L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway

V. Lefebvre*, R. R. Behringer and B. de Crombrugghe*
Department of Molecular Genetics, The University of Texas, MD Anderson Cancer Center, Houston, Texas, U.S.A.

Summary

Objective: This work was carried out to identify transcription factors controlling the differentiation of mesenchymal cells into chondrocytes.

Design: We delineated a cartilage-specific enhancer in the collagen type 2 gene (Col2a1) and identified transcription factors responsible for the activity of this enhancer in chondrocytes. We then analyzed the ability of these transcription factors to activate specific genes of the chondrocyte differentiation program and control cartilage formation in vivo.

Results: A 48-bp sequence in the first intron of Col2a1 drove gene expression specifically in cartilage in transgenic mouse embryos. The transcription factors L-Sox5, Sox6, and Sox9 bound and cooperatively activated this enhancer in vitro. They belong to the Sry-related family of HMG box DNA-binding proteins, which includes many members implicated in cell fate determination in various lineages. L-Sox5, Sox6, and Sox9 were coexpressed in all precartilaginous condensations in mouse embryos and continued to be expressed in chondrocytes until the cells underwent final hypertrophy. Whereas L-Sox5 and Sox6 are highly homologous proteins, they are totally different from Sox9 outside the HMG box domain. The three proteins cooperatively activated the Col2a1- and aggrecan genes in cultured cells. Heterozygous mutations in SOX9 in humans lead to campomelic dysplasia, a severe and generalized skeletal malformation syndrome. Embryonic cells with a homozygous Sox9 mutation were unable to form cartilage in vivo and activate essential chondrocyte marker genes. Preliminary data indicated that the mutation of Sox5 and Sox6 in the mouse led to severe skeletal malformations.

Conclusions: L-Sox5, Sox6, and Sox9 play essential roles in chondrocyte differentiation and, thereby, in cartilage formation. Their discovery will help to understand further the molecular mechanisms controlling chondrogenesis in vivo, uncover genetic mechanisms underlying cartilage diseases, and develop novel strategies for cartilage repair. © 2001 OsteoArthritis Research Society International

Key words: Chondrogenesis, Collagen 2, Sox5, Sox6, Sox9, Transcription.
Chondrogenesis is thus a multi-step cell differentiation pathway, genetically characterized by a complex cascade of specific gene activations and inactivations. Many types of molecules have been shown to play roles in the control of chondrogenesis. They are classed into two main categories. Patterning factors trigger the formation and determine the fate, shape and size of individual skeletal elements. They include a number of bone morphogenetic proteins and fibroblast growth factors, their signaling molecules, and transcription factors such as homeobox factors. They are each typically found in only a restricted number of skeletal elements. Differentiation factors, in contrast, are defined as molecules involved in the specific control of one or several steps of the chondrocyte differentiation pathway and are expressed in every cartilage. They also include diffusible growth factors, receptors and intracellular signaling mediators, as well as transcription factors. Noticeably, the secreted factor parathyroid-hormone-related peptide (PTHrP), its receptor (PTH/PTHrP receptor), and the extracellular indian hedgehog (Ihh) signaling molecule control the differentiation of prehypertrophic chondrocytes into hypertrophic cells. Because the majority of chondrogenesis regulatory factors have been identified recently, the precise mode of action of most of them is still largely unknown. It is also clear that many players remain to be identified. Our main goal over the last years has been to identify and characterize transcription factors that act as master control factors to determine the switch to chondrogenesis and control specifically and directly the activation of chondrocyte marker genes.

Our initial hypothesis of the existence of master chondrogenic differentiation factors was based on the examples of other cell lineages that are derived from the same mesenchymal precursor cells as chondrocytes, and that have been shown to be each governed by a specific set of master transcription factors. For instance, in the myogenic cell lineage, four basic helix-loop-helix muscle regulatory factors (MyoD, myogenin, Myf-5, and MRF4), the myocyte enhancer factor 2 (MEF2) group of MADS-box regulators, Pax-3 and the muscle LIM protein (MLP) cooperate to positively control transcription of muscle-specific differentiation genes. In the osteoblastic lineage, the factor Cbfα1/Osf2, a runt family member, is expressed in all osteoblastic precursor and differentiated cells, is absolutely required to form bones \( \text{in vivo} \), and directly activates a number of marker osteoblastic genes. In the adipoblast pathway, C/EBP beta and delta, ADD-1/SREBP-1, and the peroxisome proliferator activated receptor-gamma (PPAR gamma) play pivotal roles in the differentiation of adipocytes and expression of cell-specific genes.

