TRANSCRIPTIONAL REGULATION OF THE MATRILIN-1 GENE

Identification of binding sites for cartilage-specific transcription factors in the short promoter of the matrilin-1 gene

Summary of Ph.D. Thesis

Otgonchimeg Rentsendorj

Supervisor: Ibolya Kiss, PhD, D.Sc.

Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences

Szeged
2006

Introduction

Cartilage anlagen develop in the embryo before bone, and thus provide the first skeleton of the embryo. It acts as the main source of body longitudinal growth that provides structural templates and induction signals for the formation of most bones through a process called endochondral ossification. Cartilage is an elaborate network of large macromolecules, synthesised and deposited by the specialized cells, called chondrocytes. They deposit these macromolecules around themselves to form an extracellular matrix (ECM).

Chondrocytes express a number of specific genes, such as types II, VI, IX, and XI collagens, aggrecan, link protein, CD-Rap, COMP, and matrilin-1. It is clear that components of the cartilage matrix itself play important roles in either modulating or maintaining the phenotype of chondrocytes and their correct organization in the growth plate. In addition, the precise expression pattern depends on a balance of positive and negative transcription factors, proteins that control the synthesis of mRNAs from these specific genes. In cartilage, a number of studies indicate that Sox9 is an activating transcription factor, indispensable for chondrogenesis. During chondrogenesis, Sox9 is expressed in chondroprogenitors and chondrablasts, and activates the cartilage-specific genes, such as *Col2a1*, *Col11a2*, *aggrecan*, and *CD-Rap*. The study of chimeric mice that harbour Sox9−/− cells revealed that cells without Sox9 do not express chondrocyte marker genes such as *Col2a1*, *Col9a2*, *Col11a2*, and *aggrecan*. Sox9 binds to and directly activates the cartilage-specific regulatory elements of *Col2a1* as well as *Col11a2*, which can direct cartilage-specific expression in transgenic mice. Sox9 also transactivates the promoter activities of *Col9a1*, *aggrecan*, and *CD-Rap*.

Additional Sox factors, such as L-Sox5 and Sox6, are also involved in chondrogenesis. The expression of L-Sox5 and Sox6 requires Sox9. They are activated in prechondrocytes and highly expressed in chondroblasts in all developing cartilage elements of the mouse embryo. Although individual Sox5 or Sox6 null mice are born

with minor cartilage defects, double knockout mice develop a severe, generalized chondrodysplasia. In the double null mice, prechondrocytes are unable to progress along the differentiation pathway, expressing a very low level of early cartilage genes, such as *Col2a1* and *aggrecan*, but they fail to turn on stage-specific matrix genes such as *COMP* and *matrilin-1*. *In vitro* experiments have suggested that, besides Sox9, L-Sox5 and Sox6 are also involved in *Col2a1* expression. They appear to form a large complex with each other and other nuclear proteins in chondrocytes. Therefore, it is believed that the three Sox cooperatively activate the *Col2a1*. They were also able to suppress expression of markers for hypertrophic chondrocytes, thus confirming that L-Sox5, Sox6, and Sox9 constitute a master chondrogenic trio.

Matrilin-1 is tightly associated with aggrecan and type II collagen fibrils, the two major macromolecular networks of cartilage. It also forms collagen-independent pericellular filaments via the vWFA domains, and interacts with $\alpha 1\beta 1$ integrin. Therefore, it seems to perform an adapter function in the assembly of the cartilaginous matrix. Among cartilage protein genes, the matrilin-1 has a unique expression pattern, restricted to certain chondrocyte developmental stages and distinct zones of the growth plate. It has been confirmed that the *matrilin-1* gene is activated only in late proliferative chondrocytes *in vitro*. This suggests that the regulation of the gene may involve similar as well as different molecular mechanisms as compared to other cartilage protein genes, such as *Col2a1*. To get insight into the common and distinct molecular mechanisms controlling cartilage-specific gene expression, we aimed to study DNA elements and transcription factors, involved in the transcriptional regulation of the matrilin-1 gene.

