200 mg	$NaIO_3$
50 g	$KAl(SO_4)_2$

Shake till the solution turn blue-violet and then add

50 g	Chloralhydrate [$Cl_3CCH(OH)_2$]
1 g	Citric acid

HCl-alcohol (100 ml)

0.5 ml Conc. HCl
70% Ethanol

Eosin solution

1% (w/v) Eosin in ethanol

4.6 GENETIC METHODS

4.6.1 Recombineering (In vivo cloning)

The 3' homology arm of the targeting vector was cloned using recombineering. Recombineering or recombinant engineering is a powerful method for fast and effective construction of vectors for subsequent manipulation of the mouse genome. This method is based on homologous recombination in *E. coli* using recombination proteins provided from the λ phage. The bacterial strains that are used in recombineering are modified in such a way that they contain a defective λ prophage

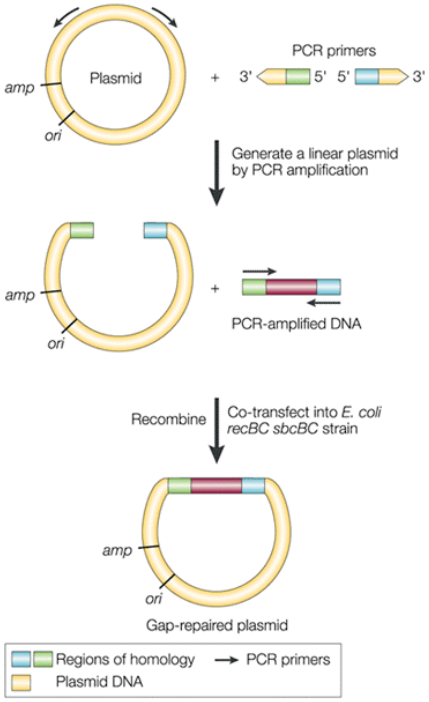


Figure 4.1 **Schematic representation of *in vivo* cloning**. *In vivo* cloning uses two linear DNAs - a vector and a target DNA - which carry stretches of homology to each other at their ends. Both linear DNAs are electroporated into competent cells to allow homologous recombination between them, thereby repairing the plasmid DNA by closing the circle. (amp, ampicillin resistance gene; ori, origin of replication). *adapted from Copeland et al., Nature Reviews Genetics; 2001.*

in their genome. The modified bacterial strains include DY380/SW102, EL250/SW105 and EL350/SW106. The phage genes of interest *exo*, *bet* and *gam* are transcribed from the λPL promoter. This promoter is repressed by the temperature sensitive repressor cI857 at 32^{o}C and derepressed (the repressor is inactive) at 42^{o}C. Therefore, when the bacterial strain containing this prophage is kept at 32^{o}C, no recombination proteins are produced. However after a brief heat shock at 42 ^{o}C, a sufficient number of recombinant proteins are produced. Exo is 5' – 3'exonuclease that can create single stranded overhangs on linear DNA. On the other hand, bet

protects these overhangs and assists in the subsequent recombination process whereas gam prevents degradation of linear DNA by inhibiting E. coli RecBCD protein.

Upon the introduction of a linear DNA (PCR product, oligo, etc.) with sufficient homology in the 5' and 3' ends to a target DNA molecule that is already present in the bacteria (plasmid, BAC, or the bacterial genome itself) by electroporation, the introduced DNA can be modified by *exo* and *bet* and undergo homologous recombination with the target molecule.

The method has several advantages over the classical genetic engineering. For instance, it is independent of restriction sites, allows high flexibility, has no size limit and causes no unwanted mutations.

4.7 BIOINFORMATIC PROGRAMS AND SERVERS

Genomic databases	Mouse (Build 37.1) Human (Build 37.1) Rhesus macaque (Build 1.1) Cattle (Build 3.1) Horse (EquCab 2.1) Dog (Build 2.1) Chicken (Build 2.1) Zebrafish (Zv7)
Signal peptide prediction	SignalP v3.1 (http://www.cbs.dtu.dk/services/SignalP)
Protein parameters	Expasy ProtParam (http://www.expasy.ch/tools/protparam.html)
Domain prediction servers	SMART (http://smart.embl-heidelberg.de/)
Alignment programs	ClustalX v1.81 Boxshade v3.21 (http://www.ch.embnet.org/software/BOX_form.html) Jalview (http://www.jalview.org/)
Phylogenetic program	Phylip v3.66
Graphical software	Vector NTI (Invitrogen)

5 Abstract

Collagen VI and collagen XXVIII are two extracellular matrix proteins that belong to the superfamily of von Willebrand Factor A (VWA) domain containing molecules. Earlier studies on collagen VI indicated that this widely distributed protein is composed of α1, α2 and α3 polypeptide chains, which form a microfibrillar network in close association with basement membranes in muscle and other tissues. In contrast, an initial study on collagen XXVIII reported that it forms a homotrimer and has a very restricted localization at specific basement membranes of peripheral nerves.

In this dissertation, the identification and characterization of three novel collagen VI chains, α4, α5 and α6, that show similarity to the collagen VI α3 chain is described. The genes coding for the new chains are arranged in tandem on mouse chromosome 9. The proteins contain seven N-terminal VWA domains followed by a collagenous domain, two C-terminal VWA domains and a unique domain. In addition the collagen VI α4 chain carries a Kunitz domain at the C-terminus whereas the collagen VI α5 chain contains an additional VWA domain and unique domain. The lengths of the collagenous domains and the positions of the structurally important cysteine residues are identical in the collagen VI α3, α4, α5 and α6 chains. In mouse, the new chains show a very restricted and differential expression mainly associated with basement membranes. They are sometimes detected in regions where the collagen VI α3 chain is not expressed, suggesting that the α3 chain is not required for their assembly. Analysis of the collagen VI α1 chain deficient mouse strain, confirmed that the new chains require the α1 chain and may substitute for the α3 chain, probably forming α1α2α4, α1α2α5 and α1α2α6 heterotrimers. In humans, only the genes coding for the collagen VI α5 and α6 chains are preserved. The *COL6A4* gene has been inactivated due to large pericentric inversion on chromosome 3 that split the gene in two pieces and transformed it into two non-processed pseudogenes. In humans, the collagen VI α5 and α6 chains are present in close association with the basement membranes of skeletal muscle and skin. Ullrich Congenital Muscular Dystrophy (UCMD) and Bethlem Myopathy (BM) patients carrying mutations

in *COL6A1*, *COL6A2* and *COL6A3* show also skin phenotypes like keloid scarring or keratosis pilaris. Immunohistochemical analysis of the new chains in the skin of UCMD and BM patients showed a disturbed staining pattern only when the *COL6A1* or *COL6A2* genes are affected. This indicates that the new chains may substitute for the collagen VI α3 chain forming α1α2α5 and α1α2α6 heterotrimers. However the exact role of new chains for the development of skin phenotypes in myopathy patients remains to be elucidated.

The functional role of collagen XXVIII is not known. Therefore, the inactivation of the *Col28a1* gene in mouse was initiated. A targeting vector disrupting the exon 2 of *Col28a1* was generated *in vitro,* followed by ES cell targeting *in vivo.* Positive ES clones were injected into blastocysts and transferred to surrogate mothers, which resulted in a chimeric mice carrying both the wild type and the targeted allele. However, the targeted allele did so far not enter the germline.

6 Zusammenfassung

Kollagen VI und Kollagen XXVIII sind Mitglieder einer Superfamilie von Proteinen der extrazellulären Matrix, die von Willebrand Faktor A (VWA) Domänen enthalten. Das weit verbreitete Kollagen VI besteht aus je einer α1, α2, und α3 Polypeptidkette und bildet ein Netzwerk aus Mikrofibrillen, das in enger Verbindung mit Basalmembranen im Muskel und anderen Geweben steht. Im Gegensatz dazu bildet Kollagen XXVIII Homotimere und zeichnet sich durch eine sehr begrenzte Lokalisierung an Basalmembranen peripherer Nerven aus.

In dieser Dissertation wird die Identifizierung und Charakterisierung von drei neuen Kollagen VI Ketten, α4, α5, und α6, beschrieben, die strukturelle Ähnlichkeit zur Kollagen VI α3 Kette aufweisen. Die für die neuen Ketten kodierenden Gene liegen im Genom der Maus hintereinander auf Chromosom 9. Die Proteine bestehen aus sieben N-terminalen VWA Domänen, gefolgt von einer kollagenen Domäne, zwei C-terminalen Domänen und einer uniquen Domäne. Die Kollagen VI α4 Kette enthält zusätzlich eine Kunitz Domäne am C-Terminus, wohingegen die Kollagen VI α5 Kette eine zusätzliche VWA Domäne und eine weitere unique Domäne enthält. Die Längen der kollagenen Domänen sowie die Positionen der strukturell wichtigen Cysteinreste sind identisch in den Kollagen VI α3, α4, α5 und α6 Ketten. In der Maus zeigen die neuen Ketten eine sehr limitierte und unterschiedliche Expression, die meist mit Basalmembranen assoziert ist. Manchmal wurden die neuen Ketten in Regionen beobachtet, in denen die Kollagen VI α3 Kette nicht exprimiert ist, was darauf schließen lässt, dass Kollagen VI auch ohne die α3 Kette aufgebaut sein kann. Die Analyse eines Kollagen VI α1 defizienten Mausstamms bestätigte, dass auch die neuen Ketten die α 1 Kette zur Assemblierung eines Trimers benötigen und eventuell die α3 Kette ersetzen, wahrscheinlich um Heterotrimere bestehend α1, α2 und entweder α4, α5 oder α6 zu bilden. Beim Menschen sind nur die Gene erhalten, die für die Kollagen VI α5 und α6 Kette kodieren. Das *COL6A4* Gen wurde durch eine große perizentrische Inversion des Chromosoms 3 inaktiviert. Hierbei wurde das Gen in zwei Hälften gespalten und in zwei nicht-

prozessierte Pseudogene umgewandelt. Beim Menschen konnten die Kollagen VI α5 und α6 Ketten in Verbindung mit Basalmembranen im Muskel und in der Haut nachgewiesen werden. Mutationen in *COL6A1*, *COL6A2* oder *COL6A3* führen beim Menschen zu kongenitaler Muskelschwäche vom Typ Ullrich (UCMD) oder zu Bethlem-Myopathie (BM). Patienten zeigen u.a. Hautphänotypen, wie überschießende Narbenbildung oder Reibeisenhaut. Die immunhistochemische Analyse der neuen Ketten in der Haut von UCMD und BM Patienten zeigte nur dann ein verändertes Färbemuster, wenn die *COL6A1* oder *COL6A2* Gene mutiert waren. Dies deutet ebenfalls darauf hin, dass die neuen Ketten die Kollagen VI α3 Kette ersetzen und dabei Heterotrimere bestehend aus α1, α2, und α5 bzw. α6 bilden könnten. Die Aufklärung der Rolle der neuen Ketten bei der Entstehung der Hautphänotypen bei Patienten mit Muskelschwäche bedarf in Zukunft weiterer Untersuchungen.