In this review, we will summarize the experimental evidences that led us and other groups to identify the first chondrogenic transcription factor, Sox9. We will also present the pieces of evidence that we have accumulated that strongly suggest that L-Sox5 and Sox6 are additional chondrogenic transcription factors, presumably working in cooperativity with Sox9 in the early stages of the chondrocyte differentiation pathway.

**Identification of candidate chondrogenic transcription factors**

In order to qualify as master chondrogenic transcription factor, a candidate molecule must meet a number of criteria. First, it has to be expressed during chondrogenesis specifically and be present in all sites of chondrogenesis. It is, nevertheless, conceivable that its expression pattern be not entirely restricted to the chondrocyte differentiation pathway as its chondrogenic action could be specified by cooperativity with other chondrogenic factors. Second, the factor has to be directly involved in the activation of chondrocyte-specific genes. Therefore, it must bind directly or indirectly to regulatory regions in chondrocyte marker genes. Third, the mutation of the gene for this factor must have dramatic consequences on the formation of cartilages \( \text{in vivo} \), proving the requirement for this factor in one or several steps of the chondrocyte differentiation program.

Following these criteria, we undertook a search for chondrogenic transcription factors. We designed our initial strategy on their postulated ability to bind to regulatory sequences in chondrocyte marker genes. Our first goal was thus to identify \( \text{cis} \)-acting regulatory elements that are responsible for the chondrocyte-specific expression of these genes. In order to identify factors acting in the early steps of chondrogenesis, we chose the gene for collagen 2 (Col2a1). Indeed, expression of this gene is highly cell-specific, starts as soon as precartilaginous mesenchymal condensations form, and remains abundant in fully differentiated chondrocytes. Using in parallel the transgenic mouse approach and transient transfection experiments in cultured cells, we first confirmed earlier reports that had suggested that the first intron of \( \text{Col2a1} \) was needed for cartilage-specific expression. We then showed that \( \text{Col2a1} \) promoter and upstream sequences were not needed for chondrocyte-specific expression, and upon progressive delineation of the first intron chondrocyte regulatory sequences, identified a 48-bp segment as the minimum DNA element needed to direct expression of a reporter gene specifically to chondrocytes in mouse embryos and in cultured cells. In agreement with our data, Krebsbach and collaborators (1996) delineated a 100-bp \( \text{Col2a1} \) chondrocyte-specific enhancer that encompassed the 48-bp sequence, and Bell and collaborators (1997) delineated a 309-bp cartilage-specific \( \text{COL2A1} \) enhancer that included the 3′ half of the 48-bp enhancer. Four tandem copies of the 48-bp enhancer were needed to obtain high levels of reporter gene expression, presumably because this short sequence lacked \( \text{cis} \)-acting elements that may be needed in the endogenous gene to amplify the signal induced by the chondrocyte-specific transcription factors binding to the minimal enhancer.

We then used the 48-bp enhancer sequence to identify the nuclear proteins that were responsible for the specific activation of the enhancer in chondrocytes. Several protein binding sites in this enhancer were shown to be required for enhancer activity in chondrocytes and also to form a large protein-DNA complex with chondrocyte-specific nuclear proteins in gel retardation assays (Fig. 1A and B). These sites all resemble a consensus binding site for proteins with a high-mobility-group (HMG) DNA-binding domain. Upon both literature and experimental screening, we found the chondrocyte-specific protein-enhancer complex to be made of Sox9, Sox6, and a new long form of Sox5, which we called L-Sox5 (Fig. 1C). These three proteins are members of the Sox subclass of HMG DNA-binding proteins. Bell and collaborators (1997) showed that an additional HMG-like protein-binding site located 50-bp downstream of the 48-bp enhancer in the \( \text{COL2A1} \) gene also bound Sox9 and contributed to the activity of the 309-bp enhancer in chondrocytes. These data led us to postulate that L-Sox5, Sox6, and Sox9 may form a trio of chondrogenic factors and further experiments were designed to test this hypothesis.
SOX PROTEINS AND THE CONTROL OF CELL FATE IN VARIOUS LINEAGES

Sox proteins are a subfamily of over 20 members within the large family of proteins that harbor one or several so-called high-mobility-group (HMG) DNA-binding domains. The HMG domain is a 79-amino-acid motif consisting of three α-helices arranged in a twisted L-shape. The HMG domain of Sox proteins are 50% identical or more to that of SRY, a testis-determining factor. The identity includes the conservation of specific amino acid residues important for DNA contact and conformational structure. Based on identity within and outside the HMG domain, Sox proteins are classified into seven subgroups, designated A to G. Members of different subgroups show differences in DNA-binding affinity and vary greatly in structural organization, transactivation capability, and ability to interact and act in synergism with other proteins.

