

Published in final edited form as:

*Int J Biochem Cell Biol.* 2010 March ; 42(3): 429–432. doi:10.1016/j.biocel.2009.07.016.

## The SoxD transcription factors – Sox5, Sox6, and Sox13 – are key cell fate modulators

Véronique Lefebvre

Department of Cell Biology and Orthopaedic Research Center, Cleveland Clinic Lerner Research Institute, 9500 Euclid Avenue (NC10), Cleveland, OH 44195

### Abstract

Sox5, Sox6, and Sox13 constitute the group D of sex-determining region (Sry)-related transcription factors. They are highly conserved in the family-specific high-mobility-group (HMG) box DNA-binding domain and in a group-specific coiled-coil domain. The latter mediates SoxD protein dimerization and thereby preferential binding to pairs of DNA recognition sites. The *SoxD* genes have overlapping expression and cell-autonomously control discrete lineages. *Sox5* and *Sox6* redundantly enhance chondrogenesis, but retard gliogenesis. *Sox5* hinders melanogenesis, promotes neural crest generation, and controls the pace of neurogenesis. *Sox6* promotes erythropoiesis, and *Sox13* modulates T cell specification and is an autoimmune antigen. SoxD proteins enhance transactivation by Sox9 in chondrocytes, but antagonize Sox9 and other SoxE proteins in oligodendrocytes and melanocytes, and also repress transcription through various mechanisms in several other lineages. While their biological and molecular functions remain incompletely understood, the SoxD proteins have thus already proven that they critically modulate cell fate in major lineages.

### Keywords

Sox5; Sox6; Sox13; cell fate; differentiation

### Introduction

The Sox family is comprised of 20 genes. These genes encode transcription factors with a high-mobility-group (HMG) box DNA-binding domain highly similar to that of the sex-determining region (Sry) protein. According to sequence identity inside and outside this domain, the *Sox* genes are classified into 8 groups, A to H. This review focuses on the *SoxD* group, which is composed of 3 genes - *Sox5*, *Sox6*, and *Sox13* - in most vertebrates, and 1 gene in the *D. melanogaster* fly (Sox102F) and other invertebrates. We summarize and evaluate current knowledge on the unique properties and biological roles of the vertebrate SoxD proteins. This review is part of a special journal issue on Sox transcription factors, and we therefore recommend readers to consult accompanying reviews for complementary information on general properties of Sox proteins and on specific properties of other Sox proteins and invertebrate SoxD proteins.

Corresponding author: Véronique Lefebvre; lefebvv@ccf.org; 9500 Euclid Avenue (NC10), Cleveland, OH 44195; telephone: (216) 445.0762; telefax: (216) 444.9404.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Structure and molecular function

Like most *Sox* genes, *Sox5*, *Sox6*, and *Sox13* were cloned in the 90's through searches for Sry-related genes (Denny et al., 1992; Connor et al., 1995; Takamatsu et al., 1995; Kido et al., 1998; Roose et al., 1998). Their gene and protein structures are highly identical to each other, but they are related to other *Sox* genes and proteins only in the HMG box. The human *SOX5* and *SOX6* genes are located in paralogous chromosomal regions on 12p12.1 and 11p15.3-15.2, respectively, and are more closely related to each other than to *SOX13*, located on 1q32. The 3 genes have 12-16 coding exons. They are spread across 300-400kb of genomic DNA in the case of *SOX5* and *SOX6*, but across 12kb only in the case of *SOX13*.

With molecular weights of 48 to 89kDa, the SoxD proteins are among the largest *Sox* proteins (Fig. 1). They harbor two highly conserved functional domains. The family-specific HMG box DNA-binding domain is located in the C-terminal half of the proteins. It is 87% identical among all human and mouse SoxD proteins, and less than 60% identical to that of other *Sox* proteins. The other domain is a group-specific coiled coil. It is located in the N-terminal half of the proteins, and is 76% identical among all human and mouse SoxD proteins. The proteins display only short stretches of identity outside these two domains. Both *Sox5* and *Sox6* are expressed as short transcripts (2 and 3kb, respectively) in adult testis and as long transcripts (6 and 8 kb, respectively) in other tissues. Both *Sox6* transcripts appear to encode the full-length protein, but the short *Sox5* transcript encodes a protein isoform that lacks the N-terminal half of the full-length protein. This protein was the first to be discovered, and was thus named Sox5. The full-length protein was originally named L-Sox5 or Sox5-L (Lefebvre et al., 1998; Hiraoka et al., 1998), but for simplicity most authors have referred to it as Sox5. We therefore propose that this latter appellation be consistently adopted in future references, not only for simplicity, but also because this long Sox5 isoform is structurally and functionally equivalent to Sox6 and Sox13.

