Heterogenous Transcript Formation During Expression of Extracellular Matrix Genes

Written by Selen Muratoglu

Supervised by Dr. Ferenc Deák

Institute of Biochemistry,
Biological Research Center
Hungarian Academy of Sciences

Szeged
1998

To my father

ACKNOWLEDGEMENTS

Above all, I would like to thank to my supervisor, Dr. Ferenc Deák, for his guidance in my scientific work and his patience and kind support throughout my stay in Hungary. I would like to express my appreciation to Dr. Kiss Ibolya for giving me the opportunity to do this Doctorate in her laboratory and for her support and kindness.

I am grateful to all past and present members of laboratory. I thank particularly to Ildikó Karcagi and Tibor Rauch, for their support and invaluable collaboration and friendship, to Irén Fekete, Anikó Simon and Katalin Kovács for excellent technical assistance. Moreover, all the members of Institute of Biochemistry who helped me in all aspects of my work are greatly appreciated. I would also like to express my appreciation to Gabriella Tick for her help in editorial work. I am particularly grateful to Ildikó Szatmári for her invaluable friendship and support.

I am indebted to Drs. Andor Udvardy and Imre Boros for their valuable criticisms on the thesis, suggestions and encouragement for this work.

I would also like to thank to Mária Tóth and András Borka for excellent artistic work in preparation of photographs and figures.

I should also like to thank all my friends from BRC for their continuous support throughout my stay in Hungary.

Last, I am indebted to my family who have patiently endured the varied phases of my studies.

This work is carried out in the Institute of Biochemistry, Biological Research Center of Hungarian Academy of Sciences, Szeged, Hungary.

TABLE OF CONTENTS

ABBREVIATIONS

AURE	AU-rich elements
bp	base pair(s)
CAT	chloramphenicol acetyl transferase
CEC	chicken embryo chondrocytes
CEF	chicken embryo fibroblasts
CF1	cleavage factor 1
CMP	cartilage matrix protein
dbEST	database of expressed sequence tags
DMEM	Dulbecco's modified Eagle's medium
ECM	extracellular matrix
EGF	epidermal growth factor
FCS	fetal calf serum
HC	hypertrophic culture
HDM	high-density mesenchyme
hnRNP	heterogeneous nuclear ribonulclear protein complex
kb	kilobase(s)
LP	link protein
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PF2	polyadenylation factor 2
PBS	phosphate buffered saline
RACE	rapid amplification of cDNA ends
RT-PCR	reverse transcription coupled polymerase chain reaction
SR	serine-arginine-rich
SMC	smooth muscle cells
snRNP	small nuclear ribonucleoproteins
vWFA	von Willebrand factor type A-like
UTR	untranslated region

INTRODUCTION

I. CARTILAGE AND CHONDROCYTE DIFFERENTIATION

Skeleton is formed of two tissues: various bones and cartilages. Cartilage is an elaborate network of large macromolecules, synthesised and deposited by the specialised cells, called chondrocytes. They deposit these macromolecules around themselves to form an extracellular matrix (ECM). Cartilage is a fundamental biological material that plays a diverse role during development. It acts as the main source of longitudinal growth as a provisional tissue before bone formation or as epithelium covering the articular surface. Then in the growing body, cartilage composes the zones called growth plates, which are found in the tips of bones. During vertebrate embryogenesis, the first elements of skeleton to be formed are cartilage templates that are subsequently replaced by a process called endochondral ossification [1]. Cartilage first appears in the embryo at a time when the limb buds have begun to extend. The precise onset in the chick embryo is after 4.5 days of incubation that is the time for the limb bud cells in the mesoderm to be irreversibly committed to one of the several developmental pathways leading to different types of limb tissues; cartilage, bone or connective tissue. The overall differentiation process for cartilage development is achieved by chondrocyte differentiation, called chondrogenesis.

ECM molecules are synthesised during chondrogenesis. These molecules characterise many morphological and biochemical changes that are associated with limb development. The maturation pathway from mesenchymal prechondrogenic cells to fully differentiated hypertrophic chondrocytes proceeds through differentiation stages that can be defined by biochemical markers and several other parameters such as cell volume, cellular phenotype and growth kinetics. The chondrocyte differentiation is divided into three basic periods. This process starts with aggregation of undifferentiated, pre-chondrogenic mesenchymal cells that are found first in the limb bud and later in the perichondrium. These cells synthesise type I collagen and fibronectin. Then they differentiate into proliferating, which is also called stage I chondrocytes. Besides the synthesis of a new set of macromolecules, changes in morphology and gene expression pattern also occur. The main components of accumulating cartilaginous ECM are aggrecan, hyaluronan complexes stabilised with link protein (LP) and a fine fibrillar network of collagens, composed of type II, IX and XI collagens. Other non-collagenous, matrix proteins, including matrilin-1 (earlier named as cartilage matrix protein) modulate the interaction between the components of the two macromolecular networks [2]. These two non-collagenous proteins will be further discussed

below for their functional importance in chondrocyte development. On the basis of tissue culture data, it was suggested recently that two substages should be distinguished within stage I. Stage Ia is defined as an early stage that is characterised by the transient increase and decrease of mRNA levels of all three type VI collagen chains. This stage is followed by stage Ib, characterised by the inactivation of collagen VI genes and transcriptional activation and synthesis of cartilage specific genes [3]. At later stages, chondrocytes at the distal ends of long bones continue to proliferate. However, the chondrocytes in the centre of the cartilage elements or in the cranial part of sterna undergo further differentiation into hypertrophic chondrocytes (stage II). This process includes exit from cell cycle, enlargement of the chondrocytes and change in their metabolic activity and in specific gene expression. Hypertrophic chondrocytes start to synthesise type X collagen, produce large amount of alkaline phosphatase and start accumulating hydroxyapatite. Recently, a third osteoblast like stage (stage III) is suggested for a fraction of hypertrophic chondrocytes, although it has long been a general view that all chondrocytes terminally differentiate to hypertrophy and die.

A tight regulation of the different steps of endochondral ossification and chondrogenesis, in particular, is critical to bone morphogenesis and growth. Several growth and transcription factors have been implicated in the control of chondrocyte differentiation through several, yet to be fully understood, pathways [4].

TWO NON-COLLAGENOUS CARTILAGE MATRIX PROTEINS AND THEIR IMPORTANCE

Link Protein

The highly organized ECM of cartilage that consists of collagen fibrils and aggregates of aggrecan and hyaluronan which are altogether providing this connective tissue resistance to large compressive and shear forces [5]. Link protein (LP) [6] is an abundant non-collagenous glycoprotein (40-48 kDa) that performs an important function in the assembly and stabilization of the aggregates via its simultaneous non-covalent interaction with both aggrecan and hyaluronan. LP has recently been also found in small quantities in a variety of non-cartilaginous tissues such as aorta and lens capsule. This suggests that it has a more general function in the assembly of the extracellular matrices by enhancing the interaction of other tissue-specific proteoglycans with hyaluronan [6].

The complete amino acid sequence has been reported for rat chondrosarcoma LP [7, 8], and derived from cDNA sequences for chick, rat, pig and human LPs [9-12]. The mature protein consists of three structural and functional domains stabilized by disulfide bonds. The NH_2-terminal immunoglobulin-like domain interacts with aggrecan, while the tandemly repeated COOH-terminal modules can independently bind to hyaluronan [13,14].

In the condensed mesenchyme cells differentiating to chondrocytes during endochondral bone formation, the genes for LP and aggrecan are turned on simultaneously, shortly following the activation of the collagen II gene [15]. The chicken LP gene is larger than 100 kb, consists of 5 exons and four introns [16, 17]. Our laboratory reported that LP gene has multiple initiation sites located 33-78 bp downstream of a TATA-like motif [18]. The first exon, which consists of a solely untranslated sequences of 503 bp, carries alternative splice sites. Therefore, the 5'-end of the LP mRNA is highly heterogeneous in chicken. The 5'-flanking region of the LP gene has also been determined from other species [19, 20]. The promoter for the LP gene in Swarm rat chondrosarcoma [19] was located at a position different from that reported for the chicken [18] and human genes [20]. Thus the size of the first, untranslated exon is 289 bp in human, but it is only 62 bp in rat.

Matrilin-1

Matrilin-1, a non-cartilaginous glycoprotein, is one of the major components of cartilage. It is a compact homotrimer of subunits, assembled via a coiled-coil α-helix [21]. It was first purified from bovine tracheal cartilage [22] and was detectable in various other cartilage types but missing from articular cartilage and intervertebral discs [23]. Monomers of chicken matrilin-1 are of 54-kDa in size and consist of two von Willebrand factor type A-like domains separated by an epidermal growth factor-like module and a C-terminal trimerization domain [24]. Apart from cartilage, matrilin-1 expression was reported only in notochord and certain structures of eye [25, 26] suggesting a very specialised function. Although exact function of matrilin-1 is unknown, it is proposed to have an assembly role of ECM since it interacts with aggrecan and collagen II fibrils [24, 27, 28]. An important bridging function between the two macromolecular networks is suggested due to the assumption of simultaneous interaction of matrilin-1 to aggrecan and collagen II fibrils.

Temporal and spatial expression of matrilin-1 gene along with other cartilage protein genes was studied in the developing mouse limb by immunostaining [29, 30] and in human by *in situ* hybridisation. These studies as well as *in situ* hybridisation in the chicken limb at early

stages of chondrogenesis [15] revealed that the genes for type II collagen, aggrecan and link protein were activated in sequential order and produced matrilin-1 gene transcription.

II. A NOVEL EXTRACELLULAR MATRIX PROTEIN: MATRILIN-2

Cross-reaction of a matrilin-1 specific antisera, showing immunostaining not only in cartilage but also in perichondrium led to the identification of a closely related gene product. Screening of a mouse cDNA library for cross-hybridising clones with the chicken matrilin-1 probe yielded the cDNA clones of a novel protein later called matrilin-2 [31].

Sequence similarity search proved that the closest relative of matrilin-2, within the von Willebrand Factor type-A like module superfamily, is matrilin-1 [31]. The protein precursor of 956 amino acids consists of a putative signal peptide, two von Willebrand Factor type A-like domains connected by 10 epidermal growth factor-like modules, a potential oligomerisation domain at the carboxyterminus and a unique, 75 amino acid long segment, which has no sequence similarity to any other protein sequenced to date, and it contains potential N-glycosylation sites. The presence of a putative secretory signal peptide and the lack of a transmembrane domain in the coding region suggested an extracellular protein, which was confirmed by immunological detection of the secreted protein in the cell culture media. Based on preliminary data, matrilin-2 gene is expressed in a variety of organs and cell lines. A 3.9 kb of mRNA was detected in the limbs of day 11 mouse embryos, calvaria, uterus and heart with a high abundance and in lower abundance in skeletal muscle, brain and skin. However, it was not detected in epiphyseal cartilage samples of new-born mice and neither in trachea, femur, lung, spleen and kidney. Mouse fibroblastic cell lines, as well as a rat osteoblast cell line were all positive for matrilin-2 message. Immunohistochemistry localised the protein to perichondrium but not to the cartilage matrix in tracheal cartilage sections. In sections of trabecular bone the osteoblast layer was positive.

Although the tissue localisation differs, the primary structure and complementary expression pattern in the developing skeletal elements of matrilin-2 compared to matrilin-1, indicate a possible similar role in the organisation of ECM in the other tissues [31].

III. HETEROGENEITY OF mRNA

Complex transriptional units produce multiple mature mRNAs that can give rise to multiple, although related, proteins. Four basic mechanisms and combinations of these can give rise to more than one mRNA from a single gene. First, several primary transcripts initiated at alternative promoters, second, differential termination or 3'-post-transcriptional processing, third, alternative splicing of the same primary transcript and fourth, recombination of genomic sequences, such as for the immunoglobulin genes. The last mechanism is out of scope of this dissertation.

REGULATION OF GENE EXPRESSION BY ALTERNATIVE PROMOTER USAGE

Promoters have been defined as modulatory DNA structures which contain a complex array of cis-acting regulatory elements required for accurate and efficient initiation of transcription of a gene. Lately, it is more evident that they are also prime target elements through which diversity and flexibility in the complex patterns of gene expression in multicellular organisms are created.

Alternative promoter usage can influence gene expression in several ways. The level of transcription initiation can vary between alternative promoters. They can have different tissue specificity and react differently to some signals. Expression of genes in more than one tissue or developmental stage may require distinct combinations of -possibly tissue specific- transcription factors, especially if the same gene is able to respond in different cell types to the same extracellular signals or when it has to respond in the same cell to different signals. Furthermore, the turnover or translational efficiency of mRNA isoforms with different leader exons can differ. Finally, it can lead to the generation of protein isoforms differing at the amino terminus [for review see ref. 32]. Therefore, usage of alternative promoters should be considered as an additional way to create regulatory diversity and provide a way for co-ordinating the synthesis of functionally related proteins that must act together to mediate a certain biological response.

Major and minor spliceosomes

Precise removal of intervening RNA sequences (introns) from the pre-messenger RNA (pre-mRNA) by RNA splicing in eukaryotic nuclei is a major step in the regulation of gene expression. Almost all mRNAs in eukaryotes are produced by splicing of intron-containing precursors, pre-mRNAs. RNA splicing provides a mechanism whereby protein isoform diversity can be generated which may restrict the expression of particular proteins with specialised functions to certain cells and tissue types during development [33].

In some cases, pre-mRNAs are alternatively spliced depending on the state of differentiation or the cell type. This type of regulated splicing will lead to the synthesis of different proteins from the same pre-mRNA or can function as an on-off switch by generating RNAs which either lack or posses an open reading frame, as in the case of sex determination in Drosophila [34].

A common feature of the mRNA precursors that introns are bordered at their 5' and 3' splice sites by to the invariant GT and AG nucleotides, respectively was described previously. Excision of these introns requires the stepwise assembly of U1, U2, U4/U6·U5 small nuclear ribonucleoprotein (snRNP) particles and a number of non-snRNP protein factors, assembling a splicesome [35]. In addition to the 5' and 3' consensus splice sites, branch point sequence and polypyrimidine tract, both in close proximity to the 3' splice site, are highly conserved intronic sequences, which are involved in the splicing process [for reviews 36-38].

Recently a growing amount of data suggested a minor class of natural pre-mRNA introns with 5'-AT and AC-3' boundaries, as an exception to the GT-AG rule [39, 40] in metazoan genes. Splicing of these introns requires a different set of snRNPs (U11, U12, U5 and U4atac/U6atac) compared with the snRNPs required for splicing the majority of pre-mRNAs. The minor snRNPs have a lower abundance in the nucleoplasm, 10^3 - 10^4 copies per cell, than the highly abundant set of major snRNPs (10^5 - 10^6 per cell) [41]. By establishing a He-La cell *in vitro* system that accurately splices a pre-mRNA substrate containing the minor class AT-AC intron, it has been shown that the low abundance U11 and U12 snRNPs act analogously with U1 and U2 snRNPs, respectively, while U5 snRNP

joins the splicesomal complex as in the major spliceosomal assembly. [42, 43]. It is also implicated that U5 snRNP is a common component of both kinds of splicesome. The two other low abundance snRNPs, U4atac and U6atac, were identified and characterised as the essential components of the AT-AC splicesome [44]. However, the complete mechanism of AT-AC splicing is yet to be revealed.

More recently, a subclass of AT-AC introns was defined by the divergent 5' splice site and presumptive branch site sequences. Intron 21 of SCN4A voltage-gated skeletal muscle sodium channel gene and the corresponding intron 25 of the SCN5A cardiac muscle sodium channel gene are the examples of this divergent subclass of AT-AC introns [45, 46]. Interestingly, in these genes, the two subclasses of AT-AC introns coexist with multiple major class GT-AG introns. It was revealed that the divergent subclass of AT-AC introns requires the same set of snRNAs as the GT-AG introns. Lately, Sharp and Burge [47] pointed out that any single intron is of one class depending upon the type of splicesome formed since no chimeric splicesome composed of a mixture of snRNPs from the two snRNP families was reported till date. Therefore, a new classification of introns was supported; they should be identified by the type of splicesome either U2-type or U12-type and not only on the basis of their terminal sequences as Dietrich et. al. suggested [46].

Alternative RNA Splicing

Alternative RNA splicing is the process that allows the selection of different combinations of splice sites within a pre-mRNA. The use of differential splicing patterns in transcripts from a single gene yields mRNAs with different primary structures [49]. The capacity to generate different, but closely related protein isoforms by alternative splicing increases significantly the phenotypic variability that can be obtained from single genes. It is a fundamental mechanism for regulating eukaryotic gene expression. In many cases, alternative RNA splicing contributes to developmentally regulated and cell type specific patterns of gene expression. This mechanism is especially important in the production of structural and functional variability in the case of the modular proteins of the extracellular matrix.

Splice site selection is the major question for explaining the mechanisms of both basic and alternative splicing. Although a great deal of information is now available concerning the general constitutive splicing reaction, the molecular basis of alternative splice site selection in vertebrates has not been revealed [for review see 49]. It is suggested that the general splicing factors participate in regulated alternative splicing. Cell-specific differences in the

concentrations or activities of general splicing factors could regulate alternative splicing [for review see 50]. Specialised proteins, like serine-arginine-rich (SR) and hnRNP proteins, that control alternative splicing of specific pre-mRNAs also mediates this regulation [for review see 51, 52]. A large variety of splicing decisions may be managed by combining cell-specific variation in the abundance of individual SR and hnRNP proteins with pre-mRNA-specific differences in the strength of splice sites and position of modulating elements. Moreover, a number of features in the pre-mRNA have been shown to influence alternative splice site selection [53]. These include the relative strength of 5' and 3' splice sites, intron size, exon size, branch point sequence or location, exon sequence and intron sequences.