Die funktionelle Bedeutung von Kollagen XXVIII ist bisher völlig unbekannt. Daher sollten in der vorliegenden Arbeit Vorraussetzungen geschaffen werde, um das *Col28a1* Gen in der Maus zu inaktivieren. Zunächst wurde ein Targeting Vektor hergestellt, in dem das Exon 2 des *Col28a1* Gens unterbrochen wurde, um transgene ES-Zellen zu generieren. Positive ES-Zellklone wurden bereits in Blastocysten injiziert, die dann in scheinschwangere Mäuse transferiert wurden. Die daraus resultierenden chimären Nachkommen trugen sowohl das Wildtypallel als auch das Transgen. Bisher konnten allerdings keine Nachkommen generiert werden, die das Transgen in der Keimbahn trugen.

7 References

Aumailley, M., Mann, K., von der Mark, H. & Timpl, R. (1989) Cell attachment properties of collagen type VI and Arg-Gly-Asp dependent binding to its alpha 2(VI) and alpha 3(VI) chains. *Exp Cell Res* **181**(2), 463-74.

Aumailley, M., Specks, U. & Timpl, R. (1991) Cell adhesion to type-VI collagen. *Biochem Soc Trans* **19**(4), 843-7.

Baker, N.L., Morgelin, M., Peat, R., Goemans, N., North, K.N., Bateman, J.F. & Lamande, S.R. (2005) Dominant collagen VI mutations are a common cause of Ullrich congenital muscular dystrophy. *Hum Mol Genet* **14**(2), 279-93.

Baldock, C., Sherratt, M.J., Shuttleworth, C.A. & Kielty, C.M. (2003) The supramolecular organization of collagen VI microfibrils. *J Mol Biol* **330**(2), 297-307.

Ball, S.G., Baldock, C., Kielty, C.M. & Shuttleworth, C.A. (2001) The role of the C1 and C2 a-domains in type VI collagen assembly. *J Biol Chem* **276**(10), 7422-30.

Bella, J. & Berman, H.M. (1996) Crystallographic evidence for C alpha-H...O=C hydrogen bonds in a collagen triple helix. *J Mol Biol* **264**(4), 734-42.

Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266**(5182), 75-81.

Berisio, R., Vitagliano, L., Mazzarella, L. & Zagari, A. (2002) Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3). *Protein Sci* **11**(2), 262-70.

Bethlem, J. & Wijngaarden, G.K. (1976) Benign myopathy, with autosomal dominant inheritance. A report on three pedigrees. *Brain* **99**(1), 91-100.

Bidanset, D.J., Guidry, C., Rosenberg, L.C., Choi, H.U., Timpl, R. & Hook, M. (1992) Binding of the proteoglycan decorin to collagen type VI. *J Biol Chem* **267**(8), 5250-6.

Bienkowska, J., Cruz, M., Atiemo, A., Handin, R. & Liddington, R. (1997) The von willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. *J Biol Chem* **272**(40), 25162-7.

Bischof, J.M., Chiang, A.P., Scheetz, T.E., Stone, E.M., Casavant, T.L., Sheffield, V.C. & Braun, T.A. (2006) Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum Mutat* **27**(6), 545-52.

Bonaldo, P., Braghetta, P., Zanetti, M., Piccolo, S., Volpin, D. & Bressan, G.M. (1998) Collagen VI deficiency induces early onset myopathy in the mouse: an animal model for Bethlem myopathy. *Hum Mol Genet* **7**(13), 2135-40.

Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R. & Colombatti, A. (1990) Structural and functional features of the alpha 3 chain indicate a bridging role for chicken collagen VI in connective tissues. *Biochemistry* **29**(5), 1245-54.

Borza, D.B., Bondar, O., Ninomiya, Y., Sado, Y., Naito, I., Todd, P. & Hudson, B.G. (2001) The NC1 domain of collagen IV encodes a novel network composed of the alpha 1, alpha 2, alpha 5, and alpha 6 chains in smooth muscle basement membranes. *J Biol Chem* **276**(30), 28532-40.

Boutaud, A., Borza, D.B., Bondar, O., Gunwar, S., Netzer, K.O., Singh, N., Ninomiya, Y., Sado, Y., Noelken, M.E. & Hudson, B.G. (2000) Type IV

collagen of the glomerular basement membrane. Evidence that the chain specificity of network assembly is encoded by the noncollagenous NC1 domains. *J Biol Chem* **275**(39), 30716-24.

Bradley, M., Soderhall, C., Luthman, H., Wahlgren, C.F., Kockum, I. & Nordenskjold, M. (2002) Susceptibility loci for atopic dermatitis on chromosomes 3, 13, 15, 17 and 18 in a Swedish population. *Hum Mol Genet* **11**(13), 1539-48.

Bu, L.M., Bradley, M., Soderhall, C., Wahlgren, C.F., Kockum, I. & Nordenskjold, M. (2006) Susceptibility loci for atopic dermatitis on chromosome 21 in a Swedish population. *Allergy* **61**(5), 617-21.

Burg, M.A., Tillet, E., Timpl, R. & Stallcup, W.B. (1996) Binding of the NG2 proteoglycan to type VI collagen and other extracellular matrix molecules. *J Biol Chem* **271**(42), 26110-6.

Camacho Vanegas, O., Bertini, E., Zhang, R.Z., Petrini, S., Minosse, C., Sabatelli, P., Giusti, B., Chu, M.L. & Pepe, G. (2001) Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proc Natl Acad Sci U S A* **98**(13), 7516-21.

Chu, M.L., Conway, D., Pan, T.C., Baldwin, C., Mann, K., Deutzmann, R. & Timpl, R. (1988) Amino acid sequence of the triple-helical domain of human collagen type VI. *J Biol Chem* **263**(35), 18601-6.

Chu, M.L., Pan, T.C., Conway, D., Kuo, H.J., Glanville, R.W., Timpl, R., Mann, K. & Deutzmann, R. (1989) Sequence analysis of alpha 1(VI) and alpha 2(VI) chains of human type VI collagen reveals internal triplication of globular domains similar to the A domains of von Willebrand factor and two alpha 2(VI) chain variants that differ in the carboxy terminus. *EMBO J* **8**(7), 1939-46.

Chu, M.L., Zhang, R.Z., Pan, T.C., Stokes, D., Conway, D., Kuo, H.J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R. & et al. (1990) Mosaic structure of globular domains in the human type VI collagen alpha 3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. *EMBO J* **9**(2), 385-93.

Chung, E., Rhodes, K. & Miller, E.J. (1976) Isolation of three collagenous components of probable basement membrane origin from several tissues. *Biochem Biophys Res Commun* **71**(4), 1167-74.

Colombatti, A., Ainger, K. & Colizzi, F. (1989) Type VI collagen: high yields of a molecule with multiple forms of alpha 3 chain from avian and human tissues. *Matrix* **9**(3), 177-85.

Colombatti, A., Bonaldo, P. & Doliana, R. (1993) Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins. *Matrix* **13**(4), 297-306.

Colombatti, A., Mucignat, M.T. & Bonaldo, P. (1995) Secretion and matrix assembly of recombinant type VI collagen. *J Biol Chem* **270**(22), 13105-11.

Copeland, N.G., Jenkins, N.A. & Court, D.L. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**(10), 769-79.

Cotterill, S.L., Jackson, G.C., Leighton, M.P., Wagener, R., Makitie, O., Cole, W.G. & Briggs, M.D. (2005) Multiple epiphyseal dysplasia mutations in MATN3

cause misfolding of the A-domain and prevent secretion of mutant matrilin-3. *Hum Mutat* **26**(6), 557-65.

Demir, E., Sabatelli, P., Allamand, V., Ferreiro, A., Moghadaszadeh, B., Makrelouf, M., Topaloglu, H., Echenne, B., Merlini, L. & Guicheney, P. (2002) Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. *Am J Hum Genet* **70**(6), 1446-58.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45.

Dumas, J.J., Kumar, R., McDonagh, T., Sullivan, F., Stahl, M.L., Somers, W.S. & Mosyak, L. (2004) Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibalpha complex reveals conformation differences with a complex bearing von Willebrand disease mutations. *J Biol Chem* **279**(22), 23327-34.

Dziadek, M., Kazenwadel, J.S., Hendrey, J.A., Pan, T.C., Zhang, R.Z. & Chu, M.L. (2002) Alternative splicing of transcripts for the alpha 3 chain of mouse collagen VI: identification of an abundant isoform lacking domains N7-N10 in mouse and human. *Matrix Biol* **21**(3), 227-41.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**(4), 953-71.

Finnis, M.L. & Gibson, M.A. (1997) Microfibril-associated glycoprotein-1 (MAGP-1) binds to the pepsin-resistant domain of the alpha3(VI) chain of type VI collagen. *J Biol Chem* **272**(36), 22817-23.

Fitzgerald, J., Morgelin, M., Selan, C., Wiberg, C., Keene, D.R., Lamande, S.R. & Bateman, J.F. (2001) The N-terminal N5 subdomain of the alpha 3(VI) chain is important for collagen VI microfibril formation. *J Biol Chem* **276**(1), 187-93.

Fitzgerald, J., Rich, C., Zhou, F.H. & Hansen, U. (2008) Three novel collagen VI chains, alpha4(VI), alpha5(VI), and alpha6(VI). *J Biol Chem* **283**(29), 20170-80.

Francomano, C.A., Cutting, G.R., McCormick, M.K., Chu, M.L., Timpl, R., Hong, H.K. & Antonarakis, S.E. (1991) The COL6A1 and COL6A2 genes exist as a gene cluster and detect highly informative DNA polymorphisms in the telomeric region of human chromosome 21q. *Hum Genet* **87**(2), 162-6.

Franzke, C.W., Tasanen, K., Schumann, H. & Bruckner-Tuderman, L. (2003) Collagenous transmembrane proteins: collagen XVII as a prototype. *Matrix Biol* **22**(4), 299-309.

Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. & Engel, J. (1983) Electron-microscopical approach to a structural model of intima collagen. *Biochem J* **211**(2), 303-11.

Furukawa, T. & Toyokura, Y. (1977) Congenital, hypotonic-sclerotic muscular dystrophy. *J Med Genet* **14**(6), 426-9.

Furuto, D.K. & Miller, E.J. (1980) Isolation of a unique collagenous fraction from limited pepsin digests of human placental tissue. Characterization of one of the constituent polypeptide chains. *J Biol Chem* **255**(1), 290-5.

Furuto, D.K. & Miller, E.J. (1981) Characterization of a unique collagenous fraction from limited pepsin digests of human placental tissue: molecular organization of the native aggregate. *Biochemistry* **20**(6), 1635-40.

Gailit, J. & Ruoslahti, E. (1988) Regulation of the fibronectin receptor affinity by divalent cations. *J Biol Chem* **263**(26), 12927-32.

Hatamochi, A., Aumailley, M., Mauch, C., Chu, M.L., Timpl, R. & Krieg, T. (1989) Regulation of collagen VI expression in fibroblasts. Effects of cell density, cell-matrix interactions, and chemical transformation. *J Biol Chem* **264**(6), 3494-9.

Heikkila, P. & Soininen, R. (1996) The type IV collagen gene family. *Contrib Nephrol* **117**, 105-29.

