A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene

Véronique Lefebvre¹, Ping Li and Benoit de Crombrughe¹

Department of Molecular Genetics, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Box 11, Houston, TX 77030, USA

¹Corresponding authors

e-mail: benoit_decrombrughe@molgen.mdacc.tmc.edu or veronique_lefebvre@molgen.mdacc.tmc.edu

Transcripts for a new form of Sox5, called L-Sox5, and Sox6 are coexpressed with Sox9 in all chondrogenic sites of mouse embryos. A coiled-coil domain located in the N-terminal part of L-Sox5, and absent in Sox5, showed >90% identity with a similar domain in Sox6 and mediated homodimerization and heterodimerization with Sox6. Dimerization of L-Sox5/Sox6 greatly increased efficiency of binding of the two Sox proteins to DNA containing adjacent HMG sites. L-Sox5, Sox6 and Sox9 cooperatively activated expression of the chondrocyte differentiation marker *Col2a1* in 10T1/2 and MC615 cells. A 48 bp chondrocyte-specific enhancer in this gene, which contains several HMG-like sites that are necessary for enhancer activity, bound the three Sox proteins and was cooperatively activated by the three Sox proteins in non-chondrogenic cells. Our data suggest that L-Sox5/Sox6 and Sox9, which belong to two different classes of Sox transcription factors, cooperate with each other in expression of *Col2a1* and possibly other genes of the chondrocytic program.

Keywords: chondrogenesis/collagen2/Sox5/Sox6/Sox9

Introduction

Sox (Sry-type HMG box) proteins, which form a subfamily of DNA-binding proteins with a high-mobility-group (HMG) domain, have critical functions in a number of developmental processes, including sex determination, neurogenesis and skeleton formation (Laudet *et al.*, 1993; Pevny and Lovell-Badge, 1997; Southard-Smith *et al.*, 1998). Individual members of the Sox family show >50% identity in their HMG domain to Sry, the testis-determining factor (Wright *et al.*, 1993). An essential role for SOX9 in skeleton formation was demonstrated with the identification of mutations in *SOX9* in patients with campomelic dysplasia (Foster *et al.*, 1994; Wagner *et al.*, 1994; Kwok *et al.*, 1995; Meyer *et al.*, 1997). This disease is characterized by severe malformations of essentially all cartilage-derived structures and is also often associated with XY sex reversal (Houston *et al.*, 1983; Mansour *et al.*, 1995).

Cartilage formation is a complex and essential process

in vertebrates. Cartilages are obligatory templates for the formation of endochondral bones during development and also constitute permanent skeletal structures in the respiratory tract, in articular joints and other organs. Chondrocytes differentiate following condensation of mesenchymal cells in different locations of the embryo, including the frontonasal mass, branchial arches, sclerotomes and limb buds. Typically, chondrocytes express a set of genes encoding cartilage-specific extracellular matrix components such as collagen II (encoded by the *Col2a1* gene), collagens IX and XI, and aggrecan. In growth plates, chondrocytes undergo further differentiation and hypertrophy, producing a matrix in which type X collagen is abundant and calcification occurs. Apoptosis follows and cartilage is replaced by bone. Our understanding of chondrogenesis at the molecular level is still limited. Several cytokines, including bone morphogenetic proteins, Indian Hedgehog, parathyroid hormone-related peptide and fibroblast growth factors, are involved in either skeletal patterning or discrete steps of the chondrogenic pathway, and several transcription factors, such as Hox and Pax family members, are involved in patterning of skeletal primordia (Cancedda *et al.*, 1995; Erlebacher *et al.*, 1995; Hall and Miyake, 1995). However, less is known about the transcription factors that control the determinative switch for chondrocyte differentiation and the activation of marker genes at each step of the chondrogenic cascade.

Our laboratory has used genes for specific cartilage matrix components to identify transcription factors that control gene expression in chondrocytes. We have shown that a multimerized 48 bp sequence in the first intron of *Col2a1* is sufficient to confer chondrocyte-specific expression both in transgenic mice and in transient transfection of cultured cells (Lefebvre *et al.*, 1996). SOX9 binds to the enhancer and activates *Col2a1* constructs in transient transfections of non-chondrocytic cells (Lefebvre *et al.*, 1997) and in transgenic mice (Bell *et al.*, 1997). SOX9 also activates the *Col2a1* gene when ectopically expressed in some non-cartilaginous sites in transgenic mice (Bell *et al.*, 1997). Moreover, *Sox9* is expressed along with *Col2a1* during chondrogenesis in mouse embryos (Wright *et al.*, 1995; Ng *et al.*, 1997; Zhao *et al.*, 1997). Therefore, direct activation of *COL2A1* is believed to be an important function of SOX9 in chondrogenesis (Lefebvre and de Crombrughe, 1998).

Several lines of evidence suggest that other transcription factors in addition to SOX9 may be needed to specify the high-level expression of *COL2A1* in chondrocytes. *Sox9* is expressed in cells that do not express *Col2a1*, such as those in genital ridges and specific areas of the embryonic heart (Ng *et al.*, 1997; Zhao *et al.*, 1997). *Sox9* is highly expressed in the Sertoli cells of the testis and is involved in male gonad differentiation (Kent *et al.*, 1996; Morais da Silva *et al.*, 1996). The target genes of *Sox9* in these

cells and in chondrocytes are not clearly defined, but the phenotypes of these two cell types are so different that it is likely that Sox9 contributes to their differentiation by controlling expression of different genes. Different functions of Sox9 in these cells must be specified by differential expression of other factors. Ectopic expression of *SOX9* in transfected cells, in which *Col2a1* was silent, did not result in *Col2a1* activation (V.Lefebvre and B.de Crombrughe, unpublished data), and ectopic expression of *SOX9* in transgenic mice led to activation of *Col2a1* only in a subset of tissues within the domain of ectopic expression of *SOX9* (Bell *et al.*, 1997). We therefore hypothesized that other factors either activate or derepress *SOX9* or cooperate with *SOX9* in *COL2A1*-expressing cells.

We recently reported that the 48 bp enhancer of *Col2a1* formed a large and abundant complex with nuclear proteins from chondrocytes but not from other cells (Zhou *et al.*, 1998). These proteins were designated CSEPs, for chondrocyte-specific enhancer-binding proteins. They included Sox9 and unidentified protein(s). CSEPs appeared to contact the 48 bp DNA at several sites homologous to a consensus for HMG-domain proteins. Mutagenesis demonstrated a good correlation between binding of CSEPs to DNA and enhancer activity in chondrocytes, both in transient transfection experiments and in transgenic mice. In addition, we showed that two chondrocyte-specific enhancer elements located in the *Col11a2* promoter contained HMG-like sites that were essential for enhancer activity and formation of an enhancer–CSEP-like complex (Bridgewater *et al.*, 1998). These results suggested that other HMG-domain proteins cooperate with Sox9 to generate *Col2a1* and *Col11a2* enhancer activity and presumably gene expression in chondrocytes.

We show here that in addition to Sox9, CSEPs are composed of a new long form of Sox5 (L-Sox5), and of Sox6, which both are members of a Sox subclass different from that of Sox9. L-Sox5 and Sox6 harbor a coiled-coil domain that mediates protein dimerization and efficient binding to adjacent HMG DNA sites. The three Sox genes are coexpressed in chondrogenesis and cooperate in *Col2a1* activation. Our data strongly suggest that L-Sox5, Sox6 and Sox9 together contribute to control *Col2a1*, and perhaps other important genes of the chondrocyte phenotype.

Results

A long form of Sox5 (L-Sox5), Sox6 and Sox9 form complexes with the 48-bp *Col2a1* enhancer

The CSEP proteins that form a chondrocyte-specific complex with the 48 bp *Col2a1* enhancer were previously shown to include a protein or proteins with an apparent M_r of 75–95 kDa (Zhou *et al.*, 1998). These proteins exhibited DNA-binding properties of HMG-domain proteins, including binding to several HMG-like sites in the *Col2a1* 48 bp enhancer, binding to a probe containing a consensus binding site for HMG-domain proteins (1HMG probe), binding to the minor groove of DNA and binding to DNA in the presence of poly(dG–dC) but not poly(dI–dC) (Zhou *et al.*, 1998). We also obtained evidence that CSEP bound with high affinity to a tandem dimer of the

1HMG probe (2HMG probe) both in EMSA and in Southwestern blots (data not shown).

On the basis of these results, the 2HMG probe was chosen to clone cDNAs for CSEPs by the Southwestern screening approach. cDNA expression libraries were made from primary chondrocytes of newborn mouse ribs. Several clones that showed stronger binding to the 2HMG probe in the presence of poly(dG–dC) than poly(dI–dC) encoded sequences of Sox5 or Sox6. Interestingly, whereas the previously reported transcript for Sox5 had a length of 2 kb and encoded a 43 kDa protein (Denny *et al.*, 1992), the Sox5 cDNA that was reconstituted from overlapping clones was 3.9 kb long and encoded a 75 kDa protein corresponding to Sox5 with an additional N-terminal sequence (data submitted to DDBJ/EMBL/GenBank; see later, in Figure 4). This long form of Sox5 was designated L-Sox5. Sox6 cDNA clones encoded Sox6 isoforms identical or very similar to those described for testis (data submitted to DDBJ/EMBL/GenBank; see later, in Figure 4).