The Sox HMG domain specifically recognizes DNA sequences with the degenerate consensus C(A/T)TTTG(A/T)(A/T). In contrast to most other types of transcription factors, which contact DNA in the major groove, Sox (and other HMG proteins) target the minor groove of DNA. Binding induces a sharp DNA bend, estimated at about 70°-90°, such that the concave binding surface of the HMG domain perfectly fits to the DNA groove. Based on this bending property, Sox proteins are believed to act as architectural proteins that, like the HMGI/Y and LEF-1 HMG proteins, organize local chromatin structure to promote assembly of macromolecular regulatory complexes involving transcription factors bound to DNA at proximal sites. In addition to this presumed accessory function, several Sox proteins, including Sox9, feature a potent transcription activation domain and are able to act as typical transcriptional activators. Several Sox proteins have also been shown to synergize with cell-specific transcription factors, including POU-domain proteins, and it has been suggested that this partner mechanism may account for the specific functions of Sox proteins in different cell types. For instance, Sox2 synergizes with Oct3/4 to activate a cell-specific fibroblast growth factor 4 gene enhancer in embryonic stem cells, but synergizes with a different factor, called αEF3, to activate a cell-specific α-crystallin gene enhancer in lens cells.

Or over the last years, Sox proteins have been increasingly demonstrated to be required for cell fate specification in various cell lineages and mutations of their genes in the mouse and in humans have been shown to lead to severe developmental defects and diseases. For instance, Sry, whose gene maps to the Y chromosome, is involved in testis and male differentiation; mutations in SRY lead to sex reversal in XY humans and expression of Sry in XX mice leads to male differentiation. Other examples include the requirement of Sox1 for lens development, Sox4 for endocardial ridge development and B cell differentiation, and Sox10 for determination of neural crest and glial cells (review in Ref. 22). These data thus add weight to our hypothesis that L-Sox5, Sox6, and Sox9 could be involved in the specification of the chondrocyte lineage.

STRUCTURAL ORGANIZATION OF L-Sox5, Sox6 AND Sox9

L-Sox5 and Sox6 belong to the subgroup D of Sox proteins. The two proteins are overall 67% identical to each other, and share more than 90% identity in the HMG domain and in a coiled-coil domain (Fig. 2). This latter domain, of 82 residues, is made of a leucine-zipper associated with a glutamine-rich protein segment. It is located in the N-terminus of L-Sox5 and Sox6, and is not present in Sox5, a spermatid-specific shorter product of the same Sox5 gene. By mediating the formation of protein homodimers and heterodimers, this domain allows L-Sox5 and Sox6 to bind with a much higher affinity to DNA containing two adjacent HMG-binding sites than to DNA harboring a single site. By comparison, Sox9, which does not homodimerize, binds with similar affinity to both types of DNA. This finding implicates that L-Sox5 and Sox6, in contrast to monomeric Sox proteins, must recognize their target genes at pairs of adjacent HMG sites. Sox9 is a single product of the Sox9 gene.

SOX5 belongs to the Sox subgroup E. Its HMG domain displays only 50% identity with that of L-Sox5 and Sox6,
and outside the HMG domain, Sox9 is a protein totally different from L-Sox5 and Sox6. As already mentioned above, Sox9 has no dimerization domain but features a potent transactivation domain at its C-terminus. It also harbors a distinctive domain, called ‘POA’ and made of 41 residues of solely proline, glutamine and alanine, whose function is still unknown.

Hence, the structural organization of L-Sox5, Sox6, and Sox9 and their functional properties in vitro suggested that L-Sox5 and Sox6 could fulfill similar roles in vivo. Their roles may be complementary to those of Sox9 but must be clearly distinct from those of Sox9.

**L-Sox5, Sox6 and Sox9 are coexpressed during chondrogenesis**

*In situ* hybridization experiments have revealed coexpression of the transcripts for L-Sox5, Sox6 and Sox9 in every site of chondrogenesis in the mouse embryo. The three types of transcripts are detectable as early as precursor mesenchymal cells start to form precartilaginous condensations and remain abundantly expressed in every fully formed cartilage onward, in parallel with Col2a1 transcripts. During endochondral ossification, expression of the three Sox genes and of Col2a1 is high in resting and proliferating chondrocytes, and rapidly turned off when chondrocytes undergo hypertrophy in the end of their differentiation pathway (Fig. 3). These data are thus consistent with a role for each of the three Sox genes in specifying chondrogenesis at early steps and maintaining the chondrocyte phenotype in fully differentiated cells. As L-Sox5 and Sox6, which are very similar proteins, are coexpressed in chondrogenesis, we postulated that they may have redundant roles in this cell differentiation pathway.