Heiskanen, M., Saitta, B., Palotie, A. & Chu, M.L. (1995) Head to tail organization of the human COL6A1 and COL6A2 genes by fiber-FISH. *Genomics* **29**(3), 801-3.

Higuchi, I., Shiraishi, T., Hashiguchi, T., Suehara, M., Niiyama, T., Nakagawa, M., Arimura, K., Maruyama, I. & Osame, M. (2001) Frameshift mutation in the collagen VI gene causes Ullrich's disease. *Ann Neurol* **50**(2), 261-5.

Hudson, B.G., Kalluri, R., Gunwar, S. & Noelken, M.E. (1994) Structure and organization of type IV collagen of renal glomerular basement membrane. *Contrib Nephrol* **107**, 163-7.

Huizinga, E.G., Martijn van der Plas, R., Kroon, J., Sixma, J.J. & Gros, P. (1997) Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. *Structure* **5**(9), 1147-56.

Illidge, C., Kielty, C. & Shuttleworth, A. (1998) The alpha1(VIII) and alpha2(VIII) chains of type VIII collagen can form stable homotrimeric molecules. *J Biol Chem* **273**(34), 22091-5.

Irwin, W.A., Bergamin, N., Sabatelli, P., Reggiani, C., Megighian, A., Merlini, L., Braghetta, P., Columbaro, M., Volpin, D., Bressan, G.M., Bernardi, P. & Bonaldo, P. (2003) Mitochondrial dysfunction and apoptosis in myopathic mice with collagen VI deficiency. *Nat Genet* **35**(4), 367-71.

Ishikawa, H., Sugie, K., Murayama, K., Ito, M., Minami, N., Nishino, I. & Nonaka, I. (2002) Ullrich disease: collagen VI deficiency: EM suggests a new basis for muscular weakness. *Neurology* **59**(6), 920-3.

Jander, R., Rauterberg, J. & Glanville, R.W. (1983) Further characterization of the three polypeptide chains of bovine and human short-chain collagen (intima collagen). *Eur J Biochem* **133**(1), 39-46.

Jander, R., Rauterberg, J., Voss, B. & von Bassewitz, D.B. (1981) A cysteine-rich collagenous protein from bovine placenta. Isolation of its constituent polypeptide chains and some properties of the non-denatured protein. *Eur J Biochem* **114**(1), 17-25.

Jobsis, G.J., Keizers, H., Vreijling, J.P., de Visser, M., Speer, M.C., Wolterman, R.A., Baas, F. & Bolhuis, P.A. (1996) Type VI collagen mutations in Bethlem myopathy, an autosomal dominant myopathy with contractures. *Nat Genet* **14**(1), 113-5.

Keene, D.R., Engvall, E. & Glanville, R.W. (1988) Ultrastructure of type VI collagen in human skin and cartilage suggests an anchoring function for this filamentous network. *J Cell Biol* **107**(5), 1995-2006.

Khoshnoodi, J., Pedchenko, V. & Hudson, B.G. (2008) Mammalian collagen IV. *Microsc Res Tech* **71**(5), 357-70.

Kielty, C.M., Boot-Handford, R.P., Ayad, S., Shuttleworth, C.A. & Grant, M.E. (1990) Molecular composition of type VI collagen. Evidence for chain heterogeneity in mammalian tissues and cultured cells. *Biochem J* **272**(3), 787-95.

Klein, G., Muller, C.A., Tillet, E., Chu, M.L. & Timpl, R. (1995) Collagen type VI in the human bone marrow microenvironment: a strong cytoadhesive component. *Blood* **86**(5), 1740-8.

Knupp, C. & Squire, J.M. (2001) A new twist in the collagen story--the type VI segmented supercoil. *EMBO J* **20**(3), 372-6.

Kuo, H.J., Maslen, C.L., Keene, D.R. & Glanville, R.W. (1997) Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. *J Biol Chem* **272**(42), 26522-9.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259), 680-5.

Lamande, S.R., Bateman, J.F., Hutchison, W., McKinlay Gardner, R.J., Bower, S.P., Byrne, E. & Dahl, H.H. (1998a) Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum Mol Genet* **7**(6), 981-9.

Lamande, S.R., Morgelin, M., Adams, N.E., Selan, C. & Allen, J.M. (2006) The C5 domain of the collagen VI alpha3(VI) chain is critical for extracellular microfibril formation and is present in the extracellular matrix of cultured cells. *J Biol Chem* **281**(24), 16607-14.

Lamande, S.R., Morgelin, M., Selan, C., Jobsis, G.J., Baas, F. & Bateman, J.F. (2002) Kinked collagen VI tetramers and reduced microfibril formation as a result of Bethlem myopathy and introduced triple helical glycine mutations. *J Biol Chem* **277**(3), 1949-56.

Lamande, S.R., Shields, K.A., Kornberg, A.J., Shield, L.K. & Bateman, J.F. (1999) Bethlem myopathy and engineered collagen VI triple helical deletions prevent intracellular multimer assembly and protein secretion. *J Biol Chem* **274**(31), 21817-22.

Lamande, S.R., Sigalas, E., Pan, T.C., Chu, M.L., Dziadek, M., Timpl, R. & Bateman, J.F. (1998b) The role of the alpha3(VI) chain in collagen VI assembly. Expression of an alpha3(VI) chain lacking N-terminal modules N10-N7 restores collagen VI assembly, secretion, and matrix deposition in an alpha3(VI)-deficient cell line. *J Biol Chem* **273**(13), 7423-30.

Lampe, A.K. & Bushby, K.M. (2005) Collagen VI related muscle disorders. *J Med Genet* **42**(9), 673-85.

Lampe, A.K., Dunn, D.M., von Niederhausern, A.C., Hamil, C., Aoyagi, A., Laval, S.H., Marie, S.K., Chu, M.L., Swoboda, K., Muntoni, F., Bonnemann, C.G., Flanigan, K.M., Bushby, K.M. & Weiss, R.B. (2005) Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J Med Genet* **42**(2), 108-20.

Lee, J.O., Rieu, P., Arnaout, M.A. & Liddington, R. (1995) Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* **80**(4), 631-8.

Loeser, R.F. (1997) Growth factor regulation of chondrocyte integrins. Differential effects of insulin-like growth factor 1 and transforming growth factor beta on

alpha 1 beta 1 integrin expression and chondrocyte adhesion to type VI collagen. *Arthritis Rheum* **40**(2), 270-6.

Lu, J., Teh, C., Kishore, U. & Reid, K.B. (2002) Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system. *Biochim Biophys Acta* **1572**(2-3), 387-400.

Lucioli, S., Giusti, B., Mercuri, E., Vanegas, O.C., Lucarini, L., Pietroni, V., Urtizberea, A., Ben Yaou, R., de Visser, M., van der Kooi, A.J., Bonnemann, C., Iannaccone, S.T., Merlini, L., Bushby, K., Muntoni, F., Bertini, E., Chu, M.L. & Pepe, G. (2005) Detection of common and private mutations in the COL6A1 gene of patients with Bethlem myopathy. *Neurology* **64**(11), 1931-7.

Marks, D.S., Gregory, C.A., Wallis, G.A., Brass, A., Kadler, K.E. & Boot-Handford, R.P. (1999) Metaphyseal chondrodysplasia type Schmid mutations are predicted to occur in two distinct three-dimensional clusters within type X collagen NC1 domains that retain the ability to trimerize. *J Biol Chem* **274**(6), 3632-41.

Mercuri, E., Lampe, A., Straub, V., Yuva, Y., Longman, C., Wright, M., Brown, S., Sewry, C., Bonnemann, C., Kinali, M., Brockington, M., Hausser, I., Hilton Jones, D., Voit, T., Bushby, K. & Muntoni, F. (2004) Congenital muscular dystrophy with short stature, proximal contractures and distal laxity. *Neuropediatrics* **35**(4), 224-9.

Mercuri, E., Yuva, Y., Brown, S.C., Brockington, M., Kinali, M., Jungbluth, H., Feng, L., Sewry, C.A. & Muntoni, F. (2002) Collagen VI involvement in Ullrich syndrome: a clinical, genetic, and immunohistochemical study. *Neurology* **58**(9), 1354-9.

Merlini, L., Morandi, L., Granata, C. & Ballestrazzi, A. (1994) Bethlem myopathy: early-onset benign autosomal dominant myopathy with contractures. Description of two new families. *Neuromuscul Disord* **4**(5-6), 503-11.

Meulenbelt, I., Chapman, K., Dieguez-Gonzalez, R., Shi, D., Tsezou, A., Dai, J., Malizos, K.N., Kloppenburg, M., Carr, A., Nakajima, M., van der Breggen, R., Lakenberg, N., Gomez-Reino, J.J., Jiang, Q., Ikegawa, S., Gonzalez, A., Loughlin, J. & Slagboom, E.P. (2009) Large replication study and meta-analyses of DVWA as an osteoarthritis susceptibility locus in European and Asian populations. *Hum Mol Genet* **18**(8), 1518-23.

Minamitani, T., Ariga, H. & Matsumoto, K. (2004a) Deficiency of tenascin-X causes a decrease in the level of expression of type VI collagen. *Exp Cell Res* **297**(1), 49-60.

Minamitani, T., Ikuta, T., Saito, Y., Takebe, G., Sato, M., Sawa, H., Nishimura, T., Nakamura, F., Takahashi, K., Ariga, H. & Matsumoto, K. (2004b) Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. *Exp Cell Res* **298**(1), 305-15.

Miyamoto, Y., Shi, D., Nakajima, M., Ozaki, K., Sudo, A., Kotani, A., Uchida, A., Tanaka, T., Fukui, N., Tsunoda, T., Takahashi, A., Nakamura, Y., Jiang, Q. & Ikegawa, S. (2008) Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat Genet* **40**(8), 994-8.

Muzny, D.M., Scherer, S.E., Kaul, R., Wang, J., Yu, J., Sudbrak, R., Buhay, C.J., Chen, R., Cree, A., Ding, Y., Dugan-Rocha, S., Gill, R., Gunaratne, P., Harris, R.A., Hawes, A.C., Hernandez, J., Hodgson, A.V., Hume, J., Jackson,

A., Khan, Z.M., Kovar-Smith, C., Lewis, L.R., Lozado, R.J., Metzker, M.L., Milosavljevic, A., Miner, G.R., Morgan, M.B., Nazareth, L.V., Scott, G., Sodergren, E., Song, X.Z., Steffen, D., Wei, S., Wheeler, D.A., Wright, M.W., Worley, K.C., Yuan, Y., Zhang, Z., Adams, C.Q., Ansari-Lari, M.A., Ayele, M., Brown, M.J., Chen, G., Chen, Z., Clendenning, J., Clerc-Blankenburg, K.P., Davis, C., Delgado, O., Dinh, H.H., Dong, W., Draper, H., Ernst, S., Fu, G., Gonzalez-Garay, M.L., Garcia, D.K., Gillett, W., Gu, J., Hao, B., Haugen, E., Havlak, P., He, X., Hennig, S., Hu, S., Huang, W., Jackson, L.R., Jacob, L.S., Kelly, S.H., Kube, M., Levy, R., Li, Z., Liu, B., Liu, J., Liu, W., Lu, J., Maheshwari, M., Nguyen, B.V., Okwuonu, G.O., Palmeiri, A., Pasternak, S., Perez, L.M., Phelps, K.A., Plopper, F.J., Qiang, B., Raymond, C., Rodriguez, R., Saenphimmachak, C., Santibanez, J., Shen, H., Shen, Y., Subramanian, S., Tabor, P.E., Verduzco, D., Waldron, L., Wang, Q., Williams, G.A., Wong, G.K., Yao, Z., Zhang, J., Zhang, X., Zhao, G., Zhou, J., Zhou, Y., Nelson, D., Lehrach, H., Reinhardt, R., Naylor, S.L., Yang, H., Olson, M., Weinstock, G. & Gibbs, R.A. (2006) The DNA sequence, annotation and analysis of human chromosome 3. *Nature* **440**(7088), 1194-8.