Antibodies were generated against the C-termini of Sox5, Sox6 and Sox9, and affinity purified. In Western blotting, each antibody species specifically recognized its Sox protein target in extracts of fibroblasts transfected with Sox expression plasmids (Figure 1A). In EMSA, each antibody preparation specifically supershifted complexes formed between DNA and its target (Figure 1B). In Western blots of RCS cell extracts, the antibodies strongly reacted with proteins at the level of L-Sox5, Sox6 and Sox9 (Figure 1C), indicating that each of these Sox proteins was indeed made by these cells. The same reactions were seen with extracts of primary chondrocytes and chondrocytic MC615 cells (data not shown), but not with extracts of BALB/3T3 (Figure 1C) or 10T1/2 fibroblasts (see control lanes in Figure 1B). Note that Sox5 antibodies showed no reaction in chondrocyte samples at the level of the 43-kDa short form of Sox5. In extracts of adult mouse testis, Sox5 antibodies recognized Sox5 but not L-Sox5 (Denny *et al.*, 1992). It appeared, therefore, that only L-Sox5 is made in chondrocytes, not Sox5, whereas only Sox5 is made in the testis. These protein data are consistent with RNA data, which show that chondrocytes express only the transcript for L-Sox5 and testis only the transcript for short Sox5 (see Figure 2A). In EMSA, L-Sox5 and Sox6 made in transfected fibroblasts formed complexes with the 48 bp *Col2a1* enhancer probe that migrated at the level of the CSEP–DNA complex (Figure 1D). Sox5 did not bind to the 48 bp element. As shown previously (Zhou *et al.*, 1998), SOX9 formed two complexes with the enhancer, a major one migrating faster than the CSEP–DNA complex and a minor one migrating at the level of the CSEP–DNA complex (Figure 1D). Sox5, Sox6 and Sox9 antibodies were each able to partially supershift the CSEP–48-bp-*Col2a1* enhancer complex formed with RCS cell nuclear extracts (Figure 1E), indicating that each of these Sox proteins contributed to complex formation. A virtually complete supershift of the CSEP–DNA complex was obtained by incubating EMSA reactions with both Sox5 and Sox6 antibodies. No further supershift was visible when the three Sox protein antibodies were included in the reactions. The same results were obtained with primary chondrocyte extracts (data

Fig. 1. L-Sox5, Sox6 and Sox9 are present in chondrocytes and form the CSEP-Col2a1-48-bp enhancer complex. (A) Antibodies (AB) against Sox5, Sox6 and SOX9 specifically recognized their target in Western blots. Samples were extracts of 10T1/2 fibroblasts transfected with expression plasmids encoding no protein (–), L-Sox5, Sox6A or SOX9. The M_r of protein standards ($\times 10^3$) is indicated. Reaction with antibodies is seen at the level expected for each Sox protein. (B) Antibodies against Sox5, Sox6 and Sox9 specifically supershifted the complex of their target with DNA. L-Sox5 and Sox6 protein samples were cell extracts from 10T1/2 fibroblasts transfected with Sox expression plasmids. The SOX9 sample was the product of *in vitro* transcription/translation with SOX9 plasmid. Samples were used in EMSA with the 2HMG probe. Two microliters of SOX crude antisera were included, as indicated. A preimmune serum was used as a control (first lane for each protein). None of the complexes seen in the figure was seen in samples containing no SOX protein (data not shown). SOX9 synthesized *in vitro* formed two distinct complexes with the 2HMG probe. The lower and upper complexes probably involved one and two molecules of SOX9 per molecule of DNA, respectively. (C) Antibodies against Sox5, Sox6 and Sox9 identified their target in Western blots of nuclear extracts of RCS cells but not BALB/3T3 fibroblasts. An intense reaction with antibodies was seen at the level of each Sox protein. (D) L-Sox5, Sox6 and SOX9 bound to the 48 bp probe. 10T1/2 fibroblasts were transfected with expression plasmids for no protein (–), L-Sox5, Sox6, SOX9 or Sox5. Extracts of these cells and nuclear extracts of RCS cells were used in EMSA with the 48 bp *Col2a1* probe. L-Sox5 and Sox6 formed a complex with the probe that ran with a mobility similar to that of the CSEP-DNA complex. As shown previously (Zhou *et al.*, 1998), SOX9 formed two complexes with the *Col2a1* enhancer probe, a minor one that migrated at the level of the CSEP-DNA complex (arrow), and a major one that migrated faster (arrowhead). (E) Antibodies against Sox5, Sox6 and Sox9 supershifted the CSEP-48-bp-*Col2a1* complex. Two or four microliters of SOX antisera (AB) were included in EMSA of RCS nuclear extracts with the 48 bp *Col2a1* probe, as indicated. Sox5 preimmune serum was used in the control reaction (first lane). None of the preimmune serums supershifted any protein-DNA complex (data not shown). As described previously (Zhou *et al.*, 1998), Sox9 present in RCS extracts did not form a fast migrating complex with DNA (as seen in D). (F) EMSA with wild-type and mutant 48 bp *Col2a1* probes. The upper strand of the wild-type 48 bp element (wt) is shown from 5' to 3'. Sites 1–4 are 7 bp HMG-like binding sites. Nucleotides corresponding to those of the heptamer consensus HMG site C[A/T]TTG[A/T][A/T] are underlined. Mutated sites are spelt out, whereas wild-type nucleotides are indicated by dots. Sites 1–3 and mutants mA1, mA3, mA4, mA6 and mA8 were described previously (Zhou *et al.*, 1998). Note that site 2 in mA3 harbors the same five consensus nucleotides as wild-type site 2, whereas, in other mutations, sites 1–4 retained only three or four consensus nucleotides. Binding of CSEP to wild-type and mutant 48 bp probes was tested using RCS nuclear extracts. Binding of L-Sox5, Sox6 and SOX9 was tested using extracts of 10T1/2 cells transfected with either one of the Sox protein expression plasmids. All probes had the same radioactivity. Arrow, migration level of CSEP. Arrowhead, fast-migrating complex of SOX9 with DNA.

Fig. 2. Long transcripts of *Sox6* and *Sox5* are expressed at high levels in chondrocytes. (A) Northern blot with total RNA. Samples were as follows: Pr. Ch., primary chondrocytes; 2d p. Ch., chondrocytes at the second passage in culture; RCS cells; newborn mouse brain and adult mouse testis. Hybridization was performed with probes that recognized both the long and short transcripts of either *Sox5* or *Sox6*. Staining of 28S and 18S rRNA is shown as the reference for RNA loading. The sizes of *Sox5* and *Sox6* transcripts were calculated by comparison with the migration of RNA standards. (B) Northern blot with total RNA. Samples were as follows: Pr. Ch., primary chondrocytes; Pr. sk. fibr., primary skin fibroblasts from newborn mice; RCS cells; MC615 mouse immortalized chondrocytic cells at an early passage; 10T1/2 and BALB/3T3 mouse embryo fibroblasts; ROS rat osteosarcoma cells; C₂C₁₂ mouse myoblastic cells; EL4 mouse lymphoma cells; HeLa human carcinoma cells; COS monkey kidney cells; and MCTs, mouse immortalized chondrocytes. Hybridization was performed with probes for the long and short transcripts of either *Sox5* or *Sox6*. Only signals at the level of the long transcripts are shown. No significant hybridization was seen at the level of the short transcripts in any sample (data not shown). Hybridization with an 18S rRNA probe is shown as the control for RNA loading. RNA samples were the same as in Lefebvre *et al.* (1997). (C) Same Northern blot as in (B), but with total RNA from newborn mouse tissues.

not shown). L-Sox5 and Sox6 therefore appeared to be predominant CSEP components.

Altogether these results indicated that three different Sox proteins, L-Sox5, Sox6 and Sox9, were present in chondrocytes and bound to the 48 bp *Col2a1* element. L-Sox5 is a long product of the *Sox5* gene that has not been identified previously.

L-Sox5, Sox6 and Sox9 contact several HMG-like sites in the 48 bp *Col2a1* enhancer

The *Col2a1* 48 bp enhancer contains four HMG-like sites, each containing 5 or 6 bp of the 7 bp consensus HMG site (Figure 1F). Previously we showed that mutations that disrupted any one of sites 1–3 inhibited enhancer activity in chondrocytes in transgenic mice and partially inhibited binding of CSEP to the enhancer (mutants mA1, mA4 and mA6 in Figure 1F; Zhou *et al.*, 1998). A mutation that disrupted site 4 also impaired binding of CSEP to the enhancer (mutant mA9 in Figure 1F). The effect of this mutation on enhancer activity has not been tested. A mutation that preserved all five HMG consensus nucleotides of site 2 did not affect the binding of CSEP (mA3 in Figure 1F) and only weakly inhibited enhancer activity in RCS cells (Zhou *et al.*, 1998). A mutation that disrupted three sites completely inhibited binding of CSEP (mA8 in Figure 1F; Zhou *et al.*, 1998). We also reported

previously that subfragments of the 48 bp element containing either one of the four HMG-like sites alone were unable to bind CSEP or to compete with the 48 bp element for binding of CSEP (Lefebvre *et al.*, 1996, 1997; Zhou *et al.*, 1998; our data, not shown). These results indicated that the four sites of the 48 bp element cooperatively contributed to formation of the CSEP–enhancer complex.

As described previously (Lefebvre *et al.*, 1997; Zhou *et al.*, 1998), the faster-migrating SOX9–DNA complex was inhibited by mutation of site 3 but not by mutation of other sites (Figure 1F). In contrast, the slower-migrating SOX9–DNA complex was partially inhibited by mutations disrupting any of the four sites and was completely inhibited by mutation of several sites. These results are consistent with the notion that the faster-migrating complex was formed by binding of one molecule of SOX9 to site 3, whereas the slower-migrating complex was formed by cooperative binding of two or more SOX9 molecules to several sites on each DNA molecule.

Consistent with the results obtained with CSEP, binding of L-Sox5 or Sox6 to the 48 bp enhancer was partially inhibited by mutations that disrupted any one of the four HMG-like sites and completely inhibited by a mutation of several sites (Figure 1F). Mutation mA1 slightly but reproducibly inhibited binding of L-Sox5 and Sox6 to DNA. As seen with CSEP, L-Sox5 and Sox6 were unable

to bind to DNA probes harboring only one of the *Col2a1* HMG-like sites (data not shown). Taken together, these data indicated that L-Sox5 and Sox6 were able to contact cooperatively all four HMG-like sites of the *Col2a1* 48 bp enhancer.

In chondrocyte nuclear extracts, the faster-migrating complex of Sox9 with the 48 bp enhancer was not seen (Zhou *et al.*, 1998; compare also extracts of RCS cells and extracts of SOX9-transfected 10T1/2 cells in Figure 1D), but antibody supershift experiments indicated that Sox9 was present in the slower migrating CSEP complex. This result strongly suggested that, in chondrocytes, Sox9 bound the 48 bp element not as a single molecule but in cooperativity with other Sox9, L-Sox5 or Sox6 molecules.

These data suggest a model whereby the four HMG-like sites of the 48 bp element and the three Sox proteins could participate *in vivo* in the formation of a large protein-enhancer complex that would activate expression of *Col2a1* in chondrocytes. Consistent with this model in which mutation of any of the HMG-like sites would dismantle the protein-enhancer complex, we have observed that disruption of any of sites 1, 2 or 3 resulted in abolition of the activity of the enhancer in chondrocytes of transgenic mice (Zhou *et al.*, 1998).