IMPORTANCE OF THE 3' UNTRANSLATED REGION

Formation and processing of the 3' end is an essential step in RNA synthesis and it reflects a variety of essentially unrelated reactions with distinct functional sequences. The overriding importance of the 3'-untranslated region (UTR) of eukaryotic mRNA is manifested in the fact that it is the repository of all signals, determining mRNA localisation, polyadenylation, mRNA stability and of controlling translation initiation [for review see 54, 55].

Polyadenylation of mRNA precursors

As it pointed above, the formation of mRNA requires precise RNA processing and one of the mechanisms is pre-mRNA splicing. The second important and just as complicated mechanism is nuclear polyadenylation of these messages. It forms the 3'-end through specific endonucleolytic cleavage of the precursor RNA and the addition of ~250 adenosine residues. As an obligatory step in mRNA biosynthesis, 3'-processing may contribute to the regulation of gene expression through the modulation of both the level and coding capacity of the mature message.

The sequences in the pre-mRNA that signal polyadenylation, as well as the factors recognising them and processing the pre-mRNA have been well studied [reviewed in; 56, 57]. Processing requires sequence elements in the RNA upstream and downstream of the site of cleavage and poly(A) addition. The upstream element is the essentially highly conserved (90%) hexamer signal AAUAAA. All mutations of this consensus motif decrease the efficiency of the element *in vitro* [58]. The downstream element consists of more diffuse, highly variable sequences that are generally rich in U or in G and U residues, and are located between 14 and 70 nucleotides downstream of poly (A) signal [59]. Some pre-

mRNAs contain, in addition to these elements, auxiliary U-rich signals, usually located upstream of AAUAAA, which act as enhancers in the reaction [60, 61].

The gene specificity of downstream sequences suggests that it may be a primary determinant of poly(A) site efficiency. It is shown that this efficiency is a direct reflection of the stability of complexes formed with factors that recognise the two sequence elements that define poly(A) site [62]. It is suggested that a crucial determinant of poly(A) site efficiency is the nature of the interaction of cleavage factor 1 (CF1) with the downstream element. Poly(A) site can only form then a stable ternary complex whereas an inefficient site forms a relatively unstable complex. In transcription units where multiple poly(A) sites are utilised, the promoter proximal site would be at an advantage over a more distal site since it would be transcribed first. A very efficient site would preclude use of a downstream poly(A) site. In this case, the important factors would be the relative strengths of the poly(A) sites and their position within the transcription unit. The distance between the two sites should determine the time in which the first site would be available for processing without competition. Taken together, the relative efficiencies of the poly(A) sites are dictated by the stability of the CF1-PF2 (polyadenylation factor II imparts AAUAAA specificity to both cleavage and polyadenylation and interacts with the RNA dependening on the AAUAAA element) complex that forms at the poly(A) site as defined by the nature of the downstream GU-rich sequence element. The more stable the complex, the greater is the probability that the RNA will be processed at that location.

In the cell nucleus, transcription, splicing and 3'-end formation are coupled processes and they influence each other. Coupling of 3' processing and the transcription termination is thought to prevent premature termination and thus help to ensure the synthesis of full-size pre-mRNAs. The precise mechanism of coupling between splicing and 3'-end processing has not been resolved. Nevertheless, some progress has been made in resolving these mechanisms. According to the exon definition model of splicing, the 3'-end processing signals act as the functional equivalent of a downstream 5' splice site that in non-terminal introns communicates across the exon with 3' splice site of the upstream intron [63]. The removal of the last intron by splicing from a pre-mRNA containing multiple introns depends on the presence of a 3'-processing signal downstream. Likewise, the efficiency of 3'-end formation is increased by the presence of an upstream intron. The U1 snRNP has been proposed to mediate this interaction [61].

With the recent data and with the points discussed above it is becoming clear that there is a significant degree of regulation at the ends of mRNA that plays an important role in the control of gene expression. Furthermore, a number of genes utilise alternate poly(A) signals in the 3'-UTR regions of their mRNAs. This may result in a further level of regulation, as although the same protein product will be produced, different 3'-UTRs will be formed which may alter the stability, translation or transport of mRNA.

mRNA Stability

mRNA degradation in eukaryotic cells is a regulated process that provides powerful means for controlling gene expression. The mRNA decay rate is a major determinant of mRNA abundance in all organisms. The change in the half life of an mRNA can affect their abundance without any change in their transcription. Half-lives of most mRNAs are probably determined by mRNA binding proteins that influence the mRNA stability [64]. The half-life of an mRNA also affects how rapidly it reaches its new steady-state following an increase in transcription [65].

The 3'-UTRs of many unstable mammalian mRNAs contain regions rich in adenosine and uridine (AU-rich elements or AUREs) [64, 66] . AUREs range in size from 50 to 150 nucleotides. Generally they contain multiple copies of pentanucleotide, AUUUA, and have a high content of U and sometimes also A residue. However, by use of synthetic AU-rich sequences it is shown that the presence of AUUUA motif(s), even in an AU-rich region, does not guarantee a destabilising function of an AURE. A nanomer with an AUUUA pentanucleotide core UU**AUUUA**(U/A)(U/A), gives a modest destabilising effect. Moreover, the importance of a high percentage of U residues, capable of enhancing the destabilising ability of AUUUA motif, is also emphasised.

SPECIFIC AIMS

- Establishing culture systems which represent stages of *in vivo* chondrogenesis. Isolation of RNA samples from these cultures for gene expression analysis of cartilage specific proteins.

- Characterization of LP and matrilin-1 gene expression during cartilage development. Analysis of their gene expression patterns throughout *in vitro* chondrocyte development.

- Analyses of the heterogeneity of the matrilin-1 and LP transcripts and the importance of this heterogeneity for the regulation of gene expression.

- Cloning and sequencing the full length human matrilin-2 cDNA and studying the gene expression pattern of both mouse and human matrilin-2 in various cell cultures. Investigation of the origin of heterogeneity detected also in matrilin-2 transcripts.

MATERIALS AND METHODS

Oligonucleotide primers

Table 1: *Oligonucleotide Primers*. The numbers within the names of oligonucleotide primers denote the positions in the related cDNA sequences, except the ones for 5'/3' RACE

Name	5'	3'	Tm °C
CMP			
CMP552N	cacgctacggcaaatcg		50.7
CMP1114N	ggggttcccaaggtgggcatagt		62.7
CMP1349N	cagaggaagatccatgcg		49.7
CMP1354N	gaggatccatgcgaatg		45.8
CMP1477R	tgatcttattctccagggc		46.4
CMP1689R	agctctcagccatgcgg		51.7
CMP2316R	ggcaccgctttcgtcag		52.1
CMP2364R	gggctgaggaaggagctggtggt		63.5
LP			
L1221N	gccacaaagagacgctgggac		61.7
MAT-2			
mM425N	gtgtttgtcgtcgacagctct		51.8
mM2345N	tattccacacaggtccgaaca		52.7
mM2555N	ggacgggctcaggatga		50.9
mM2635R	cctaccccaacggcataca		53.1
mM2677N	cattgacaagcatctcttct		44.0
mM2939R	tttgtgtagaccgtgaaaga		45.4
mM3029R	cgttctggaacagtataagg		44.1
mM1001N	cacttctgcctcaacacacc		50.2
mM3117R	ctgtatctcaggcgattttc		46.6
hM906N	tggccaatttcagccagat		52.8
hM2316R	ggggaaattgtcaaggaatcta		51.7
hMTR875/20R	tcccaatggacttcaaggtg		52.1
hMTR965/20R	ccgtgcacaacttcttctgg		53.1
hMMTR1027/22R	tgagccagggatgttgatgcag		59.5
CRPBbv/20R *	cttgaccgtgctgccatatt		52.7
5'/3' RACE			
dT_{17} adapter	gactcgagtcgacatcgt_{17}		
XSC	gactcgagtcgacatcgt		42.6
TLT	ccagcgagcagagtgacgaggactcgagctcaagct_{17}		
TAG	ccagcgagcagagtgacg		54.2
LIN	cgaggactcgagctcaagc		51.4

*CRPBbv/20R: position in mouse cDNA: 568

Chondrocyte cell culture

Primary cultures were made from Rhode Island White chicken embryos. Primary chicken embryo chondrocytes (CEC) were obtained from sterna of day 14 embryos using 0.1% collagenase treatment (type I, SIGMA Chemical Company, St. Louis, MO/USA), as described [38], and cultured adherent to the petri dish in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS; Gibco/BRL, Gaithersburg, MD/USA) for two days, unless indicated otherwise. In some cases CEC cultures were made from whole sterna of 18-day-old chicken embryos, or from sterna dissected into the caudal one third, the middle one third and the core region of the cranial one third, to achieve cell populations enriched in resting, proliferating and hypertrophic chondrocytes, respectively [67, 68]. Hypertrophic chondrocytes originating from the core region of the sterna were cultured in suspension at a starting density of 1.5×10^6 cells in 10-cm plastic plates in DMEM supplemented with 10% FCS for 4 days, then for additional 5 days in DMEM:Coon's modified F12 (SIGMA Chemical Company, St. Louis, MO/USA) 1:1 medium.

To obtain high-density mesenchyme (HDM) cultures undergoing chondrogenesis "*in vitro*", limb buds of stage 24 [69] chicken embryos were prepared and cultivated in Ham's F12:DMEM 6:4 medium supplemented with 10 % FCS at a density of 5×10^6 cells per 35-mm plates as described [70, 71].

Mouse and Rat Cell Cultures

The mouse fibroblastic cell lines WEHI 164 and NIH 3T3, the rat osteogenic sarcoma UMR-106 and the small intestine epithelial IEC-6 cell lines were obtained from the American Type Culture Collection (Rockville, MD). The mouse C2/7 cells with skeletal muscle characteristics [72], the smooth muscle-like cell 9E11G [73], the keratinocyte carcinoma PVD(A)I [74] the rat Schwann-cell line RN22 [75] and the Swarm rat chondrosarcoma cell line (RCSC) [76] were obtained from the laboratories of origin. The mouse immortalized endothelial cells m1END, derived from mesenteric lymph nodes [77] and the SV40-transformed lymphoid vascular endothelial cell SVEC [78] were provided by L. Sorokin and R. Hallmann (Erlangen), smooth muscle cells (SMC) from rat aorta were cultured by F. Michaelsen (Cologne), using standard methods [79]. Unless recommended otherwise by the supplier, the cell lines were cultivated in DMEM supplemented with 10% FSC (Life Technologies), and utilised for RNA analysis.

RNA preparation and Northern blot analysis

Total RNA was prepared from guanidinium thiocyanate lysates of chicken cell cultures, human and mouse cell lines by acid-phenol-chloroform extraction [80], or from sterna of 14.5-day-old chick embryos by repeated ethanol precipitation from guanidinium chloride extracts [81]. RNA samples were electrophoresed in 0.8% or 1% agarose gel in the presence of 2.2 M formaldehyde, and blotted to Hybond N membranes (Amersham, Little Chalfont/UK). Filters were hybridized, washed and reprobed as recommended by the supplier.

Hybridisation probes

The following chicken cDNA fragments were used as hybridization probes: 1.3-kb fragment complementary to the translated region of matrilin-1 mRNA and 0.9-kb *Eco*RI fragment of 3'-UTR [82], the 1.5-kb insert of pLPG2 coding for link protein [9], a 0.7-kb *Bam*HI-*Eco*RI fragment of pCgII-SO1 [83] coding for 3'-untranslated region and 90 amino acids of the C-propeptide of α1(II) collagen, 0.55-kb cDNA for α1(X) collagen [84], 1.2-kb *Eco*RI fragment for α2(VI) collagen [85], 675-bp cDNA for α1(I) collagen [86], 0.3-kb cDNA copy of the 27S rRNA and a 1.8-kb cDNA coding for β-actin [87].

Polymerase Chain Reactions (PCR) and Analysis of the Amplification Products

The oligonucleotide primers used in PCR reactions are summerised in Table 1. By using the appropriate annealing temperature according to the primer pair used in a given reaction, all PCR reactions were performed in the following reaction cycles;

98°C 5min	Tann 1min	72°C 30min	1x
95°C 1min	Tann 1min	72°C 2min	30x
72°C 10min			1x

The Taq polymerase was added in the first cycle's 30 minutes elongation period to avoid false priming below the annealing temperature and decrease in the efficiency of the enzyme due to long initial denaturation incubation. The PCR products were separated on a 1.2 % agarose gel and visualised by ethidium bromide staining.

All the PCR products were cloned either directly or after fragment isolation in pT7Blue(R) vector (Novagen) unless otherwise indicated. Both cDNA strands were sequenced with Sequenase version 2.0 (U.S. Biochemical Corp.) either directly in plasmids using gene specific or universal primers or after subcloning in pBluescript SK II +.

Reverse transcription coupled PCR (RT-PCR)

3 μg of total RNA was reverse transcribed with MMLV-RT (Gibco-BRL) in a 50 μl reaction volume with TAG-Linker-dT_{17} (TLT) primer unless otherwise indicated. After denaturating at 65°C the total RNA sample and the primer, they were annealed by 5 minutes incubation on ice. The elongation was performed with an incubation cycle of 60 minutes at 37°C, 15 minutes at 42°C and 5 minutes at 95°C. PCR amplifications were performed by *Thermus aquaticus* polymerase (Sigma, Perkin Elmer, Amersham, Pharmacia) on 1μl (higher dilutions are indicated) of the reverse transcribed ss cDNA with the adaptor and/or gene specific primers indicated.

3' RACE analysis of RNA samples and analysis of amplification products

The RACE reactions were carried out as suggested by Frohman [88]. 5 or 3 μg of total RNA samples prepared from cultured chondrocytes of juvenile chick sternum were reverse transcribed with MMLV-RT (Gibco-BRL) in 50 μl using dT_{17}-adapter primer. In the first RACE reactions, PCR amplification was performed with XSC and the gene specific primer CMP 1354N at annealing temperature of 45°C.

Later, we performed modified 3' RACE reaction using primers of higher Tm values and applied a second amplification reaction with gene specific nested primers to increase the specificity. Reverse transcription of the same RNA sample was primed with TLT primer. The first amplification was made between TAG and a gene specific primer CMP 1114N (exon 5) at the annealing temperature of 58°C. The second amplification reaction was performed at 53°C between LIN primer and a more downstream gene specific primer CMP 1349N lying on exon 6.

5' RACE analysis of human matrilin-2

First strand cDNA was synthesised from 1 μg of total RNA sample of HEp-2, using 10pmoles of hMTR 1027R gene specific primer with 50 units of Expand MMLV-RT (Boehringer-Mannheim), in the presence of $\alpha[^{32}P]$ dCTP. The primer was extended with a cycle of 60 minutes at 42°C, 15 minutes 53°C and 5 minutes at 95°C. The unincorporated dNTPs were separated on a spun-column [89] G50. An anchor sequence, which was a homopolymer tail of A, was then added to the 3'-end of the cDNA using terminal deoxynucleotidyl transferase (Boehringer Mannheim) and dATP by an incubation at 37°C for 5 minutes and at 65°C for 5 minutes. The dA tailed ss cDNA was first annealed with

TLT primer at 48°C and then the two consecutive PCRs were performed. The first amplification was done on the 1/50th of cDNA annealed using TLT primer with 10 pmoles of TAG and 10 pmoles of gene specific primer hMTR965/20R in a final volume of 50 μl. Then 1 μl of this reaction was further amplified with 10 pmoles each of LIN primer and gene specific primer hMTR 875/20R in 50 μl. The product of the last PCR amplification was directly cloned in pT7Blue(R) and sequenced.

The same dA tailed ss cDNA was utilised for a second round of 5' RACE for cloning the more upstream sequences. The reaction was performed as described above with the gene specific primers hMTR965/20R and CRPBbv/20R. With this reaction the translation initiation site was cloned and sequenced.

Analysis of the relative activity of the two LP gene promoters during chondrocyte differentiation

Plasmid construction and sequence analysis

The chicken LP gene sequence is numbered from the start point of the longest mRNA. The *Eco*RI-*Bgl*II fragment from +394 to +676 and the *Pvu*II-*Hin*dIII fragment from +741 to +1073 of the cDNA clone pLPF4 [9] were inserted into the *Eco*RI-*Bam*HI or *Sma*I- *Hin*dIII sites of M13mp19, respectively, to construct plasmids D241 and D242. Structure and integrity of the constructs were checked by nucleotide sequencing using Sequenase version 2.0 (U.S. Biochemical Corp.)

Primer extension assay

Total RNA samples of CEC, HDM and HC were used in this assay. Poly(A)$^+$ RNA was purified on oligo d(T) linked to magnetic beads (Dynal AS, Oslo, Norway). Transcription start sites were mapped by extending a gene specific primer with T4 DNA polymerase on a DNA template annealed to the mRNA as described previously [90]. Single stranded DNA was purified from an M13mp19 subclone carrying the insert of pLP5S3 [18] and was used as a template. The LP gene-specific primer L1221N was extended. For the estimation of the relative promoter usage, single stranded DNA templates were purified from D241 and D242 M13mp19 clones. 40 fmoles of each of the ss DNA template were simultaneously annealed to RNA sample and 200 fmoles of –40 sequencing primer [32P]-labelled at the 5'-end with polynucleotide kinase. The primer was elongated by T4 DNA polymerase. A 1760-nt LP RNA made *in vitro* by T3 RNA polymerase was used in the

assay as an external quantification standard. Band intensities were compared by phosphorimage analysis (Molecular Dynamics Phosphorimager 445SI).