Myllyharju, J. & Kivirikko, K.I. (2001) Collagens and collagen-related diseases. *Ann Med* **33**(1), 7-21.

Myllyharju, J. & Kivirikko, K.I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* **20**(1), 33-43.

Nerlich, A.G., Schleicher, E.D., Wiest, I., Specks, U. & Timpl, R. (1994) Immunohistochemical localization of collagen VI in diabetic glomeruli. *Kidney Int* **45**(6), 1648-56.

Ober, C., Cox, N.J., Abney, M., Di Rienzo, A., Lander, E.S., Changyaleket, B., Gidley, H., Kurtz, B., Lee, J., Nance, M., Pettersson, A., Prescott, J., Richardson, A., Schlenker, E., Summerhill, E., Willadsen, S. & Parry, R. (1998) Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. *Hum Mol Genet* **7**(9), 1393-8.

Odermatt, E., Risteli, J., van Delden, V. & Timpl, R. (1983) Structural diversity and domain composition of a unique collagenous fragment (intima collagen) obtained from human placenta. *Biochem J* **211**(2), 295-302.

Oohashi, T., Sugimoto, M., Mattei, M.G. & Ninomiya, Y. (1994) Identification of a new collagen IV chain, alpha 6(IV), by cDNA isolation and assignment of the gene to chromosome Xq22, which is the same locus for COL4A5. *J Biol Chem* **269**(10), 7520-6.

Otten, C., Wagener, R., Paulsson, M. & Zaucke, F. (2005) Matrilin-3 mutations that cause chondrodysplasias interfere with protein trafficking while a mutation associated with hand osteoarthritis does not. *J Med Genet* **42**(10), 774-9.

Pan, T.C., Zhang, R.Z., Sudano, D.G., Marie, S.K., Bonnemann, C.G. & Chu, M.L. (2003) New molecular mechanism for Ullrich congenital muscular dystrophy: a heterozygous in-frame deletion in the COL6A1 gene causes a severe phenotype. *Am J Hum Genet* **73**(2), 355-69.

Pepe, G., Bertini, E., Bonaldo, P., Bushby, K., Giusti, B., de Visser, M., Guicheney, P., Lattanzi, G., Merlini, L., Muntoni, F., Nishino, I., Nonaka, I., Yaou, R.B.,

Sabatelli, P., Sewry, C., Topaloglu, H. & van der Kooi, A. (2002a) Bethlem myopathy (BETHLEM) and Ullrich scleroatonic muscular dystrophy: 100th ENMC international workshop, 23-24 November 2001, Naarden, The Netherlands. *Neuromuscul Disord* **12**(10), 984-93.

Pepe, G., Bertini, E., Giusti, B., Brunelli, T., Comeglio, P., Saitta, B., Merlini, L., Chu, M.L., Federici, G. & Abbate, R. (1999a) A novel de novo mutation in the triple helix of the COL6A3 gene in a two-generation Italian family affected by Bethlem myopathy. A diagnostic approach in the mutations' screening of type VI collagen. *Neuromuscul Disord* **9**(4), 264-71.

Pepe, G., de Visser, M., Bertini, E., Bushby, K., Vanegas, O.C., Chu, M.L., Lattanzi, G., Merlini, L., Muntoni, F. & Urtizberea, A. (2002b) Bethlem myopathy (BETHLEM) 86th ENMC international workshop, 10-11 November 2000, Naarden, The Netherlands. *Neuromuscul Disord* **12**(3), 296-305.

Pepe, G., Giusti, B., Bertini, E., Brunelli, T., Saitta, B., Comeglio, P., Bolognese, A., Merlini, L., Federici, G., Abbate, R. & Chu, M.L. (1999b) A heterozygous splice site mutation in COL6A1 leading to an in-frame deletion of the alpha1(VI) collagen chain in an italian family affected by bethlem myopathy. *Biochem Biophys Res Commun* **258**(3), 802-7.

Perkins, S.J., Smith, K.F., Williams, S.C., Haris, P.I., Chapman, D. & Sim, R.B. (1994) The secondary structure of the von Willebrand factor type A domain in factor B of human complement by Fourier transform infrared spectroscopy. Its occurrence in collagen types VI, VII, XII and XIV, the integrins and other proteins by averaged structure predictions. *J Mol Biol* **238**(1), 104-19.

Perona, J.J., Tsu, C.A., Craik, C.S. & Fletterick, R.J. (1993) Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J Mol Biol* **230**(3), 919-33.

Persikov, A.V., Ramshaw, J.A. & Brodsky, B. (2005) Prediction of collagen stability from amino acid sequence. *J Biol Chem* **280**(19), 19343-9.

Pfaff, M., Aumailley, M., Specks, U., Knolle, J., Zerwes, H.G. & Timpl, R. (1993) Integrin and Arg-Gly-Asp dependence of cell adhesion to the native and unfolded triple helix of collagen type VI. *Exp Cell Res* **206**(1), 167-76.

Pullig, O., Weseloh, G. & Swoboda, B. (1999) Expression of type VI collagen in normal and osteoarthritic human cartilage. *Osteoarthritis Cartilage* **7**(2), 191-202.

Ramachandran, G.N. & Kartha, G. (1954) Structure of collagen. *Nature* **174**(4423), 269-70.

Sadler, J.E. (1998) Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem* **67**, 395-424.

Saitta, B., Wang, Y.M., Renkart, L., Zhang, R.Z., Pan, T.C., Timpl, R. & Chu, M.L. (1991) The exon organization of the triple-helical coding regions of the human alpha 1(VI) and alpha 2(VI) collagen genes is highly similar. *Genomics* **11**(1), 145-53.

Scacheri, P.C., Gillanders, E.M., Subramony, S.H., Vedanarayanan, V., Crowe, C.A., Thakore, N., Bingler, M. & Hoffman, E.P. (2002) Novel mutations in collagen VI genes: expansion of the Bethlem myopathy phenotype. *Neurology* **58**(4), 593-602.

Shoulders, M.D. & Raines, R.T. (2009) Collagen structure and stability. *Annu Rev Biochem* **78**, 929-58.

Söderhall, C., Marenholz, I., Kerscher, T., Ruschendorf, F., Esparza-Gordillo, J., Worm, M., Gruber, C., Mayr, G., Albrecht, M., Rohde, K., Schulz, H., Wahn, U., Hubner, N. & Lee, Y.A. (2007) Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol* **5**(9), e242.

Specks, U., Mayer, U., Nischt, R., Spissinger, T., Mann, K., Timpl, R., Engel, J. & Chu, M.L. (1992) Structure of recombinant N-terminal globule of type VI collagen alpha 3 chain and its binding to heparin and hyaluronan. *EMBO J* **11**(12), 4281-90.

Stedman, H.H., Sweeney, H.L., Shrager, J.B., Maguire, H.C., Panettieri, R.A., Petrof, B., Narusawa, M., Leferovich, J.M., Sladky, J.T. & Kelly, A.M. (1991) The mdx mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy. *Nature* **352**(6335), 536-9.

Sugimoto, M., Oohashi, T. & Ninomiya, Y. (1994) The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. *Proc Natl Acad Sci U S A* **91**(24), 11679-83.

Swoboda, B., Pullig, O., Kirsch, T., Kladny, B., Steinhauser, B. & Weseloh, G. (1998) Increased content of type-VI collagen epitopes in human osteoarthritic cartilage: quantitation by inhibition ELISA. *J Orthop Res* **16**(1), 96-9.

Swoboda, B., Pullig, O., Kladny, B., Pfander, D. & Weseloh, G. (1999) [Collagen type VI content in healthy and arthritis knee joint cartilage]. *Z Orthop Ihre Grenzgeb* **137**(6), 540-4.

Tanaka, T., Ikari, K., Furushima, K., Okada, A., Tanaka, H., Furukawa, K., Yoshida, K., Ikeda, T., Ikegawa, S., Hunt, S.C., Takeda, J., Toh, S., Harata, S., Nakajima, T. & Inoue, I. (2003) Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. *Am J Hum Genet* **73**(4), 812-22.

Tillet, E., Gential, B., Garrone, R. & Stallcup, W.B. (2002) NG2 proteoglycan mediates beta1 integrin-independent cell adhesion and spreading on collagen VI. *J Cell Biochem* **86**(4), 726-36.

Tillet, E., Wiedemann, H., Golbik, R., Pan, T.C., Zhang, R.Z., Mann, K., Chu, M.L. & Timpl, R. (1994) Recombinant expression and structural and binding properties of alpha 1(VI) and alpha 2(VI) chains of human collagen type VI. *Eur J Biochem* **221**(1), 177-85.

Tsukahara, S., Miyazawa, N., Akagawa, H., Forejtova, S., Pavelka, K., Tanaka, T., Toh, S., Tajima, A., Akiyama, I. & Inoue, I. (2005) COL6A1, the candidate gene for ossification of the posterior longitudinal ligament, is associated with diffuse idiopathic skeletal hyperostosis in Japanese. *Spine (Phila Pa 1976)* **30**(20), 2321-4.

Valdes, A.M., Spector, T.D., Doherty, S., Wheeler, M., Hart, D.J. & Doherty, M. (2009) Association of the DVWA and GDF5 polymorphisms with osteoarthritis in UK populations. *Ann Rheum Dis* **68**(12), 1916-20.

Vanegas, O.C., Zhang, R.Z., Sabatelli, P., Lattanzi, G., Bencivenga, P., Giusti, B., Columbaro, M., Chu, M.L., Merlini, L. & Pepe, G. (2002) Novel COL6A1 splicing mutation in a family affected by mild Bethlem myopathy. *Muscle Nerve* **25**(4), 513-9.

Veit, G., Kobbe, B., Keene, D.R., Paulsson, M., Koch, M. & Wagener, R. (2006) Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *J Biol Chem* **281**(6), 3494-504.

Voit, T. (1998) Congenital muscular dystrophies: 1997 update. *Brain Dev* **20**(2), 65-74.

Weil, D., Mattei, M.G., Passage, E., N'Guyen, V.C., Pribula-Conway, D., Mann, K., Deutzmann, R., Timpl, R. & Chu, M.L. (1988) Cloning and chromosomal localization of human genes encoding the three chains of type VI collagen. *Am J Hum Genet* **42**(3), 435-45.

Whittaker, C.A. & Hynes, R.O. (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* **13**(10), 3369-87.