Aims of study

Functional analysis of the main regulatory regions of the gene revealed that the long promoter with and without the intronic

enhancer region, as well as the short promoter with the intronic enhancer region equally direct the transgene expression to distinct zones of the growth plate in transgenic mice. Based on this observation, we raised the question, whether the short promoter may have a role in the tissue-specific regulation and if it involves cartilage-specific and/or developmental stage-specific control elements. The following specific aims were set out:

- to delineate *cis*-regulatory elements that are important for the transcriptional regulation of the matrilin-1 gene and to identify proteins binding to these sites, especially focusing on those sites, which possibly function via interacting with Sox-family proteins in the short promoter region.
- to verify Sox protein-binding sites in the short promoter using *in vitro* gelshift and supershift assays.
- to reveal tissue-specific binding of transcription factors to the short promoter region using *in vivo* footprinting.
- to study the role of these putative elements in the transcriptional regulation of the matrilin-1 gene.

Methods

In order to accomplish our aims we performed the following experimental methods:

- ✓ Sequence analysis
- ✓ Isolation of genomic DNA from tissue culture
- ✓ *In vivo* footprinting with dimethyl sulphate and UV light
- ✓ Preparation of cell extracts for EMSA
- ✓ Synthesis and purification of bacterially expressed GST-SOX9
- ✓ EMSA and supershift experiments
- ✓ Luciferase reporter gene constructs for transient expression assay
- ✓ Site directed mutagenesis
- ✓ Transient expression assay

Results and Discussion

In this study, we have shown that the short promoter of chicken matrilin-1 gene contains at least two cis-acting Sox-specific islands labelled as Pe1 and Ine. Both elements interact with chondrogenic transcription factors of Sox9, L-Sox5 and Sox6 *in vitro*. Point mutations in the Sox motifs of either element interfered with or altered the formation of nucleoprotein complexes *in vitro* and significantly decreased the reporter gene activity in CEC in transient expression assays, stressing the importance of these elements in the transcriptional upregulation of the matrilin-1 gene. Transient expression studies confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. Based on the data, we hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements. In addition, we propose that distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

We were able to show *in vivo* occupancy of the Sox motifs in genomic footprinting in the expressing cell type, but not in the non-expressing, which support the involvement of Pe1 and Ine in the tissue-specific regulation of the gene. Apart from Pe1 and Ine, in chondrocytes, footprints were also observed at the NFI contact points of the SI and SII elements that were identified previously. As opposed to this, the complete absence of footprints in the short promoter in fibroblasts in repeated experiments indicated no factors bound to their recognition sequences in the non-expressing cell type. This suggests that regulation at the chromatin level can be involved in the activation of the gene in chondrocytes.

Although Pe1 can bind to Sox9, L-Sox5 and Sox6 *in vitro*, it may show a preference for Sox9 *in vivo*, as the inverted Sox motifs carried by Pe1 are more similar to the preferred binding sites of Sox9, than to those of L-Sox5/Sox6. In our experiments, GST–Sox9 formed only a single complex with Pe1 and Ine. Even though Sox9 is

not capable of forming homodimers in solution, consistent with observations from other laboratories, we found that it could bind only to intact inverted pairs of Sox motifs, thus supporting the conclusion that Sox9 dimerization might have occurred upon DNA binding. As opposed to this, Sox9 was reported to bind as a monomer to *cis* elements involved in sex determination.

Collateral analysis based on computer searches revealed that Pe1 element of the chicken matrilin-1 gene is highly conserved under evolutionary pressure between chicken and mammals implies that the element performs a very important function in the transcriptional regulation of the matrilin-1 gene in amniotes.

This is the first report demonstrating that paired Sox-specific sites mapped within Ine, that is close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. The close proximity of the Sox-binding sites of Ine to the TATA box raises questions regarding the possible interactions of Sox proteins with general transcription machinery. An involvement of HMG1 domain proteins in the regulation of gene transcription has been suspected for a long time. It has been reported for example that the abundant HMG1 protein, which binds angled structures in the DNA without any sequence specificity, can interact in a species-specific manner with the core domain of the TATA box-binding protein (TBP), and block the formation of the preinitiation complex (PIC) by preventing the binding of TFIIB to TBP. Differing from the HMG1 protein, Sox9 binds to the DNA in a sequence-specific manner and has a transactivation domain. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription machinery assembled on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. It is yet to be confirmed that this interaction or combination of general transcription machinery really function on the natural Ine element of the matrilin-1 gene.