In addition to their expression in chondrogenesis, Col2a1 and the three Sox genes are also expressed in some areas of the brain and neural tube of mouse developing embryos. Each of the three Sox genes is also found at a generally low level in a restricted number of other tissues in embryonic and adult mice, although never altogether as in chondrogenic tissues. Each of the three Sox proteins is thus believed to have also specific functions in some non-cartilaginous tissues. For instance, Sox9 is expressed in the Sertoli cells of testis where it contributes to cell fate specification and male differentiation. Short transcripts of Sox6 and Sox5 (encoding the short Sox5 protein) are found in mature testis, where they may play a role in the latest stages of the differentiation pathway of spermatid cells.

**L-Sox5, Sox6 and Sox9 cooperate to activate chondrocyte marker genes**

Binding of L-Sox5, Sox6, and Sox9 to the Col2a1 enhancer was shown to mediate high levels of activation of this enhancer in nonchondrogenic cells. Indeed, when Col2a1 reporter constructs, which harbored the 48-bp enhancer or a larger enhancer segment, were transfected in fibroblasts together with expression plasmids for the three Sox proteins, activation of the constructs was found to be much higher when the three Sox proteins were coexpressed than when either one was expressed alone (Fig. 4A). Moreover, the three Sox proteins were able not only to act cooperatively in the activation of these Col2a1 reporter constructs, but they were also able to cooperate to stimulate at high levels the expression of the endogenous Col2a1 gene in 10T1/2 fibroblasts and in the poorly differentiated MC615 chondrocytic cell line (Fig. 4B). These experiments thus strongly suggest that the three Sox proteins act in concert as direct activators of the collagen 2 gene in chondrocytes.

Moreover, it appears likely that the three Sox proteins may also target other chondrocyte marker genes that are coexpressed with Col2a1 in the early stages of the chondrocyte differentiation program. Indeed, upon overexpression in MC615 cells, the three Sox proteins also cooperated to upregulate expression of the aggrecan gene. Two chondrocyte-specific enhancer-elements in the promoter of the collagen 11 Col11a2 gene contain a cluster of HMG-like binding sites, similar to that present in the 48-bp Col2a1 enhancer. These Col11a2 HMG-like sites were also shown to bind L-Sox5, Sox6, and Sox9 and were needed for the expression of these enhancers in the cartilages of mouse embryos. Altogether, these data suggest that the three Sox proteins may be involved in the activation of the most essential cartilage matrix genes and thus control a critical part of the chondrogenesis program.

**L-Sox5, Sox6, and Sox9 are essential for cartilage formation in vivo**

In 1994, Foster and collaborators, and Wagner and collaborators reported that heterozygous mutations in
the human SOX9 gene caused the very severe form of chondrodysplasia that is called campomelic dysplasia, and that is often associated with XY sex reversal. Campomelic dysplasia is characterized by bowing of long bones, and malformations in many cartilages and cartilage-derived skeletal structures. Death most often occurs at or soon after birth due to respiratory distress. It is an autosomal dominant disorder, believed to be due to SOX9 haplo-insufficiency in most cases. It has, however, been suggested that in some cases the SOX9 mutations could lead to expression of a SOX9 mutant protein that could interfere in a dominant-negative manner with wild-type SOX9, increasing the severity of the disease. 