A search for other HMG consensus and HMG-like sites (with six nucleotides of the heptamer consensus) in 1 kb of *Col2a1* promoter sequence and in a 468 bp element of the first intron (+1878/+2345), which acts as a strong chondrocyte-specific enhancer in transgenic mice (Zhou *et al.*, 1995), revealed the presence of a few scattered sites, but no cluster of two or more binding sites for Sox9, L-Sox5 and Sox6 was found that resembled that of the 48 bp element (data not shown).

Long transcripts of the Sox6 and Sox5 genes are expressed in chondrocytes

Sox5 and Sox6 were previously shown to be highly expressed in the testis of adult mice (Denny *et al.*, 1992; Connor *et al.*, 1995; Takamatsu *et al.*, 1995). The transcripts were ~2.0 and ~3.2 kb long, respectively. Traces of longer transcripts (~10 kb) were described for Sox6 in immature mouse testis (Takamatsu *et al.*, 1995) and several tissues of adult mice (Connor *et al.*, 1995), and also for SOX5 in human fetal brain (Wunderle *et al.*, 1996).

In Northern blots of total RNA, the short transcripts of Sox6 (3.2 kb) and Sox5 (2 kb) were abundant in adult mouse

testis, as expected (Figure 2A). These short transcripts were not expressed (or were expressed at a very low level in the case of Sox6) in chondrocytes or in any cell line or newborn mouse tissue examined (Figure 2A; data not shown). A 6.3 kb Sox5 transcript and a 7.7 kb Sox6 transcript were found in similar relative abundance in primary chondrocytes from ribs of newborn mice, as well as in RCS and early-passage MC615 cells (Figure 2A and B). These three chondrocyte cells were previously shown to be well differentiated, expressing *Col2a1* at a high level in parallel with Sox9 (Lefebvre *et al.*, 1997). When rib chondrocytes were allowed to dedifferentiate by two passages in monolayer culture, Sox5 and Sox6 expression sharply declined (Figure 2A), as did *Col2a1* and Sox9 expression (Lefebvre *et al.*, 1997). Transcripts of Sox5, Sox6 and Sox9 were found in some non-chondrocytic cell types, but in contrast to chondrocytes, none of the non-chondrocytic cells coexpressed the three Sox genes or expressed *Col2a1* at high levels (Figure 2B; Lefebvre *et al.*, 1997).

The long transcripts of Sox5 and Sox6 were present in the brain and some other non-cartilaginous tissues of newborn mice (Figure 2A and C). However, no non-cartilaginous tissue, besides brain and testis, was found to coexpress Sox9, Sox5 or Sox6 (Figure 2A and C; Lefebvre *et al.*, 1997). In testis, Sox9 expression is restricted to the somatic Sertoli cells (Kent *et al.*, 1996; Morais da Silva *et al.*, 1996); Sox5 is expressed as a short transcript in post-meiotic germ cells, mostly in round spermatids (Denny *et al.*, 1992); and, since expression of Sox6 correlates with that of the protamine gene (Takamatsu *et al.*, 1995), a marker of spermatid differentiation, it is possible that Sox6 is expressed in testis in the same cells and at the same time as Sox5. The three Sox proteins, therefore, do not appear to be expressed in the same cells in testis.

In conclusion, chondrocytes, and perhaps some brain cells, appear to express together long transcripts for Sox5 and Sox6 and transcripts for Sox9. Expression of these transcripts correlates with expression of *Col2a1* in chondrocytes.

The long transcripts of Sox5 and Sox6 are coexpressed with Sox9 and Col2a1 during chondrogenesis in mouse embryos

Previous whole-mount *in situ* experiments showed that Sox6 was expressed in the developing nervous system

Fig. 3. Transcripts for L-Sox5, Sox6 and Sox9 are coexpressed with *Col2a1* in chondrogenesis in mouse embryos. (A) *In situ* hybridization of sections through 10.5-day-old mouse embryos. A dark-field picture of a section hybridized with *Col2a1* probe was inverted to show the following: ba1-3, first to third branchial arches; fb, forelimb bud; hb, hindlimb bud; hm, head mesenchyme; ne, neural epithelium; no, notochord; nt, neural tube; ot, otic vesicle; and sc, sclerotome. All these sites expressed transcripts of type II collagen, L-Sox5, Sox6 and Sox9, as shown in dark-field pictures. Expression of *Col2a1* in hindlimb buds is not yet seen in 10.5-day-old embryos, in contrast to expression of the three Sox genes. (B) *In situ* hybridization of sagittal sections through 12.5-day-old mouse embryos. Dark-field pictures of sections hybridized with *Col2a1* and Sox6 probes were inverted to show skeletal and non-skeletal structures; h.t.c., cartilage primordium of hyoid bone, thyroid and cricoid cartilages. Dark-field pictures show expression of transcripts for type II collagen, L-Sox5, Sox6 and Sox9 in these areas. (C) *In situ* hybridization of longitudinal sections through a forelimb of a 15.5-day-old mouse embryo. Adjacent sections were hybridized with collagen and Sox RNA probes. *Col10a1*, gene for $\alpha 1(X)$ collagen; *Colla1*, gene for pro- $\alpha 1(I)$ collagen. Arrows point to the diaphysis of cartilages in the metacarpus, where chondrocytes become hypertrophic before giving way to osteoblasts. In these areas, chondrocytes are downregulating expression of the Sox genes and activating expression of *Col10a1*. A bracket underlines the diaphysis of the ulna, where ossification is more advanced than in the metacarpals: at each extremity, hypertrophic chondrocytes have turned off expression of the Sox genes and are switching from *Col2a1* to *Col10a1* expression; in the center, osteoblasts are expressing *Colla1*. (D) *In situ* hybridization of longitudinal sections through a hindlimb of a 17.5-day-old mouse embryo. Adjacent sections were hybridized with collagen and Sox RNA probes. Brackets show different zones of cartilage and bone in the tibia. Chondrocytes in hyaline cartilage and proliferating chondrocytes in growth plates actively express *Col2a1* and transcripts for L-Sox5, Sox6 and Sox9; at a later stage of differentiation, hypertrophic chondrocytes no longer express the Sox genes, and *Col2a1* expression is being progressively replaced by *Col10a1* expression. Osteoblasts and cells in surrounding tissues are expressing *Colla1*.

of mouse embryos; expression was high in early-stage embryos, but disappeared by embryonic day 12.5 (Connor *et al.*, 1995). To obtain information on *Sox6* and *Sox5* expression during chondrogenesis in mouse embryos, we performed a series of *in situ* experiments and compared the expression patterns of these two *Sox* genes with those of *Sox9* and *Col2a1*. For *Sox5*, we chose a probe that recognized the long transcript of *Sox5* but not the short

transcript, since we were interested in sites of expression of the L-*Sox5* protein only. For *Sox6*, two probes were tested that recognized either the long transcript only or both the long and short transcripts, since both transcripts appear to encode the same protein (Connor *et al.*, 1995; Takamatsu *et al.*, 1995); the two probes gave identical results (data not shown).

At about mid-stage embryogenesis (day 10.5 post-

coitum), mesenchymal cells form prechondrocytic condensations at different sites in the embryo. The sclerotomal components of somites contain precursor cells for the axial skeleton, the head mesenchyme and first and second branchial arches will generate part of the craniofacial skeleton, and the lateral plate mesoderm and limb buds will give rise to the appendicular skeleton. Transcripts for L-Sox5 and Sox6 were found in all these sites together with *Col2a1* and *Sox9* RNAs (Figure 3A; Cheah *et al.*, 1991; Wright *et al.*, 1995; Ng *et al.*, 1997; Zhao *et al.*, 1997). The four RNAs also co-localized in some non-chondrogenic sites such as brain, neural tube, otic vesicles and notochord.

At day 12.5 of embryonic development, cartilage is actively forming in all future cartilaginous and endochondral skeletal structures, including cartilages of the nose and ear, the thyroid, cricoid and hyoid cartilages, Meckel's cartilage, cartilages of the base of the skull, ribs, vertebrae, forelimbs and hindlimbs (Figure 3B). High levels of transcripts for L-Sox5 and Sox6 were found in all these structures, together with *Col2a1* and *Sox9* RNAs. Expression of the four genes was also seen in some areas of the brain and spinal cord. Transcripts for Sox6 and L-Sox5, but not Sox9, were also visible in liver and a few other non-cartilaginous areas.

In the limbs of 15.5- and 17.5-day-old embryos (Figure 3C and D), expression of *Col2a1* and of the three *Sox* genes was high in the cartilaginous templates of the radius and ulna, carpals, metacarpals and tibia, and in proliferating chondrocytes of growth plates. When chondrocytes became hypertrophic in growth plates, they activated expression of *Col10a1* whereas RNAs, for all three *Sox* proteins disappeared rapidly and simultaneously and RNAs for *Col2a1* disappeared more slowly. Only traces of *Col2a1* and *Sox* RNAs were found after cartilage was replaced by bone and expression of *Col1a1* was activated in osteoblasts.

In conclusion, RNAs for collagen II, L-Sox5, Sox6 and Sox9 were expressed simultaneously and at high levels from early stages of chondrogenesis in all cartilaginous sites in mouse embryos. Expression of the three *Sox* genes

appeared to be inhibited just before *Col2a1* expression in hypertrophic chondrocytes. These results are consistent with a role for all three *Sox* proteins in the activation of *Col2a1* in chondrogenesis, and possibly also in the activation of other genes of the chondrogenic program.

Comparison of Sox5, L-Sox5 and Sox6 cDNAs and polypeptides

Connor *et al.* (1995) and Takamatsu *et al.* (1995) previously isolated Sox6 cDNA species from adult mouse testis cDNA libraries. The cDNAs reported by the two groups encoded slightly different proteins, called Sox6 and Sox-LZ, respectively. For convenience, we have renamed them Sox6A and Sox6B. The proteins differ in two segments, most likely encoded by alternatively spliced exons (Figure 4A). Segment 1 encodes a MAAAA amino acid sequence (S1A) in Sox6A and a SSAAA sequence (S1B) in Sox6B, with different A codons being used in S1A and S1B. Segment 2 encodes a 41-amino-acid sequence and is present only in Sox6A. When total cDNA from primary chondrocytes was used in the polymerase chain reaction (PCR) with primers flanking the region of the S1 and S2 segments, the amplified products had the size expected for Sox6A (478 bp) and Sox6B (355 bp) products (data not shown). Sequencing of the PCR products indicated that the longer product contained S1A and S2, and was identical to Sox6A. The shorter product contained S1A and lacked S2. It therefore corresponded to a new form of Sox6, which we have named Sox6C (Figure 4A). In DNA-binding and transactivation assays, the three forms of Sox6 displayed identical activities (data not shown). These PCR data and additional sequencing of Sox6 cDNA clones from chondrocyte cDNA libraries suggested that the short and long transcripts of Sox6 both encoded protein isoforms with minor differences. Sox6 cDNA clones obtained from chondrocyte libraries contained an additional 3' untranslated sequence at the 3' end of the previously reported cDNAs from testis (data submitted to DDBJ/EMBL/GenBank). The longest sequence was 1.5 kb and accounts in part for the difference between the long and short transcripts of *Sox6*.