Sequence analysis

The following programmes (with the given www site addresses when necessary) were used for the analyses of the nucleotide sequences:

1. www.ibc.wustl.edu/~zuker/rna; for prediction of RNA secondary structure and Gibbs free energy of a given RNA sequence
2. http://expasy.hcuge.ch/ch2d/PI_tool.html; for prediction of molecular weight (M_r)of a protein of a given amino acid sequence.
3. WISCONSIN PACKAGE Version 8.1-UNIX Genetics Computer Group, Inc. for deduction of amino acid sequences of a given nucleotide sequence and alignments of nucleotide and amino acid sequences.
4. Oligo 3.1 for designing the oligonucleotide primers of a given nucleotide sequence.

RESULTS

I. Characterisation of chondrogenic culture systems representing consecutive stages of cartilage development

Different chondrogenic culture systems were established to study temporal and spatial regulation of gene expression during chondrogenic development in the chicken embryogenic sterna. Application of several culturing conditions rendered us to establish in our laboratory those systems which represent and cover the stages of chondrocyte development. The morphological changes during the days of culturing were monitored by phase contrast microscopy (data not shown).

The onset of chondrogenesis was represented by culturing limb bud mesenchymal cells prepared from stage 24 embryos. Nodule formation and accumulation of glycosaminoglycans (data not shown) were detected by the third day in high density cultures of committed mesenchymal cells (HDM). To study the molecular events at later stages, proliferating and hypertrophic chondrocyte cultures were utilised. Chondrocyte from sterna of day 14.5 embryos were cultured (CEC). Based on morphological observations, the CEC cultures predominantly consisted of proliferating (stage I) chondrocytes; 2-3 small, round cells staying together in territorial matrix. HC cultures were formed by separated, large individual cells. Sterna of day 18 embryos were divided into three parts as described in the materials and methods section. Differentiating hypertrophic chondrocytes were cultured for a longer time, in suspension favouring differentiation towards hyperthrophy. After the fourth day in culture, a large number of cells was having hypertrophic morphology.

Kinetics of LP and matrilin-1 gene expressions compared to expression of cartilage proteins as the markers of chondrogenic developmental stages

After reproducing culturing conditions under which the cells represented one stage of chondrogenesis, we examined the gene expression pattern in *in vitro* cultures. For this purpose, we isolated total RNA from the cells on various days of culturing and performed Northern hybridisation analyses to monitor the steady state levels of cartilage-specific mRNAs. Blots of HDM and CEC samples were consecutively hybridised with cDNA fragments for α1(II) collagen, matrilin-1, LP as well as β-actin and α2(VI) collagen to measure the temporal changes in the relative abundance of mRNAs (fig. 1). The hybridisation with 27S rRNA probe served as a control for the amount of RNA blotted on the filter, which renders the comparison of the intensities of hybridisation signal possible.

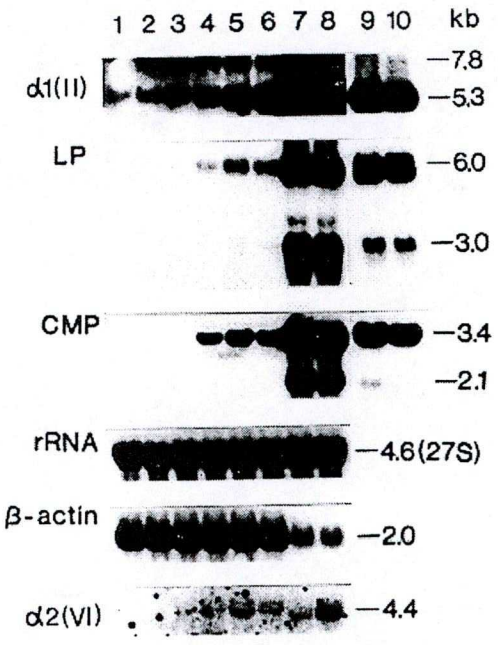

Figure 1: *Accumulation of cartilage specific mRNAs during the early stage of in vitro chondrogenesis.* 10 μg aliquots of total RNA were electrophoresed, blotted and hybridised consecutively with the probes indicated on the left. In lanes 1 to 6 samples were taken at day 1, 2, 3, 4, 5 and 6, respectively, from HDM cultures undergoing *in vitro* chondrogenesis. In lanes 7 and 8 RNA was isolated from 2-day CEC cultures obtained from the sterna of day 14 and 18 chicken embryos, respectively. Lanes 9 and 10 are shorter exposures of lanes 7 and 8, respectively. Sizes of messages are indicated on the right.

The steady state levels of cartilage specific matrilin-1, LP and α1(II) collagen mRNAs were increasing in the days of culture in HDM. Amongst those, the highest level of hybridisation signal was detected for α1(II) collagen as it was already present from the first day of culture and increased through the days of culturing. Hybridisation pattern of LP was similar, though the relative abundance was much lower than collagen II message. By longer exposure of the filter, hybridisation signal for LP message was detectable on the first days of HDM culture (data not shown). However, the matrilin-1 mRNA level reached detectability later than the former ones. It appeared on the fourth day of HDM culture and accumulated with a higher rate than the LP transcripts. This sequential activation of the cartilage-specific genes recapitulated the *in vivo* activation of these genes. Yet, the steady state levels of all three cartilage specific mRNAs were more than an order of magnitude higher in cultures of sternal chondrocytes isolated from 14 and 18-day chicken embryos, than in the HDM cultures.

Multiple mRNA variants were detected for these transcripts in the early stages of chondrogenesis. In the case of α1(II) collagen, two transcripts of 7.8 and 5.3-kb were detectable in all HDM and CEC samples, but with different relative abundance (fig. 1, lanes 7 and 8). In the case of LP, the shorter and less abundant 3.0-kb mRNA variant reached the level of detection only on the latest days of HDM culture although the 6.0-kb major variant was present even from the first day. The LP gene is also transcribed at all developmental stages during *in vitro* chondrogenesis, although the early proliferative stage (stage Ia) is characterised by a lower steady state LP mRNA level than the later stages [69]. We have also observed and reported [17] the presence of the multiple transcripts for matrilin-1. The shorter 2.1-kb mRNA variant appeared in the CEC cultures as a minor transcript along with the 3.4-kb major transcript. The origin of this minor variant will be discussed later. The hybridisation pattern of α2(VI) collagen mRNA was different from the preceding hybridisations. The steady state level of α2(VI) collagen mRNA reached the peak value at day 5 in HDM culture but, unlike cartilage specific transcripts, it was relatively low in CEC cultures. No α1(X) collagen message was produced (data not shown) in the HDM and CEC cultures, indicating that cells have not reached hypertrophy and represent proliferating, stage I chondrocytes. The inverse relationship in HDM and CEC cultures between the α2(VI) and matrilin-1 messages let us conclude that HDM cells reached chondrocyte differentiation stage Ia and CEC cultures are predominated by stage Ib cells. Subsequently, the same filter was probed with β-actin. High steady state level of mRNA was detected in HDM cultures whereas the lower abundance in CEC suggested a high remodelling in the HDM cultures. Finally, we were able to conclude that the HDM and CEC cultures are representing the early stages of chondrogenesis.

Parallely, in collaboration with the laboratory of R. Cancedda, kinetic of expression of the same genes was studied in another chondrocyte culture system. Investigation of temporal changes in gene expression in suspension culture of dedifferentiated tibial chondrocytes supported the previous observations [3]. It had also elucidated the parallel onset of matrilin-1 and stage II specific marker α1(X) collagen gene [17].

To confirm the presence of matrilin-1 mRNA in hypertrophic cells, we also employed the culture system described by Lu Valle et. al. [91]. Sterna of day 18 embryos were divided into three parts [67] and RNA was isolated from cell populations enriched in resting, proliferating and hypertrophic chondrocytes. Northern hybridisation revealed the presence of α1(X) collagen mRNA in samples of hypertrophic chondrocytes cultured for 2 days, but

not in cultures of proliferating and resting cells (fig 2. lanes 1-3). Upon prolonged culturing of hypertrophic chondrocytes in suspension facilitating hypertrophy, the α1(X) collagen mRNA accumulated.

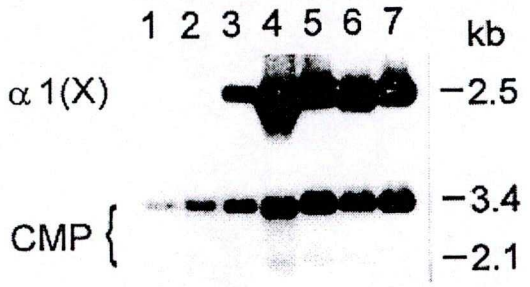

Figure 2: *Northern blot analysis of RNA samples from suspension cultures of sternal chondrocytes.* Sterna of day 18 embryos were dissected into regions enriched in resting (lane 1); proliferating (lane 2); and hypertrophic (lane 3) chondrocytes; the cells were cultured in suspension for 2 days and used for RNA isolation. The cells originating from the hypertrophic core region sample were cultured further in suspension for 4, 7 and 9 days (lanes 4, 5, 6, respectively). Lane 7 sample is equivalent to lane 6, but an independent preparation was used. 3 ìg RNA aliquots were hybridised.

Rehybridisation of the same filter with matrilin-1 probe showed a low level of expression of matrilin-1 in resting chondrocytes (figure 2, lane1) and a higher level in proliferating and hypertrophic chondrocytes (figure 2, lanes 2 and 3). Signal intensities for matrilin-1 and α1(X) collagen mRNAs were parallel in hypertrophic cell samples (lanes 4-7). Therefore, we concluded that matrilin-1 mRNA is produced by hypertrophic cells of sternal origin as well.

II. Heterogeneity in matrilin-1 mRNA lies in the 3'-untranslated region (UTR)

The Northern hybridisation analysis performed to unveil the temporal changes in gene expression pattern of matrilin-1 gene revealed two sizes of transcript. Chondrocyte cultures representing the proliferating stage of chondrogenesis yielded a minor, 2.1 kb mRNA variant besides the 3.4 kb major form (fig. 1). Knowing that the cDNA fragment used as hybridisation probe covered a large part of the translated region of matrilin-1 (probe **a** in fig. 3), we suspected that the two transcripts differ in the 3'-UTR. Thus, a non-overlapping second probe lying in the 3'-UTR of matrilin-1 was designed (see probe in fig.3). The consecutive probing with the two cDNA fragments clearly showed (fig. 3) that the difference is indeed due to 3'-UTR of matrilin-1 gene.

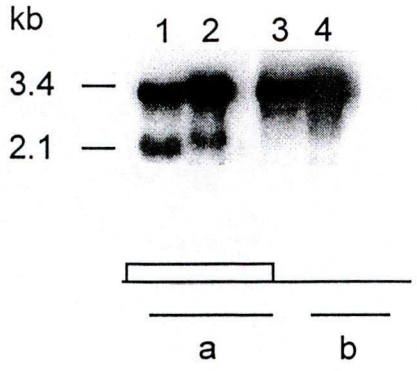

Figure 3: *Northern hybridisation analysis of the two matrilin-1 mRNA species.* Total RNA samples isolated from cultured chondrocytes of juvenile chicken sternum (**lanes 1 and 3)** and cartilage from embryonic chicken sternum (**lanes 2 and 4**) were blotted and hybridised consecutively with probe **a** (lanes 1, 2) or **b** (lanes 3, 4). The sizes of transcripts are shown on the left-hand side of the figure. Note the disappearance of the shorter mRNA variant in the second hybridisation of the same filter with the probe **b**. The size difference observed for the shorter transcript by the hybridisation with probe a should be an ambiguity in the migration of mRNA.

Therefore, the 3'-UTR was analysed in detail. 3' RACE analysis was performed as described in Materials and Methods. Usage of the oligo-dT adapter primer (see table 1) in the reverse transcription reaction enabled us to choose all the functional mRNA variants that can be created by the post-transcriptional modifications. The amplification products using a forward gene specific primer CMP 1354N and the reverse XSC (see table 1) primer pair, should reflect the full-size and the shorter mRNA variants as described schematically in figure 4. Heterogeneous low molecular weight products were obtained (not shown). The most intense bands were cloned into pT7Blue(R) vector and sequenced. Two kinds of products, shorter than the full-size mRNA, were identified. The first group of minor RNAs utilised a more upstream polyadenylation signal. Two independent clones, differing in the usage of either of the two alternative polyadenylation sites were obtained. Both of these variants were utilising the same polyadenylation signal (AATATA) at the position of 1772, which was distinct from the major product's poly(A) signal in the position of 2821 [71]. The first variant utilised a poly(A) addition site 16 bp downstream of the poly(A) signal, whereas the second variant's poly(A) addition site was 33 bp downstream (fig. 4). Yet, both alternative polyadenylation products, represented as form B in figure 4, yielded only a shorter 3'-UTR, rendering no change in the translated region. However, a second shorter form was detected too, which selected of an alternative 3' splice site in exon 8 (figure 4 form C). Sequence data of this product revealed that, in this mRNA variant, translational stop signal (fig. 4) was spliced out, giving rise to an altered protein sequence in the trimerisation domain. Moreover, the

alternative splicing generating the minor class form also utilised an AC-3' splice site and with a high probability the same U12 AT-AC [47] splicesome, as the splicing of the major mRNA.

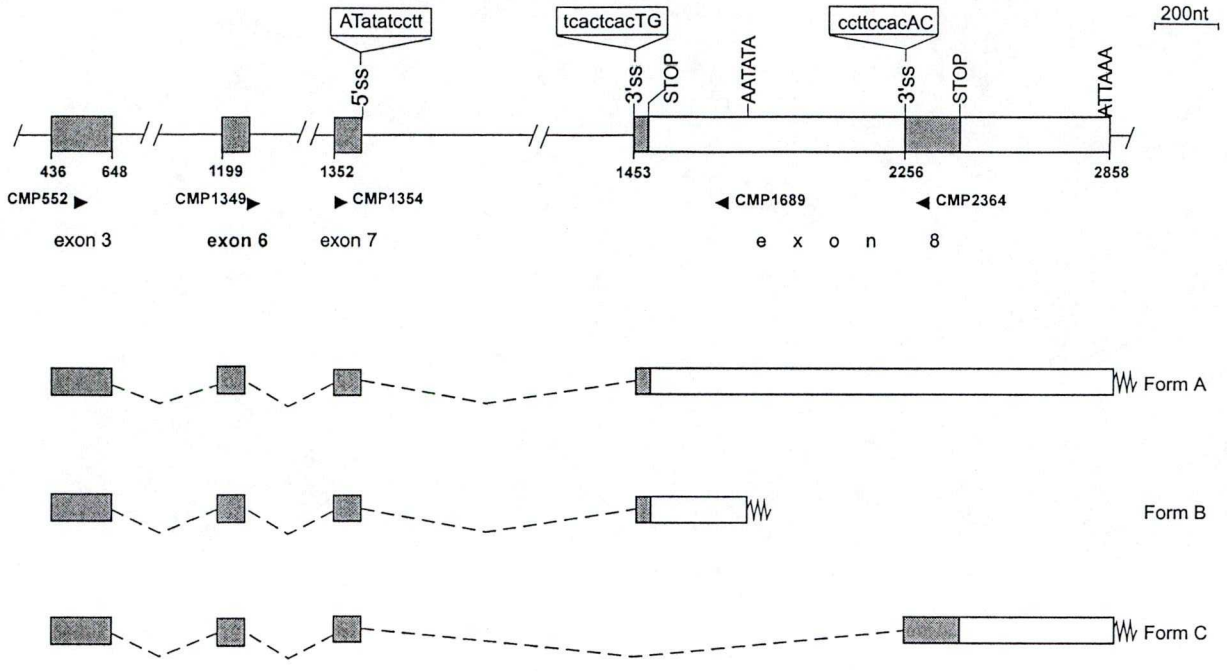

Fig. 4: *Schematic representation of the mRNA isoforms of matrilin-1, differing in their 3' end.* Open boxes, exons; shaded parts are the translated regions; dashed lines connecting the exons, introns. Form A, full-size, Form B, alternative polyadenylation site usage and Form C, alternative splicing mRNA isoforms. The sequences of splice sites and polyadenylation signals are indicated. Arrowheads depict the oligonucleotide primers.

To obtain a reproducible reaction, which would amplify all possible mRNA isoforms, the 3' RACE conditions were modified. New gene specific oligonucleotide primers with higher Tm values (CMP 1114N and CMP 1349N) were designed and used in the consecutive PCR amplifications. Both of the consecutive PCR amplification reactions yielded two distinct bands (fig. 5, lane 1 and 4) which were verified to be the full-size and the alternatively polyadenylated forms according to their sizes indicated in the figure 6. By Southern hybridisation the identities of the bands were further supported. The same samples were blotted and hybridised with two non-overlapping radiolabelled cDNA fragments lying upstream and downstream of the major 3' splice site. The two intense products, visualised

by ethidium-bromide staining, were positive with both probes.