Wiberg, C., Hedbom, E., Khairullina, A., Lamande, S.R., Oldberg, A., Timpl, R., Morgelin, M. & Heinegard, D. (2001) Biglycan and decorin bind close to the n-terminal region of the collagen VI triple helix. *J Biol Chem* **276**(22), 18947-52.

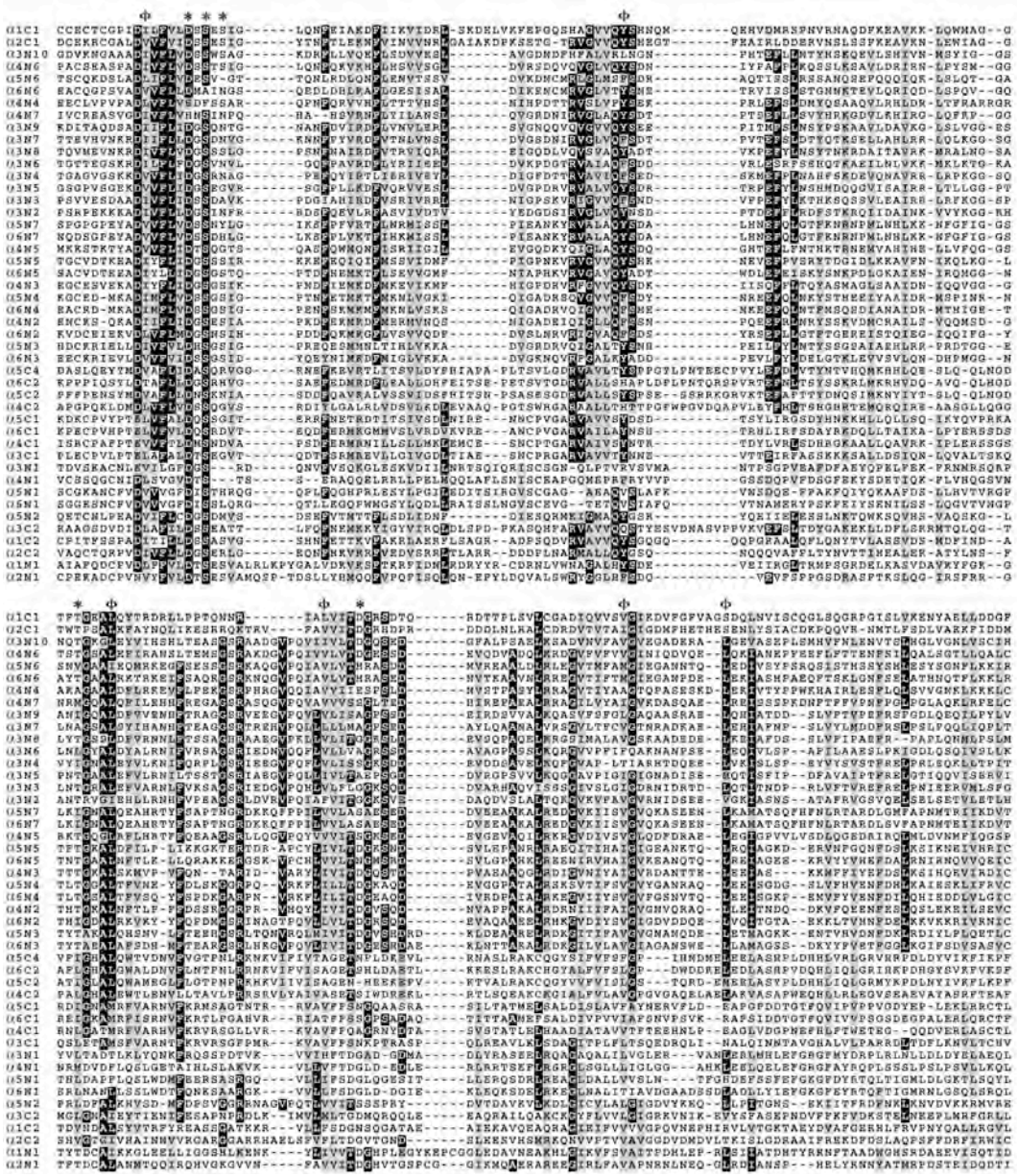
Wiberg, C., Klatt, A.R., Wagener, R., Paulsson, M., Bateman, J.F., Heinegard, D. & Morgelin, M. (2003) Complexes of matrilin-1 and biglycan or decorin connect collagen VI microfibrils to both collagen II and aggrecan. *J Biol Chem* **278**(39), 37698-704.

Zhang, R.Z., Sabatelli, P., Pan, T.C., Squarzoni, S., Mattioli, E., Bertini, E., Pepe, G. & Chu, M.L. (2002) Effects on collagen VI mRNA stability and microfibrillar assembly of three COL6A2 mutations in two families with Ullrich congenital muscular dystrophy. *J Biol Chem* **277**(46), 43557-64.

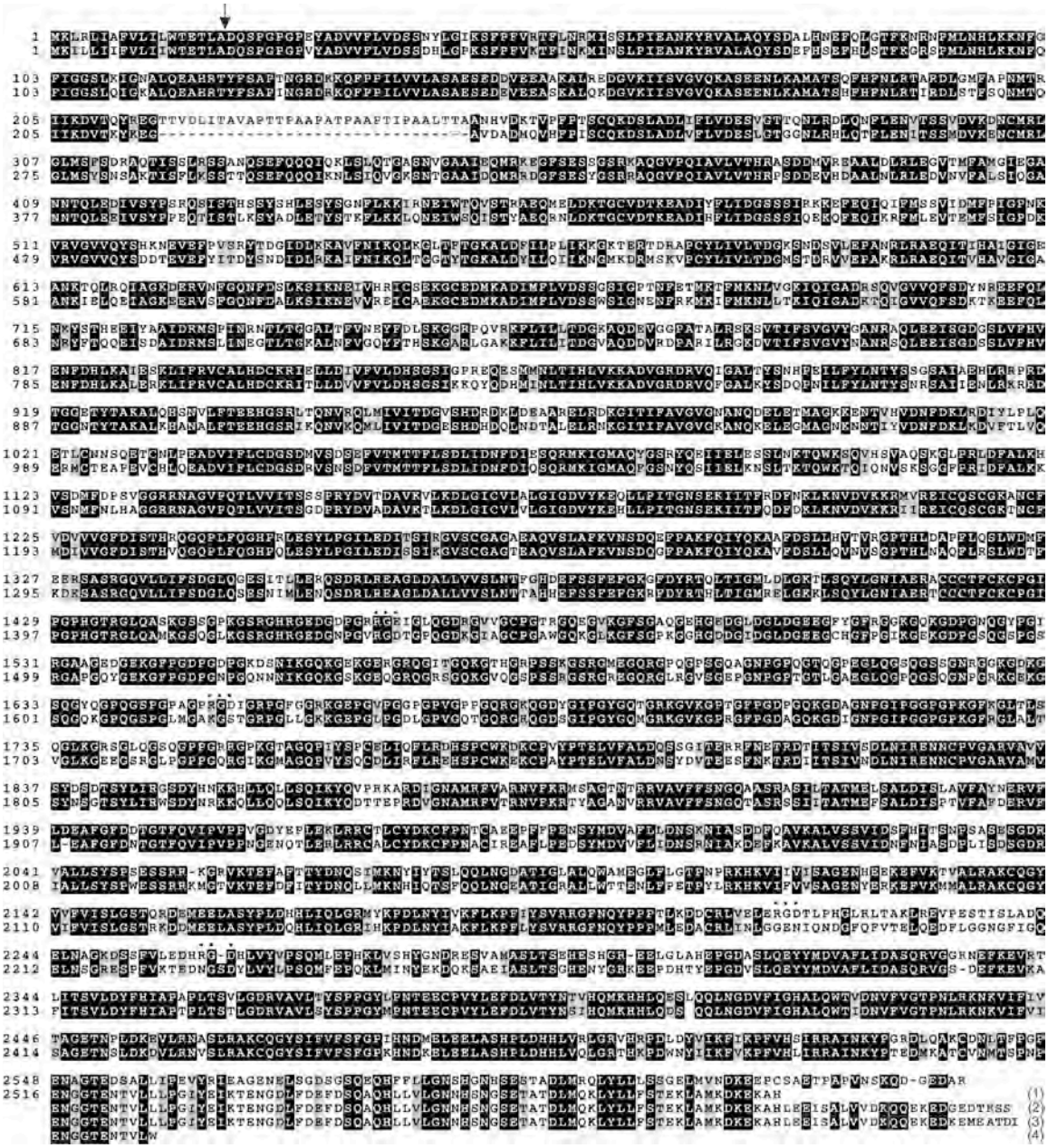
Zhou, J., Mochizuki, T., Smeets, H., Antignac, C., Laurila, P., de Paepe, A., Tryggvason, K. & Reeders, S.T. (1993) Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors. *Science* **261**(5125), 1167-9.

Zhou, J. & Reeders, S.T. (1996) The alpha chains of type IV collagen. *Contrib Nephrol* **117**, 80-104.

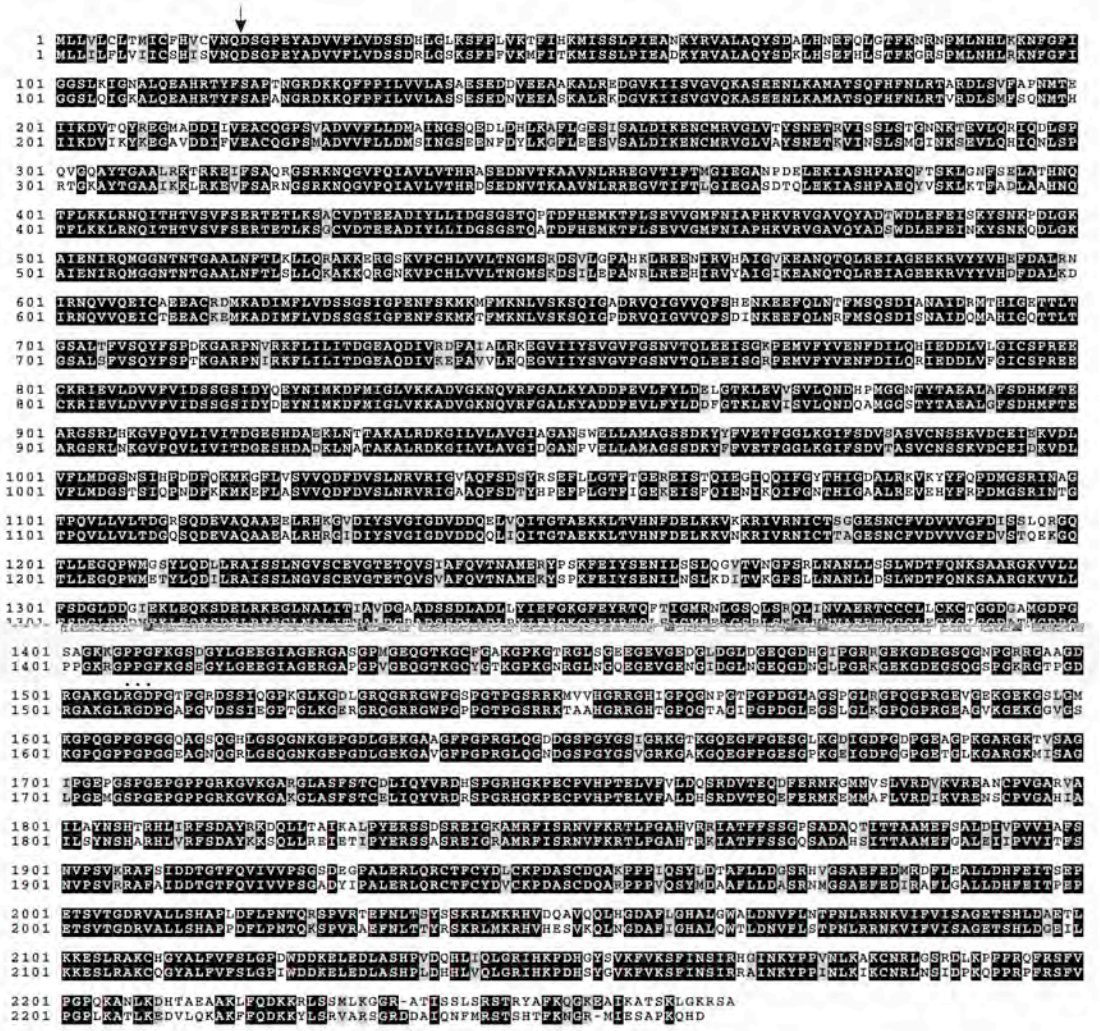
APPENDIX



Appendix Fig. 1 **Amino acid alignment of murine collagen VI VWA domains**. The sequences were aligned by CLUSTAL X using the default parameters. The conserved metal ion dependent adhesion site and the conserved hydrophobic moieties are denoted with (*) and (Φ), respectively.