Fig. 4. Comparison of Sox5, L-Sox5 and Sox6 cDNAs and polypeptides. (A) Schematic comparison of Sox6A, Sox6B and Sox6C polypeptides. Three forms of Sox6 differed in segments S1 and S2. Sox6A and Sox6B are schematized according to the sequences reported by Connor *et al.* (1995) and Takamatsu *et al.* (1995), respectively. Sox6C, whose cDNA sequencing has not been completed, is schematized based on the assumption that it is identical to Sox6A and Sox6B outside of segments S1 and S2. Two coiled-coil domains (1st cc and 2d cc) and the HMG DNA-binding domain, are indicated. S1A and S1B segments are compared at the nucleotide level and, in parentheses, at the amino acid level. (B) Comparison of Sox5 and L-Sox5 cDNAs. Sequences available for Sox5 and L-Sox5 cDNAs are presented as blocks. Shaded areas indicate coding sequences, between the first in-frame ATG codon and the next stop codon. Sox5 and L-Sox5 cDNAs are identical in the segment delineated by the two dotted lines, but they differ totally on either sides of this segment. Numbers refer to nucleotide positions relative to the 5' end. The Sox5 cDNA representation follows the sequence published by Denny *et al.* (1992). L-Sox5 cDNA is schematized according to a 3881 bp sequence obtained from overlapping clones. Both cDNAs must lack 5' and/or 3' untranslated sequences, as RNA transcript lengths were estimated to be ~2 kb for Sox5 and ~6.3 kb for L-Sox5. (C) Alignment of L-Sox5 and Sox6A amino acid sequences. The sequence of L-Sox5 (upper sequence), derived from its cDNA sequence, is compared with that of Sox6A (lower sequence) (Connor *et al.*, 1995). Numbers refer to amino acid positions in the proteins. Sequence alignment was generated using the GAP program from the Genetics Computer Group (GCG, Madison, WI). Dots were introduced within sequences to maximize alignment. Vertical lines denote amino acid identities. Double and single dots between sequences indicate amino acid changes with high and low degrees of conservation, respectively. The methionine translation initiation codon of the short form of Sox5 is circled in the L-Sox5 sequence (residue 288). Boxes outline two potential coiled-coil domains and the HMG DNA-binding domain. The sequence of Sox6A spanning residues 330–379 was omitted because it has no counterpart in L-Sox5. (D) Schematic comparison of Sox5, L-Sox5 and Sox6A polypeptides. Each protein is represented as a block between its N- and C-termini. HMG, HMG DNA-binding domain; cc, potential coiled-coil domain. Hatched boxes represent regions of >10 residues in Sox6A that do not exist in L-Sox5. Dotted lines link regions of similarity between the proteins. Numbers refer to the position of the residues marking domain limits. The predicted molecular weight of each protein is given in parentheses. (E) Delineation of potential coiled-coil domains in L-Sox5 and Sox6A. The amino acid sequences of L-Sox5 and Sox6A were analyzed with the Coilscan program from GCG using an unweighted matrix at a window size of 28 residues. The probability of each residue participating in coiled-coil formation is plotted against its position in the sequences. A score of 1.0 indicates a maximum probability. The first potential coiled coil involves residues 158–240 in L-Sox5 and residues 181–262 in Sox6A; the second potential coiled coil spans residues 364–403 in L-Sox5 and 488–515 in Sox6A.

The L-Sox5 cDNA that was reconstituted from overlapping clones contained a 2037 bp open reading frame that encodes a 679-amino-acid protein, with 287 amino acids N-terminal to, and in frame with, the 392-amino-acid Sox5 protein sequence (Denny *et al.*, 1992) (Figure 4B–D). It contained 189 bp of 5' untranslated sequence, with an in-frame stop codon 120 bp upstream of the first ATG (data not shown). The first three in-frame methionine codons are surrounded by sequences homologous to the Kozak consensus and are therefore putative translation initiation codons (data not shown). Sox5 (Denny *et al.*, 1992) and L-Sox5 cDNAs share an uninterrupted 1571 bp sequence that includes the 1179 bp coding sequence of

Sox5 (Figure 4B). Upstream and downstream of this region, the sequences available for the two cDNAs are totally different. Sox5 and L-Sox5 mature RNAs must therefore arise from differential splicing of primary transcripts in the 5' and 3' regions, and different promoter usage for the two RNAs cannot be excluded.

Previous comparison of the short Sox5 protein with Sox6A (Connor *et al.*, 1995) indicated an overall high degree of identity (61%), with a maximum of 93% in the HMG DNA-binding domain (Figure 4C and D). By comparison, the HMG domains of Sox5 and Sox6 are only 54 and 53% identical to that of Sry, respectively, and 50% and 49% identical to that of Sox9. The N-terminal

segment of L-Sox5 also has a high degree of sequence identity (72%) with the N-terminus of Sox6 (Figure 4C and D). Computer analysis of the protein sequences revealed two regions compatible with formation of coiled coils in L-Sox5 and Sox6 (Figure 4E) (Lupas, 1996). The more upstream coiled-coil domain is highly conserved between L-Sox5 and Sox6 (91% identity) and has a maximum probability score for coiled-coil formation of 1.0 in both proteins. It is particularly long, with 83 residues in L-Sox5 and 82 in Sox6, which implies that the coiled coil may complete almost 24 helical turns. It includes in its N-terminal part the leucine zipper previously identified in Sox6 (Connor *et al.*, 1995; Takamatsu *et al.*, 1995) and a glutamine-rich domain in its C-terminal part, referred to as a Q box (Kido *et al.*, 1998) (Figure 4C). The more downstream coiled-coil domain is located in the sequence of L-Sox5 that is shared with Sox5. This domain is shorter, with 40 residues in Sox5 and 29 in Sox6. It has a maximum probability score of 1.0 in Sox5 but a low-probability score of 0.39 in Sox6. It is only moderately conserved between the two proteins (59% identity). No region of homology with Sox9 was found outside the HMG domain.

In conclusion, the striking identity of L-Sox5 and Sox6 in the first coiled-coil domain and in the HMG domain strongly suggests that each of these domains must serve one or several important functions. The presence of these domains in both proteins and the otherwise overall high degree of identity between the two proteins (67%) suggest that L-Sox5 and Sox6 may play similar roles *in vivo*.

Dimerization of L-Sox5 and Sox6 highly stabilizes binding to adjacent HMG sites on DNA

L-Sox5 and Sox6 bound to the 2HMG probe much more efficiently than to the 1HMG probe (Figure 5A). In contrast, SOX9 bound with similar efficiency to both probes. This observation, the presence of potential coiled coils in L-Sox5/Sox6 and the ability of Sox6 to homodimerize (Takamatsu *et al.*, 1995) led to the hypothesis that dimerization of the proteins through their coiled coils might stabilize their binding to adjacent DNA-binding sites.

To test this hypothesis, various deletion forms of L-Sox5 were synthesized *in vitro* and treated with glutaraldehyde (Figure 5B), which has the ability to cross-link interacting polypeptides (Wong, 1991). After separation by SDS-PAGE (Figure 5C), polypeptides with the expected M_r were seen for each full-length and deleted protein species in samples treated with no glutaraldehyde. Upon treatment with glutaraldehyde, monomeric forms of Sox6 and L-Sox5 were less abundant, and polypeptide species appeared that had the apparent M_r of homodimers. Similarly, the 151/679 and 32/437 deletions formed cross-linked species with an M_r consistent with dimerization. In these deletions, the two coiled-coil domains were intact. Deletion 213/679, short Sox5 and deletion 32/221–304/437, all of which had the first coiled-coil domain partially or totally deleted, did not form specific cross-linked species, nor did SOX9. These results indicated that the first coiled-coil domain of L-Sox5, like that of Sox6 (Takamatsu *et al.*, 1995), was involved in specific protein-protein interactions, most likely protein homodimerization. The second coiled-coil domain of L-Sox5, which is present in 213/679, short Sox5 and deletion 32/221–304/437,

appeared to be unable to mediate protein dimerization by itself. It might, however, have a role in stabilizing protein dimerization through the first coiled-coil domain.

In EMSA (Figure 5D), deletion 151/679 bound DNA as efficiently as L-Sox5 (lanes 2 and 3), but deletion 213/679 did not (lane 4), nor did Sox5 (lane 5). The first coiled-coil domain of L-Sox5 (residues 158–240) was still intact in 151/679 but largely deleted in 213/679. It appeared therefore to be involved in the high affinity of L-Sox5 for the 2HMG probe. A protein truncated in the C-terminus (32/437) did not bind DNA (lane 6), an expected result given that the HMG domain was deleted.

When L-Sox5 was mixed with deletion 151/679, the respective complexes of the two proteins formed, but a third abundant complex also formed whose migration level indicated that it was likely to involve one molecule of each of the two proteins (lane 9). A similar intermediate complex also formed with Sox6 and 151/679 (lane 15). This result strongly suggested that L-Sox5 binds the 2HMG probe as a homodimer or heterodimer with Sox6. On the basis of its DNA-binding properties, a similar conclusion can be made for Sox6.

When L-Sox5 was mixed with deletion 213/679 (lane 10), only the complex of L-Sox5 homodimer with DNA formed, indicating that the coiled-coil domains of two molecules of L-Sox5 were needed for efficient binding to DNA. Also, when L-Sox5 was mixed with deletion 32/437 (lane 12), no heterodimers bound to the 2HMG probe, indicating that the two C-termini of L-Sox5 molecules (which included the HMG domain) were required to bind DNA.

Together, these results suggest a model (Figure 5E) in which binding of two molecules of L-Sox5 and Sox6 to adjacent HMG sites on DNA is highly stabilized by protein dimerization through the coiled-coil domains.

Cooperative activation of chondrocyte-specific Col2a1 constructs by L-Sox5, Sox6 and SOX9

The ability of L-Sox5, Sox6 and SOX9 to activate *Col2a1* reporter constructs in non-chondrogenic cells was tested using constructs in which *Col2a1* enhancer segments were placed in an intron downstream of the promoter, as is the case in the endogenous *Col2a1* gene.

