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

Introduction

Gene expression is regulated at multiple levels. The most general mechanism is regulation of the inititation of transcription. However, additional regulatory mechanisms may lead to generation of multiple transcripts from the same gene which may encode distinct protein isoforms, or may have the same coding capacity but differ in translational efficiency, stability or transport. These mechanisms including alternative promoter usage, alternative splicing and alternative polyadenylation site usage are important to create diversity and flexibility in gene regulation. By use of alternative promoters, expression of genes in more than one cell type or developmental stage can be achieved, since each promoter can have different tissue specificity and react differently to some signals. In the case of formation of mRNA, precise RNA processing is achieved by pre-mRNA splicing and polyadenylation. Almost all mRNAs in eukaryotes are produced by splicing of intron-containing pre-mRNAs and it is of great importance in the regulation of gene expression. In some cases, pre-mRNA is alternatively spliced. Alternative splicing is another powerful mechanism in generating closely related protein isoforms from a single gene. Alternative splicing creates mRNA isoforms by allowing selection of different combinations of splice sites within a pre-mRNA. These mRNA isoforms contribute to developmentally regulated and cell type specific patterns of gene expression. Another important mechanism in RNA processing is polyadenylation, by which the 3'-end of the mRNA is formed. 3'-end processing may contribute to the fine tuning of the level of gene expression since the 3'-untranslated region is the repository of signals determining mRNA localisation, polyadenylation, mRNA stability and controlling translation initiation. All these mechanisms and combinations of them enable a single gene to function as a complex transcriptional unit via formation of multiple mature mRNAs and thereby contribute to fine tuning and flexibility of gene expression.

In the case of the extracellular matrix (ECM) proteins, generation of multiple size transcripts is rather a rule then exception. ECM proteins are usually large, multimodular components and various module combinations can be generated by alternative splicing. The mRNAs of matrix molecules very often posses long 3'-untranslated region, where utilisation of alternative polyadenylation signals can be

seen. In the cartilaginous ECM there are several examples of this kind of proteins whose gene expression patterns are thoroughly analysed because of the overriding biological importance of cartilage in skeletal development.

Cartilage has diverse roles during development. It is a network of macromolecules, synthesised and deposited by the specialised cells called chondrocytes. Cartilage development, called chondrogenesis, is achieved by chondrocyte differentiation. During differentiation of growth plate chondrocytes, distinct stages can be observed that are characterised by changes in cell morphology and gene expression pattern. First, undifferentiated mesenchymal (prechondrogenic) cells aggregate to form chondrogenic nodules then they differentiate into proliferating (stage I) chondrocytes. This stage is characterised by expression of cartilage specific genes; type II collagen, link protein (LP) and matrilin-1. By stabilising aggrecan-hyaluronan complexes that are major components of the accumulating cartilaginous ECM, LP acts in formation of a macromolecular network. The second molecular network is the fine fibrillar one composed of type II, IX and XI collagens. Matrilin-1 modulates the interaction of the two networks. Proliferating chondrocytes further differentiate into hypertrophic cells (stage II) which synthesise the highest level of type X collagen.

Most recently isolation of cDNA clones encoding a new extracellular protein, matrilin-2, has been reported in our laboratory. Although matrilin-2 is closely related to matrilin-1, the two genes show distinct expression patterns. The similar primary structure and complementary expression pattern in developing skeletal elements of matrilin-2 compared to matrilin-1 indicate a possible analogous role in the organisation of ECM in other tissues.

Specific Aims

Our laboratory has been studying transcriptional regulation of the LP and matrilin-1 genes. Previously, heterogeneity of the 5'-untranslated region of the LP mRNAs was observed. To be able to see, whether the pattern of LP transcripts generated by multiple alternative splicing shows variation during chondrocyte development, chondrocyte populations of distinct developmental stages had to be prepared. During characterisation of the cell population, size heterogeneity of the

chicken matrilin-1 transcript was observed. Finally, the nature and cell type specificity of matrilin-2 transcripts had to be elucidated. In this direction our specific aim was to

- Establish cell culture systems representing the stages of cartilage development *in vitro*. Then to study gene expression patterns of LP and matrilin-1 in comparison to collagen genes through the stages of development utilising the RNA samples isolated from the established cell cultures.
- Clone and sequence the full length human matrilin-2 cDNA and study the gene expression patterns of both mouse and human matrilin-2 in various cell cultures.
- Reveal the molecular nature of heterogeneous transcript formation of the LP, matrilin-1 and matrilin-2 genes and provide explanation for the potential significance in regulation of gene expression.

Experimental Procedures

Cell lines and cultures: High-density mesenchyme (HDM) cultures were prepared from limb buds of 4.5 day old chicken embryos. Primary chicken embryo chondrocytes (CEC) were obtained either from sterna of day 14 embryos or from the middle part (enriched in proliferating cells) of sterna of day 18 embryos. Hypertrophic cells (HC) were also obtained from sterna of day 18 embryos but from the core part where hypertrophic cells are predominant and cultured in suspension, which favours differentiation.