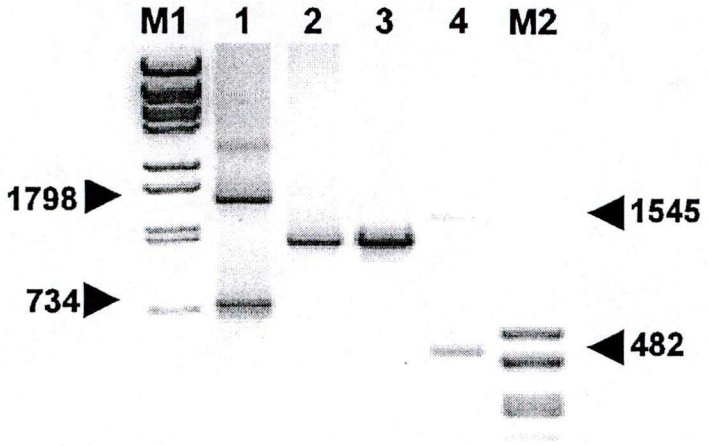

Figure 5: *3' RACE analysis of chicken matrilin-1 gene*. Total RNA of cultured juvenile chicken sternum chondrocytes was reverse transcribed using TLT primer. Various regions of cDNA were amplified, products were separated on a 1.2% agarose gel and stained with ethidium bromide for visualisation. λ Hind III DNA size marker (**M1**), amplification with TAG and CMP 1114N primer pair on total RNA of (**lane 1**), amplification with CMP 2364R and CMP 1114N primer pair on ss cDNA (**lane2**), control amplification reaction with CMP 2364R and CMP 1114N primer pair on the cDNA plasmid (D179/11) harbouring the 3'-end of matrilin-1 gene (**lane 3**), second amplification done on the first amplification product (lane 1) with LIN and CMP 1349N (**lane 4**), pUC *Bsp*I DNA size marker (**M2**). The sizes of 3' RACE prodocts (lane 1 and 4) are given on either sides of the figure.

However, the alternatively spliced form was not detectable with 3' RACE. Therefore, to clarify the mRNA isoform varieties and to control if the sequenced 3' RACE clones represent the only alternatively spliced form, RT-PCR was performed using a gene specific primer pair. The predominance of the full-sized mRNA competing with the shorter mRNA variants was overcome by the following modifications in the RT-PCR reaction condition. By a ten fold increase in the concentration of TLT primer used for reverse transcription, the less abundant RNA transcripts were represented in the cDNA with a higher ratio to the abundant ones (fig.7).

Figure 6 : *Nucleotide sequence of 3' UTR of chicken matrilin-1 gene.* Exon 8 is shown with the upper case letters while the part of intronic region and the 3'-flanking region is denoted by lower case letters. The coding sequence is numbered from the start codon in exon 1 (not included). Both proximal and distal poly (A) signals are indicated by boldface letters while diamonds (♦) denote the cleavage sites. The downstream U-rich elements are underlined. The termination codons are shown by underlined boldface letters. Alternative 3' ss is denoted by an (↓). Deduced amino acid sequences of form A and form C of figure 4 are denoted under the corresponding nucleotide sequence.

We have further supported the occurrence of alternative splicing by depletion of the full-size mRNA and amplifying only the minor variant produced by alternative splicing. An oligonucletide primer (CMP1689R) complementary to the spliced out region was annealed to the total RNA. The RNase-H digestion was performed [92] to selectively digest only the full size mRNA template. This enrichment of shorter transcript rendered efficient amplification as shown in figure 7.

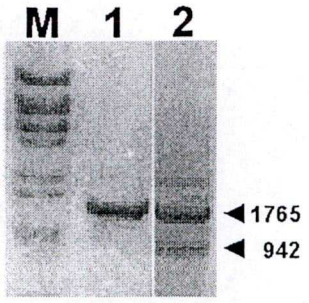

Figure 7: *RT-PCR analysis with modified conditions.* Amplification of the ss cDNA primed with 10μmoles of TLT primer **(lane 1).** Amplification of cDNA primed with 100 μmmoles of TLT primer on 1/100th fraction of ss cDNA **(lane 2)**. The sizes of amplification products are shown on the right-hand-side of the figure.

The RT-PCR amplification done after RNase-H digestion using the gene specific primers amplified only the shorter products, while the full-size product had disappeared (fig. 8). Ampilification product performed on the templated which was digested with RNase-H, migrated anomalously (fig. 8, lane 2). However, cloning and sequencing of this PCR amplification product reproducibly gave the same result demonstrating the presence of the alternatively spliced mRNA isoform, in addition to the alternatively polyadenylated and the full-size ones.

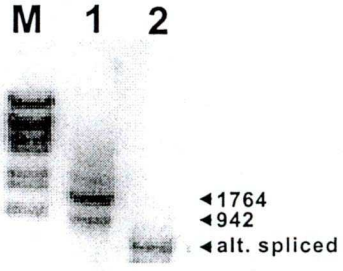

Figure 8: *Selective RT-PCR amplification of the short mRNA isoform created by alternative splicing.* PCR amplification after RNase-H treatment of the mRNA (**lane 1**), PCR amplification of the untreated template (**lane 2**). Note the disappearance of the longer product using the full-size mRNA isoform as a template. The sizes of the products are denoted on the right-hand-side of the figure. Selectively amplified, alternatively spliced product on lane 2 is also indicated.

The size difference between the full size and the alternatively polyadenylated products was around 1.1 kb, which agreed well with the size difference revealed by the Northern hybridisation analysis (fig. 5). Therefore, we concluded that the size difference detected in the northern hybridisation analysis could be attributed to the alternative poly(A) site usage.

III. The chicken LP gene is transcribed from two promoters

Previous studies of our laboratory to map the cis control elements in the 5'-flanking region of the chicken LP gene revealed that fragments of the first exon of the gene show promoter activity in transient expression assays. Along with the analysis of the first, upstream promoter, the potential downstream promoter was also analysed by inserting fragments of interest into pEMBLCAT, upstream of the promoterless gene for chloramphenicol acetyltransferase (CAT). A part of the first exon, from +183 to +495, lacking the TATA motif and initiation sites identified previously [18], had shown relative high activity in CEC, isolated from sterna of 14-day-old embryos expressing theLP gene at high level, and compared with a chicken embryo fibroblastic culture which is a non-expressing control. Thus, this region from exon 1 was shown to function as a second promoter in transient expression assays in CEC. Primer extension with T4 DNA polymerase as shown in figure 9 revealed that the downstream promoter is capable of initiating RNA chains *in vivo* as well (Deák F. unpublished data).

A 5'-end-labelled oligonucleotide lying downstream to the first promoter, and RNA sample isolated from the sterna of chicken embryos were annealed simultaneously to a single stranded DNA carrying exons 1 and 2 of the LP gene, as shown schematically in figure 9. The primer was extended by T4 DNA polymerase, the DNA synthesis was halted by the 5'-end of LP mRNA transcribed from downstream promoter. Comparison of primer extension in the presence and absence of poly(A)+ RNA added to the template-primer pair revealed that the synthesis of 75, 78-80 and 96-97-nt products was specifically terminated by RNA (fig. 9). Therefore, the transcription initiation sites were mapped immediately upstream and within the CACTTC motif as well as CAAACT motif matching fully the consensus start sites identified for eukaryotic mRNAs (CANYYY or YYCANT/AYY) [93, 94]. A TATA-like ACTTAA motif is located 30 bp upstream of the CACTTC sequence. Thus, the downstream promoter was shown to be functional *in vivo* in chicken embryo chondrocytes.

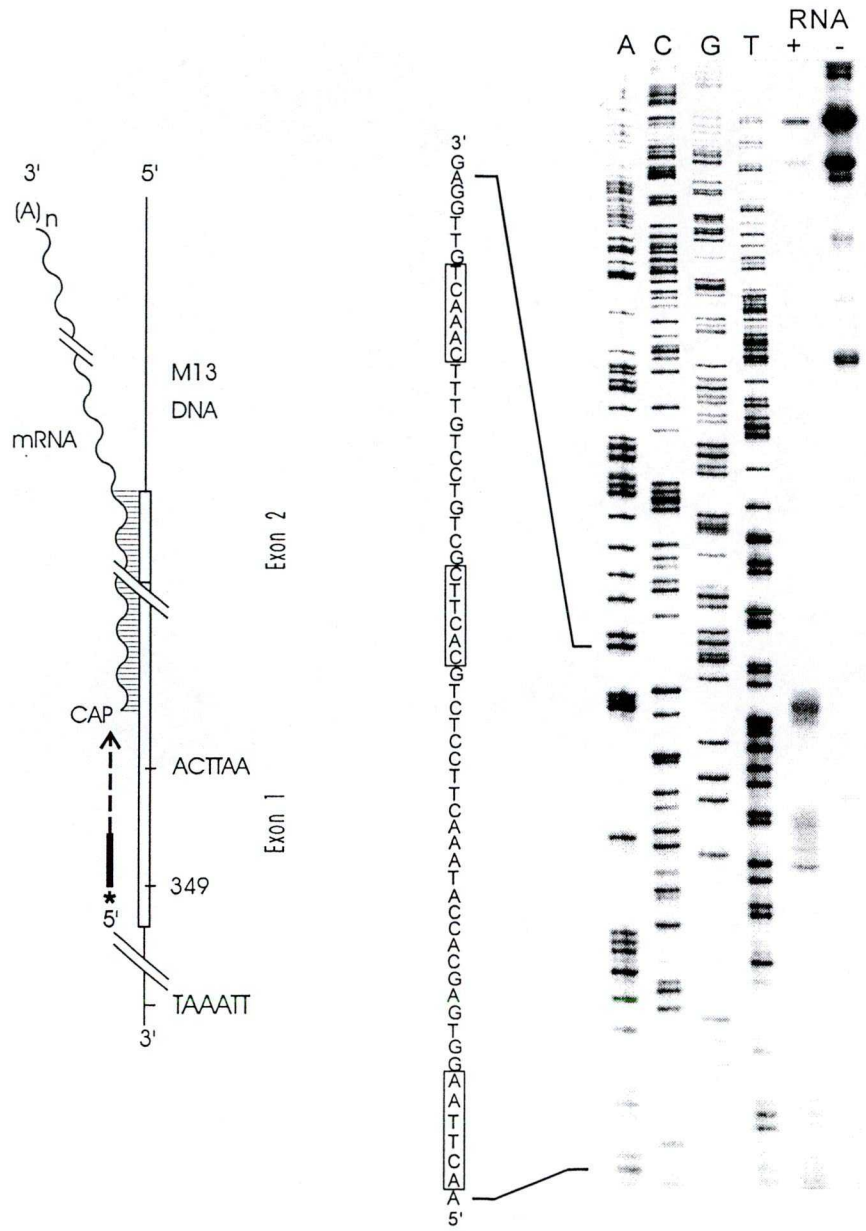

Figure 9 *Mapping of transcription start points by T4 DNA polymerase.* The diagram to the left depicts the experimental strategy. Solid bar with asterisk represents the end-labeled LP gene specific primer. 4 mg aliquots of poly(A)$^+$ RNA from sterna of day 14 chicken embryos (lane +) or tRNA (lane -) were annealed to the single-stranded DNA template before primer extension. The extension products were applied to a denaturing polyacrylamide gel simultaneously with a dideoxy sequencing reaction of the same DNA template and nonphosphorylated primer.

The relative activity of the two alternative promoters was analysed with a modified T4 DNA polymerase analysis. The modification was necessary because of multiple transcription start points and alternative splice site usage in the first exon creating a high level of heterogeneity. These restrained extension of gene specific primers as well as S1 mapping by creating multiple bands. Thus we developed a modified T4 DNA polymerase primer extension assay as shown schematically in figure 10.

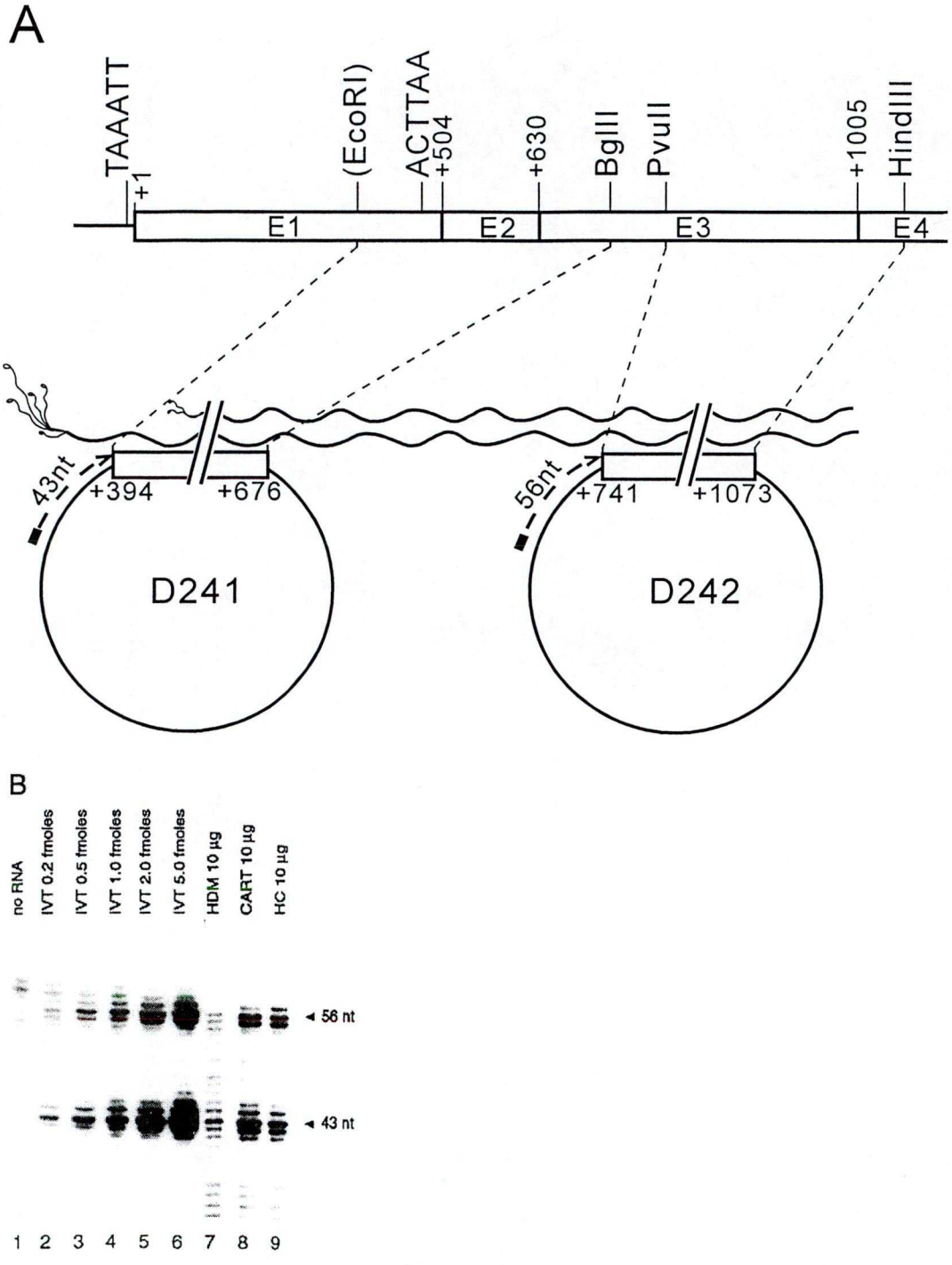

Figure 10: *Comparison of the relative activity of the two promoters in chondrocytes of different developmental stages.* (**A**) Schematic representation of the experimental design for the modified T4 DNA polymerase primer extension assay. Location of the promoters and the structure of the coding region of the chicken LP gene is shown for comparison on the top. Exons are numbered from E1 to E4. LP mRNA species (wavy lines) hybridise to the complementary regions carried by subclones D241 and D242, thereby blocking the extension of the –40 sequencing primer (solid bar) as shown at the bottom. While LP mRNA species transcribed from either of the two distinct promoters can anneal to D242 and result in the 56-nt primer product, only those transcribed from the upstream

promoter can anneal to subclone D241 and lead to synthesis of the 43-nt product. (**B**) Comparative analysis of the primer extension products. Indicated amounts of the *in vitro* transcript (IVT) (lanes 2-6), as well as total RNA samples, isolated from HDM culture (lane 7); embryo sternum cartilage (CART) (lane 8) and hypertrophic culture (HC) (lane 9) were subjected to primer extension assay using equimolar amounts of D241 and D242 templates. The primer extension products were separated on a sequencing gel along with a sequencing ladder and autoradiographed.

Two templates of single stranded M13mp19 were prepared harbouring non-overlapping fragments of similar size of the LP gene (fig. 10A). The D241 subclone carries a DNA fragment that is complementary to LP mRNA from +394, a position located 3' of the alternatively spliced region but 5' of the downstream promoter. The annealed LP mRNA blocks the elongation on the vector-insert border. Therefore, elongation of the -40 sequencing primer on this template would reflect only the mRNAs created from the first promoter. On the other hand, the second subclone D242 harbours a more downstream part of LP gene that would reflect the abundance of transcripts produced from both promoters when the same -40 sequencing primer is extended on this second template. Elongation products on both templates result in a single extension product. To serve as a quantification standard, RNA covering the LP gene from +342 to +1948 was synthesised *in vitro* and used parallelly.

We used phosphorimage analysis to quantify the extension. The assay was shown to be linear up to 5 fmoles of the LP mRNA standard for both 43nt and 56nt extension products (lanes 2-6). 10 μg of total RNA samples isolated from HDM, CEC and hypertrophic chondrocytes used for this analysis yielded 0.35 fmoles, 1.44 fmoles and 1.25 fmoles, respectively of LP mRNA transcribed from both promoters. Whereas, the amount of LP mRNA transcribed from the upstream promoter was 0.33 fmoles, 1.41 fmoles and 0.76 fmoles in the above mentioned samples, respectively. We calculated in the light of these values that in hypertrophic chondrocytes the activity of the second promoter was considerably increased. Therefore, we concluded that although the first promoter is more active throughout chondrogenesis, the relative utilisation of the downstream promoter increased substantially in hypertrophy.

