

SURVEY AND SUMMARY

From head to toes: the multiple facets of Sox proteins

Michael Wegner*

Zentrum für Molekulare Neurobiologie, Universität Hamburg, Martinistraße 52, D-20246 Hamburg, Germany

Received September 18, 1998; Revised and Accepted November 6, 1998

ABSTRACT

Sox proteins belong to the HMG box superfamily of DNA-binding proteins and are found throughout the animal kingdom. They are involved in the regulation of such diverse developmental processes as germ layer formation, organ development and cell type specification. Hence, deletion or mutation of Sox proteins often results in developmental defects and congenital disease in humans. Sox proteins perform their function in a complex interplay with other transcription factors in a manner highly dependent on cell type and promoter context. They exhibit a remarkable crosstalk and functional redundancy among each other.

INTRODUCTION

Sox proteins entered the scientific stage less than a decade ago with the identification of Sry as the long elusive testis-determining factor Tdy/TDF, located on the Y chromosomes of mouse and man (1,2). Sry was found to contain a domain with similarity to the DNA-binding domains of the abundant non-histone chromosomal proteins HMG-1 and HMG-2. This so-called HMG domain is present in a large number of proteins which all belong to the HMG box superfamily. There are few amino acid positions within this 70–80 amino acid long domain which are conserved throughout the HMG box superfamily (3). Thus HMG domains can be highly diverse. However, in subgroups of the superfamily, strong conservation of HMG box sequences is observed.

Such a subgroup is the Sox protein family. Proteins are grouped into this family if they contain an HMG domain with strong amino acid similarity (usually >50%) to the HMG domain of Sry, which is also known as the Sry box. It is this box that gave the Sox protein family its name.

Few reviews have in the past addressed the Sox protein family (4,5). Here, I focus on recent advances in our understanding of Sox protein function.

DNA-BINDING AND THE HMG DOMAIN OF SOX PROTEINS

The HMG domain mediates DNA-binding of Sox proteins. Like other HMG box-containing proteins, Sox proteins have the ability to recognize specific DNA structures *in vitro* such as four-way

junction DNAs of variable sequence (6,7), (CA)_n repeats that can adopt such structures (8) or 1,2-d(GpG) cisplatin adducts (9). It is unclear whether this capacity is important for Sox protein function *in vivo*.

Sox proteins also bind to specific DNA sequences. Among HMG box proteins, this ability for sequence-specific DNA recognition is unique to Sox proteins and the distantly related TCF/LEF family (3). The consensus motif for Sox proteins has been defined as the heptameric sequence 5'-(A/T)(A/T)CAA(A/T)G-3' (10).

The structure of the Sry-type HMG domain has been solved both on and off DNA (11,12). Like other HMG domains, it consists of three α -helices (I–III) that are arranged in a twisted L-shape, with the antiparallel helices I and II forming the long and helix III and its associated N-terminal extension forming the short arm of the L-shape. The overall structure is maintained by a hydrophobic core. The amino acids that constitute this core are as highly conserved among Sox proteins as the ones which provide the base-specific DNA contacts. Whereas the overall conformation of the HMG domain remains unaltered upon DNA binding, a large conformational change is induced into target DNA such that its minor groove follows the concave binding surface of the HMG domain perfectly. As a consequence, DNA bound by Sry or other Sox proteins has an overall 70–85° bend, a widened minor groove and is helically unwound relative to classical B-DNA (6,13–16).

The characteristic minor groove binding activity of Sox proteins (17) is different from DNA binding of most other transcription factors which mainly target the major groove. Thus binding of Sox proteins to DNA in close proximity to other transcription factors is sterically feasible. This and the aforementioned ability to bend DNA has led to the hypothesis that Sox proteins might perform part of their function as architectural proteins by organizing local chromatin structure and assembling other DNA-bound transcription factors into biologically active, sterically defined multiprotein complexes (18,19).

Sox proteins might also recruit proteins into these complexes which by themselves do not bind to DNA, but are linked to signal transduction pathways. Such a mechanism is suggested by analogy to the related TCF/LEF proteins which bind β -catenin upon Wnt signaling and thereby translate these extracellular signals into changes of gene expression (for a review see 20,21). Involvement in the Wnt signaling pathway is, however, specific for TCF/LEF proteins, as the β -catenin interaction domain is not conserved between TCF/LEF and Sox proteins.