A construct with a 309 bp *Col2a1* promoter (p309; Figure 6A) was previously shown to be inactive in chondrocytes of transgenic mice (Zhou *et al.*, 1995). In transient transfection of 10T1/2 cells, this construct was slightly but reproducibly activated by L-Sox5/Sox6 but not by SOX9, and no cooperation occurred among the three Sox proteins (Figure 6B). A similar construct that included four tandem copies of the 48 bp enhancer [p309–(4x48); Figure 6A] was specifically expressed in chondrocytes both in transgenic mice (Lefebvre *et al.*, 1996) and in transient transfection (data not shown). In non-chondrogenic cells, this construct was not activated by L-Sox5, Sox6 or a combination of L-Sox5 and Sox6 at a higher level than the p309 construct, and it was moderately activated by SOX9 (Figure 6C). Interestingly, when SOX9 was co-expressed with L-Sox5, Sox6, or both L-Sox5 and Sox6, a higher activation occurred than with any Sox protein alone, demonstrating cooperativity between SOX9 and L-Sox5/Sox6.

Fig. 5. Dimerization of L-Sox5 and Sox6 stabilizes binding to adjacent HMG sites. (A) L-Sox5 and Sox6 bind the 2HMG probe more efficiently than the 1HMG probe. Extracts of 10T1/2 fibroblasts transfected with empty (–), SOX9, L-Sox5 or Sox6 expression plasmids were incubated with the 1HMG probe or the 2HMG probe. The two probes had the same radioactivity. Arrow, L-Sox5–Sox6 complexes with DNA and slow-migrating SOX9–DNA complex. Arrowhead, fast-migrating SOX9–DNA complex; this complex was often seen as a doublet. (B) Deletions in L-Sox5. L-Sox5 full-length protein is schematized as described in Figure 4D. Truncated proteins were named by the first and last residues of L-Sox5 that they contain. In 32/221–304/437, the sequence 222/303 in 32/437 was deleted. (C) The first coiled-coil domain of L-Sox5 and Sox6 is involved in protein–protein interactions. Sox6, L-Sox5, truncated L-Sox5 polypeptides and SOX9 were synthesized *in vitro* with [³⁵S]methionine and then incubated with (+) or without (–) glutaraldehyde. Autoradiographs are shown after polypeptide separation by SDS–PAGE. Arrowheads, glutaraldehyde-induced cross-linked polypeptides. The *M_r* of protein standards is indicated ($\times 10^3$). (D) L-Sox5–Sox6 bound the 2HMG probe as dimers. Extracts of 10T1/2 fibroblasts transfected with empty (–), L-Sox5 or Sox6 expression plasmids were preincubated with products of *in vitro* transcription/translation obtained with plasmids encoding no protein (–) or one of the truncated L-Sox5 proteins (indicated on top of the lanes). EMSA was performed with the 2HMG probe. Arrows, complexes of L-Sox5 and Sox6 with DNA. Arrowhead, complex of 151/679 with DNA. Star, heterocomplexes of 151/679, and L-Sox5 or Sox6 with DNA. (E) Model. Two molecules of L-Sox5/Sox6 (ovals) form a highly stable complex with DNA upon binding to two adjacent recognition sites on DNA (thick curved line). The two protein molecules dimerize through their coiled-coil domains and may induce a strong bend of DNA at the sites of binding.

Similarly, a p309–(2 $\times$ 231) construct, which harbored two copies of a 231 bp *Col2a1* chondrocyte-specific enhancer (Lefebvre *et al.*, 1996), including the 48 bp element, was activated at a higher level by the three Sox factors together than by each Sox protein individually (Figure 6D). Deletion of a 10 bp sequence in the 231 bp enhancer, which corresponded to the 3' end of the 48 bp element, abolished enhancer activity in chondrocytes (Lefebvre *et al.*, 1997) and also transactivation by the

three Sox factors (Figure 6D). In contrast to L-Sox5/Sox6, Sox4 and the short form of Sox5 did not activate the p309–(2 $\times$ 231) construct, nor did they cooperate with SOX9 to generate a high-level transactivation of the construct in fibroblasts (Figure 6E).

Together these results indicated that L-Sox5 and Sox6 significantly enhanced transactivation of *Col2a1* constructs by SOX9. L-Sox5 and Sox6 appeared to function similarly, whereas the short form of Sox5 was inactive.

Fig. 6. Cooperative activation of chondrocyte-specific *Col2a1* promoter-enhancer constructs by L-Sox5, Sox6 and SOX9. (A) Schematic of *Col2a1* constructs. The β geo reporter gene was driven by a 309 bp (p309; -309/+308) *Col2a1* promoter. *Col2a1* enhancer segments were four tandem copies of the 48 bp intron-1 element (4x48), two tandem copies of a 231 bp element (2x231) or two tandem copies of a 231 bp element in which 10 bp (triangles) were deleted (2x221). The deletion corresponded to the 3' end of the 48 bp element. Enhancers were cloned downstream of the *Col2a1* exon 1 in an intron delimited by a splice donor (SD), the proximal 70 bp of the *Col2a1* intron 1 and a splice acceptor (SA). The '-' and '+' symbols indicate whether the constructs were inactive or active, respectively, in chondrocytes. (B) Activation of p309. 10T1/2 fibroblasts were transfected with p309 and 900 ng of expression plasmids. These included empty vector (-) and, as indicated, 300 ng of L-Sox5 (L5), Sox6 (S6) and SOX9 (S9) plasmids. (C) Activation of p309-(4x48). Similar experiment as in (B) but with p309-(4x48) instead of p309. (D) Activation of p309-(2x231) and p309-(2x221). 10T1/2 fibroblasts were transfected with p309-(2x231) or p309-(2x221), and either 600 ng of empty (-) L-Sox5, Sox6 or SOX9 plasmid, or 200 ng of each of the L-Sox5, Sox6 and SOX9 plasmids. (E) No activation of p309-(2x231) by Sox4 and Sox5. 10T1/2 fibroblasts were transfected with p309 or p309-(2x231), and either 600 ng of empty (-), Sox4 (S4), Sox5 (S5), L-Sox5, Sox6 or SOX9 plasmid, or 300 ng of SOX9 plasmid and 300 ng of empty or SOX plasmid.

Cooperative activation of chondrocyte-specific genes by L-Sox5, Sox6 and SOX9

The ability of L-Sox5/Sox6 and SOX9 to activate the endogenous *Col2a1* gene was tested in several cell types by Northern blot analysis after transient transfection with Sox expression plasmids. As described previously (Lefebvre *et al.*, 1997), 10T1/2 cells spontaneously expressed *Col2a1*, but at a much lower level than did differentiated chondrocytes (Figure 7A). Following transfection of Sox plasmids, Sox RNAs and proteins accumulated in large amounts in the cells within 24 h, but rapidly disappeared during the next 48 h (Figure 7A). A significant increase in the *Col2a1* mRNA level was observed 24 h after transfection of either SOX9 or L-Sox5/Sox6 plasmids (Figure 7A and B). Depending on the experiments, SOX9 was either as potent as, or more potent than, L-Sox5/Sox6. The increase in *Col2a1* RNA was transient and no longer detectable after 48 h (Figure 7A), and was therefore concomitant with high levels of Sox proteins. The same results were observed whether L-Sox5 and Sox6 were tested alone or together, indicating that they had redundant activities (data not shown). No significant increase in *Col2a1* expression was observed when the cells were transfected with an expression plasmid for Oct-1, a POU-domain transcription factor capable of binding to the 48 bp *Col2a1* enhancer (data not shown). This result demonstrated the specificity of Sox protein activity.

Interestingly, coexpression of SOX9 and L-Sox5/Sox6 resulted in a much higher activation of *Col2a1* expression than when each Sox protein was expressed individually. Similar results were obtained with MC615 cells (Figure 7B), which were tested after repeated passages in culture when expression of chondrocyte markers was severely reduced or lost. Their low level of *Col2a1* expression was slightly stimulated upon expression of SOX9 alone or L-Sox5/Sox6, but it was strongly stimulated by coexpression of the three Sox proteins. When the cell culture medium was supplemented with bone morphogenetic protein 2 (BMP-2), a cytokine known to promote chondrogenesis (Reddi, 1998), *Col2a1* RNA level also increased. In this condition also, overexpression of the three Sox factors highly stimulated *Col2a1* expression. When cells that did not spontaneously express *Col2a1* were tested, such as skin primary fibroblasts from newborn mice, BALB/3T3 fibroblasts and COS cells, no induction of *Col2a1* expression was detected following transfection of Sox protein expression plasmids (data not shown).

Activation by L-Sox5/Sox6 and SOX9 of other genes expressed in chondrocytes along with *Col2a1* was also examined (Figure 7B). The *aggrecan* gene, which encodes the protein core of a large aggregating proteoglycan found in abundance in cartilage, was weakly expressed by dedifferentiated MC615 cells, but overexpression of the three Sox proteins led to a clear increase in its RNA level.

Fig. 7. Cooperative activation of chondrocyte-specific genes by L-Sox5, Sox6 and SOX9. **(A)** Transient transfection of 10T1/2 cells. L-Sox5, Sox6 and/or SOX9 expression plasmids were transfected as indicated. Cells were harvested 24, 48 or 72 h after the start of transfection. Total cell extracts were used in EMSA with the 2HMG probe. Single-tailed arrow, L-Sox5–Sox6–DNA complexes; double-tailed arrow, fast-migrating SOX9–DNA complexes. The weak intensity of SOX9–DNA complexes compared with L-Sox5–Sox6–DNA complexes reflects a less-efficient binding of SOX9 to the 2HMG probe. Total RNA was hybridized in Northern blots with probes for 18S rRNA (reference for RNA loading), *Sox9* and *Col2a1*. Differentiated MC615 cells at an early passage are shown in the left lane as a reference for chondrocytic cells. Endogenous *Sox9* RNA forms a band at the top of the panel; SOX9 RNAs expressed from the transfected plasmid run as an intense band migrating faster. The more intense signal seen at the level of endogenous *Sox9* RNA in cells transfected with SOX9 may be due either to increased expression of *Sox9* or to larger transcripts from SOX9 plasmid. The exposure of the autoradiograph of the *Col2a1* blot was eight times as short (3 h) for MC615 cells as for 10T1/2 cells. The intensity of hybridization signal for *Col2a1* RNA in the different culture conditions is plotted as fold increase over the signal given by the control culture 24 h after transfection. **(B)** Transfection of MC615 cells. Cells that had essentially lost their chondrocytic phenotype by repeated passaging were transfected with Sox expression plasmids, as indicated. BMP-2 was added at the start of transfection. Cells were harvested 24 h later and RNA analyzed in Northern blot with various probes, as indicated. RNA from RCS cells was used as a reference for differentiated chondrocytes. The intensity of the *Col2a1* hybridization signal in the different conditions is plotted as fold increase over the signal given by the control culture in the absence of BMP-2. Hybridization signals obtained with an 18S rRNA probe are shown as a reference for RNA loading. The exposure of the autoradiograph of the *aggrecan* blot was five times as short for RCS cells (3 h) as for 10T1/2 cells.