Various mouse, rat and human permanent cell lines were used as sources to study matrilin-2 gene expression.

RNA preparation and analyses: Total RNA was prepared from guanidinium thiocyanate lysates of all cell cultures and cell lines by acid-phenol-chloroform extraction or from sterna of 14.5-day-old chicken embryos by repeated ethanol precipitation from guanidinium chloride extracts. RNA samples were analysed by Northern blot analysis using the appropriate cDNA fragments as hybridisation probes.

Plasmid constructs, necessary for primer extension assays, were made by basic cloning techniques and they are checked for their integrity by nucleotide

sequencing. Total RNA samples representing the stages of chondrogenesis were used for the assay. Transcription start sites were mapped by extending a 5'-end labelled gene specific primer with T4 DNA polymerase on a ss DNA template annealed to the mRNA. For the estimation of relative promoter usage the assay was modified. Two ss DNA templates were designed. One, representing the transcripts initiated from the first promoter and the second represents transcripts initiated from both promoters. ss DNA templates were simultaneously annealed to each RNA sample and -40 sequencing primer. The primer was elongated by T4 DNA polymerase. An *in vitro* transcript of LP was created and used in the assay as an external quantification standard. Band intensities were compared by Phosphorimage analysis.

Reverse transcription, polymerase chain reaction (PCR) and analysis of amplification products: Total RNA was reverse transcribed by MMLV-RT using either oligo dT tailed adapters or gene-specific oligonucleotide primers. Fraction of this reaction was amplified with gene specific oligonucleotide primers. Conditions of all PCR reactions were optimised according to given template and appropriate primer pair. An additional first cycle with a prolonged elongation time was added. All PCR products were cloned and when necessary sequenced for both cDNA strands.

3' Rapid amplification of cDNA ends (RACE): For amplification of the 3'-end, total RNA samples were reverse transcribed with a Tag-Linker-dT_{17} primer. ss cDNA was amplified by two consecutive amplifications with nested primer pairs of Tag+gene-specific primer 1 and linker+gene-specific primer 2.

5' RACE: Total RNA sample was used for first strand cDNA synthesis, with a gene specific primer. Homopolymer tail of dA was added to the 3'-end of cDNA by terminal deoxynucleotidyl transferase. dA tailed single stranded cDNA was first annealed with Tag-Linker-dT_{17} primer and then amplified in the same fashion with 3'RACE.

Results and Discussion

In our laboratory we established systems representing and covering the stages of chondrocyte development by application of several culturing conditions. The cell stages were verified by monitoring cell morphology and collagen gene

expression pattern during the days of cultures. Nodule formation was observed in HDM cultures representing the onset of chondrogenesis. CEC cultures were made of predominantly proliferating (stage I) while HC cultures were separated, individual cells (stage II). Study of expression patterns of several cartilage specific genes revealed the following results.

- The same sequential activation of the genes for type II collagen, LP and matrilin-1 proved the *in vitro* cultures were capable of reproducing *in vivo* chondrogenesis.

- Steady state levels of cartilage specific mRNAs were lower in mesenchymal cultures than in proliferating cultures. A transient peak in the steady state level of type VI collagen mRNA was observed, specifying the substage Ia. When type VI collagen mRNA production ceased, matrilin-1 gene expression appeared and increased continuously in cultures of chondrocytes producing type X collagen, a marker for hypertrophy. Thus, matrilin-1 can be defined as a marker for stages Ib and II of chondrocyte development.

  The cell cultures were utilised to test the activity of transcriptional regulatory elements responsible for development stage-specific expression of the chicken matrilin-1 gene.

- During this study, a minor, 2.1 kb matrilin-1 mRNA was discovered, in addition to the major, 3.4 kb one. Consecutive Northern hybridisation revealed that the multiple transcripts differ in their 3'-end. Amplification of 3'-end of matrilin-1 mRNA with RT-PCR and 3' RACE yielded multiple products. Cloning and nucleotide sequencing of these products revealed that the shorter mRNA itself is heterogeneous, generated by either usage of an alternative 3' splice site for the terminal intron or usage of an upstream alternative polyadenylation site. Alternative polyadenylation leads to a shorter 3'-UTR without any change in the protein product. Sequencing data of the 3'-UTR revealed a U-rich element right upstream to the 3'-end of the longer mRNA. This element apparently increases the efficiency of the downstream polyadenylation site by facilitating the formation of a more stable polyadenylation. Moreover, the usage of downstream site

provides a considerably long 3'-UTR that may harbour many important signals determining mRNA localisation, stability and signals controlling translation initiation. We need to keep in mind that the relative amount of the transcripts with shorter and longer 3'-UTR deos not automatically mean that those are utilised by the same relative efficiency in translation.