To further study the function of Sox9 in vivo, Bi and collaborators generated both heterozygous and homozygous Sox9 mutant mouse embryonic stem (ES) cells. Because humans with a SOX9 heterozygous mutation die at or soon after birth, the behavior of these mutant ES cells was analyzed in mouse embryo chimeras. In mouse chimeras derived from wild-type blastocysts injected with Sox9 heterozygous mutant cells, cartilages were found to contain both mutant and wild-type cells and both types of cells were able to express chondrocyte-specific markers such as Col2a1 and aggrecan. In contrast, in mouse chimeras generated using Sox9 homozygous mutant ES cells, cartilages were always made of wild-type cells only. Sox9 homozygous null cells were excluded from cartilages in a cell-autonomous fashion at the condensing stage of prechondrocytic mesenchymal cells, and as cartilages formed, Sox9 null cells accumulated at the periphery of cartilages, forming patches of ectopic mesenchymal tissue. These cells expressed the Sox9 null alleles but were unable to express chondrocyte-specific markers such as Col2a1, Col9a2, Col11a2 and aggrecan. In teratomas generated by injecting wild-type or Sox9 heterozygous null embryonic stem cells subcutaneously into syngeneic mice, many different tissue types developed, including cartilages. In contrast, when Sox9 homozygous null cells were used, teratomas formed that contained multiple tissue types but no cartilage could be detected. Altogether, these results thus identified Sox9 as the first transcription factor that is essential for chondrocyte early differentiation and cartilage formation.

To determine the roles that L-Sox5 and Sox6 play in chondrogenesis in vivo, we have recently inactivated their genes in the mouse using standard procedures of homologous DNA recombination into embryonic stem cells (our unpublished data). Preliminary analysis of mutant mice indicates severe abnormalities in essentially all endochondral skeletal elements and proves genetic interaction between Sox5 and Sox6. Overall, the skeletal defects are similar to those seen in mice with mutations in the Col2a1, aggrecan, or link protein genes. The preliminary results thus support the notion that L-Sox5 and Sox6 control critical genetic events in the chondrocyte differentiation pathway. Further analysis of Sox5/Sox6 double mutant mice is in progress to define the precise roles of L-Sox5 and Sox6 in vivo and determine whether they are also absolutely required for chondrocyte differentiation and cartilage formation, as is Sox9.

CONCLUSIONS AND PERSPECTIVES

Human, mouse, and molecular genetic approaches have recently converged to successfully identify the first master chondrogenic factor, Sox9, and two additional master candidates, L-Sox5 and Sox6. Sox9’s role in cartilage formation was first recognized upon detection of mutations in and around the SOX9 gene in humans with campomelic dysplasia. Sox9, L-Sox5 and Sox6 were then found to be expressed concomitantly with the cartilage marker Col2a1 in all precartilaginous condensations and fully formed cartilages of mouse embryos, and to cooperatively activate Col2a1 through direct binding to a cartilage-specific intron-1 enhancer. Gene inactivation experiments in the mouse then demonstrated the master role of Sox9 in the conversion of mesenchymal precursor cells into differentiated chondrocytes. Indeed, without Sox9, cells committed to the chondrocytic lineage were unable to form precartilaginous condensations and to activate a number of cartilage markers, including Col2a1 and aggrecan. Preliminary
data in mouse mutants have indicated that L-Sox5 and Sox6 are also essential for proper chondrogenesis in vivo. The identification of these first three chondrogenic differentiation factors constitutes an important step in unraveling genetic mechanisms controlling chondrogenesis but it also prompts a number of new questions. Concerning, for instance, the precise mode of action of each of L-Sox5, Sox6, and Sox9. Do they really form a trio of chondrogenic factors? How do they co-operate with each other at the molecular level? Are they involved at the same step of chondrocyte differentiation or at successive steps? What is their panel of target genes? Questions can also be raised concerning the mechanisms whereby their genes are turned on in prechondrocytic cells and off in hypertrophic chondrocytes. What are the roles of bone morphogenetic proteins, fibroblast growth factors, sonic and indian hedgehog, and many other regulatory and signaling factors in controlling or modulating the expression of the three Sox genes or the activity of the three Sox proteins? As L-Sox5, Sox6, and Sox9 are essential during chondrogenesis in the embryo, questions can also be raised concerning their implication in the maintenance of the chondrocyte phenotype in permanent cartilages, and in cartilage degeneration and regeneration processes in the adult. The answer to all these questions will provide new important insights into our understanding of how L-Sox5, Sox6 and Sox9 control essential steps in the chondrogenesis pathway. In parallel, it is to hope that additional factors will soon be identified that have important roles in chondrogenesis, in parallel with the three Sox proteins or upstream and downstream of them. Piece by piece, we will then be able to progressively reconstitute the whole network of genetic cascades that specify the complex and multistep differentiation of chondrocytes.

Acknowledgments

Studies in the authors’ laboratories were supported by NIH grants AR46249 to VL, HD30284 to RRB, AR40335 and AR42909 to BdC, and AR42919 to RRB and BdC. VL was the recipient of a Postdoctoral fellowship and an Arthritis Investigator award from the Arthritis Foundation. We thank our collaborators for important contributions to published and unpublished studies presented here.

References