*Tel: +49 40 42803 6274; Fax: +49 40 42803 6602; Email: wegner@uke.uni-hamburg.de

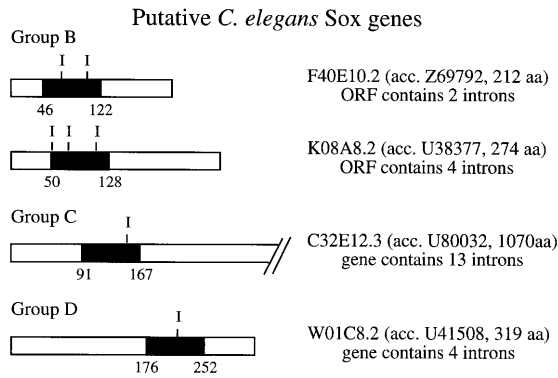


Figure 1. Sox proteins in *C.elegans*. Four putative translation products corresponding to Sox proteins were identified in genomic sequences from *C.elegans* using the BLAST search program. The location of the HMG domain within the respective protein is indicated by a black box. The position of introns within the HMG domain is indicated (I).

NMR analyses of the HMG domains of Sox proteins have also shown that domain structure as well as sequence-specific DNA binding and bending activities impose severe constraints, resulting in a very limited number of permissive amino acid choices at numerous positions. These constraints are largely responsible for the high degree of sequence conservation within the HMG domain of Sox proteins which in turn allowed the rapid isolation of partial sequences for many members of the Sox family by a degenerate PCR approach (2,22,23).

NOMENCLATURE OF SOX PROTEINS

Contrary to other transcription factor families, such as the POU family (24,25) the names of individual members of the Sox protein family conform to general rules. All Sox proteins except Sry carry the term Sox in their name, an additional suffix and sometimes a prefix. For human proteins, all three letters are capitalized, for other species only the first. The suffix usually consists of a number, less frequently of a letter. Numbers are assigned to Sox proteins consecutively in the order of their identification, with the count being now at 24. If a prefix is present in the name of a Sox protein, it generally indicates the species from which it was isolated. The use of such a prefix is not very common for Sox proteins identified in mammals, but is often used for other vertebrates. Most Sox proteins so far have been cloned from man, rodents, chicken, *Xenopus* and rainbow trout. Therefore, the most common prefixes used are 'x' for *Xenopus laevis*, 'c' for chicken (*Gallus gallus*) and 'rt' for rainbow trout (*Oncorhynchus mykiss*).

It should, however, be pointed out that Sox proteins have additionally been identified in a number of vertebrates and invertebrates, including various marsupials, reptiles, ascidians and *Drosophila melanogaster*. A survey of genomic sequences from the nematode *Caenorhabditis elegans* furthermore revealed the presence of at least four putative genes for Sox proteins (Fig. 1), although experimental evidence for their expression is presently lacking. This indicates that Sox proteins are widely distributed throughout the animal kingdom. I was, however, unable to detect a Sox family member among HMG domain-containing proteins encoded in the yeast genome.

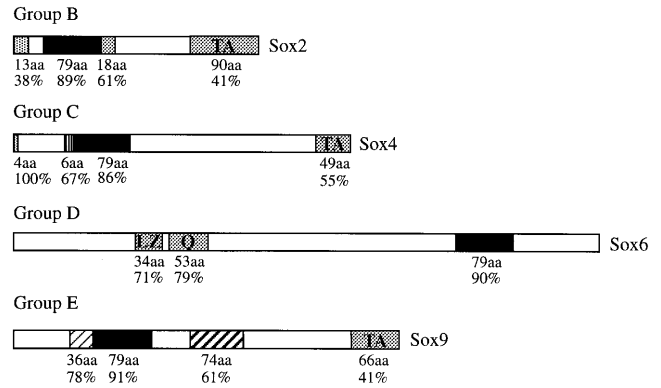


Figure 2. Conserved regions within groups of Sox proteins. In addition to the HMG domain (black box), further regions are conserved between several members of the same group of Sox proteins as indicated by differently shaded and striped boxes. One member of group B-E is shown. Sox2 was compared with Sox1 and Sox3 (99), Sox4 with Sox11, Sox22 and rtSox24 (26), Sox6 with L-Sox5, xSox12 and Sox13 (116), Sox10 with Sox9 and SoxP1 (44). The number of amino acids (aa) and the percentage of amino acid identity (%) between compared Sox proteins is listed below the conserved regions. LZ, leucine zipper; Q, glutamine-rich domain; S, serine-rich domain; TA, trans-activation domain

CLASSIFICATION OF SOX PROTEINS

All Sox proteins that have been identified so far (including the ones from *Drosophila* and *C.elegans*) can be placed into one of seven subgroups, A–G (Table 1). Members of the same subgroup usually share a >80% amino acid identity within the HMG domain; members of different classes fall below this mark. Additionally, vertebrate members of the same subgroup share regions of significant homology even outside the HMG domain (Fig. 2). In some subgroups, regions flanking the HMG domain are highly conserved. This is for instance the case in classes B, C and E. Whereas the conserved region within group B is located C-terminal to the HMG box and consists of a short amino acid sequence motif, conserved amino acids in classes C and E are N-terminal of the HMG domain and in the case of subgroup E span 36 amino acids. Other conserved regions in several subgroups comprise the transactivation domains, which in the case of classes C and E are located at the extreme C-terminus (17,26–31).