The SoxD HMG box domain preferentially binds DNA sequences featuring an AACAAAT motif in electrophoretic mobility shift assay (EMSA) in vitro (Connor et al., 1994). The coiled-coil domain mediates homodimerization as well as heterodimerization of the SoxD proteins with each other (Lefebvre et al., 1998). It blocks binding of the full-length proteins to single recognition sites (Takamatsu et al., 1995; Roose et al., 1998), but strongly enhances binding to pairs of recognition sites (Lefebvre et al., 1998). The proteins efficiently bind in vitro and in vivo to sites harboring one or two mismatches in the preferred site. Moreover, they have little predilection for the relative orientation of the paired sites and for the length of the intervening sequence, from 0 to at least 19bp (Han and Lefebvre, 2008). They are thereby more flexible than Sox9 and other *Sox* proteins in choosing DNA sequences. Taking such flexibility into account, putative SoxD binding sites can be found in virtually any promoter or DNA regulatory region. It is therefore mandatory that solid experimentation be performed to ascertain such sites as actual targets of SoxD proteins in vivo.

The SoxD proteins have no known transactivation or transrepression domain, but they do participate in transcriptional activation and repression (Fig. 2). For instance, Sox5 and Sox6 synergize with Sox9 in activating many chondrocyte-specific extracellular matrix genes (Lefebvre et al., 1998; Han and Lefebvre, 2008). They bind to sites distinct from those of Sox9 on enhancers in these genes, and facilitate Sox9 DNA binding through an as yet unknown mechanism. In contrast, Sox5 and Sox6 interfere with the activation of myelin genes by Sox9 and Sox10 in oligodendrocytes and with the activation of the *Mitf* and *Dct* marker genes by Sox10 in melanocytes (Stolt et al., 2006 and 2008). They do so by competing with these proteins for binding to *Sox* recognition sites in the promoter of differentiation markers, and they thereby block transactivation. Sox6 also represses expression of embryonic globin genes in erythroid cells by binding to consensus sites in their proximal promoter (Yi et al., 2006). Sox5 and Sox6

have been shown in vitro to be able to induce transrepression by directly binding to the CtBP2 co-repressor and histone deacetylase Hdac1 (Murakami et al., 2001; Iguchi et al., 2007; Stolt et al., 2008). Sox6 and Sox13 have also been shown in vitro to be able to block canonical Wnt signaling in T cells and pancreatic beta-cells, respectively, but through different mechanisms (Iguchi et al., 2007; Melichar et al., 2007). While Sox6 can physically interact with beta-catenin, Sox13 can physically interact with the Tcf1 transcription factor. These data thus show that the SoxD proteins can act as either positive or negative modulators of transcription and might do so through various mechanisms. In vivo confirmation of some of these mechanisms and additional biochemical studies are needed to fully uncover the mechanisms underlying these activities. Additional studies are also needed to elucidate whether the differential activities of the SoxD proteins are dictated primarily by the nature of each specific target gene or by the molecular context characterizing each specific cell lineage.

## Expression and regulation

Like most *Sox* genes, each *SoxD* gene is expressed in a limited subset of cell types. Both *Sox5* and *Sox6* are highly expressed in spermatids, neurons, oligodendrocytes, and chondrocytes (Denny et al., 1992; Connor et al., 1995; Lefebvre et al., 1998; Stolt et al., 2006). *Sox5* and *Sox13* are co-expressed in pancreatic epithelial cells (Lioubinski et al., 2003). *Sox5* alone is expressed in melanoblasts (Stolt et al., 2008), *Sox6* in erythroid cells (Dumitriu et al., 2006; Yi et al., 2006) and skeletal myoblasts (Hagiwara et al., 2005), and *Sox13* in arterial walls, kidney, and liver (Roose et al., 1998). The mechanisms underlying these specific expression patterns of the *SoxD* genes are virtually unknown. *Sox5* and *Sox6* are co-expressed with *SoxE* genes in chondrocytes, oligodendrocytes, and melanocytes, and require these genes for expression, but it is not known whether the SoxE proteins directly activate the *SoxD* genes. Similarly, data regarding the possibility that the SoxD proteins are controlled at the translational and post-translational levels are still missing. The proteins are nuclear in all cells that expressed their RNA, ruling out all but little regulation at the translational and nuclear translocation levels. Important directions of future research are thus to uncover the mechanisms whereby the *SoxD* genes and their protein products are regulated.