RNA for *Col9a2*, which codes for the $\alpha 2$ chain of type IX collagen, was undetectable in MC615 cells and not induced by any of the three Sox protein expression plasmids. RNAs for *aggrecan* and *Col9a2* were not detectable in 10T1/2 cells, even after transfection of the three Sox protein expression plasmids (data not shown).

RNA levels of genes that are expressed in chondrocytes but not in parallel with *Col2a1* were examined in order to test the specificity of gene transactivation by the three Sox proteins. *Col10a1*, a characteristic marker of hypertrophic chondrocytes, was expressed at low levels in MC615-differentiated cells, but its expression was not affected by transfection of Sox protein expression plasmids. The genes for matrix Gla protein (MGP) and osteopontin, two extracellular matrix proteins produced by chondrocytes and also some other cell types, were expressed in MC615 cells, but transfection of the three Sox factors did not significantly affect their RNA levels (Figure 7B).

In conclusion, L-Sox5/Sox6 and SOX9 were found to stimulate cooperatively expression of *Col2a1* and also *aggrecan*, two major markers of chondrocytes. These data strongly suggest that the three Sox proteins together control important aspects of the chondrocyte phenotype.

Discussion

We have shown here that a new form of Sox5 (L-Sox5) and Sox6, which both are dimeric Sox proteins that preferentially bind to adjacent HMG sites rather than to isolated sites, are coexpressed with Sox9 during chondrogenesis, efficiently bind to several HMG-like sites in the 48 bp chondrocyte-specific *Col2a1* enhancer, and cooperate with SOX9 in activating the chondrocyte marker gene *Col2a1*.

L-Sox5 differs from Sox5 by an additional 287 amino-acid sequence at the N-terminus. L-Sox5 and Sox6 show a striking degree of identity in the HMG DNA-binding

domain located in the C-terminal part and in a coiled-coil domain located in the N-terminal part. This coiled-coil domain has been shown previously to mediate homodimerization of Sox6 (Takamatsu *et al.*, 1995), and we have shown that L-Sox5 also dimerizes through this domain, either with itself or with Sox6. Sox12, Sox13 and Sox23 (Komatsu *et al.*, 1996; Kido *et al.*, 1998; Yamashita *et al.*, 1998) also feature a similar coiled-coil domain and their HMG domain is much more similar to those of L-Sox5 and Sox6 than to those of Sry, Sox9 and other Sox family members. Based on these similarities, the five Sox proteins are classified as Sox subgroup D (Wright *et al.*, 1993; Yamashita *et al.*, 1998). We have presented strong evidence that dimerization of L-Sox5/Sox6 highly stabilizes binding to DNA at adjacent recognition sites. Indeed, L-Sox5 and Sox6 bound to an element harboring two HMG sites much more efficiently than to an element harboring a single site. Two molecules of Sox polypeptides were binding to one molecule of 2HMG probe, and both the HMG and coiled-coil domains of the two polypeptide molecules were required for efficient DNA binding. SOX9, which does not homodimerize, did not bind to the 2HMG probe more efficiently than to the 1HMG probe. The complementary roles of the coiled-coil and HMG domains of L-Sox5/Sox6 in DNA binding probably account for the high degree of conservation of both domains in D-Sox family members. It also implies that target genes for these Sox proteins must harbor pairs of HMG binding sites with a configuration compatible with binding of D-Sox protein dimers.

Transcripts for L-Sox5 and Sox6 were expressed along with Sox9 in all prechondrogenic areas and cartilages during mouse embryonic development. Expression of all three Sox genes was inhibited when chondrocytes became hypertrophic in growth-plate cartilages; in these cells expression of the chondrocyte marker gene *Col2a1* is also downregulated. Sox9, Sox5 (transcript for L-Sox5) and Sox6 were also expressed in some non-cartilaginous sites, such as notochord, otic vesicles and some areas of the brain, but they were not coexpressed where *Col2a1* was not expressed. In cell cultures, the three Sox genes were coexpressed only in primary chondrocytes and chondrocytic cell lines, both of which highly expressed *Col2a1*, and a sharp decrease in the three Sox RNA levels accompanied loss of *Col2a1* expression during chondrocyte dedifferentiation. We have also shown that the three Sox proteins were present in chondrocytes.

Our data indicate that when SOX9 and L-Sox5/Sox6 were transfected individually in 10T1/2 and MC615 cells, they only produced a modest increase in expression of the endogenous *Col2a1* gene. However, upon cotransfection, cooperativity occurred among the three Sox proteins, leading to expression levels of *Col2a1* that were an order of magnitude or more higher than in control cells. Expression of *aggrecan* was also significantly increased in MC615 cells transfected with the three Sox protein expression plasmids. The two classes of Sox proteins, L-Sox5/Sox6 and Sox9, are therefore able to cooperate functionally with each other in the activation of both *Col2a1* and *aggrecan*. These results, together with the strong correlation that exists between expression of *Col2a1*, *aggrecan* (Glumoff *et al.*, 1994) and the three Sox genes during chondrogenesis in mouse embryos,

support the notion that the three Sox proteins may also play a role in the activation of *Col2a1*, *aggrecan* and possibly other chondrocyte marker genes *in vivo*.

A 48 bp element in the first intron of *Col2a1* was previously shown to be sufficient to direct expression of a reporter gene in cartilage of transgenic mice and to contain sites essential for the activity of longer *Col2a1* enhancer segments in chondrocytes (Lefebvre *et al.*, 1996; Bell *et al.*, 1997; Zhou *et al.*, 1998). This element is therefore likely to be involved in *Col2a1* expression in chondrocytes. Sox9, L-Sox5 and Sox6 bound to the 48 bp enhancer at four HMG-like sites, three of which were demonstrated to be necessary for enhancer activity in chondrocytes (Lefebvre *et al.*, 1996; Bell *et al.*, 1997; Zhou *et al.*, 1998), whereas the role of the fourth has not been tested. On their own, L-Sox5/Sox6 were weak activators of *Col2a1* enhancer constructs. However, they efficiently cooperated with SOX9 in the activation of the enhancer, leading to a level of activation that was several-fold higher than when each Sox protein was transfected individually. The promoter-enhancer configuration in these constructs was similar to that of the endogenous *Col2a1* gene, in which the chondrocyte-specific enhancer is located in the first intron downstream of the transcription start site. L-Sox5/Sox6 were able to contact all four HMG-like recognition sites of the 48 bp *Col2a1* enhancer. They probably bound as dimers, as their complexes with the enhancer and with the 2HMG probe migrated at the same level. Even though the binding sites for L-Sox5/Sox6 in the 48 bp enhancer are different from the preferred binding sites of the short Sox5 (ACAAT and AACAAAG; Denny *et al.*, 1992), our experiments indicate that L-Sox5/Sox6, as well as SOX9, efficiently bound to the enhancer. It is possible that the proximity of several HMG-like sites in the enhancer was favorable to cooperativity between the two types of Sox proteins in achieving transcriptional activation.

Together the three Sox factors enhanced *Col2a1* and *aggrecan* expression in cells that expressed low levels of these genes, but they did not induce *Col2a1* expression in cells in which the gene was silent, even though they activated *Col2a1* constructs at high levels in transient transfection of these same cells. It is possible that additional factors or coactivators may be needed to open the chromatin of chondrocyte-specific genes that were silent in the cells that were tested or that these genes were inactivated by other epigenetic mechanisms. Bell *et al.* (1997) reported that Sox9 was capable of activating *Col2a1* in some ectopic sites of transgenic mouse embryos. It is possible that the *Col2a1* gene is not repressed in mouse embryos as it might be in tissue culture cells or that the ectopic sites in which Sox9 was able to activate *Col2a1* contained factors that allowed Sox9 to activate *Col2a1*.

Denny *et al.* (1992) reported that Sox5 was unable to activate transcription from a minimal promoter linked to multimerized Sox5 binding sites. In our experiments, L-Sox5/Sox6 were able to activate *Col2a1* constructs whereas Sox5 was not. The more efficient binding of L-Sox5/Sox6 to DNA, compared with that of Sox5, is sufficient to explain why L-Sox5/Sox6 transactivated whereas Sox5 did not. But it is also possible that the N-terminus of L-Sox5/Sox6 harbors one or more domains

involved in transactivation. Takamatsu *et al.* (1995) reported that full-length Sox6 was unable to transactivate a reporter construct containing four copies of an HMG-binding site. Transactivation and DNA binding occurred upon deletion of the leucine zipper of the protein. In our experiments, L-Sox5/Sox6 bound to the *Col2a1* enhancer and transactivated *Col2a1* constructs as full-length proteins. Differences between the two studies may be due to differences in DNA targets. It may be that the sequence of the *Col2a1* enhancer, which is a potential target of L-Sox5/Sox6, the distance between HMG-like sites and eventually the presence of binding sites for additional activators are essential for L-Sox5/Sox6 binding to DNA and transactivation function.

HMG-domain proteins have been shown to participate as architectural factors in transcriptional activation of several genes (Grosschedl *et al.*, 1994; Werner and Burley, 1997). By their ability to bend DNA, a property also demonstrated for Sox5 and SOX9 (Connor *et al.*, 1994; Lefebvre *et al.*, 1997), they facilitate interactions between proteins bound at non-adjacent DNA sites and thereby promote the assembly of multiprotein-enhancer complexes (Giese *et al.*, 1995; Pevny and Lovell-Badge, 1997). It is possible that L-Sox5/Sox6 have such a role, eventually with Sox9, in organizing an enhancer-protein complex and also in bringing this complex close to the basal transcriptional machinery, which in *Col2a1* is 2.2 kb upstream of the enhancer. But the function of Sox9 is probably not limited to an architectural role, as it was shown to have a potent transactivation domain (Südbeck *et al.*, 1996; Lefebvre *et al.*, 1997; Ng *et al.*, 1997).