Appendix Fig. 2 **Alignment of amino acid sequences of human and murine collagen VI α5 chains.** The mouse sequence is shown in the upper lines. The different C-termini of the human collagen VI α5 chain are given in the last four lines. Sequence (1) contains a premature stop codon due to SNP rs1497312. Sequence (2) is based on the AM906083 sequence. Sequence (3) represents EU085556 and sequence (4) is based on the human lung cDNA clone AK123718. The *arrow* marks the potential signal peptide cleavage sites and the *dots* mark the RGD sequence.



Appendix Fig. 3 **Alignment of amino acid sequences of human and murine collagen VI α6 chains**. The mouse sequence is shown in the upper lines. The *arrow* marks the potential signal peptide cleavage sites and the *dots* mark the RGD sequence

.....CREA SVGDIVFLVH NSINPQHAHS

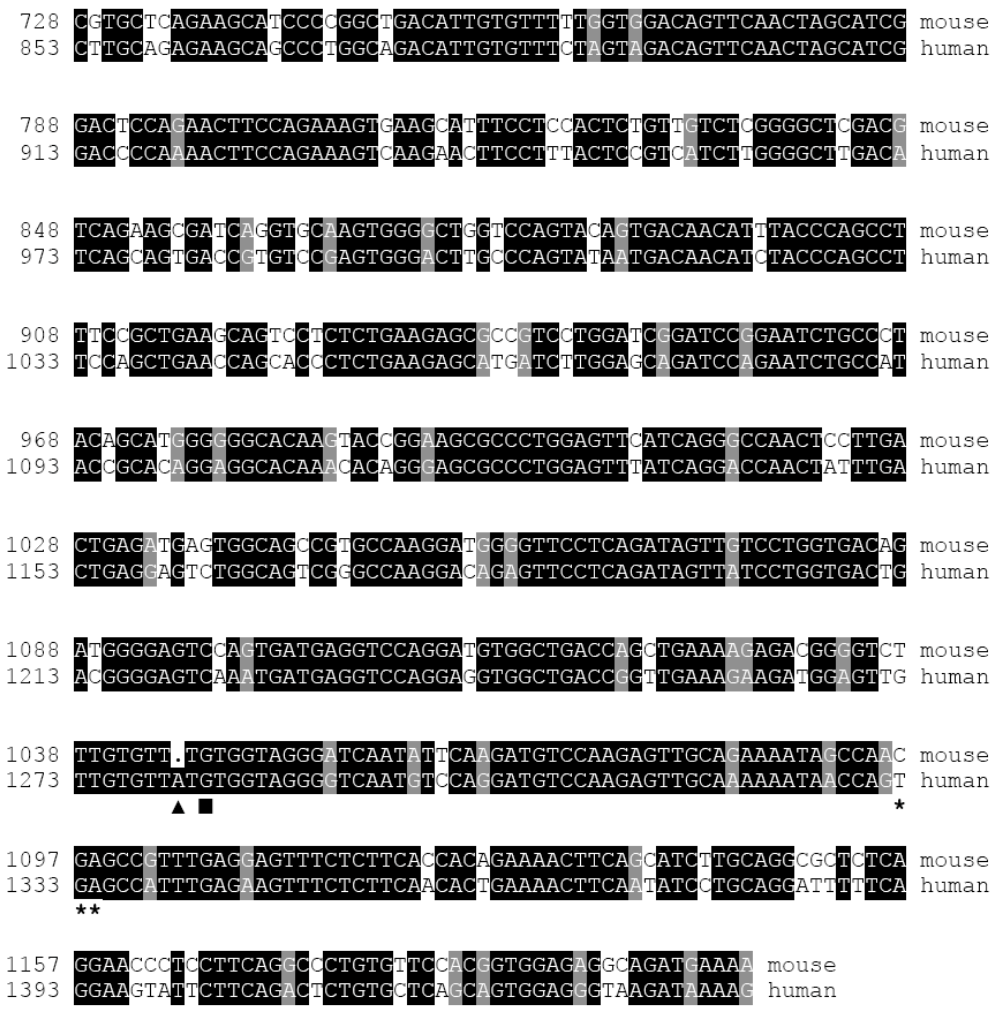
VRNFLYILAN SLQVGRDNIR VGLAQYSDTP TSEFLLSVYH RKGDVLKHIR

GLQFKPGGNR **MGQALQFILE HHFREGAGSR ASQGVPQVAV VVSSGLTEDH**

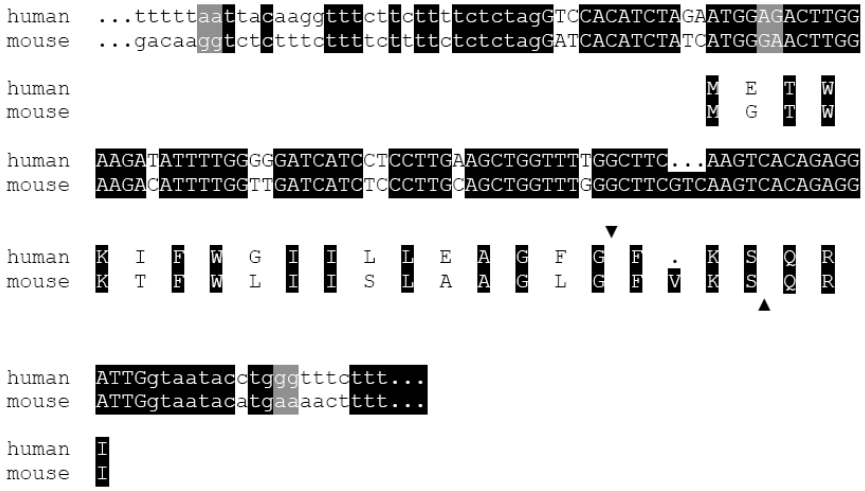
IREPAEALRR AGILVYAIGV KDASQAELRE ISSSPKDNFT FFVPNFPGLP

GLAQKLRPEL CSTLGKAAQY TERESP....

Appendix Fig. 4 **Open reading frame of the second exon of the human DVWA gene**. The proposed amino acid sequence of DVWA is given in bold face.



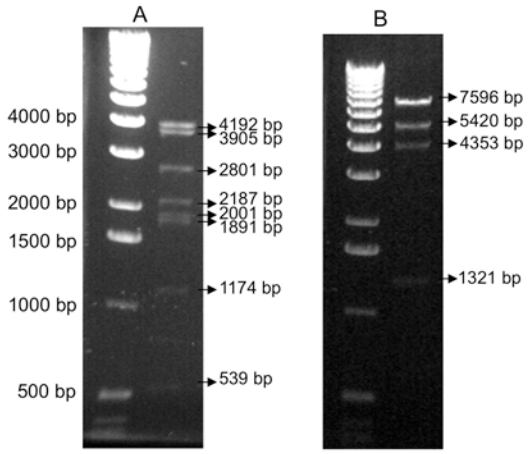
Appendix Fig. 5 **Alignment of DNA sequences of the third exon of the human DVWA gene and the corresponding fourth exon of the murine *Col6a4* gene**. The inserted adenine is marked by an arrowhead. A filled square indicates the position of the nucleotide that is involved in the cysteine to tyrosine polymorphism. The premature stop codon of DVWA is marked by asterisks. The numbers are according to accession numbers AM231151 (mouse) and AB299979 (human).