By bending the DNA, HMG-box proteins are known to promote the binding of other transcription factors to the DNA. Lining

up with these observations, our results suggest that, in addition to Sox proteins, other transcription factors may also be involved in the activity of the short matrilin-1 promoter, thereby contributing to the developmental stage-specific activity of the matrilin-1 gene. Therefore, future studies will be needed for mapping the putative cofactor-binding sites and identifying the interacting Sox partner factors on the Pe1 and Ine elements.

These findings imply for the first time an important role of cis-regulatory elements that are functioning via Sox family proteins in the transcriptional upregulation of the chicken matrilin-1 gene.

List of novel findings

1. Throughout 1.2-kb promoter region tested between positions -1137 and +64, we identified two potential Sox-specific islands labelled as Pe1 and Ine in the short promoter of chicken matrilin-1 gene.
   a. Pe1 element includes a pair of inverted motifs highly similar to the AGAACAATGG motif, which was shown to be the preferred binding site of Sox9 *in vitro*.
   b. Ine1 harbors two pairs of inverted putative Sox motifs with 5/10, 6/10, 7/10 and 4/9 identity with the consensus Sox-binding sites CA/TTTGA/TA/T *in vitro*.
2. We observed *in vivo* occupancy of the Sox motifs in genomic footprinting in the expressing cell type, but not in the non-expressing, which support the involvement of Pe1 and Ine in the tissue-specific regulation of the gene. Apart from Pe1 and Ine, in chondrocytes, footprints were also visible at the NFI contact points of the SI and SII elements that were identified previously.
3. We provided evidence that both elements (Pe1 and Ine) can interact with chondrogenic transcription factors of Sox9, L-Sox5 and Sox6 *in vitro*.

   Eventhough, Pe1 contain pair of inverted Sox9 motifs, we noticed that the element form only a single nucleoprotein complex with GST–Sox9. Point mutations in either Sox

motifs interfered with this complex, suggesting that GST-Sox9 could bind only to intact inverted pairs of Sox motifs and that Sox9 dimerization might have occurred upon DNA binding.

4. Using transient expression studies, we confirmed that both Pe1 and Ine elements significantly contribute to the moderate activity of the short promoter in chondrocytes. We hypothesize that the element does not drive the high cartilage-specific expression of the promoter as an enhancer, but may rather act by modulating the promoter activity and mediating the effect of distal promoter and intronic enhancer elements.
5. This is the first report demonstrating that paired Sox-specific sites mapped within Ine, that is close to the TATA motif can play a functional role in the transcriptional activation of a cartilage protein gene. We hypothesize that, according to the intrinsic flexibility of DNA bending and cooperative binding ability with multiple transcription factors, Sox proteins (L-Sox5, Sox6 and Sox9) may interact with the components of the general transcription machinery assembled on Ine, which may generate a surface complementary to the RNA polymerase II transcription machinery. In addition, distal promoter and intronic elements may also function by forming multi-protein complexes via interacting with Pe1 and Ine.

These findings imply for the first time an important role of cis-regulatory elements that are functioning via Sox family proteins in the transcriptional upregulation of the chicken matrilin-1 gene.

Acknowledgements

This work was carried out in the Institute of Biochemistry, Biological Research Center of Hungarian Academy of Sciences, Szeged, Hungary, during the years 1999-2005. I wish to express my deepest gratitude to my supervisor, Dr. Ibolya Kiss, for giving me the opportunity to work in her excellent group and for introducing me to the field of molecular biology and matrix biology. The depth of her

contribution to these studies has been widespread and is too expansive to list here.

I wish to express my appreciation and respect for Dr. Ferenc Deák for his never-ending enthusiasm and for putting to good use a sharp, inquiring mind.

I would also like to acknowledge the Szeged Biological Research Center, especially past and present Directors of the Institute of Biochemistry, for providing financial support as well as for arranging and conducting useful courses and seminars.

I also wish to express my thanks to all the members of the laboratory, past and present. In particular, I wish to thank to Éva Korpos, Ildikó Karcagi, and Oana Sicora, for their invaluable support, love and friendship during these years. I would like to extend special thanks for the excellent technical support that I have received from Anikó Simon, Ildikó Kravjár, Katalin Kovács and Irén Fekete. Their unequalled skills at the bench have been an integral part of much of the work presented in this thesis. Furthermore, Ildiko Karcagi, Tibor Rauch, Andrea Nagy, Ildikó Sinkó, Andreea Daraba, and Endre Barta are warmly acknowledged for their contributions to exploring the importance of the matrilin-1, which formed the essential basis of the work presented here. To Tibor Rauch and András Blastyák I owe the most gratitude, who has enthusiastically taught their experience and expertise in the protein-DNA interaction techniques. Enormous thanks to Mária Tóth for providing of quality art work. I am also especially indebted to Piroska Szabó who introduced me into Genomic footprinting and Gene manipulation via Homologous recombination teqniques.