The function of L-Sox5/Sox6 *in vivo* is not known. The ability of L-Sox5/Sox6 to cooperate with SOX9 in *Col2a1* activation and the strong correlation between expression of L-Sox5/Sox6 and chondrogenesis suggest that mutations in the *SOX5* and *SOX6* genes might result in cartilage malformation diseases of still unknown causes. However, because of a possible redundancy of these two highly similar factors, a mutation in only one of their genes might cause only mild or no skeletal abnormalities. Although L-Sox5 and Sox6 belong to the same family of DNA-binding proteins as Sox9, and present the same expression pattern as Sox9 in chondrogenesis, it is unlikely that they play the same role as Sox9 because they differ substantially from Sox9 in DNA-binding and transactivation properties. The severe phenotype of patients with campomelic dysplasia, in which mutations in *SOX9* are heterozygous, also strongly suggests that no other protein with a function that overlaps that of *SOX9* exists in chondrocytes. Our data strongly suggest that Sox9 and L-Sox5/Sox6 represent two different subclasses of Sox proteins that are coexpressed during chondrogenesis, where they have distinct, complementary roles in the activation of important chondrocyte phenotype markers such as *Col2a1*.

Materials and methods

cDNA cloning

cDNA libraries were made from primary chondrocytes of ribs of newborn mice. Total RNA was isolated from cells cultured for 2–3 days (Lefebvre *et al.*, 1994). Poly(A)⁺ RNA was purified using the Poly(A) Quick mRNA isolation kit from Stratagene (La Jolla, CA). Synthesis of double-stranded cDNA and ligation of adaptors were performed using the Superscript Choice system from Gibco-BRL (Gaithersburg, MD). Prim-

ing was performed with a mixture of oligo-dT and random hexamers. One expression library was made in the λ gt11 phage vector and another in the λ Triplex phagemid vector from Clontech (Paolo Alto, CA). λ DNA was packaged with Gigapack III Gold Packaging extract (Stratagene). Southwestern screening was performed according to the method of Vinson *et al.* (1988) and Singh *et al.* (1989), and Clontech's instructions, after denaturation of filters with guanidine hydrochloride. Filters were preincubated for 30 min in incubation buffer (20 mM HEPES pH 7.9, 10% glycerol, 0.1% Nonidet P-40, 0.5 mM EDTA, 1 mM PMSF, 2 μ g/ml leupeptin, 2 μ g/ml pepstatin and 0.5 mM DTT) and further incubated for 2 h in new buffer, supplemented with 50 fmol/ml of ³²P-labeled 2HMG probe and 1 μ g/ml poly(dG-dC) or poly(dI-dC). Filters were washed four times for 1 min in incubation buffer and autoradiographed. New segments of L-Sox5 and Sox6 cDNAs were sequenced on both strands. A fragment of *Sox6* cDNA was amplified by PCR using primers encompassing two *HincII* restriction sites in the coding sequence and first-strand cDNA from mouse chondrocytes or adult testis. PCR products were electrophoresed in agarose gel, eluted and sequenced.

Electrophoretic mobility shift assay (EMSA)

EMSAs were performed as described previously (Lefebvre *et al.*, 1997) using a 1HMG or *Col2a1* 48 bp probe (Lefebvre *et al.*, 1997), or a 2HMG probe. The latter probe consisted of two tandem repeats of the 1HMG probe with *Bam*HI and *Bgl*II-cleaved sites at the 5' and 3' ends, respectively, and with the two repeats linked by the sequence GGATCT. Extracts from transfected fibroblasts were made in 100 mM potassium phosphate buffer, pH 7.2, containing 0.2% Triton X-100. Proteins were synthesized *in vitro* using the TNT T7 Quick Coupled Transcription/translation system (Promega, Madison, WI) and the expression plasmids described below.

Antibody preparation and Western blotting

Antisera were raised in rabbits (Genosys Biotechnologies, The Woodlands, TX) using keyhole limpet hemocyanin conjugated to a peptide homologous to the C-terminus of Sox9 (Bridgewater *et al.*, 1998) or corresponding to a sequence in the C-terminus of either Sox5 (PDVDYGSSENHIAG) or Sox6 (PKSDYSSSENAPEPV). Specific antibodies were purified as described previously (Bridgewater *et al.*, 1998). Purified antibodies (~0.1 mg/ml) were dialyzed against phosphate-buffered saline. Western blots were prepared as described previously (Lefebvre *et al.*, 1997) with Sox protein antibodies diluted 1:1000.

Probes for Northern blotting and in situ hybridization

A *NcoI*-*NruI* 552 bp fragment of mouse *L-Sox5* cDNA, located in the 5' end of the coding sequence, specifically hybridized with the long transcript of *Sox5*. A *NcoI* 448 bp fragment of *Sox5* cDNA, which included the HMG box, specifically hybridized with the short and long transcripts of *Sox5*. A *Bgl*II-*Hind*III 454 bp fragment of *Sox6* cDNA, located 1.3 kb downstream of the stop codon, specifically hybridized with the long transcript of *Sox6*. An *AccI*-*PstI* 478 bp fragment of *Sox6* cDNA, located in the 5' end of the coding sequence, specifically hybridized with the short and long transcripts of *Sox6*. The *aggrecan* probe was a 650 bp fragment of the mouse cDNA, encoding the C-terminal half of the G2 domain and most of the KS adjacent domain (Walcz *et al.*, 1994). Other probes were as described previously (Lefebvre *et al.*, 1995, 1997).

Northern blotting

Total RNA was isolated and analyzed in Northern blots as described previously (Lefebvre *et al.*, 1997). To facilitate transfer of large RNA species to nylon membrane, agarose gels were treated with 50 mM NaOH for 15 min before blotting. RNA standards were from Gibco-BRL. Hybridization signals were quantified on autoradiograms using the Intelligent Quantifier software program from Bio Image (Ann Arbor, MI).

In situ hybridization

Preparation of mouse embryo sections and hybridization with sense and antisense RNA probes labeled with [α -³⁵S]UTP were performed as described previously (Zhao *et al.*, 1997). Autoradiograms with collagen and Sox RNA probes were developed after 1 and 6–7 days, respectively. Sense probes showed no detectable signal over background.

Expression plasmids

Full-length mouse *L-Sox5* and *Sox6* coding sequences and deletions 151/679 and 213/679 were amplified by PCR and cloned into the pcDNA-5'UT and pcDNA-5'UT-FLAG mammalian expression plasmids, as described previously for human *SOX9* and mouse *Sox5* and *Sox4*

(Lefebvre *et al.*, 1997). The FLAG epitope did not affect DNA binding and transactivation properties of any Sox protein (data not shown). PCR products were verified by DNA sequencing. To obtain deletion 32/437, an *NcoI* fragment of L-Sox5 cDNA was blunted and cloned into blunt-ended *BamHI* and *EcoRI* sites of pcDNA-5'UT-FLAG; a translation termination codon was located in the *XbaI* site located downstream of the *EcoRI* site. Deletion 32/221–304/437 in L5 was obtained by cutting off a *PstI* fragment in deletion 32/437. Sox6B (previously called SoxLZ) cDNA was a gift from Tadayoshi Shiba and Shinya Yamashita (Takamatsu *et al.*, 1995). Sox6A and Sox6C coding sequences were reconstituted by replacing the *HincII* fragment of Sox6B cDNA with Sox6A- and Sox6C-specific fragments, which were obtained by PCR.

Synthesis of protein in vitro and cross-linking

Protein was synthesized *in vitro* with [³⁵S]methionine. For cross-linking, 2 µl of protein sample was preincubated for 30 min in 7 µl of DNA-binding buffer, with or without 5 fmol of 2HMG oligonucleotide, and incubated for 10 min with 1 µl of 0.1% glutaraldehyde. Polypeptide species were separated by SDS-PAGE in a polyacrylamide gradient (5–12%) gel and revealed by autoradiography.

Cell cultures

All cell types were cultured as described previously (Lefebvre *et al.*, 1997). Primary chondrocytes from ribs of newborn mice were used, unless otherwise indicated, after 2–3 days in culture, when they were essentially fully differentiated (Lefebvre *et al.*, 1994). RCS cells were previously shown to display a highly stable phenotype of early-stage chondrocytes (Mukhopadhyay *et al.*, 1995). MC615 cells were used after early passage when they exhibited a fairly differentiated chondrocytic phenotype (Mallein-Gerin *et al.*, 1993) or after repeated passages when they had essentially lost their chondrocytic phenotype. Human recombinant bone morphogenetic protein-2 (Genetics Institute, Cambridge, MA) was added to the culture medium at a concentration of 150 ng/ml.

Transient transfection

The p309 *Col2a1*–*βgeo* reporter construct has been described previously (Zhou *et al.*, 1995). The *βgeo* gene encodes a fusion protein with *E.coli* *β*-galactosidase and neomycin-resistance activities. Constructs p309–(4x48), p309–(2x231) and p309–(2x221) were obtained by cloning wild-type and mutant enhancer fragments (Lefebvre *et al.*, 1996) into p309 as described previously for other constructs (Zhou *et al.*, 1995). Reporter constructs were cotransfected with Sox protein expression plasmids using lipofectamine (Gibco-BRL) (Lefebvre *et al.*, 1997). A plasmid, pSV2 β gal or pGL2 (Promega), was included in all transfections as an internal control for transfection efficiency. Reporter activities were determined after normalization for transfection efficiency. Reporter and control plasmids were transfected in a 3:1 ratio, and expression plasmids were included in various amounts, as indicated. Empty expression plasmid was added, whenever necessary, to transfect the same total amount of DNA in all samples. To assess expression of endogenous genes, cells were transfected using either lipofectamine or FuGENE 6 (Boehringer Mannheim, Indianapolis, IN) according to the manufacturer's instructions.

DDBJ/EMBL/GenBank accession numbers

The accession Nos are: L-Sox5, AJ010604; Sox6, AJ010605.

Nomenclature

SOX refers to human proteins; Sox refers to mouse proteins. Genes are italicized.