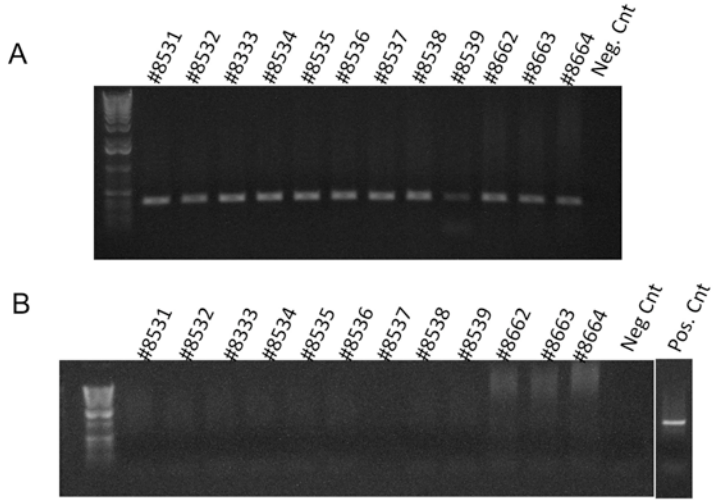
Appendix Fig. 6 Alignment of DNA and protein sequences of the proposed exon of the human DVWA gene coding for the signal peptide sequence together with neighboring intron sequences and the respective second exon of the murine *Col6a4* gene. Arrowheads indicate the signal peptide cleavage sites. The signal peptide sequences were predicted by a method using neural networks or hidden Markov models. It should be noted that the position of the cleavage site is different between man and mouse and that the human sequence had a lower prediction score. This could be the consequence of a three base pair deletion in the vicinity of the proposed cleavage site.

```
CACATCTAGAATGGAGACTTGGAAGATATTTTGGGGGATCATCCTCCTTGAAGCTGGTTTTGGCTTCAAGTCACAGAGGATTGTCTGCAGGGAAGCCTCT  100
        M  E  T  W  K  I  F  W  G  I  I  L  L  E  A  G  F  G  F  K  S  Q  R  I  V  C  R  E  A  S    30

CTGGGAGATGTTGTGTTTCTGGTGGACACCAACATCAACCCCCAACATGCACGCAGTGTGCGGAACTTCTTGTACATTCTGGTAAACAGCTTCAACGTCA  200
L  G  D  V  V  F  L  V  D  T  N  I  N  P  Q  H  A  R  S  V  R  N  F  L  Y  I  L  V  N  S  F  N  V  S   64

GCAGTGAGACCATCCGTGTGGGTCTGGCCAAGTATAGCGATGTGCCCCATTCAGAGTTCTTGCTTTCCACCTACCACCGCAAAGGTGATGTGTTGAGACA  300
  S  E  T  I  R  V  G  L  A  K  Y  S  D  V  P  H  S  E  F  L  L  S  T  Y  H  R  K  G  D  V  L  R  H   97

CATTCGGCAGTTTCAATTTAAGCCTGGGGGCAAGAAGATGGGCCTGGCCCTGAAGTTCATTCTAGACCACCACTTCCAGGAAGCATCAGGGAGCCGGGCA  400
 I  R  Q  F  Q  F  K  P  G  G  K  K  M  G  L  A  L  K  F  I  L  D  H  H  F  Q  E  A  S  G  S  R  A   130

AGCCAAGAGGTACCTCAGATCGCCGTGGTGATCAGCAGTGGGCCAGTTGAGGACCATGTCCATGGACCTGCGAAGGCATTGAGGAAGGCAGGCATCCTGC  500
S  Q  E  V  P  Q  I  A  V  V  I  S  S  G  P  V  E  D  H  V  H  G  P  A  K  A  L  R  K  A  G  I  L  L  164

TTTATGCTATTGGCGTCAGAGATGCAGTTTGGGCAGAGCTCAGGGAAATTGCCAGTAGTCCTCAGGAGAATTTTACCTCCTTTGTTCCTAACTTCTCTGG  600
  Y  A  I  G  V  R  D  A  V  W  A  E  L  R  E  I  A  S  S  P  Q  E  N  F  T  S  F  V  P  N  F  S  G   197

TCTGAGCAATCTCGCCCAGAAGCTGCGGCAGGAACTCTGTGATACGTTGGCAAAGGCAGCTCCACATGTTGACCACGTCTCTCCAGCTTGCAGAGAAGCA  700
 L  S  N  L  A  Q  K  L  R  Q  E  L  C  D  T  L  A  K  A  A  P  H  V  D  H  V  S  P  A  C  R  E  A   230

GCCCTGGCAGACATTGTGTTTCTAGTAGACAGTTCAACTAGCATCGGACCCCAAAACTTCCAGAAAGTCAAGAACTTCCTTTACTCCGTCATCTTGGGGC  800
A  L  A  D  I  V  F  L  V  D  S  S  T  S  I  G  P  Q  N  F  Q  K  V  K  N  F  L  Y  S  V  I  L  G  L  264

TTGACATCAGCAGTGACCGTGTCCGAGTGGGACTTGCCCAGTATAATGACAACATCTACCCAGCCTTCCAGCTGAACCAGCACCCTCTGAAGAGCATGAT  900
  D  I  S  S  D  R  V  R  V  G  L  A  Q  Y  N  D  N  I  Y  P  A  F  Q  L  N  Q  H  P  L  K  S  M  I   297

CTTGGAGCAGATCCAGAATCTGCCATACCGCACAGGAGGCACAAACACAGGGAGCGCCCTGGAGTTTATCAGGACCAACTATTTGACTGAGGAGTCTGGC 1000
 L  E  Q  I  Q  N  L  P  Y  R  T  G  G  T  N  T  G  S  A  L  E  F  I  R  T  N  Y  L  T  E  E  S  G   330

AGTCGGGCCAAGGACAGAGTTCCTCAGATAGTTATCCTGGTGACTGACGGGGAGTCAAATGATGAGGTCCAGGAGGTGGCTGACCGGTTGAAAGAAGATG 1100
S  R  A  K  D  R  V  P  Q  I  V  I  L  V  T  D  G  E  S  N  D  E  V  Q  E  V  A  D  R  L  K  E  D  G  364

GAGTTGTTGTGTTAGTGGTAGGGGTCAATGTCCAGGATGTCCAAGAGTTGCAAAAAATAACCAGTGAGCCATTTGAGAAGTTTCTCTTCAACACTGAAA 1200
  V  V  V  L  C  G  R  G  Q  C  P  G  C  P  R  V  A  K  N  N  Q  *                                    385

ACTTCAATATCCTGCAGGATTTTTCAGGAAGTATTCTTCAGACTCTGTGCTCAGCAGTGGAGGGTAAGA                                1269
```

Appendix Fig. 7 **Sequence of the human collagen VI α4 chain cDNA**. The amino acid sequence is given below. The inserted adenine (nt 1114) that leads to the frameshift is shaded in black. The sequence was deposited in the database under accession number FN394065.



Appendix Fig. 8 **Restriction enzyme digest of the targeting vector with NheI (A) and EcoRI (B)**



Appendix Fig. 9 **Screening the progeny of chimeras**. A) Control for the quality of template DNA with COMP primers. B) Screening with the 5' homology arm primers (*see Table 4.3 in Materials and methods*). Neg. Cnt, negative control and Pos. Cnt, positive control.

Exon		Size[a]	Splice donor	Intron	Size	Splice acceptor	Codon phase	Amino acid interrupted
1	5' UTR		TAGGGgtaggtg	1	13701	ttctctctagGAT		
2	5' UTR/SigPep	89	GATTGgtaatac	2	2319	tcttttgaagTCT	1	Val
3	N7	603	TCCAGgtaacg	3	2485	ttctttacagCGT	1	Ala
4	N6	588	GAAAAgtaagc	4	1441	ttacttttagAGT	1	Lys
5	N5	609	AGAAGgtaagt	5	2263	tggttttcagAAT	1	Glu
6	N4	645	GGAAGgtagga	6	3521	cattggccagGCT	1	Gly
7	N3	546	AGAGAgtaaga	7	640	cgcccccagACT	1	Asn
8	N2	555	CCAAGgtgagt	8	3544	attttcaaagGCT	1	Gly
9	N1	335	TGAAGgtactg	9	1212	tgttccccagGTC	0	Lys/Val
10	N1	79	CAGAGgtaagc	10	1012	cattttgaagGGC	1	Gly
11	N1	146	AACTCgtgagt	11	190	tgtgtttcagGAC	0	Leu/Asp
12	col1	93	AGCAGgtaaga	12	2606	gttttctcagGGA	0	Gln/Gly
13	col2	54	ACCACgtgagt	13	5676	ttcttcacagGGA	0	His/Gly
14	col3	72	CTAAGgtaagt	14	1774	ataatttcagGGA	0	Lys/Gly
15	col4	27	AGAAGgtagga	15	2072	cctttccgcagGGC	0	Lys/Gly
16	col5	45	AACAGgtatgg	16	3436	cttttcctagGGC	0	Gln/Gly
17	col6	54	GTCGGgtaaaa	17	3070	gtaatttaagGGC	0	Arg/Gly
18	col7	63	ATCCGgtgagt	18	454	gttttctcagGGG	0	Pro/Gly
19	col8	66	ATCAGgtaaga	19	6589	tttatttcagGGA	0	Gln/Gly
20	col9	54	CTCGGgtaggt	20	137	gtgctcttagGGG	0	Arg/Gly
21	col10	36	TGAAGgtcagt	21	3726	acttctgcagGGT	0	Lys/Gly
22	col11	63	CACAGgttgta	22	1714	tttggaacagGGA	0	Gln/Gly
23	col12	63	CCCAGgtaggt	23	1671	atcattgcagGGG	0	Gln/Gly
24	col13	63	TGAAG**gcgagt**	24	902	**ctattaaaag**GGA	0	Lys/Gly
25	col14	63	AGCCTgtaagt	25	2566	atttttttagGGT	0	Pro/Gly
26	col15	51	GGAAGgtgagc	26	2072	ctgaccctagGGT	0	Lys/Gly
27	col16	36	TGCAAgtaaga	27	101	ccttgttaagGGA	0	Gln/Gly
28	col17	63	AGAGGgtaaga	28	1135	ttcattatagGGC	0	Arg/Gly
29	col18	63	AAATGgtaatg	29	1514	ttctttccagGGC	0	Met/Gly
30	col19	36	GGACGgtaggt	30	1321	ctctctacagCCT	0	Thr/Pro
31	spacer	37	CTGCCgtgagt	31	2602	tcccttccagCTT	1	Pro
32	spacer	12	AACAGgtaaat	32	84	ccgtgtcaagGTA	1	Gly
33	C1	494	GGCTGgtatgt	33	10034	gttgaactagGTG	0	Leu/Val
34	C1	91	CTATGgtaaga	34	2283	ttcttgccagACA	1	Asp
35	C2	684	AAATAgtgagt	35	2,990	ttactttcagGTG	1	Ser
36	C3	97	AAGAGgtgagc	36	992	tttcctgcagGTT	2	Arg
37	C3	212	CCACGgtgagc	37	5086	ttgcttgcagGTC	1	Gly
38	C4	53						

Appendix Tab. 1 **Splice junction sites in the murine *Col6a4* gene.** Exon sequences are in upper case and intron sequences are in lower case letters. The GC-AG intron is in bold. [a]The lengths of the 5' and 3' UTRs are not included.