I would not have enjoyed life in Szeged so much without my dear friends from ITC. Their companionship has provided the most important and vital meaning in the midst of this furious scientific whirlwind.

Finally, I would like to express my thoughts and deepest gratitude to all of my family for their unconditional support and encouragement. My mother Myadagmaa contributed greatly on this. I thank her for everything from A to Z. My children, Belgutei and Enkhluun, have been my greatest source of inspiration. Being tired of

missing them, missing their fantastic stories, smiles, their little voices, being tired of not being able to fulfill their requests are the best motivations I know for finishing this work. This thesis would not have been possible without them.

List of original Publications

1. Rentsendorj, O., Nagy, A., Sinko, I., Daraba, A., Barta, E. and Kiss. I. (2005) Highly conserved proximal promoter element harboring paired Sox9-binding sites contributes to the tissue- and developmental stage-specific activity of the matrilin-1 gene. Biochem J. **389**, 705-16

2. Karcagi, I., Rauch, T., Hiripi, L., Rentsendorj, O., Nagy, A., Bősze, Zs. and Kiss, I. (2004) Functional analysis of the regulatory regions of the matrilin-1 gene in transgenic mice reveals modular arrangement of tissue-specific control elements. Matrix Biol. **22**, 605-618

3. Rentsendorj, O. (2002) The chondrogenic master transcription factor Sox9 binds to the regulatory region of the matrilin-1 gene. Dissertation Summary, Acta Biologica Szegediensis. **46** (1-2): 45

Presentations and Posters

1. Rentsendorj, O., Rauch, T., Karcagi, I. and Kiss, I. (2003) The role of *cis*-regulatory elements in the transcriptional regulation of the matrilin-1 gene, 8th Workshop of the Hungarian Biochemical Society, Molecular Biology Section, Tihany, Hungary

2. Rentsendorj, O., Rauch, T., Karcagi, I. and Kiss, I. (2002) Identification of Sox protein recognition sites in the regulatory region of the matrilin-1 gene using *in vivo* and *in vitro* assays, Straub-days, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

3. Rentsendorj, O., Rauch, T., Karcagi, I. and Kiss, I. (2002) Mapping of cartilage specific control elements in the regulatory region of the matrilin-1 gene using in vivo footprinting, XVIIIth FECTS Meeting, Brighton, UK

4. Rentsendorj, O., Rauch, T., Karcagi, I. and Kiss, I. (2002) In vivo footprinting reveals specific control elements in the regulatory region of matrilin-1 gene, 7th Workshop of the Hungarian Biochemical Society, Molecular Biology Section, Keszthely, Hungary

5. Rauch, T., Rentsendorj, O., Karcagi, I. and Kiss, I. (2002) Direct interaction between Sox9 and CTF1/NFI modulates the transcriptional activity of the matrilin-1 gene, XVIIIth FECTS Meeting, Brighton, UK

Unrelated Publications and Posters

1. Rentsendorj, O., Deák, F. and Kiss I (2001) The mouse link protein gene (Crtl1): structure, expression and function. Biokémia, 25: 32-35

2. Rentsendorj, O., Szabó, P. and Kiss, I. (2001) Desigh and construction of targeting vector for the conditional inactivation of Crtl1 gene in embryonic stem cell, 6th Workshop of the Hungarian Biochemical Society, Molecular Biology Section, Sárospatak, Hungary

3. Rentsendorj, O. and Garcia J. G. N. (2005) Cloning and characterization of a human non-muscle Myosin Light Chain Kinase promoter, ATS Meeting, San Diego, CA, USA

4. Rentsendorj, O., Welsh-Servinsky, L. E., Shimoda, L. A., Moldobaeva, A., Mirzapoiazova, T., Verin, A. D. and Pearse, D. B. (2006) Protein Kinase G and p38 MAPK in H_2O_2-induced pulmonary endothelial barrier dysfunction, FASEB, Experimental Biology, San Francisco, CA, USA