This finding supported further the observation, that the LP gene can also be transcribed *in vivo* from the second promoter. Furthermore, there can be distinct regulatory mechanisms explaining the difference in promoter usage during chondrogenesis.

IV. Heterogeneity in mouse matrilin-2 gene transcripts

After proving that the closest relative of matrilin-2 is matrilin-1 within the von Willebrand Factor type-A like module superfamiliy, we made further database search. The search made in database of expressed sequence tags (dbEST) revealed that there is size heterogeneity in the human dbEST sequences reported. Moreover, the cDNA clones that were isolated by our laboratory for the mouse matrilin-2 showed the same size heterogeneity in their 3'-end. The sequence comparison of both dbEST data and our cDNA clones showed that the heterogeneity affected the unique module of matrilin-2. This region was further analysed by RT-PCR for the possible tissue specificity. Total RNA samples isolated from several permanent mouse cell lines (indicated on the top of figure 11, panel c) were reverse transcribed followed by a PCR amplification utilising nested primer pairs.

Two consecutive nested PCR reactions using two gene-specific primer pairs (mM2555N-mM3117R and mM2677N-mM2939R, fig. 11a), flanking the unique region were performed. Amplification of cDNA between the later primer pair yielded two distinct DNA fragments of 262 and 205 bp. Both splice forms were found in all cell types tested (Fig. 11, panel c). The detectable levels of matrilin-2 gene expression in all the samples analysed showed that matrilin-2 expression was not confined only to the connective tissue cell types, but it was observed also in myoblasts, and the epithelial and endothelial cell lines tested.

Possible mRNA heterogeneity within other parts of the translated region was further tested by RT-PCR analysis. In the SVEC endothelial and the rat chondrosarcoma cell lines, alternative splicing affected only the middle third of the unique module, but not the region encoding the coiled-coil, the vWFA and EGF-like modules (fig. 11, panel b).

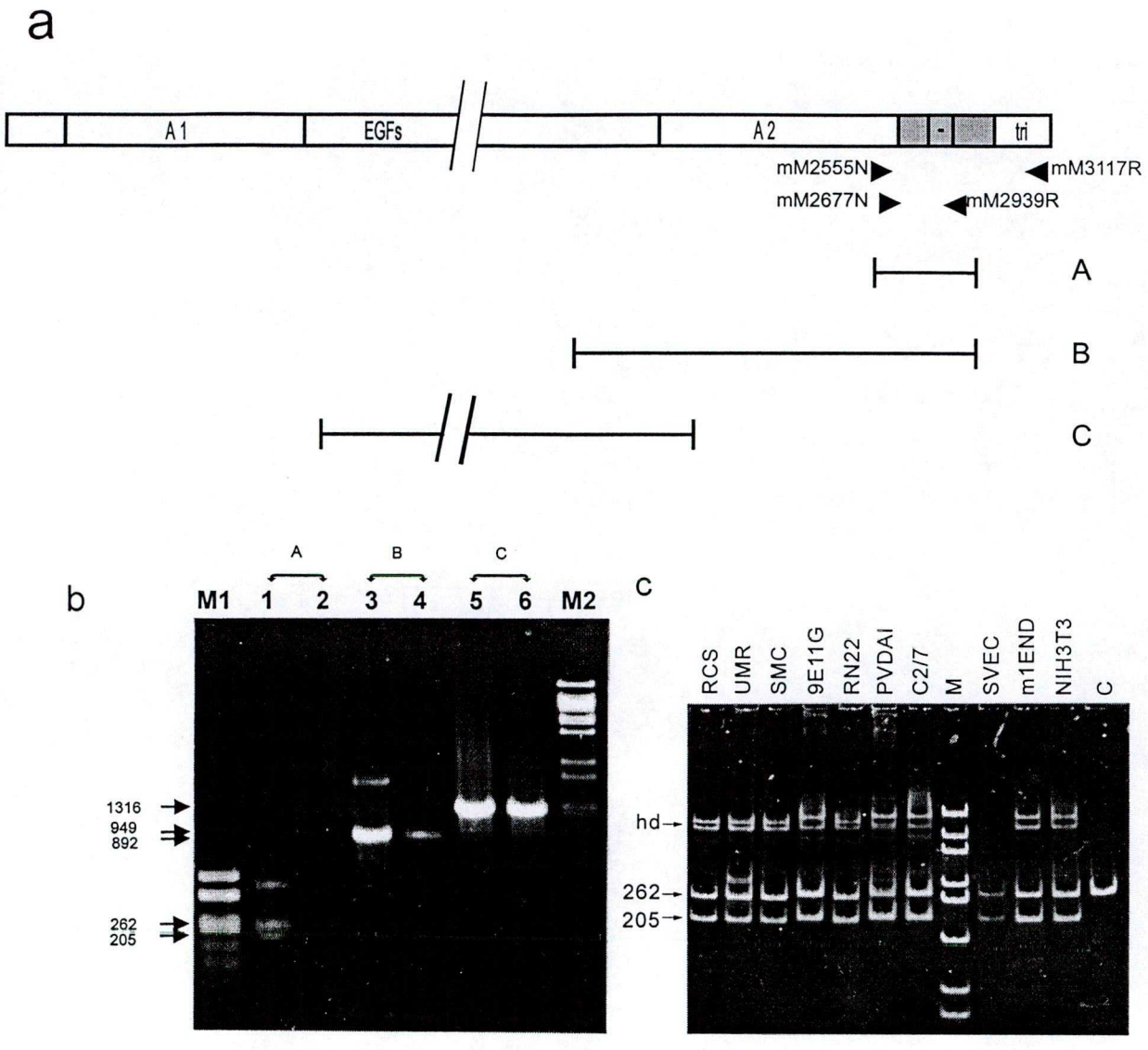

Figure 11 a. *RT-PCR analysis of alternative splicing of the mouse matrilin-2 mRNA.* Schematic representation of the amplified regions with respect to protein modules. A1 and A2 stand for the vWFA modules 1 and 2, respectively. **b.** *Detection of amplification products in the regions shown in panel (a).* Analysis of the possible heterogeneity in more upstream modules of mouse matrilin-2 by RT-PCR. Amplification products shown on panel (a) for the unique, A2 and EGF modules, done on the total RNA from samples SVEC lanes 1, 3 and 5 or rat chondrosarcoma cell lines lanes 2, 4, and 6 and separated electrophoretically in 1.2 % agarose gel. Product sizes are denoted on the left-hand side of the figure. **c.** *Comparison of various cell lines (origins explained in Materials and Methods) for the heterogeneity in the unique module.* Amplification products using the primer pairs shown in panel (a), were separated by elecrophoresis in a 5% polyacrylamide gel on the total RNA samples denoted on the top, and stained with ethidium-bromide for visualisation. The sizes of the alternatively spliced products are indicated with arrows and heteroduplexes as hd on the left-hand side of the figure.

V. Cloning the full-length cDNA for human matrilin-2

Previously, our laboratory identified several expressed sequence tags, encoding short regions of human matrilin-2. From the I.M.A.G.E. consortium depository one cDNA clone was acquired and sequenced (clone 1 in fig.12).

The clone encoded the vWFA2 module, the unique sequence and the α-helical coiled coil domain of human matrilin-2 [31]. To be able to clone the complete cDNA several human cell lines were compared, as possible sources for matrilin-2 mRNA. Using the I.M.A.G.E. clone cDNA as hybridisation probe in Northern hybridisation, RNA from a human larynx carcinoma cell line HEp-2 was selected for further cloning. The total RNA of this cell line was used as a template for RT-PCR amplification of EGF modules and flanking sequences using appropriate gene specific primer pairs designed on the basis of sequence 1 and 2, a short human EST entry. A 1.4 kb cDNA, covering the 10 EGF modules and flanking regions was amplified, cloned and sequenced (fig. 12, sequence 3). Mouse matrilin-2 specific oligonucleotide primers were selected, which map to a region highly conserved within the matrilin family, and the vWFA1 module sequence was amplified (fig. 12, sequence 4). The most 5'-end sequence (fig. 12, sequence 5) was generated by 5' RACE (as explained in Materials and Methods) , utilising primers designed on the basis of sequence 4. At least two clones were sequenced for each amplified region.

A composite nucleotide sequence of 3497 bp was obtained from the overlapping clones (fig. 13). The translation start site was assigned according to the more upstream one from two in-frame ATG triplets at positions 127-129 since flanking sequence of this triplet matched better to the consensus motif of translation initiation site [97] and moreover the functional translation start sites were very seldom preceded by in frame ATG triplets [97]. The 126 base pair cDNA located upstream of the first ATG most likely represents an untranslated sequence. The 3'-UTR of 503 nucleotides includes four putative polyadenylation signals but none of them are identical with AATAAA. The nucleotide sequence thus defines an open reading frame of 956 amino acids (fig. 13) and a protein precursor with predicted [101] M_w of 106,858. The first 23 residues correspond to a putative signal peptide and its cleavage would result in a secreted protein with a minimum M_w of 104,402. The modular structure of human matrilin-2 includes a pair of vWFA-like modules, 10 EGF repeats, a unique segment the α-helical COOH-end and a group of positively charged amino acids which are conserved in mouse (fig. 13). The deduced amino acid sequence contains only one N*X*(S/T) consensus sequence for potential N-

glycosylation at the end of first vWFA-like domain, unlike in mouse which has a second one in the unique segment [31]. Apart from this, the chondroitin sulphate attachment site, found in the unique segment of mouse matrilin-2, was not conserved.

A data bank search indicated that this protein was identified only in mouse. However, several human-expressed sequence tags were found in the GenBank EST division. We compared these sequences with sequences of our clones. There were discrepancies only in three positions (2398, 2399 and 2931) of the complete nucleotide sequence.

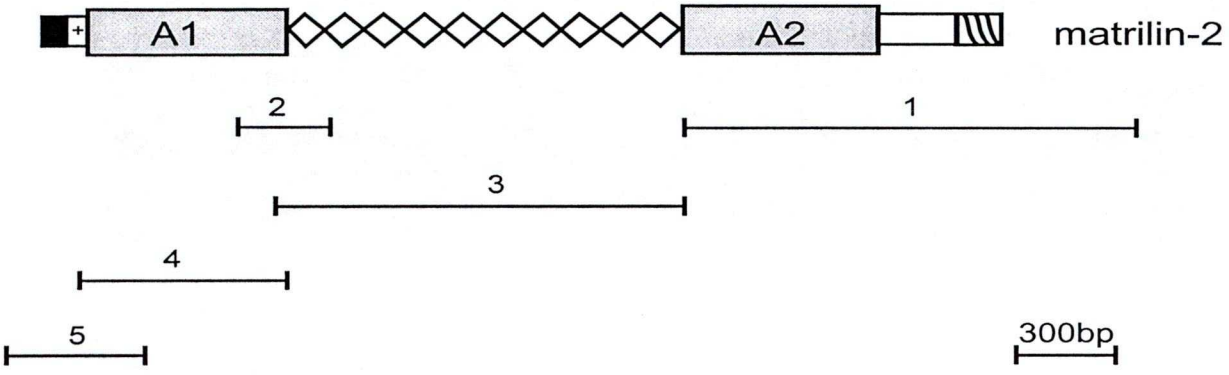

Figure 12: *Schematic strategy of cloning the complete cDNA sequence of human matrilin-2.* VWFA1 and 2 modules are represented as A1 and A2, respectively. **1,** acquired and sequenced cDNA clone from I.M.A.G.E. consortium depository. **2**, a short human EST entry sequence. **3**, an RT-PCR amplified, cloned and sequenced 1.4 kb cDNA, covering the 10 EGF modules and flanking regions. **4**, a 5' RACE cDNA clone of vWFA1 module sequence. **5**, a 5' RACE cDNA clone covering the most 5'-end sequence.