Exon		Size[a]	Splice donor	Intron	Size	Splice acceptor	Codon phase	Amino acid interrupted
1	5' UTR		TCTTGgtaagt	1	12125	tccttttcagAAT		
2	5' UTR/SigPep	67	CCCAGgtacca	2	2236	tattgtgcagGGC	1	Gly
3	N7	696	GACAGgtagga	3	5046	gtctccacagTTC	1	Val
4	N6	633	AACCGgtatgt	4	2299	tccttttagGCT	1	Gly
5	N5	561	AAAAGgtaaga	5	2493	tcccaaacagGAT	1	Gly
6	N4	556	GCATGgtaagt	6	2471	ttcttttcagATT	1	Asp
7	N3	576	AGAGAgtaagc	7	2142	tctgttccagCTT	1	Thr
8	N2	576	AGCCAgtgagt	8	1941	tggtctccagATT	1	Asn
9	N1	341	GCCAGgtatac	9	99	ctttctgtagGTT	0	Gln/Val
10	N1	82	AGCAGgtaatg	10	2241	tattaactagGAC	1	Gly
11	N1	143	ACCTGgtgagt	11	2186	ctctctctagGGA	0	Leu/Gly
12	col1	93	CAAAGgtgctg	12	2676	ttcctcctagGGT	0	Lys/Gly
13	col2	54	ACCCTgtaagt	13	2473	cacttcttagGGA	0	Pro/Gly
14	col3	72	AAAAGgtaagc	14	301	tcttttgcagGGA	0	Glu/Gly
15	col4	27	CTCAGgtattg	15	129	catcccttagGGG	0	Gln/Gly
16	col5	45	AAGAAgtgagc	16	91	gtttgtttagGGA	0	Glu/Gly
17	col6	54	ATCAGgtaact	17	2476	tctttttcagGGT	0	Gln/Gly
18	col7	63	ATCCCgtaagt	18	1293	tatctcctagGGT	0	Pro/Gly
19	col8	66	GACAAgtaatt	19	396	gtggttctagGGG	0	Gln/Gly
20	col9	54	GCAGGgtaagt	20	3484	ggtgtttcagGGA	0	Arg/Gly
21	col10	36	CCTCAgtaagt	21	1000	tgttcattagGGG	0	Ser/Gly
22	col11	62	CACAGgtatgc	22	1496	caacaaaaagGGA	0	Gln/Gly
23	col12	63	CTCAGgtgagt	23	950	ttattttcagGGT	0	Gln/Gly
24	col13	63	GAAAGgtaaac	24	1814	ctctgtctagGGG	0	Lys/Gly
25	col14	63	AGCAGgtatga	25	2705	gtacttgcagGGA	0	Gln/Gly
26	col15	51	TTAAGgtaaat	26	101	aaaatttcagGGC	0	Lys/Gly
27	col16	36	AGAAGgtgagg	27	1679	atgtgatcagGGT	0	Lys/Gly
28	col17	63	TGACTgtaagt	28	1675	ttcttttcagCTC	0	Thr/Leu
29	col18	63	GGAGAgtgagt	29	2937	ccccccttagGGT	0	Arg/Gly
30	col19	36	ATTCCgtatgt	30	105	aactttccagCCC	0	Ser/Pro
31	spacer	31	CAGCCgtgagt	31	1749	ctctccccagCCT	1	Pro
32	spacer	12	GAAAGgtgagt	32	990	attttcacagACA	1	Asp
33	C1	488	TTGGGgtaaga	33	6200	ttcactgcagTTT	0	Gly/Phe
34	C1	90	CTATGgtaaga	34	518	gcatttttagACA	1	Asp
35	C2	657	CAGACgtaagt	35	2782	ccctccgtagGCG	1	Arg
36	C3	91	CTCAGgtgggt	36	2552	cttcatctagATT	2	Arg
37	C3	209	TTCTGgtaagc	37	10953	tctgttgcagAAC	1	Glu
38	C4	690	CAGACgtaagt	38	1300	ctgtctccagGTG	1	Arg
39	C5	97	GCATTgtgagt	39	686	ttcatcccagACT	2	Leu
40	C5	233	ACAAGgtagtc	40	4492	tgctttacagATG	1	Asp
41	spacer/3' UTR	17						

Appendix Tab. 2 **Splice junction sites in the murine *Col6a5* gene.** Exon sequences are in upper case and intron sequences are in lower case letters. [a]The lengths of the 5' and 3' UTRs are not included.

Exon		Size[a]	Splice donor	Intron	Size	Splice acceptor	Codon phase	Amino acid interrupted
1	5' UTR		GCACGgtaagt	1		tgcccttcagCTT		
2	5' UTR/SigPep	61	TTCTGgtaaga	2	2618	tgattctatagGCC	1	Gly
3	N7	598	AGAAGgtgggt	3	1428	ccccacacagCCT	1	Ala
4	N6	621	ATCTGgtaaag	4	1164	tccatttcagCCT	1	Ala
5	N5	561	AGAAGgtagga	5	733	tcccacacagCCT	1	Ala
6	N4	558	TGAAGgtgagc	6	2491	tccttttcagAAT	1	Glu
7	N3	576	AGTTGgtaagg	7	2710	ctctctgtagATT	1	Asp
8	N2	573	GAGCAgttaag	8	6720	tggatcctagATT	1	Asn
9	N1	344	GGAAGgtaatg	9	98	cctaccctagGTG	0	Lys/Val
10	N1	79	GGAAGgtactg	10	4110	tgttgaacagGCC	1	Gly
11	N1	155	AACTAgtaagt	11	1242	ttccccgaagATC	0	Leu/Ile
12	col1	93	AAAAGgtgact	12	1319	ttttttcagGGA	0	Lys/Gly
13	col2	54	TTGCTgtaagt	13	1687	ttttcctcagGGA	0	Ala/Gly
14	col3	72	CTAAGgtaagg	14	102	ccattttcagGGA	0	Lys/Gly
15	col4	27	AAGAGgtaatg	15	2467	gtctctgcagGGT	0	Glu/Gly
16	col5	45	AACAGgtacaa	16	91	cattttctagGGT	0	Gln/Gly
17	col6	54	CTCAGgtaaag	17	1270	ttaactttagGGA	0	Gln/Gly
18	col7	63	ATCCGgtaagt	18	5250	ctctttttagGGA	0	Pro/Gly
19	col8	66	GACAAgtaatt	19	1543	cctctggcagGGC	0	Gln/Gly
20	col9	54	AGATGgtaaga	20	8036	tttgttttagGTA	0	Met/Val
21	col10	36	CACAGgtcagt	21	926	ctcatcacagGGA	0	Gln/Gly
22	col11	63	CGCAGgtatcc	22	1788	atgcatttagGGT	0	Gln/Gly
23	col12	63	CTCAGgtacgt	23	4072	ttcattccagGGG	0	Gln/Gly
24	col13	63	ACAAAgtaggt	24	3903	tgttggacagGGA	0	Lys/Gly
25	col14	63	TGCAGgtgggt	25	1024	ttgcttgcagGGT	0	Gln/Gly
26	col15	51	CAAAGgtaagt	26	2811	ttgatggcagGGG	0	Lys/Gly
27	col16	36	TGAAGgtacgt	27	1406	attcatttagGGT	0	Lys/Gly
28	col17	63	AAACGgtaagc	28	6119	ctgctttgagGTA	0	Thr/Val
29	col18	63	GGAAGgtaagt	29	1316	cctttcctagGGT	0	Lys/Gly
30	col19	36	TTTCCgtatgt	30	110	tgaactgcagACA	0	Ser/Thr
31	spacer	37	CAGTCgtaagt	31	5590	tcctccccagCTG	1	Pro
32	spacer	12	ACATGgtgagt	32	2912	tgtctttcagGAA	1	Gly
33	C1	494	TTTCAgtaagc	33	6796	cttcttccagATT	0	Ser/Ile
34	C1	97	CTATGgtgaga	34	2865	tggtttgcagATC	1	Asp
35	C2	672	CAGGCgtaagt	35	2854	ctctttctagATG	1	His
36	C3	94	CGAAGgtatgt	36	6011	tgttgtgcagCTT	2	Ser
37	C3/3' UTR	202						

Appendix Tab. 3 **Splice junction sites in the murine *Col6a6* gene.** Exon sequences are in upper case and intron sequences are in lower case letters. [a]The lengths of the 5' and 3' UTRs are not included.

ACKNOWLEDGEMENTS

I wish to express my deep sense of gratitude and heart-felt thanks to Prof. Dr. Mats Paulsson for giving me an opportunity to work in his laboratory. His expertise guidance, valuable suggestions, innovative ideas and constant support during the course of my work and in particular while editing my dissertation are highly memorable.

I owe my deepest heart-felt gratitude to PD. Dr. Raimund Wagener for designing and offering me a very interesting and promising project. His immense patience, sustained interest and constant encouragement assisted me to remain motivated throughout the project. I am highly grateful for his constructive criticism in editing the dissertation. I am also very grateful to him for taking time to help me outside the curricular activities.

My sincere thanks to Prof. Manuel Koch for his technical guidance and moral support during my work.

I take this opportunity to thank our collaborators Paolo Grumati, Anna Urciuolo, Patrizia Sabatelli and Paolo Bonaldo for providing the mouse model and data on human muscular dystrophy patients.

I am very thankful to Birgit Kobbe for her excellent technical assistance and also for sharing the material, particularly in the initial period.

I am also grateful to Jan Gebauer, Ann-Katrin Becker, Cristian Neacsu, Tobias Maaß, Sandra Lettmann and Christiane Otten for providing a very nice and lively environment in the lab.

My special thanks to Christian Frie, Judith Seul, Katrin Blumbach and Wieslaw Krzyzak for killing and dissecting mice whenever I approached them.

I am grateful to Frank, Bent, Lars, Geri, Ivan, Jacek, Alex and Andreas for their encouragement and also for joining me in the football games which will certainly remain as some of the most memorable moments in my life.

I am indebted to all the members of Biochemistry II for their friendly nature and helping attitude throughout my stay.

My special thanks to Ms. Shambhavi Mishra for providing moral support all through out my PhD work.

I acknowledge the International Graduate School in Genetics and Functional Genomics (IGS-GFG) for providing the financial support for this work.

Last but not the least I thank all my family members without whose moral support and affection, I wouldn't have seen this day.

Sudheer Kumar Gara

1) Sudheer Kumar Gara, Paolo Grumati, Anna Urciuolo, Paolo Bonaldo, Birgit Kobbe, Manuel Koch, Mats Paulsson and Raimund Wagener
Three novel collagen VI chains with high homology to the alpha 3 chain.
The Journal of Biological Chemistry 2008 Apr 18; 283(16):10658-70

2) Raimund Wagener, Sudheer Kumar Gara, Birgit Kobbe, Mats Paulsson and Frank Zaucke. The knee osteoarthritis susceptibility locus *DVWA* on chromosome 3p24.3 is the 5' part of the split *COL6A4* Gene. *Matrix Biology 2009 Volume 28, Issue 6*, *July 2009*, *Pages 307-310*

3) Patrizia Sabatelli, Sudheer Kumar Gara, Paolo Grumati, Anna Urciuolo, Francesca Gualandi, Rosa Curci, Stefano Squarzoni, Alessandra Zamparelli, Elena Martoni, Luciano Merlini, Mats Paulsson, Paolo Bonaldo, and Raimund Wagener.
Expression of the collagen VI α5 and α6 chains in human skin: implications for the pathogenesis of collagen VI related disorders.
Journal of Investigative Dermatology 2010, Epub 2010 Sep 30

4) Sudheer Kumar Gara, Paolo Grumati, Anna Urciuolo, Paolo Bonaldo, Patrizia Sabatelli, Mats Paulsson and Raimund Wagener.
Differential and restricted tissue distribution of novel collagen VI chains.
Matrix Biology 2011, In Review

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen beantragten Teilpublikationen - noch nicht veröffentlicht ist, sowie, dass ich eine Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. med. Mats Paulsson betreut worden.

Köln, 2. März 2010

Sudheer Kumar Gara

LEBENSLAUF

Name	**Sudheer Kumar Gara**
Adresse	Escher Straße 125
	50739 Köln
	Deutschland
Geburtsdatum	11.07.1978
Staatsangehörigkeit	Indisch

Akademische Qualifikation

Schulausbildung

(1994-1996) **Board of Secondary Education**
Andhra Pradesh Residential Junior College,
Rampachodavaram, East Godavari Dist. A.P, INDIA

Studium

(1996-1999) **Bachelor of Science**
Chemistry, Botany and Zoology
Mrs. A.V.N. College, Visakhapatnam, A.P, INDIA

(2000-2002) **Master of Science**
Biotechnology
Pondicherry University, Pondicherry, INDIA

(2002-2005) **Master of Technology**
Biological Sciences and Bioengineering
Indian Institute of Technology- Kanpur, U.P, INDIA

Promotionsstudium

(10/2005 - 02/2010) **Betreuer: Prof. Dr. Mats Paulsson**

Institut für Biochemie II, Medizinische Fakultät der

Universität zu Köln

Köln, 02.03.2010

(Sudheer Kumar Gara)