## Biological function

Key biological functions for the *SoxD* genes have been revealed through gene inactivation in the mouse. The first function to be discovered was that of *Sox5* and *Sox6* in chondrogenesis (Smits et al., 2001). Inactivation of *Sox5* causes respiratory distress leading to death upon birth due to a cleft secondary palate and small thoracic cage, whereas inactivation of *Sox6* is only occasionally lethal at birth, and skeletal defects are limited to a short sternum. Double inactivation of the genes, in contrast, is lethal 3 days before birth, apparently due to circulatory failure. The embryos have chondrocytes, but the cells fail to overtly differentiate and proliferate. Skeletal growth and ossification are thus severely impaired. As mentioned earlier, a main function for *Sox5* and *Sox6* in chondrocytes is to boost the ability of *Sox9* to activate major chondrocyte markers. *Sox5* and *Sox6* also have redundant roles in oligodendrocytes, but their roles are very different from those in chondrocytes, as they repress specification and terminal differentiation, and influence migration patterns (Stolt et al., 2006). Their mechanism of action is also different, as they directly interfere with, rather than enhance, SoxE-mediated gene activation.

As expected from their expression pattern, each *SoxD* gene also has non-redundant functions. *Sox5* is dispensable for melanogenesis, but its loss partially rescues the strongly reduced melanoblast generation and marker gene expression occurring in *Sox10* heterozygous mice (Stolt et al., 2008). *Sox5* recruits CtBP2 and HDAC1 and binds to the regulatory regions of *Sox10* target genes, thereby directly inhibiting *Sox10* activity. *Sox5* also ensures proper

development of specific neuronal cell types by controlling the timing of critical cell fate and differentiation decisions (Kwan et al., 2008; Lai et al., 2008). It acts at least in part by directly binding and downregulating such genes as *Fezf2* and *Bcl11b*. Finally, *Sox5* overexpression studies in the chick embryo have suggested that *Sox5* may also promote generation of the neural crest (Perez-Alcala et al., 2004).

*Sox6* has key roles in definitive erythropoiesis. It directly contributes to repress embryonic globin genes (Yi et al., 2006). It also promotes cell survival and proliferation in synergy with erythropoietin signaling and acts beyond erythropoietin signaling to facilitate erythroid maturation (Dumitriu et al., 2006). *Sox6*-null mice have been shown to develop cardiac conduction problems soon after birth (Hagiwara et al., 2000). This phenotype is likely responsible for their failure to thrive and their death in the second or third week of age. *Sox6* facilitates cardiac and skeletal muscle differentiation, namely by ensuring proper switch from slow to fast skeletal muscle fibers in late fetuses (Hagiwara et al., 2005).

*Sox13*-null mice are born alive and show no overt defects, but they rapidly develop severe growth abnormalities (Melichar et al., 2007). The reason is yet unknown. The analysis of fetuses with *Sox13* gain-of-function and loss-of-function mutations has revealed that *Sox13* plays a critical role in the emergence of gammadelta T cells in the thymus, while opposing alphabeta T cell differentiation. It acts at least in part by inhibiting canonical Wnt signaling.

## Medical relevance

The essential roles that have already been identified for the mouse *SoxD* genes in multiple biological processes strongly suggest that their orthologues must have similar functions in humans. Gain-of-function or loss-of-function mutations in *SOXD* genes in humans could thus cause complex disease syndromes, as is the case for several other *SOX* genes. No such mutations, however, have yet been identified and linked to human congenital diseases. *SOX13*, nevertheless, was identified as the islet cell auto-antigen 12 that causes autoimmune diseases, including type I diabetes and primary biliary cirrhosis (Kasimiotis et al., 2000; Fida et al., 2002). The *SOXD* genes are expressed in glioma, prostate, and several other types of tumors (Ueda et al., 2004; Ma et al., 2009). Their contribution to cancer development or progression remains generally unknown, but it has been suggested that *SOX5* may enhance nasopharyngeal carcinoma progression by down-regulating *SPARC* gene expression (Huang et al., 2008).