The human matrilin-2 precursor amino acid sequence, derived from the nucleotide sequences (fig. 13), shares a 86.5 % sequence identity and 93 % similarity with the mouse protein.

```
h NGGAACTCTGAAAAGGCGGGGCAGCGGGCCTGCAGCTCCTGGAGTTNAGGGAGACCCGGAAATCTCACCCTGCCCTCTTCTTGTGTTGTG 90

h TTTGTCACAGCCTTGCCCCTCTTGCTCGCCTTGAAAATGGAAAAGATGCTCGCAGGCTGCTTTCTGCTGATCCTCGGACAGATCGTCCTC 180
                                      M  E  K  M  L  A  G  C  F  L  L  I  L  G  Q  I  V  L  19
                                      .  .  .  .  .  V  .  .  L  .  M  L  G  Q  L  F  L  V
                  ↓
h CTCCCTGCCGAGGCCAGGGAGCGGTCACGTGGGAGGTCCATCTCTAGGGGCAGACACGCTCGGACCCACCCGCAGACGGCCCTTCTGGAG 270
h L  P  A  E  A  R  E  R  S  R  G  R  S  I  S  R  G  R  H  A  R  T  H  P  Q  T  A  L  L  E  48
m .  .  V  D  G  .  .  .  P  Q  A  .  F  P  .  .  .  .  .  V  .  M  Y  .  .  .  .  .  .  .  .

h AGTTCCTGTGAGAACAAGCGGGCAGACCTGGTTTTCATCATAGACAGCTCTCGCAGTGTCAACACCCATGACTATGCAAAGGTCAAGGAG 360
h S  S  C  E  N  K  R  A  D  L  V  F  I  I  D  S  S  R  S  V  N  T  H  D  Y  A  K  V  K  E  78
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  Y  .  .  .  .  .  .  .  .

h TTCATCGTGGACATCTTGCAATTCTTGGACATTGGTCCTGATGTCACCCGAGTGGGCCTGCTCCAATATGGCAGCACTGTCAAGAATGAG 450
h F  I  V  D  I  L  Q  F  L  D  I  G  P  D  V  T  R  V  G  L  L  Q  Y  G  S  T  V  K  N  E  108
m .  .  L  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h TTCTCCCTCAAGACCTTCAAGAGGAAGTCCGAGGTGGAGCGTGCTGTCAAGAGGATGCGGCATCTGTCCACGGGCACCATGACCGGGCTG 540
h F  S  L  K  T  F  K  R  K  S  E  V  E  R  A  V  K  R  M  R  H  L  S  T  G  T  M  T  G  L  138
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
h GCCATCCAGTATGCCCTGAACATCGCATTCTCAGAAGCAGAGGGGGCCCGGCCCCTGAGGGAGAATGTGCCACGGGTCATAATGATCGTG 630
h A  I  Q  Y  A  L  N  I  A  F  S  E  A  E  G  A  R  P  L  R  E  N  V  P  R  V  I  M  I  V  168
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  I  .  .  .  .  .

h ACGGATGGGAGACCTCAGGACTCCGTGGCCGAGGTGGCTGCTAAGGCACGGGACACGGGCATCCTAATCTTTGCCATTGGTGTGGGCCAG 720
h T  D  G  R  P  Q  D  S  V  A  E  V  A  A  K  A  R  D  T  G  I  L  I  F  A  I  G  V  G  Q  198
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  N  .  .  .  .  .  .  .  .  .  .  .  .

h GTAGACTTCAACACCTTGAAGTCCATTGGGAGTGAGCCCCATGAGGACCATGTNTTCCTTGTGGCCAATTTCAGCCAGATTGAGACGCTG 810
h V  D  F  N  T  L  K  S  I  G  S  E  P  H  E  D  H  V  F  L  V  A  N  F  S  Q  I  E  T  L  228
m .  .  L  .  .  .  .  .  A  .  .  .  .  .  .  .  K  .  .  .  .  .  .  .  .  .  .  S  .

h ACCTCCGTGTTCCAGAAGAAGTTGTGCACGGCCCACATGTGCAGCACCCTGGAGCATAACTGTGCCCACTTCTGCATCAACATCCCTGGC 900
h T  S  V  F  Q  K  K  L  C  T  A  H  M  C  S  T  L  E  H  N  C  A  H  F  C  I  N  I  P  G  258
m .  .  .  .  .  N  .  .  .  .  V  .  .  .  .  V  .  .  .  .  .  .  .  .  .  L  .  T  .  .

h TCATACGTCTGCAGGTGCAAACAAGGCTACATTCTCAACTCGGATCAGACGACTTGCAGAATCCAGGATCTGTGTGCCATGGAGGACCAC 990
h S  Y  V  C  R  C  K  Q  G  Y  I  L  N  S  D  Q  T  T  C  R  I  Q  D  L  C  A  M  E  D  H  288
m .  .  I  .  K  .  .  .  .  .  .  .  .  S  T  .  .  K  .  .  .  .  .  .  .  .  .  T  .  .  .

h AACTGTGAGCAGCTCTGTGTGAATGTGCCGGGCTCCTTCGTCTGCCAGTGCTACAGTGGCTACGCCCTGGCTGAGGATGGGAAGAGGTGT 1080
h N  C  E  Q  L  C  V  N  V  P  G  S  F  V  C  Q  C  Y  S  G  Y  A  L  A  E  D  G  K  R  C  318
m G  .  .  .  .  .  .  .  M  L  .  .  .  .  .  .  .  .  .  .  T  .  .  .  .  .  .  .  .  .  .

h GTGGCTGTGGACTACTGTGCCTCAGAAAACCACGGATGTGAACATGAGTGTGTAAATGCTGATGGCTCCTACCTTTGCCAGTGCCATGAA 1170
m V  A  V  D  Y  C  A  S  E  N  H  G  C  E  H  E  C  V  N  A  D  G  S  Y  L  C  Q  C  H  E  348
m T  .  M  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  E  S  .  .  .  .  R  .  .  .

h GGATTTGCTCTTAACCCAGATAAAAAAACGTGCACAAAGATAGACTACTGTGCCTCATCTAATCACGGATGTCAGCACGAGTGTGTTAAC 1260
h G  F  A  L  N  P  D  K  K  T  C  T  K  I  D  Y  C  A  S  S  N  H  G  C  Q  H  E  C  V  N  378
m .  .  .  .  .  S  .  .  .  .  .  S  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h ACAGATGATTCCTATTCCTGCCACTGCCTGAAAGGCTTTACCCTGAATCCAGATAAGAAAACCTGCAGAAGGATCAACTACTGTGCACTG 1350
m T  D  D  S  Y  S  C  H  C  L  K  G  F  T  L  N  P  D  K  K  T  C  R  R  I  N  Y  C  A  L  408
m A  Q  T  .  A  L  .  R  .  .  .  .  .  M  .  .  .  .  R  .  .  .  .  .  .  .  .  .  .  .  .

h AACAAACCGGGCTGTGAGCATGAGTGCGTCAACATGGAGGAGAGCTACTACTGCCGCTGCCACCGTGGCTACACTCTGGACCCCAATGGC 1440
h N  K  P  G  C  E  H  E  C  V  N  M  E  E  S  Y  Y  C  R  C  H  R  G  Y  T  L  D  P  N  G  438
m .  .  .  .  .  .  .  .  .  .  .  T  .  .  G  H  .  .  .  .  R  Q  .  .  N  .  .  .  .  .  .

h AAAACCTGCAGCCGAGTGGACCACTGTGCACAGCAGGACCATGGCTGTGAGCAGCTGTGTCTGAACACGGAGGATTCCTTCGTCTGCCAG 1530
h K  T  C  S  R  V  D  H  C  A  Q  Q  D  H  G  C  E  Q  L  C  L  N  T  E  D  S  F  V  C  Q  468
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  E  .  .  .  .  .  .

h TGCTCAGAAGGCTTCCTCATCAACGAGGACCTCAAGACCTGCTCCCGGGTGGATTACTGCCTGCTGAGTGACCATGGTTGTGAATACTCC 1620
h C  S  E  G  F  L  I  N  E  D  L  K  T  C  S  R  V  D  Y  C  L  L  S  D  H  G  C  E  Y  S  498
m .  .  .  .  .  .  .  .  D  .  .  .  .  .  .  .  A  .  .  .  .  .  .  N  .  .  .  .  .  .  .

h TGTGTCAACATGGACAGATCCTTTGCCTGTCAGTGTCCTGAGGGACACGTGCTCCGCAGCGATGGGAAGACGTGTGCAAAATTGGACTCT 1710
m C  V  N  M  D  R  S  F  A  C  Q  C  P  E  G  H  V  L  R  S  D  G  K  T  C  A  K  L  D  S  528
m .  .  .  T  .  K  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h TGTGCTCTGGGGGACCACGGTTGTGAACATTCGTGTGTAAGCAGTGAAGATTCGTTTGTGTGCCAGTGCTTTGAAGGTTATATACTCCGT 1800
h C  A  L  G  D  H  G  C  E  H  S  C  V  S  S  E  D  S  F  V  C  Q  C  F  E  G  Y  I  L  R  558
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h GAAGATGGAAAAACCTGCAGAAGGAAAGATGTCTGCCAAGCTATAGACCATGGCTGTGAACACATTTGTGTGAACAGTGATGACTCATAC 1890
h E  D  G  K  T  C  R  R  K  D  V  C  Q  A  I  D  H  G  C  E  H  I  C  V  N  S  D  D  S  Y  588
m D  .  .  .  .  .  .  .  .  .  .  .  D  V  N  .  .  .  .  .  L  .  .  .  .  G  E  .  .
```

```
h ACGTGCGAGTGCTTGGTGGGATTCCGGCTCGCTGAGGATGGGAAACGCTGCCGAAGGAAGGATGTCTGCAAATCAACCCACCATGGCTGC 1980
h T  C  E  C  L  V  G  F  R  L  A  E  D  G  K  R  C  R  R  K  D  V  C  K  S  T  H  H  G  C   618
m V  .  K  .  .  E  .  .  .  .  .  .  .  .  .  .  .  .  .  N  .  .  .  .  .  Q  .  .  .

h GAACACATTTGTGTTAATAATGGGAATTCCTACATCTGCAAATGCTCAGAGGGATTTGTTCTAGCTGAGGACGGAAGACGGTGCAAGAAA 2070
h E  H  I  C  V  N  N  G  N  S  Y  I  C  K  C  S  E  G  F  V  L  A  E  D  G  R  R  C  K  K   648
m .  .  M  .  .  .  .  .  .  .  .  L  .  R  .  .  .  .  .  .  .  .  .  .  .  K  H  .  .  R

h TGCACTGAAGGCCCAATTGACCTGGTCTTTGTGATCGATGGATCCAAGAGTCTTGGAGAAGAGAATTTTGAGGTCGTGAAGCAGTTTGTC 2160
h C  T  E  G  P  I  D  L  V  F  V  I  D  G  S  K  S  L  G  E  E  N  F  E  V  V  K  Q  F  V   678
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  .  .  H  .  .

h ACTGGAATTATAGATTCCTTGACAATTTCCCCCAAAGCCGCTCGAGTGGGGCTGCTCCAGTATTCCACACAGGTCCACACAGAGTTCACT 2250
h T  G  I  I  D  S  L  T  I  S  P  K  A  A  R  V  G  L  L  Q  Y  S  T  Q  V  H  T  E  F  T   708
m .  .  .  .  .  .  .  A  V  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  R  .  .  .  .

h CTGAGAAACTTCAACTCAGCCAAAGACATGAAAAAAGCCGTGGCCCACATGAAATACATGGGAAAGGGCTCTATGACTGGGCTGGCCCTG 2340
h L  R  N  F  N  S  A  K  D  M  K  K  A  V  A  H  M  K  Y  M  G  K  G  S  M  T  G  L  A  L   738
m .  .  G  .  S  .  .  .  E  .  .  .  .  .  T  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h AAACACATGTTTGAGAGAAGTTTTACCCAAGGAGAAGGGGCCAGGCCTTTTTCCACAAGGGTGCCCAGAGCAGCCATTGTGTTCACCGAC 2430
h K  H  M  F  E  R  S  F  T  Q  G  E  G  A  R  P  F  S  T  R  V  P  R  A  A  I  V  F  T  D   768
m .  .  .  .  .  .  .  .  .  .  .  V  .  .  .  .  .  L  .  .  Q  .  .  .  V  .  .  .  .  .

h GGACGGGCTCAGGATGACGTCTCCGAGTGGGCCAGTAAAGCCAAGGCCAATGGTATCACTATGTATGCTGTTGGGGTAGGAAAAGCCATT 2520
h G  R  A  Q  D  D  V  S  E  W  A  S  K  A  K  A  N  G  I  T  M  Y  A  V  G  V  G  K  A  I   798
m .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h GAGGAGGAACTACAAGAGATTGCCTCTGAGCCCACAAACAAGCATCTCTTCTATGCCGAAGACTTCAGCACAATGGATGAGATAAGTGAA 2610
h E  E  E  L  Q  E  I  A  S  E  P  T  N  K  H  L  F  Y  A  E  D  F  S  T  M  D  E  I  S  E   828
m .  .  .  .  .  .  .  .  .  .  .  I  D  .  .  .  .  .  .  .  .  .  .  .  G  .  .  .  .

h AAACTCAAGAAAGGCATCTGTGAAGCTCTAGAAGACTCCGATGGAAGACAGGACTCTCCAGCAGGGGAACTGCCAAAAACGGTCCAACAG 2700
h K  L  K  K  G  I  C  E  A  L  E  D  S  D  G  R  Q  D  S  P  A  G  E  L  P  K  T  V  Q  Q   858
m .  .  .  E  .  .  .  .  .  .  .  .  G  .  .  .  .  .  A  .  W  D  .  .  Q  Q  A  H  .

h CCAACAGAATCTGAGCCAGTCACCATAAATATCCAAGACCTACTTTCCTGTTCTAATTTTGCAGTGCAACACAGATATCTGTTTGAAGAA 2790
h P  T  E  S  E  P  V  T  I  N  I  Q  D  L  L  S  C  S  N  F  A  V  Q  H  R  Y  L  F  E  E   888
m .  .  .  P  .  .  .  .  .  K  .  K  .  .  .  .  .  .  .  .  .  .  .  .  F  .  .  .  .  .

h GACAATCTTTTACGGTCTACACAAAAGCTTTCCCATTCAACAAAACCTTCAGGAAGCCCTTTGGAAGAAAAACACGATCAATGCAAATGT 2880
h D  N  L  L  R  S  T  Q  K  L  S  H  S  T  K  P  S  G  S  P  L  E  E  K  H  D  Q  C  K  C   918
m .  .  .  S  .  .  .  .  .  .  F  .  .  .  .  S  .  .  N  .  .  .  .  S  Q  .  .  .  .  .

h GAAAACCTTATAATGTTCCAGAACCTTGCAAACGAAGAAGTAAGAAAATTTACACAGCGCTTAGAAGAAATGACACAGAGAATGGAAGCC 2970
h E  N  L  I  M  F  Q  N  L  A  N  E  E  V  R  K  F  T  Q  R  L  E  E  M  T  Q  R  M  E  A   948
m .  .  .  .  L  .  .  .  V  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

h CTGGAAAATCGCCTGAGATACAGATGAAGATTAGAAATCGCGACACATTTGTAGTCATTGTATCACGGATTACAATGAACGCAGTGCAGA 3060
h L  E  N  R  L  R  Y  R  *
m .  .  .  .  .  K  .  .  *

h GCCCCAAAGCTCAGGCTATTGTTAAATCAATAATGTTGTGAAGTAAAACAATCAGTACTGAGAAACCTGGTTTGCCACAGAACAAAGACA 3150
h AGAAGTATACACTAACTTGTATAAATTTATCTAGGAAAAAAATCCTTCAGAATTCTAAGATGAATTTACCAGGTGAGAATGAATAAGCTA 3240
h TGCAAGGTATTTTGTAATATACTGTGGACACAACTTGCTTCTGCCTCATCCTGCCTTAGTGTGCAATCTCATTTGACTATACGATAAAGT 3330
h TTGCACAGTCTTACTTCTGTAGAACACTGGCCATAGGAAATGCTGTTTTTTTGTATTGGACTTTACCTTGATATATGTATATGGATGTAT 3420
h GCATAAAATCATAGGACATATGTACTTGTGGAACAAGTTGGATTTTTTATACAATATTAAAATTCACCACTTCAGAG
```

Figure 13: *Nucleotide and deduced amino acid sequences of human matrilin-2.* Complete sequence of human matrilin-2 precursor as determined from overlapping cDNA clones of Fig. 12. *Arrowhead,* predicted propeptidase cleavage site. Positively charged amino acids at the terminus of the secreted protein, *bold face*; potential *N*-linked glycosylation site, *bold and underlined;* putative polyadenylation signals, *underlined.*

VI. Expression of matrilin-2 gene in various human cell lines and heterogeneity in its transcript

Due to the high homology between human and mouse, we suspected the presence of the same alternative splicing detected for the mouse matrilin-2. Application of the same RT-PCR analysis strategy as in the case of mouse matrilin-2, revealed both presence and heterogeneity of the gene transcript.

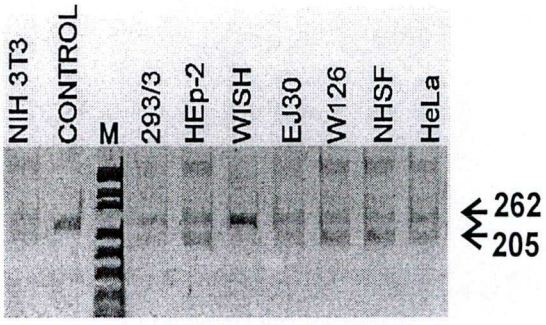

Figure 14: *RT-PCR analysis of heterogeneity within the unique region of human matrilin-2.* Total RNA samples of several human cell lines are indicated on the top. NIH 3T3 mouse fibroblast sample was run side-by-side for comparison. Control amplification of the longer isoform was done on the pCRP12 cDNA clone [31] as a template. M, pUC 12 Msp I DNA size marker. Same sizes of products with similar relative abundance were obtained as in the mouse matrilin-2 RT-PCR analysis. Only in the WISH sample, the longer form has a higher relative abundance than the shorter.

In all the cell lines examined, expression of the gene was at a detectable level. Fibroblastic cell lines EJ30, WI26, NHSF, as well as the epithelioid cell line WISH, embryonic kidney (239/3) and epidermoid carcinoma (HEp-2), all had expressed the matrilin-2 gene, as in case of the mouse cell lines and rat osteogenic sarcoma UMR-106 examined previously. In the present experiment, mRNA in the NIH 3T3, mouse fibroblastic cell line sample (fig. 14, lane 1) was detected and compared to human cell lines. In all samples, the same heterogeneity was detected as in the case of mouse. The relative abundance of the two splice forms was similar with the exception of WISH sample. In this case, the longer form had considerably higher abundance over the shorter.

DISCUSSION

In the view of gene expression patterns of several cartilage protein genes in various chondrogenic cultures representing the chondrocyte development, we concluded that these *in vitro* cultures were capable of reproducing the *in vivo* chondrogenesis. Matrilin-1 gene expression was more restricted as compared to the cartilage specific α1(II) collagen gene as it was reported in *in vivo* systems [25, 26, 30]. Activation of these genes during *in vitro* chondrogenesis in HDM culture followed the same sequential order, as reported previously by other laboratories based on immunostaining and *in situ* hybridisation in developing mouse and chicken limb [15, 30]. These data were further supported by application of a different system defined by Quarto et. al. [3]. The system is described on that chondrocytes from tibia of day 17 chicken embryos flatten and cease expression of cartilage markers when cultured in monolayers. The cells, however, resume the chondrocytic phenotype if cultured further in suspension, preventing attachment to the culture dish. Csanád Bachrati utilised this culturing system for monitoring gene expression pattern of cartilage proteins and especially the expression pattern of matrilin-1 in stage II hypertrophic cells. As well as confirming the same sequential onset of the cartilage specific genes, parallel expression of matrilin-1 with α1(X) collagen were also detected when the α1(X) collagen was predominant [17].

Although it is shown that both HDM and CEC cultures consisted of stage I proliferative chondrocytes, significant differences between their expression pattern were detected. The steady state levels of cartilage specific mRNAs were more than an order of magnitude lower in HDM than in CEC cultures, while the type VI collagen mRNA reached a plateau in HDM cultures. Differences were also observed between the two cultures regarding the relative abundance of the β-actin mRNA as well as the type II collagen mRNA variants synthesised [17]. Therefore, we conclude that stage Ia chondrocytes predominate in HDM culture, whereas stage Ib chondrocytes are the major cell components of CEC cultures.

Ultimately, these cultures representing different stages of chondrogenesis were utilised as *in vitro* models for investigating the expression pattern of the genes encoding matrilin-1 and link protein.

Heterogeneity of matrilin-1 transcript

Elaborate analyses of the 3'-UTR of the matrilin-1 gene revealed the reason of heterogeneity in its transcript, which was detected first with the Northern hybridisation analysis. By performing consecutive hybridisations with two non-overlapping probes covering either only the translated region or a part of 3'-UTR, we showed that the heterogeneity must lie in the 3'-UTR. Amplification of the 3'-end of the matrilin-1 mRNA with RT-PCR and 3' RACE yielded multiple products due to the presence of multiple mRNA isoforms. Cloning and sequencing of these products unveiled the identities of them. It is shown that both alternative polyadenylation site usage and alternative splicing play a role in the generation of several mRNA isoforms. Usage of an alternative, upstream poly(A) site gives rise to mRNA with shorter 3'-UTR without any change in the protein product. However, these alternative isoforms may have importance in gene expression. As it is discussed [61], 3'- UTR has a great importance in the regulation of gene expression since it may alter the stability, translation or transport of mRNA. In the transcription units where multiple poly(A) sites are utilised, the promoter-proximal site would be preferred to a more distal one since it is transcribed first. However, a very efficient downstream polyadenylation site would overcome the favourable position of the upstream one. In this case, the relative strengths of the two sites have a decisive role for selecting the site at which the RNA will be processed. The choice between poly(A) addition sites is defined by the number and sequence of polyadenylation signals, the distance, nucleotide sequence and length of the downstream GU-rich sequence element and, although non-essential for processing the poly(A) site, a U-rich upstream element acting as an enhancer for polyadenylation. In the case of matrilin-1, neither of the sites has the downstream GU-rich sequence element immediately after the cleavage sites. However, the distal one has the U-rich upstream element that must be increasing the efficiency of this site as shown in figure 4. Apparently this is one of the mechanisms which renders the formation of a more stable complex at the downstream location and as a result, precludes the use of the upstream site. Nevertheless, this does not allow a fine discrimination between the two sites but rather allows for site specific factors that might developmentally or in a cell type specific manner modulate the use of the inefficient site. Therefore, it is tempting to speculate that the expression of the shorter mRNA isoform, utilising the upstream poly(A) site which was detected in the proliferating stage (stage I), might have a developmental stage specificity. Moreover, the usage of the downstream site provides a considerably long 3'-UTR that may harbour many important signals controlling translation initiation. Thus the relative amount of the transcripts with shorter

and longer 3'-UTR does not automatically mean that those isoforms are utilised by the same relative efficiency in translation.