Unlike usage of alternative polyadenylation site, the alternatively spliced mRNA potentially encodes a different protein product, altered in its carboxyterminus. This region contains the coiled-coil domain crucial for formation of homotrimer structure of the protein. Alternative usage of the 3' splice site for the last intron leads to a predicted protein sequence, of which the first half is unchanged, but probably too short for coiled-coil formation and the second half cannot form α-helix. Therefore, it cannot form a homotrimer structure. However, interaction with other molecules may enable this isoform to be included into a heterooligomer structure.

The alternative splicing event observed has theoretical importance, because in both (the major and minor) mRNA form the U12-type AT-AC splice site are utilised.

- During our studies it was found out that in addition to multiple splicing events alternative promoter usage contributes to heterogeneity of the 5-UTR of the chicken LP gene. Transcription start sites for the downstream promoter were mapped by primer extension with T4 DNA polymerase. By a modified T4 DNA polymerase primer extension assay, we were able to determine relative activities of the two promoters. We detected preferential usage of the upstream promoter in all three stages of chondrocyte development. Relatively increased activity of the less active downstream promoter at the advanced stages was also observed.

  For discussion of the biological significance of the heterogeneity at 5'-end of LP mRNA, alternative splicing and usage of the two promoters should be taken into consideration together. Alternative splicing shortens the 5'-UTR by the region having high tendency to form secondary structure. Moreover, it splices out a stop and two start codons thereby removing two upstream short open reading frames. Both features have the potential of decreasing the efficiency of translation.

Therefore, this mechanism provides possibility for efficient translation of the transcript initiated from the first promoter. In rat the downstream promoter is used exclusively. Therefore, no alternative splicing in the upstream region has been reported. We might speculate that the upstream promoter has evolved to be more active in birds.

- Heterogeneity detected in the mouse matrilin-2 transcript was shown to be due to alternative splicing. All the cell lines analysed by RT-PCR showed that the heterogeneity lies in the unique module and it is due to alternative splicing of a 57 nucleotide long region. No further heterogeneity was detected in the upstream translated regions analysed. Since neither the function of matrilin-2 nor the importance of the unique region has been fully understood, it is difficult to speculate on the importance of heterogeneity detected in unique region. Nevertheless, the unpaired cysteine present in this region may play a role in forming structural variability of the protein.

- The full length cDNA for human matrilin-2 was cloned from a permanent cell line. Total RNA isolated from this cell line was used for RT-PCR and 5' RACE amplification of EGF modules, vWFA1 module and a part of 5'-flanking region. A composite sequence of 3497 bp was obtained defining an open reading frame of 956 amino acids. Deduced amino acid sequences of mouse and human matrilin-2 share a 86.5 % sequence identity and 93 % similarity. The same heterogeneity in the unique module of human matrilin-2 was detected as in mouse matrilin-2.

Publications

This thesis is based on the following publications:

1. Muratoglu, S., Bachrati, C., Malpeli, M., Szabo, 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.
2. Piechea, D., Muratoglu, S., Mörgelin, M., Hauser, N., Studer, D., Kiss, I., Paulsson, M. and Deák, F. Matrilin-2, a large, oligomeric matrix protein, is expressed by a great variety of cells and forms fibrillar networks. (submitted)

Unrelated publication:

1. Szűts, V., Möllers, U., Bittner, K., Schürmann, G., Muratoglu, S., Deák, F., Kiss, I. and Bruckner, P. (1998) Terminal differentiation of chondrocytes is arrested at distinct stages identified by their expression repertoire of marker genes. *Matrix Biol.* **17**, 435-448.

Presentations in scientific meetings

1. Muratoglu, S., Bachrati, C., Deák, F. and Kiss, I. "Analysis of Cartilage Specific Transcripts During Chondrogenesis". Fourth European Congress of Cell Biology, Prague, Czech Republic, June 26-July 1, 1994. Poster presentation.
2. Muratoglu, S., Bachrati, C., Deák, F. and Kiss, I. "Multiple RNA Species for Matrix Molecules in Cartilage". International Summer School on 'A World of RNA: Structure and Function'. NATO-FEBS-EMBO Lecture Course, Island of Spetsai, Greece. August 28-September 10, 1994. Poster presentation.
3. Moitra, J., Rákhely, G., Muratoglu, S. and Kiss, I. "Preparation of Nuclear Extract from Chondrogenic Sources". The XIVth FECTS Meeting, Lyon, France, August 30- September 10, 1994 and 2nd European Research Conference on Biology of Cartilage and Bone, Giens, France, October 1-6, 1994. Poster presentation.

4. Muratoglu, S. "Porcfehérje gének expressziója a porcdifferenciálódás különböző stádiumaiban". A Magyar Biokémiai Egyesület Molekuláris Biológiai Szyakosztálya 1. Munkaértekezlete, Seregélyes, April 16-18, 1996. Oral presentation.
5. Deák, F., Szűts, V., Muratoglu, S., Bachrati, C. and Kiss, I."The chicken link protein gene is transcribed from two distinct promoters". 8th IMP Spring Conference 'Genes', Vienna, Austria, May 22-24, 1997. Poster presentation.