## Conclusions

Our current understanding of the SoxD proteins is still in its infancy, but the impressive amount of data that has already been published by many groups is already sufficient to conclude that these structurally unique Sox proteins are biologically very important. They are highly flexible in selecting DNA binding sequences and are capable of using various mechanisms to either enhance or repress transcription. They thereby modulate such varied processes as cell proliferation, survival, differentiation, and terminal maturation in a number of cell lineages. It is our hope that this review will generate even greater enthusiasm in many groups to further uncovering the biological functions of these proteins, as well as their modes of action and regulation.

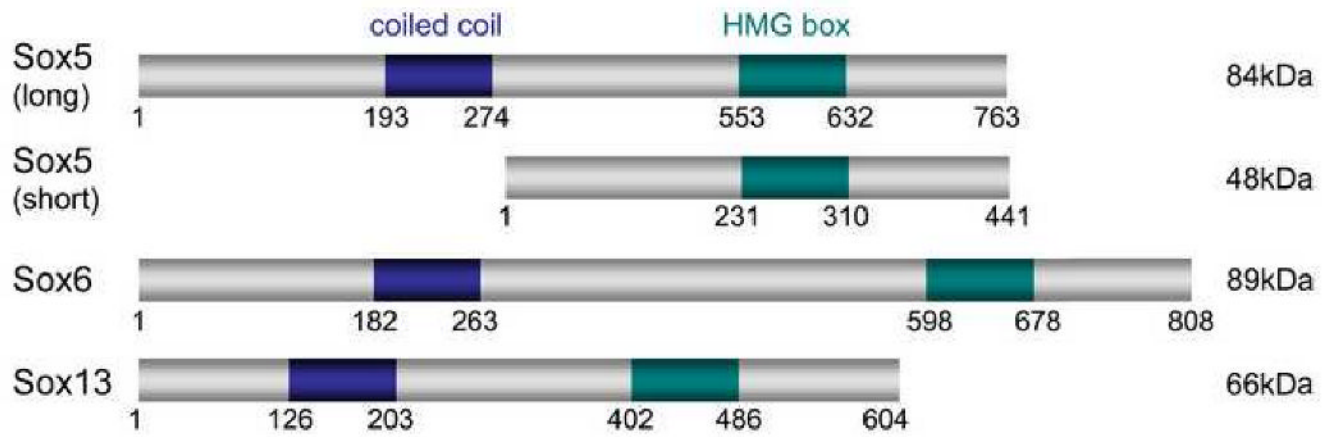
## Acknowledgments

Work on *SoxD* genes in the author's laboratory was supported by grants from NIH/NIAMS (R01 AR46249) and RoFAR. We thank Alfredo Penzo-Méndez, Bhattaram Pallavi, and Peter Dy for precious advice on the manuscript. We apologize to all authors whose important work on SoxD genes could not be cited due to space constraints.