Acknowledgements

This work was funded by NIH grants R01 AR42909 and P01 AR42919–02 to B.d.C. V.L. is the recipient of an Arthritis Investigator Award from the Arthritis Foundation. We thank James H.Kimura for RCS cells, Bjorn Olsen for MC615 cells, the Genetics Institute for BMP2, Tadayoshi Shiba and Shinya Yamashita for Sox-LZ cDNA, Kurt J.Doege for the mouse *aggrekan* probe, Gérard Karsenty for the MGP probe and William T.Butler for the osteopontin probe and ROS 17/2.8 cells. We are grateful to Sankar N.Maity for precious advice, Heidi Eberspaecher, Xin Zhou, Ni Lu and Laura C.Bridgewater for help in experimental work, and Shane Zhao for help in analysis of cDNA and protein sequences. DNA sequencing was performed by The University of Texas M.D. Anderson Cancer Center core sequencing facility, which is supported by NCI grant

CA16672. We thank William H.Klein, Richard R.Behringer, Randy L.Johnson and Sankar N.Maity for critical reviewing of the manuscript.

References

- Bell,D.M. *et al.* (1997) SOX9 directly regulates the type-II collagen gene. *Nature Genet.*, **16**, 174–178.
- Bridgewater,L.C., Lefebvre,V. and de Crombrughe,B. (1998) Chondrocyte-specific enhancer elements in the *Col11a2* gene resemble the *Col2a1* tissue-specific enhancer. *J. Biol. Chem.*, **273**, 14998–15006.
- Cancedda,R., Descalzi Cancedda,F. and Castagnola,P. (1995) Chondrocyte differentiation. *Int. Rev. Cytol.*, **159**, 265–358.
- Cheah,K.S.E., Lau,E.T., Au,P.K.C. and Tam,P.P.L. (1991) Expression of the mouse $\alpha 1$ (II) collagen gene is not restricted to cartilage during development. *Development*, **111**, 945–953.
- Connor,F., Cary,P.D., Read,C.M., Preston,N.S., Driscoll,P.C., Denny,P., Crane-Robinson,C. and Ashworth,A. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res.*, **22**, 3339–3346.
- Connor,F., Wright,E., Denny,P., Koopman,P. and Ashworth,A. (1995) The Sry-related HMG-box containing gene *Sox6* is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res.*, **23**, 3365–3372.
- Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992) An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J.*, **11**, 3705–3712.
- Erlebacher,A., Filvaroff,E.H., Gitelman,S.E. and Derynck,R. (1995) Toward a molecular understanding of skeletal development. *Cell*, **80**, 271–278.
- Foster,J.W. *et al.* (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, **372**, 525–530.
- Giese,K., Kingsley,C., Kirshner,J.R. and Grosschedl,R. (1995) Assembly and function of a TCR α enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein–protein interactions. *Genes Dev.*, **9**, 995–1008.
- Glumoff,V., Savontaus,M., Vehanen,J. and Vuorio,E. (1994) Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by Northern and *in situ* hybridization using species specific cDNA probes. *Biochim. Biophys. Acta*, **1219**, 613–622.
- Grosschedl,R., Giese,K. and Pagel,J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet.*, **10**, 94–100.
- Hall,B.K. and Miyake,T. (1995) Divide, accumulate, differentiate: cell condensation in skeletal development revisited. *Int. J. Dev. Biol.*, **39**, 881–893.
- Houston,C.S., Opitz,J.M., Spranger,J.W., Macpherson,R.I., Reed,M.H., Gilbert,E.F., Herrman,J. and Schinzel,A. (1983) The campomelic syndrome: review, report of 17 cases and follow-up on the currently 17-year old boy first reported by Maroteaux *et al.* in 1971. *Am. J. Med. Genet.*, **15**, 2–28.
- Kent,J., Wheatley,S.C., Andrews,J.E., Sinclair,A.H. and Koopman,P. (1996) A male-specific role for SOX9 in vertebrate sex determination. *Development*, **122**, 2813–2822.
- Kido,S., Hiraoka,Y., Ogawa,M., Sakai,Y., Yoshimura,Y. and Aiso,S. (1998) Cloning and characterization of mouse *mSox13* cDNA. *Gene*, **208**, 201–206.
- Komatsu,N., Hiraoka,Y., Shiozawa,M., Ogawa,M. and Aiso,S. (1996) Cloning and expression of *Xenopus laevis* *xSox12* cDNA. *Biochim. Biophys. Acta*, **1305**, 117–119.
- Kwok,C. *et al.* (1995) Mutations in SOX9, the gene responsible for campomelic dysplasia and autosomal sex reversal. *Am. J. Hum. Genet.*, **57**, 1028–1036.
- Laudet,V., Stehelin,D. and Clevers,H. (1993) Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res.*, **21**, 2493–2501.
- Lefebvre,V. and de Crombrughe,B. (1998) Toward understanding SOX9 function in chondrocyte differentiation. *Matrix Biol.*, **16**, 529–540.
- Lefebvre,V., Garofalo,S., Zhou,G., Metsäranta,M., Vuorio,E. and de Crombrughe,B. (1994) Characterization of primary cultures of chondrocytes from type II collagen/ β -galactosidase transgenic mice. *Matrix Biol.*, **14**, 329–335.
- Lefebvre,V., Garofalo,S. and de Crombrughe,B. (1995) Type X collagen gene expression in mouse chondrocytes immortalized by a temperature-sensitive simian virus 40 large tumor antigen. *J. Cell Biol.*, **128**, 239–245.

- Lefebvre, V. *et al.* (1996) An 18-base-pair sequence in the mouse pro α 1 (II) collagen gene is sufficient for cartilage expression and binds nuclear proteins that are selectively expressed in chondrocytes. *Mol. Cell. Biol.*, **16**, 4512–4523.
- Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N. and de Crombrughe, B. (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro α 1 (II) collagen gene. *Mol. Cell. Biol.*, **17**, 2336–2346.
- Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.*, **21**, 375–382.
- Mallein-Gerin, F. and Olsen, B.R. (1993) Expression of simian virus 40 large T (tumor) oncogene in mouse chondrocytes induces cell proliferation without loss of the differentiated phenotype. *Proc. Natl Acad. Sci. USA*, **90**, 3289–3293.
- Mansour, S., Hall, C.M., Pembrey, M.E. and Young, I.D. (1995) A clinical and genetic study of campomelic dysplasia. *J. Med. Genet.*, **32**, 415–420.
- Meyer, J. *et al.* (1997) Mutational analysis of the SOX9 gene in campomelic dysplasia and autosomal sex reversal: lack of genotype/phenotype correlations. *Hum. Mol. Genet.*, **6**, 91–98.
- Morais da Silva, S., Hacker, A., Harley, V., Goodfellow, P., Swain, A. and Lovell-Badge, R. (1996) Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature Genet.*, **13**, 62–68.
- Mukhopadhyay, K., Lefebvre, V., Zhou, G., Garofalo, S., Kimura, J.H. and de Crombrughe, 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- α 1 (II) collagen gene. *J. Biol. Chem.*, **270**, 27711–27719.
- Ng, L.-J. *et al.* (1997) SOX9 binds DNA, activates transcription and coexpressed with type II collagen during chondrogenesis in the mouse. *Dev. Biol.*, **183**, 108–121.
- Pevny, L.H. and Lovell-Badge, R. (1997) Sox genes find their feet. *Curr. Opin. Gen. Dev.*, **7**, 338–344.
- Reddi, A.H. (1998) Role of morphogenetic proteins in skeletal tissue engineering and regeneration. *Nature Biotech.*, **16**, 247–252.
- Singh, H., Clerc, R.G. and LeBowitz, J.H. (1989) Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. *BioTechniques*, **7**, 252–261.
- Southard-Smith, E.M., Kos, L. and Pavan, W.J. (1998) Sox10 mutation disrupts neural crest development in *Dom* Hirschsprung mouse model. *Nature Genet.*, **18**, 60–64.
- Südbeck, P., Schmitz, L., Baeuerle, P.A. and Scherer, G. (1996) Sex reversal by loss of the C-terminal transactivation domain of human SOX9. *Nature Genet.*, **13**, 230–232.
- Takamatsu, N., Kanda, H., Tsuchiya, I., Yamada, S., Ito, M., Kabeno, S., Shiba, T. and Yamashita, S. (1995) A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein. *Mol. Cell. Biol.*, **15**, 3759–3766.
- Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988) *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.*, **2**, 801–806.
- Wagner, T. *et al.* (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX-9. *Cell*, **79**, 1111–1120.
- Walcz, E., Deak, F., Erhardt, P., Coulter, S.N., Fülöp, C., Horvath, P., Doege, K.J. and Glant, T.T. (1994) Complete coding sequence, deduced primary structure, chromosomal localization and structural analysis of murine aggrecan. *Genome*, **22**, 364–371.
- Werner, M.H. and Burley, S.K. (1997) Architectural transcription factors: proteins that remodel DNA. *Cell*, **88**, 733–736.
- Wong, S.S. (1991) *Chemistry of Protein Conjugation and Cross-linking*. CRC Press, Boca Raton, FL, pp. 75–145.
- Wright, E.M., Snopek, B. and Koopman, P. (1993) Seven new members of the Sox gene family expressed during mouse development. *Nucleic Acids Res.*, **21**, 744.
- Wright E., Hargrave, M.R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A. and Koopman, P. (1995) The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genet.*, **9**, 15–20.
- Wunderle, V.M., Critcher, R., Ashworth, A. and Goodfellow, P.N. (1996) Cloning and characterization of SOX5, a new member of the human SOX gene family. *Genomics*, **36**, 354–358.
- Yamashita, A., Suzuki, S., Fujitani, K., Kojima, M., Kanda, H., Ito, M., Takamatsu, N., Yamashita, S. and Shiba, T. (1998) cDNA cloning of a novel rainbow trout SRY-type HMG box protein, rt23 and its functional analysis. *Gene*, **209**, 193–200.
- Zhao, Q., Eberspaecher, H., Lefebvre, V. and de Crombrughe, B. (1997) Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. *Dev. Dynam.*, **209**, 377–386.
- Zhou, G., Garofalo, S., Mukhopadhyay, K., Lefebvre, V., Smith, C.N., Eberspaecher, H. and de Crombrughe, B. (1995) A 182 bp fragment of the mouse pro α 1 (II) collagen gene is sufficient to direct chondrocyte expression in transgenic mice. *J. Cell Sci.*, **108**, 3677–3684.
- Zhou, G., Lefebvre, V., Zhang, Z., Eberspaecher, H. and de Crombrughe, B. (1998) Three HMG-like sequences within a 48-bp enhancer of the Col2a1 gene are required for cartilage-specific expression *in vivo*. *J. Biol. Chem.*, **273**, 14989–14997.

Received May 26, 1998; revised July 27, 1998;
accepted August 3, 1998