The mRNA stability of the two matrilin-1 mRNA isoforms created by alternative polyadenylation should also be discussed since the upstream poly(A) site shortens the 3'-UTR by almost 1 kb. In this 1 kb, several AURE destabilising elements are located (fig. 4), which might affect the stability of the longer form of mRNA [66]. AUREs provide opportunity for exogenous factors including cell growth and differentiation ones to affect gene expression. The AUREs present in the 3'-UTR of matrilin-1, might target the mRNA for rapid and selective degradation. This selective degradation in turn might regulate the gene expression by influencing the steady-state level of the mRNA and the rates at which the mRNA disappears following transcriptional repression and accumulates following transcriptional induction.

The alternative splicing detected also in the matrilin-1 3'-UTR, gives rise to a different protein product as it is shown schematically in figure 6. Since the translation stop signal was spliced out, the protein was altered in its carboxyterminus, which contains the coiled-coil domain crucial for the homotrimer structure of matrilin-1 protein. The assembly of the three identical polypeptide chains determines homotrimer structure of matrilin-1 by forming a three-stranded α-helical coiled-coil. Similarly to the alternatively polyadenylated form, the expression of the other minor mRNA isoform generated by alternative splicing was also detectable in stage I. However, alternative usage of the 3' splice site for the last intron leads to a predicted protein sequence, of which the first half coiled-coil domain is unchanged, but probably too short for coiled-coil formation and the second half cannot form α-helix due to absence of apolar residues in the conserved a and d positions of the heptads (fig. 15) where they are essential for the stability of the helix association [95]. Thus, this isoform cannot form a homotrimer structure. However, interaction with other molecules may enable this isoform to be included into a heterooligomer. Therefore, it is questionable whether the protein product encoded by the alternatively spliced mRNA form can be functional and thus, has a biological importance.
Therefore, it cannot form a homotrimer structure. However, interaction with other molecules may enable this isoform a heteromer structure.

FORM A and B

a	b	c	d	e	f	g
	Glu	Glu	Asp	Pro	Cys	Glu
Cys	Lys	Ser	**Ile**	Val	Lys	Phe
Gln	Thr	Lys	Val	Glu	Glu	Leu
Ile	Asn	Thr	**Leu**	Gln	Gln	Lys
Leu	Glu	Ala	**Val**	Ala	Lys	Arg
Ile	Glu	Ala	**Leu**	Glu	Asn	Lys
Ile	Ile					

FORM C

a	b	c	d	e	f	g
	Glu	Glu	Asp	Pro	Cys	Glu
Cys	Lys	Ser	**Ile**	Val	Lys	Phe
Gln	Thr	Lys	Val	Glu	Glu	Leu
Ie	Asn	Thr	**Leu**	Gln	Gln	Lys
His	Pro	Gln	**Leu**	Lys	Leu	Leu
Pro	Arg	Thr	Gln	Glu	Thr	Ala
Ala	Glu	Val	**Val**	Ala	Asp	Glu
Ser	Gly	Ala	**Phe**	Ile	Thr	Ala
Val	Thr	Thr	Ser	Ser	Phe	Leu
Ser	Pro	Pro	**Leu**	Ser	Thr	Leu
Thr	Ala	Ser	Gln	Pro	Ile	Ser
Thr	Ser	Phe	Arg	His	Pro	Phe
Leu	Asp	Ser	Gln	Ser	Trp	Gln
Glu	Lys					

Figure 15: *The end of matrilin-1 versus the end of open reading frame in the alternatively spliced transcript.* Deduced amino acid sequences of mRNA isoforms (forms A, B and C as explained schematically in figure 4) of matrilin-1, differing in their 3' end. Hydrophobic amino acids in positions a and d are printed in boldface, charged and polar residues in positions e and g are underlined.

However, from a mechanistic point of view, the alternative splicing process might be quite interesting since the terminal intron of matrilin-1 is a minor AT-AC-type intron, which is spliced out by the U-12 type splicesome [47]. As the only terminal AT-AC intron identified to date, it is already interesting for elucidation of the factors playing role in the U-12 splicesome mechanism, which is yet to be fully resolved. The low efficiency of utilisation detected for this alternative 3' splice site can be due to its position as it is localised in a more downstream position than the first one. Secondly, the putative branch site deviates from the consensus of the AT-AC type introns (table 2), because the totally conserved nucleotide A, which is crucial for the splicing mechanism, was shifted upstream by one basepair, adding to the inefficiency of the splice site. Besides, the distance of the branchpoint was longer than the reported ones for U-12 type AT-AC introns. All these factors render the downstream splice site unfavourable for the splicing mechanism. Nevertheless, since alternative splicing which involves the minor splicesome has not been reported, it is worthwhile to make further analysis on this process. As it is discussed by Tarn and Steitz [96] the low abundance of splicing apparatus for the removal of AT-AC introns can be rate-limiting for the maturation of their host mRNAs and thereby may contribute importantly to the regulation of these genes. Therefore, the 3'-end of matrilin-1

provides a versatile tool for future studies to reveal not only the mechanisms of splicing but also its influence on transcription and 3'-end formation.

Table 2.*Comparison of splice sites between the last intron n chicken matrilin-1 gene and other U-12 type AT-AC introns*

Intron	5' splice site	Putative branch site	3' splice site	Dist. nt[a]
chicken mtr-1 intron G	aat|atatccttt	ctccttaactct	tcactcac|tgg	12
alt.spliced form intron H[b]	aat|atatccttt	ctcctcagctct	ccttccac|ac	17
consensus[c]	ooo|atatccttt	ttccttracycy	oooooyac|ooo	10-15
conservation[d]	ooo|++++++	+++	+++|ooo	_

[a]: distance between branch point and 3' splice site.

[b]. shows 5' splice site with intron G

[c]: Sharp and Burge, 1997 Cell, 91, 875-879, Consensus sequences of U-12 type AT-AC introns

[d] o, no conservation ; +, conserved

HETEROGENEITY IN LP 5'UTR

Alternative promoter usage described in this work at the 5'-end of the chicken LP gene, contributes to the high level of heterogeneity of the transcript in addition to alternative splicing in the first exon. The modulation of gene expression can be enhanced with the usage of alternative promoters. Moreover, utilisation of alternative promoters, in combination with alternative splicing, leads to a higher structural and functional diversity of mRNA isoforms [32].

The preferential usage of the upstream promoter, detected by modified T4 DNA polymerase analysis, raises the question which possible mechanisms play a role in this regulation process. 5'-UTR mRNA isoforms may regulate LP gene expression by two possible mechanisms. First, several structural features in the 5'-UTR may exert a negative effect on the efficiency of translation. Kozak has examined factors in the 5'-UTRs that may promote efficient translation [97] in which it is stated that: 1) most

eukaryotic mRNAs have a short 5'-UTR and 2) there are no AUGs upstream of the translation initiation site of the major ORF. Between the upstream promoter and the translation start site of LP, the distance is long and the region contains two short ORFs. As reported in other genes [98], these short upstream ORFs could lead to reduced protein output through a mechanism of abortive translation. Taking this fact into consideration, usage of the alternative splice sites in the transcript initiated from the upstream promoter increases the translational efficiency by splicing out the two initiation and a termination codon precluding formation of short, upstream ORFs.

Second mechanism includes the possibility that differences in the secondary structures of the 5'-UTRs of the distinct LP transcripts may influence translation efficiency. It is known that secondary structure formation of the 5'-UTR with a Gibbs free energy of formation (ΔG) of less than -50 kcal/mole can impair the translation [99]. By using the algorithms developed by Dr. M. Zuker [100] the 5'-UTRs of LP mRNA isoforms were analysed for their tendency to form secondary structure, thus, inhibiting the translation by rendering the sequence to be stable which is upstream of AUG. Three mRNA isoforms were analysed, varying in their 5'-UTR. The first was the unspliced and the longest mRNA initiated from the upstream promoter, which had free energy of -141.2 kcal/mole. The second 5'-UTR initiated also from the upstream promoter, but the middle half removed by splicing, had a secondary structure with -45.2 kcal/mole energy. Lastly, the transcript initiated form the downstream promoter had -16.4 kcal/mole of free energy. The alternative splicing considerably decreases the stability of the transcript initiated from the upstream promoter, therefore, increasing the translational efficiency. However, among the three RNA isoforms compared, the one, starting from the downstream promoter should have the highest translational efficiency, because, 5'-UTR formed the least stable secondary structure.

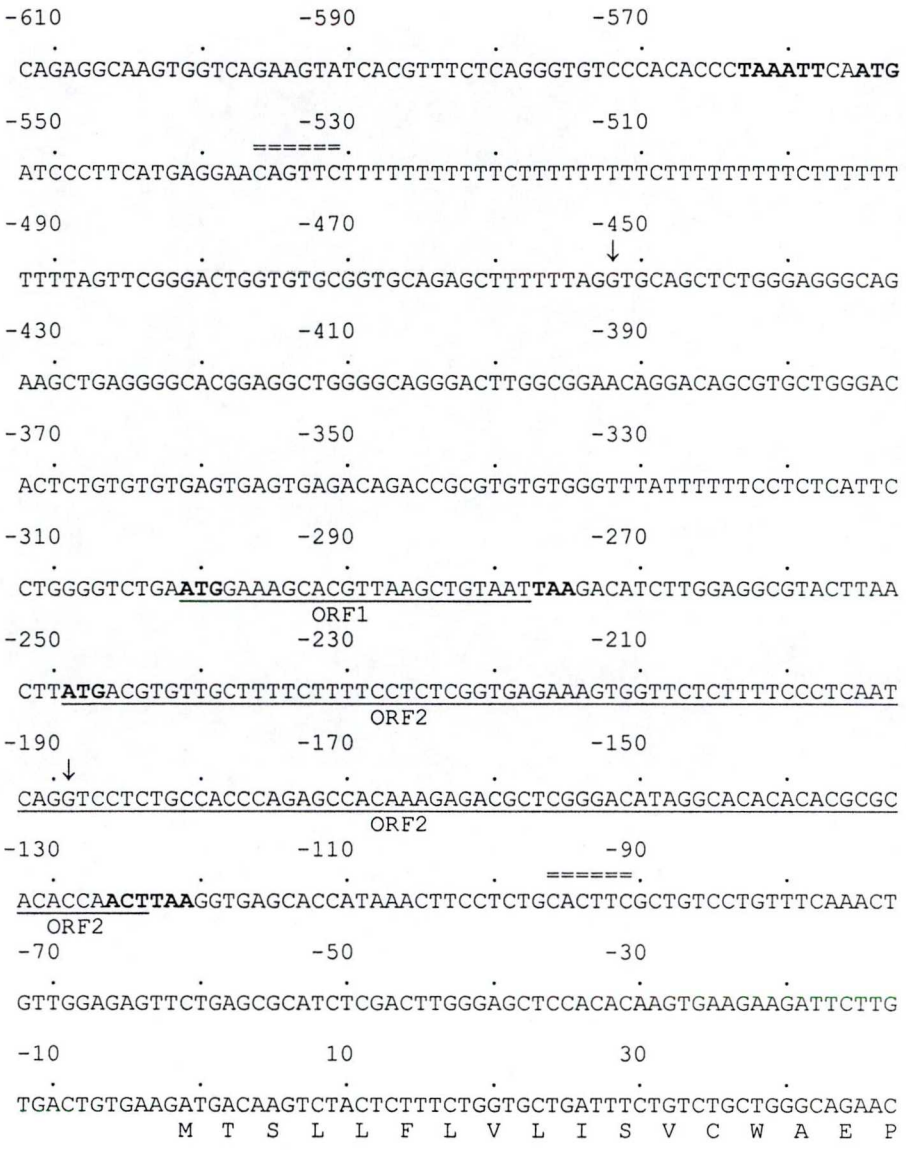

Figure 15: *Nucleotide sequence of 5' flanking region, the first exon and part of the second exon of the chicken LP gene harbouring the two distinct promoters.* Positions are given from the translation start site within exon 2. Translated region is denoted with one letter amino acid sequence. The TATA-like boxes are bold and the transcription start sites are overlined (==). Upstream ATG and TAA codons are marked as bold and the two open reading frames defined by these codons are underlined and denoted as ORF1 and ORF2. Alternatively spliced out region is marked with arrows (↓).

With the identification of the downstream promoter region, we were able to test their relative activity in various developmental stages of chondrogenesis with a modified T4 DNA polymerase analysis. This assay also proved the *in vitro* transcriptional activity of the downstream promoter. The upstream promoter was shown to be more active through the chondrogenic stages analysed. However, the low activity of the downstream promoter increased in the more advanced stage. It is probable that the *cis*-acting elements within

the region upstream of the upstream promoter support strong promoter activity even though the conformation of the 5'-UTR of mRNA from the downstream promoter renders a higher probability for efficient translation.

Studies of a number of diverse gene systems showed that alternative promoter usage resulting in alternative transcripts is an important evolutionary mechanism creating diversity in the regulatory control of genes [32]. In these systems, alternative promoter usage has been shown to be an important mechanism for regulating either tissue- or cell type specific expression. The observed high heterogeneity created in the 5'-UTR of LP in combination with alternative splicing, can provide an increased capacity in response to particular cellular or metabolic conditions.

Heterogeneity of mouse and human matrilin-2 transcripts in cell lines

Gene expression studies of mouse and human matrilin-2 genes revealed that they were not expressed only in connective tissue cell types. Mouse matrilin-2 gene was expressed in myoblasts, epithelial and endothelial cell lines tested (submitted data). Besides, matrilin-2 gene transcript was detected in fibroblastic, epitheloid and embryogenic human cell lines. Heterogeneity detected during this study in mouse and human matrilin-2 gene was due to alternative splicing in the unique module. Mapping of mouse matrilin-2 gene performed in our laboratory revealed alternative 3'-splice site selection in this process (unpublished data by L. Mátés and F. Deák). All the cell lines analysed and compared showed that the heterogeneity lies in the unique module and it is due to alternative splicing of a 57 nucleotide long region. Further RT-PCR analysis within the translated region of matrilin-2 gene in the SVEC endothelial cells and rat chondrosarcoma did not reveal heterogeneity in the coiled coil, the vWFA and EGF modules (fig. 11b). Since the function of matrilin-2 and the importance of the unique module have not been revealed yet, further studies are necessary to reveal the possible functional importance of the heterogeneity detected in its transcript. Nevertheless, the unpaired cysteine present in the region of unique module which was affected by alternative splicing may play a role in forming structural variability of the protein.

The differential expression detected for matrilin-1 and matrilin-2 genes in chicken and mouse tissues and cell lines is also noteworthy of further investigation. The transcript and protein product of matrilin-1 is confined to cartilage tissues except the articular cartilage [23] while highly homologous matrilin-2 is found in various tissues, as well as in articular cartilage but not in any other cartilaginous tissues [31]. Although this is a very preliminary result and conclusion, it is also quite interesting to define the function of these genes which would serve as markers of specific tissues.

CONCLUSIONS

Various cell cultures were established, representing consecutive stages of chondrocyte differentiation. The steady state mRNA levels for link protein and matrilin-1 were determined and compared to those of collagens α1(II), α2 (VI) and α(X), as standards. During cell differentiation the embryo limb bud mesenchyme cells cultured in high density reached predominantly the stage Ia of chondrogenesis, characterised by relatively high steady state level for α2 (VI) mRNA, but low level of mRNA for α1 (II), LP and matrilin-1. Cultured cells from sterna of 14-day-old embryos consisted predominantly of stage Ib chondrocytes, showing high levels for cartilage proteins, but low amount of α2 (VI) collagen mRNA. Both in HDM culture and suspension culture of dedifferentiated tibial chondrocytes, expression of the matrilin-1 gene followed collagen II and LP mRNA synthesis. During transition from stage Ib to stage II, matrilin-1 mRNA synthesis continued after expression of the gene for collagen α1 (X), a marker of hypertrophic chondrocytes.

In addition to the major, 3.4 kb mRNA for chicken matrilin-1, a minor, 2.1 kb species was found. Detailed analysis showed that in generation of the shorter mRNA species both alternative polyadenylation site usage and alternative splicing are involved. Two shorter polyadenylated froms were identified which were utilising an alternative poly(A) signal 1050 bp upstream of the previously identified one. This leads to a shorter 3'-untranslated region length without changing the protein sequence. In the latter event another 3' splice site (of AC type) was selected, instead of the 3' splice site of the last intron. Both introns are bordered by AT and AC sequences and spliced by U11-U12 type splicesome. The alternatively spliced product potentially codes for a protein chain with altered COOH-end of the protein.

Approximately 450 bp downstream of the transcription start site identified previously, a second, downstream promoter of the chicken LP gene was identified. By a modified T4 DNA polymerase assay, the upstream promoter was shown to be favoured throughout chondrocyte differentation, but activity of the downstream promoter relatively increased in the hypertrophic chondrocytes. A hypothesis is given to explain the biological role of two promoters together with extensive alternative splicing within the 5'-untranslated region.

The cDNA for the complete human matrilin-2 gene was cloned and sequenced. Deduced amino acid sequences of mouse and human matrilin-2 share a high homology. RT-PCR analysis of the human and mouse RNA samples showed alternative splicing, affecting the middle portion of the unique sequence of matrilin-2. Both mRNAs splice variants were found in all cell lines studied.