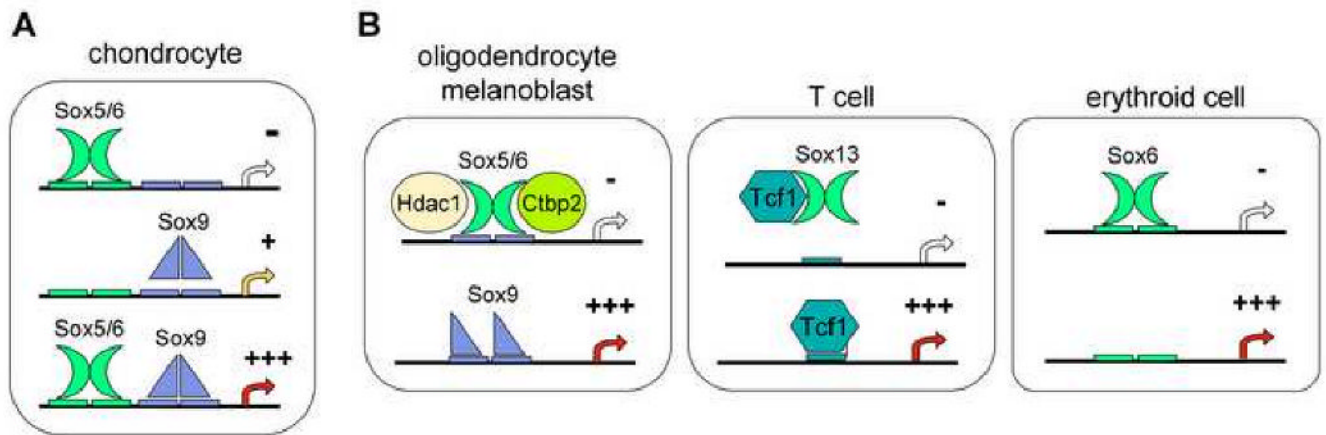
## References

- Connor F, Cary PD, Read CM, Preston NS, Driscoll PC, Denny P, Crane-Robinson C, Ashworth A. DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res* 1994;22:3339–46. [PubMed: 8078769]
- Connor F, Wright E, Denny P, Koopman P, Ashworth A. The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res* 1995;23:3365–72. [PubMed: 7567444]
- Denny P, Swift S, Connor F, Ashworth A. An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO J* 1992;11:3705–12. [PubMed: 1396566]
- Dumitriu B, Patrick MR, Petschek JP, Cherukuri S, Klingmuller U, Fox PL, Lefebvre V. Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood* 2006;108:1198–207. [PubMed: 16627753]
- Fida S, Myers MA, Whittingham S, Rowley MJ, Ozaki S, Mackay IR. Autoantibodies to the transcriptional factor SOX13 in primary biliary cirrhosis compared with other diseases. *J Autoimmun* 2002;19:251–7. [PubMed: 12473246]
- Hagiwara N, Klewer SE, Samson RA, Erickson DT, Lyon MF, Brilliant MH. Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death. *Proc Natl Acad Sci U S A* 2000;97:4180–5. [PubMed: 10760285]
- Hagiwara N, Ma B, Ly A. Slow and fast fiber isoform gene expression is systematically altered in skeletal muscle of the Sox6 mutant, p100H. *Dev Dyn* 2005;234:301–11. [PubMed: 16124007]
- Han Y, Lefebvre V. L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol Cell Biol* 2008;28:4999–5013. [PubMed: 18559420]
- Hiraoka Y, Ogawa M, Sakai Y, Kido S, Aiso S. The mouse Sox5 gene encodes a protein containing the leucine zipper and the Q box. *Biochim Biophys Acta* 1998;1399:40–6. [PubMed: 9714725]
- Huang DY, Lin YT, Jan PS, Hwang YC, Liang ST, Peng Y, Huang CY, Wu HC, Lin CT. Transcription factor SOX-5 enhances nasopharyngeal carcinoma progression by down-regulating SPARC gene expression. *J Pathol* 2008;214:445–55. [PubMed: 18085523]
- Iguchi H, Urashima Y, Inagaki Y, Ikeda Y, Okamura M, Tanaka T, Uchida A, Yamamoto TT, Kodama T, Sakai J. SOX6 suppresses cyclin D1 promoter activity by interacting with beta-catenin and histone deacetylase 1, and its down-regulation induces pancreatic beta-cell proliferation. *J Biol Chem* 2007;282:19052–61. [PubMed: 17412698]
- Kasimiotis H, Myers MA, Argentaro A, Mertin S, Fida S, Ferraro T, Olsson J, Rowley MJ, Harley VR. Sex-determining region Y-related protein SOX13 is a diabetes autoantigen expressed in pancreatic islets. *Diabetes* 2000;49:555–61. [PubMed: 10871192]
- Kido S, Hiraoka Y, Ogawa M, Sakai Y, Yoshimura Y, Aiso S. Cloning and characterization of mouse mSox13 cDNA. *Gene* 1998;208:201–6. [PubMed: 9524265]
- Kwan KY, Lam MM, Krsnik Z, Kawasaki YI, Lefebvre V, Sestan N. SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. *Proc Natl Acad Sci U S A* 2008;105:16021–6. [PubMed: 18840685]
- Lai T, Jabaudon D, Molyneaux BJ, Azim E, Arlotta P, Menezes JR, Macklis JD. SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron* 2008;57:232–47. [PubMed: 18215621]
- Lefebvre V, Li P, de Crombrughe B. A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J* 1998;17:5718–33. [PubMed: 9755172]
- Lioubinski O, Müller M, Wegner M, Sander M. Expression of Sox transcription factors in the developing mouse pancreas. *Dev Dyn* 2003;227:402–8. [PubMed: 12815626]
- Ma S, Chan YP, Woolcock B, Hu L, Wong KY, Ling MT, Bainbridge T, Webber D, Chan TH, Guan XY, Lam W, Vielkind J, Chan KW. DNA fingerprinting tags novel altered chromosomal regions and identifies the involvement of SOX5 in the progression of prostate cancer. *Int J Cancer* 2009;124:2323–32. [PubMed: 19173284]

- Melichar HJ, Narayan K, Der SD, Hiraoka Y, Gardiol N, Jeannot G, Held W, Chambers CA, Kang J. Regulation of gammadelta versus alphabeta T lymphocyte differentiation by the transcription factor SOX13. *Science* 2007;315:230–3. [PubMed: 17218525]
- Murakami A, Ishida S, Thurlow J, Revest JM, Dickson C. SOX6 binds CtBP2 to repress transcription from the Fgf-3 promoter. *Nucleic Acids Res* 2001;29:3347–55. [PubMed: 11504872]
- Perez-Alcala S, Nieto MA, Barbas JA. LSox5 regulates RhoB expression in the neural tube and promotes generation of the neural crest. *Development* 2004;131:4455–65. [PubMed: 15306568]
- Roose J, Korver W, Oving E, Wilson A, Wagenaar G, Markman M, Lamers W, Clevers. High expression of the HMG box factor sox-13 in arterial walls during embryonic development. *Nucleic Acids Res* 1998;26:469–76. [PubMed: 9421502]
- Smits P, Li P, Mandel J, Zhang Z, Deng JM, Behringer RR, de Crombrughe B, Lefebvre V. The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev Cell* 2001;1:277–90. [PubMed: 11702786]
- Stolt CC, Schlierf A, Lommes P, Hillgärtner S, Werner T, Kosian T, Sock E, Kessar N, Richardson WD, Lefebvre V, Wegner M. SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. *Dev Cell* 2006;11:697–709. [PubMed: 17084361]
- Stolt CC, Lommes P, Hillgärtner S, Wegner M. The transcription factor Sox5 modulates Sox10 function during melanocyte development. *Nucleic Acids Res* 2008;36:5427–40. [PubMed: 18703590]
- Takamatsu N, Kanda H, Tsuchiya I, Yamada S, Ito M, Kabeno S, Shiba T, Yamashita S. A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein. *Mol Cell Biol* 1995;15:3759–66. [PubMed: 7791783]
- Ueda R, Yoshida K, Kawakami Y, Kawase T, Toda M. Immunohistochemical analysis of SOX6 expression in human brain tumors. *Brain Tumor Pathol* 2004;21:117–20. [PubMed: 15696972]
- Yi Z, Cohen-Barak O, Hagiwara N, Kingsley PD, Fuchs DA, Erickson DT, Epner EM, Palis J, Brilliant MH. Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet* 2006;2:e14. [PubMed: 16462943]



**Figure 1.** Schematic of the human SoxD proteins. The coiled-coil dimerization domain and HMG box DNA-binding domain are shown, as well as the amino acid positions at the beginning and end of the proteins and internal domains. Predicted molecular weights are indicated on the right.

**Figure 2.**

Schematic illustrating various modes of action of the SoxD proteins. (A) Sox5 and Sox6 cooperate with Sox9 in transactivating chondrocyte-specific genes. They bind to recognition sites nearby, but distinct from those of Sox9. (B) SoxD proteins repress transcription through various mechanisms. They compete with SoxE proteins in binding to recognition sites on oligodendrocyte- and melanocyte-specific genes, and recruit co-repressors. Sox13 interacts with Tcf1 in T cells, thereby preventing it from binding to target genes. Sox6 blocks expression of embryonic globin genes in erythroid cells by binding to recognition sites in the proximal promoters.