REFERENCES

1. Erlebacher, A., Filvaroff, E. H., Gitelman, S. E., Derynck, R. (1995) Toward a molecular understanding of skeletal development. *Cell* **80**, 371-378.
2. Heinegård, D., Oldberg, A. (1989) Structure and biology of cartilage and bone matrix non-collagenous proteins. *FASEB J.* **3**, 2042-2051.
3. Quarto, R., Dozin, B., Bonaldo, P., Cancedda, R., Colombatti, A. (1993) Type VI collagen expression is upregulated in the early events of chondrocyte differentiation. *Development* **117**, 245-251.
4. Karsenty, G. (1998) Genetics of skeletogenesis. *Dev Genet.* **22**, 301-313.
5. Hardingham, T. E. and Fosang, A. J. (1992) Proteoglycans: many forms and many functions. *FASEB J.* **6**, 861-870.
6. Neame, P. J. and Barry, F. P. (1993) The link proteins. *Experientia* **49**, 393-402.
7. Binette, F., Cravens, J., Kahoussi, B., Haudenschiled, D. R. and Goetinck, P. F. (1994) Link protein is ubiquitously expressed in non-cartilaginous tissues where it enhances and staabilizes the interaction of proteoglycans with hyaluronic acid. *J. Biol. Chem* **269**, 19116-19122.
8. Neame, P. J., Christner, J. E. and Baker. J. R. (1986) The primary structure of link protein from rat chondrosarcoma proteoglycan aggregate. *J. Biol. Chem.* **261**, 3519-3535.
9. Deák, F., Kiss, I., Sparks, K. J., Argraves, W. S., Hampikian, G. and Goetinck, P. F. (1986) Complete amino acid sequence sequence of chicken cartilage link protein deduced from cDNA clones. *Proc. Natl. Acad. Sci.* USA, **83**, 3766-3770.
10. Doege, K., Hassell, J. R., Caterson, B. and Yamada, Y. (1986) Link protein cDNA sequence reveals a tandemly repeated protein structure. *Proc. Natl. Acad. Sci.* USA, **83**, 3761-3765.
11. Perkins, S. J., Nealis, A. S., Dudhia, J. and Hardingham, T. E. (1989) Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein. *J. Mol. Biol.* **206**, 737-753.
12. Osborne-Lawrence, S. L., Sinclair, A. K., Hicks, R. C., Lacey, S. W., Eddy, R. L., Jr., Byers, M. G., Shows, T. B. and Duby, A. D. (1990) Complete amino acid sequence of human cartilage link protein (CRTL1) deduced from cDNA clones and chromosomal assignment of the gene. *Genomics,* **8**, 562-567.

13. Périn, J. –P., Bonnet, F., Thurieau, C. and Jollès, P. (1987) Link protein interactions with hyaluronate and proteoglycans. *J. Biol. Chem.* **262**, 13269-13272.
14. Grover, J. and Roughley, P. J. (1994) The expression of functional link protein in a baculovirus system: analysis of mutants lacking the A, B and B' domains. *Biochem. J.* **300**, 317-324.
15. Stirpe, N. S. and Goetinck, P. F. (1989) Gene regulation during cartilage differentiation: teporal and spatial expression of link protein and cartilage matrix protein in the developing limb. *Development,* **107**, 23-33.
16. Kiss, I., Deák, F., Mestric, S., Delius, H., Soós, J., Dékány, K., Argraves, W. S., Sparks, K. J. and Goetinck, P. F. (1987) Structure of the chicken link protein gene: exons correlate with the protein domains. *Proc. Natl. Acad. Sci.* USA, **84**, 6399-6403.
17. Muratoglu, S., Bachrati, C., Malpeli, M., Szabó, P., Neri, M., Dozin, B., Deák, F., Cancedda, R. and Kiss, I. (1995) Expression of the cartilage matrix protein gene at different chondrocyte developmental stages. *Eur. J. Cell Biol.* **68**, 411-418.
18. Deák, F., Barta, E., Mestric, S., Biesold, M. and Kiss, I. (1991) Complex pattern of alternative splicing unusual diversity in the leader sequence of the chicken link protein mRNA. *Nucleic Acids Res.* **19**, 4983-4990.
19. Rhodes, C., Savanger, P., Line, S., Sasaki, M., Chirigos, M., Doedge. K. and Yamada, Y. (1991) Characterization of the promoter for the rat and human link protein gene. *Nucleic Acids Res.* **19**, 1933-1939.
20. Dudhia, J., Bayliss, M. T. and Hardingham, T. E. (1994) Human link protein gene: structure and transcription pattern in chondrocytes. *Biochem. J.* **303**, 329-333.
21. Hauser, N. and Paulsson, M. (1994) Native cartilage matrix protein (CMP). *J. Biol. Chem.* **269**, 25747-25753.
22. Paulsson, M., Heinegård, D., (1981) Purification and structural characterization of a cartilage matrix protein. *Biochem. J.* **197**, 367-375.
23. Paulsson, M. and Heinegård, D. (1982) Radioimmunoessay of the 148-kilodalton cartilage protein. *Biochem. J.* **207**, 207-213.
24. Argraves, W. S., Deák, F., Sparks, K. J., Kiss, I. and Goetinck, P. F.(1987) Structural features of carilage matrix protein deduced from cDNA. *Proc. Natl. Acad. Sci.* USA, **84**, 464-468.
25. Mundlos, S. and Zabel, B. (1994) Developmental expression of human cartilage matrix protein. *Dev. Dynam.* **199**, 241-254.

26. Aszódi, A., Hauser, N., Struder, D. Paulsson, M., Hiripi, L. and Bösze, Z. (1996). Cloning, sequencing and expression analysis of mouse cartilage matrix protein cDNA. *Eur. J. Biochem.* **236**, 970-977.
27. Paulsson M. and Heinegård D. (1979) Matrix proteins bound to associatively prepared proteoglycans from bovine cartilage. *Biochem. J.* **183**, 539-545.
28. Winterbottom, N., Tondravi, M. M., Harrington, T. L., Klier, F. G., Vertel, B. M. and Goetinck, P. F. (1992) Cartilage matrix protein is a component of the collagen fibril of cartilage. *Dev. Dynam.* **193**, 266-276.
29. Aszódi, A., Módis, L., Páldi, A., Recendorj, A., Kiss, I. and Bösze, Z. (1994) The zonal expression of chicken cartilage matrix protein gene in the developing skeleton of transgenic mice. *Matrix Biol.* **14,** 181-190.
30. Franzen, A., Heinegård D. and Solursh, M. (1987) Evidence of sequential apperance of cartilage matrix proteins in developing mouse limbs and in cultures of mouse mesenchymal cells. *Differentiation* **36**, 199-210.
31. Deák, F., Piecha, D., Bachrati, C., Paulsson, M. and Kiss, I. (1997) Primary structure and expression of matrilin-2, the closest relative of cartilage matrix protein within the von Willebrand factor type A-like module. *J. Biol. Chem.* **272**, 9268-9274.
32. Ayoubi, T. A., Van De Ven, W.J. (1996) Regulation of gene expression by alternative promoters. *FASEB* **10**, 453-460.
33. Moore, M. J, Query, C. C. and Sharp, P. A. (1993) Splicing of precursors to mRNA by the splicesome in *The RNA World* (Gesteland, R. F. and Atkins, J. F. eds), pp.303-355, Cold Spring Harbor Laboratory Press.
34. Keyes, L. N., Cline, T.W., Schedl, P. (1992) The primary sex determination signal of Drosophila acts at the level of transcription. *Cell.* **68**, 933-43.
35. Baserga, S. J. and Steitz, J. A. (1993) The diverse world of small ribonucleoproteins in *The RNA World* (Gesteland, R. F. and Atkins, J. F. eds), pp.359-381 Cold Spring Harbor Laboratory Press.
36. Madhani, H. D., Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet.* **28**, 1-26
37. Reed, R.(1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Current Opp. G & Dev.* **6**, 215-220.
38. Nilsen, T. W. (1994) RNA-RNA interactions in the splicesome: unrevaling the ties that bind. *Cell* **78**, 1-4.

39. Krainer, A. R., Conway, G. C., Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell.* **62**, 35-42.

40. Hall, S. L. and Padget, R. A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357-365.

41. Montzka, K. A. and Steitz, J. A. (1988) Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci.* USA, **85**, 8885-8889.

42. Tarn W-Y. and Steitz, J. A. (1996) A novel splicesome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**, 801-811

43. Yu, Y-T. and Steitz, J. A. (1997) Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci.* USA, **94**,

44. Tarn, W-Y. and Steitz, J. A. (1996) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824-1832.

45. Dietrich, R. C., Incorvaia, R., Padgett, R.A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell.* **1**, 151-60.

46. Wu, Q. and Krainer, A. R. (1997) Splicing of a divergent subclass of AT-AC introns requires the major splicesomal snRNAs. *RNA* **3**, 586-601.

47. Sharp, P. A. and Burge, C. B. (1997) Classification of introns: U2-type or U12-type. *Cell*, **91**, 875-879.

48. Breitbart, R. E., Andreadis, A. and Nadal-GInard, B. (1987) Alternative splicing: a ubiquitous mechanism for the genaration of multiple protein isoforms. *Ann. Rev. Biochem.* **56**, 467-495.

49. Sharp, P. A. (1994) Split genes and RNA splicing. *Cell* **77**:805 -815.

50. Maniatis, T. (1991) Mechanisms of alternative pre-mRNA splicing. *Science* **251**, 33-34

51. Smith, C. W., Patton, J. G., Nadal-Ginard, B. (1989) Alternative splicing in the control of gene expression. *Annu. Rev. Gen.* **23**,527-577.

52. Chabot, B. (1996) Directing alternative splicing: cast and scenarios. *TIG* **12**, 472-478

53. Wu, J. Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

54. Jackson, R. J. (1993) Cytoplasmic regulation if mRNA function: the importance of the 3' untranslated region. *Cell* **74**, 9-14.

55. Manley, J. L., Proudfoot, N. J. (1994) RNA 3' ends: formation and function: meeting review. *Genes & Dev.* **8**, 259-264

56. Proudfoot N. (1996) Ending the message is not so simple. *Cell* **87**, 779-781.

57. Keller, W. (1995) No end yet to messenger RNA 3' processing! *Cell* **81**:829-832.
58. Wahle, E. and Keller, W. (1992) The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu Rev Biochem*, **61**, 419-440.
59. Gil, A. and Proudfoot, N. J. (1984) A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. *Nature* **312**, 473-474.
60. Proudfoot, N. (1991), Poly(A) signals. *Cell.* **64**, 671-4.
61. Wahle E. (1995) 3'-end cleavage and polyadenylation of mRNA precursors. *Biochim Biophys Acta.*;**1261**, 183-94.
62. Weiss, E. A., Gilmartin, G. M., Nevins, J. R. (1991) Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. *EMBO J.* **10**,215-219
63. Robberson, B. L., Cote, G. J., Berget, S. M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol.* **10**, 84-94.
64. Ross, J. (1996) Control of messenger RNA stability in higher eukaryotes. *TIG* **12**,171-175.
65. Hargrove, J. L., Hulsey, M. G., Beale, E. G. (1995) The kinetics of mammalian gene expression. *BioEssays* **13,** 667-674.
66. Chen, C. Y., Shyu, A. B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci.* **20**, 465-470.
67. Iwamoto, M., Yagami, K., Lu Valle, P., Olsen, B. R., Petropoulos, C. J., Ewert, D. L., Pacifici, M. (1993) Expression and role of c-myc in chondrocytes undergoing endochondral ossification. *J Biol Chem.* **268,** 9645-52.
68. Leboy, P.S., Vaias, L., Uschmann, B., Golub, E., Adams, S. L., Pacifici, M. (1989) Ascorbic acid induces alkaline phosphatase, type X collagen, and calcium deposition incultured chick chondrocytes. *J Biol Chem.* **264**, 17281-6.
69. Hamburger, V., Hamilton, H. L. (1951) A series of normal stages in the development of the chicken embryo. *J. Morphol.* **88**, 49-92.
70. Kujawa, M. J., Lennon, D.P., Caplan, A. I. (1989) Growth and differentiation of stage 24 limb mesenchyme cells in a serum-free chemicallydefined medium. *Exp Cell Res.* **183,** 45-61.
71. Szabo, P., Moitra, J., Rencendorj, A., Rakhely, G., Rauch, T., Kiss, I. (1995) Identification of a nuclear factor-I family protein-binding site in the silencer region of the cartilage matrix protein gene. *J Biol Chem.* **270**, 10212-21.

72. Yaffe, D., and. Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* **270**:725-727.

73. Blank, R. S., Swartz, E. A., Thompson, M. M., Olson, E. N, and. Owens, G. K. (1995). A retinoic acid-induced clonal cell line derived from multipotential P19 embryonic carcinoma cells expresses smooth muscle characteristics. *Circ. Res.* **76:**42-749.

74. Fusenig, N. E., Amer, S. M,. Boukamp, P., and Worst, P. K.(1978). Characteristics of chemically transformed mouse epidermal cells. *Bulletin du Cancer* **65**:271-280.

75. Pfeiffer, S.E., and. Wechsler, W. (1972). Biochemically differentiated neoplastic clone of Schwann cells. *Proc. Natl. Acad. Sci. U.S.A.* 69:2885-2889.

76. Mukhopadhyay, K., Lefebre, V., Zhou, G., Garofalo, S., Kimura, J.H. and de Crombrugghe, B.(1995). Use of a new rat chondrosarcoma cell line to delineate a 119-base pair chondrocyte-specific enhancer element and to define active promoter segments in the mouse pro-alpha 1(II) collagen gene. *J. Biol. Chem.* **270**:27711-27719.

77. Sorokin, L.M., Gopfert, T.,. Hallmann, R. and Deutzmann R.. (1994). Expression of novel 400 kD laminin chains by mouse and bovine endothelial cells. *Eur. J. Biochem.* **223**:603-610.

78. O'Connell, K. A., and. Edidin, M. (1990). A mouse lymphoid endothelial cell line immortalized by simian virus SV40 binds to lymphocytes and retains functional characteristics of normal endothelial cells. *J. Immunol.* **144**:521-525.

79. Akita, M., Murata E., Schroter-Kermani C., and Merker H. J. (1993). Type I collagen synthesis in cultured aortic smooth muscle cells from spontaneously hypertensive rats and normotensive control, Wistar-Kyoto rats. *Pathobiology* 61:98-103.

80. Chomczynski, P., Sacchi N., (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159

81. Ahrens, P., Solursh, M., Reiter, R. (1977) Stage-related capacity for limb chondrogenesis in cell culture. *Dev Biol* **60**, 69-72.

82. Kiss, I., F. Deák, R. G. Holloway, Jr., H. Delius, K. A. Mebust, E. Frimberger, W. S. Argraves, P. A. Tsonis, N. Winterbottom, P. F. Goetinck: (1989) Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. *J. Biol. Chem.* **264**, 8126-8134.

83. Deák, F., Argraves W. S., Kiss I., Sparks K. J., Goetinck, P. F: (1985) Primary structure of the telopeptide and a portion of the helical domain of chicken type II procollagen as determined by DNA sequence analysis. *Biochem.* J. **229**, 189-196.

84. Castagnola, P., Torella G., Cancedda R. (1987) Type X collagen synthesis by cultured chondrocytes derived from the permanent cartilaginous region of chick embryo sternum. *J. Cell Biol.* **123**, 332-337.

85. Koller. E., Winterhalter, K. H., Trueb, B. (1989) The globular domains of type VI collagen are related to the collagen-binding domains of cartilage matrix protein and von Willebrand factor. *EMBO J.* **8,** 1073-7.

86. Vuorio, E. I., Schaefer, I. M., Vuorio, T. K, Dorfman, A., Upholt, W. B. (1984) Construction and partial characterization of recombinant cDNA clones for chicken type I collagen messenger RNAs. *Acta Chem Scand [B]*. **38,** 237-41.

87. Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J., Kirschner, M. W. (1980) Number and evolutionary conservation of alpha- and beta-tubulin and cytoplasmic beta- and gamma-actin genes using specific cloned cDNA probes. *Cell* **20**, 95-105.

88. Frohman, M. A. (1993), Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE *Methods of Enzymol.* **218**, 340-356.

89. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Moleclualr Cloning. A Laboratory Manual. 2nd edition. Cold Spring Harbor Labortory Press.

90. Hu, M. C., Davidson, N., (1986) Mapping transcription start points on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique. *Gene.*;**42**,:21-9

91. Lu Valle, P., Iwamoto, M., Fanning, P., Pacifici, M., Olsen, B. R. (1993) Multiple negative elements in a gene that codes for an extracellular matrix protein, collagen X, restrict expression to hypertrophic chondrocytes. *J Cell Biol.* **121,** 1173-9.

92. Erster, S. H., Finn, L. A., Frendewey, D. A. and Helfman D. A. (1988) *Nucleic Acids Res*, Use of RNase-H and primer extension to analyze RNA splicing. **16**, 5999-6014.

93. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Promoter sequences of eukaryotic protein-coding genes. *Science* **209,** 1406-1414.

94. Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes & Dev.* **10,** 1657-1683.

95. Cohen C, Parry DA, (1990) Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. *Proteins* 7, 1-15.

96. Tarn, W. Y. and Steitz, J. A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci.* **22**, 132-137.

97. Kozak, M. (1989) The scanning model for translation: an update. *J Cell Biol.* **108**, 229-241.

98. Oliveira, C. C., McCarthy, J. E. (1995) The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae. *J Biol Chem.* **270**, 8936-43.

99. Kozak, M.(1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci* U S A. **83**, 2850-4.

100. Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133-148.

101. Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* **15**, 529-539.