Members of group D contain a leucine zipper motif in addition to an HMG domain. The distance between HMG domain and leucine zipper is ~250 amino acids. It has been shown for Sox6/SoxLZ, rtSox23 and the long variant of Sox5 (L-Sox5) that the leucine zipper is capable of mediating homodimerization (32–34). For Sox6/SoxLZ and L-Sox5, heterodimerization has also been demonstrated (34). Unlike the leucine zipper of classic bZip transcription factors, the leucine zipper of group D Sox proteins is not adjacent to a basic region. Rather, the leucine zipper is always followed by a glutamine-rich region with which it forms a contiguous coiled-coil domain. As a consequence, dimerization does not lead to the creation of a new DNA-binding interface as is the case for bZip proteins. Rather, the influence of the leucine zipper on DNA binding seems to be indirect as leucine zipper-mediated dimerization strongly reduced the affinity of the dimer for single Sox consensus binding sites (32,33), while simultaneously increasing the ability to recognize two adjacent

Table 1. Classification of Sox proteins

Group	Factor	Species	Mapping	Size	Remarks	Accession numbers	Reference	
A	SRY	Hs Mm	Hs Yp11 Mm Y	204 aa 1.3 kb, 216aa	C-term. TA	L10101, X53772 etc. X60687 etc.	1, 61, 62 2, 68	
B	Sox1	Hs	Hs 13q34	3.9 kb, 387aa	intronless	Y13436	110	
		Mm	Mm 8	4.0 kb, 391aa		X94126	55	
		Gg	(D8Mit218)	2.5 kb, 373aa		AB011802	99	
	Sox2	Hs	Hs 3q26-27		2.4 kb, 319aa	intronless, weak	Z31560	111
		Mm			2.5 kb, 315aa	C-term. TA	X94125, U31967 D50603, U12532	55, 81 53, 100
		Gg			311 aa		AF022928, AF005476	91
		XI						
	Sox3 (SoxA)	Hs	Hs Xq26-27		2.3 kb, 443aa	intronless	X71135	112
		Mm			2.3 kb, 375aa		X94125	55
		Gg			2.1 kb, 316aa		U12467, AB011803	53, 99
XI		309aa			Y07542		113	
ZfSox19	Dr		319aa		X79821	50		
Sox21	Hs			278aa		AA040785	52	
	Gg					not deposited	52	
Sox70D (Dichaete, fishhook)	Dm	chrom III, 70D1-2		2.0 kb, 382aa		X96419 U68056	85, 86	
C	Sox4 (IRE-ABP)	Hs	Hs 6p23	5.2 kb, 474aa	intronless	X70683	17	
		Mm		5.0 kb, 440aa	C-term. TA	X70298	17	
	Sox11 (XLS13A, XLS13B)	Hs	Hs 2p25		3.5 kb, 441aa	intronless	U23752	95
		Mm			3.5 kb, 395aa	C-term. TA	AF009414	96
		Rn			3.5 kb, 395aa		AJ004858	27
		Gg			3.4 kb, 396aa		U12534, AB012237	53, 99
	Sox22	Hs	Hs 20p13		2.8 kb, 382aa	(2 <i>Xenopus</i> orthologs)	D83650, D86076	79, 115
2.8 kb, 375aa					D87209			
Sox24 ^a	Om			2.3 kb, 367aa	intronless C-term. TA	AB010741	26	
D	Sox5 & L-Sox5	Hs	Hs 12p12	1.8 kb, 392 aa	alternative splicing	X65663	38	
		Mm	Hs 8q21 (Ψ) Mm 6	2.0 kb, 392aa; 6.3 kb, 763 aa (L-Sox5)	leucine zipper in L-Sox5	S83308 X65657 AB006330, AJ010604	38 78 34, 116	
			Mm	Mm 7 (D7Mit40)	7.7 & 3.2 kb, 827 & 786aa	leucine zipper alternative splicing	U32614, D61689, AJ010605 D61688	33, 34, 77 33
	xSox12	XI		3.2 kb, 470aa	leucine zipper	D50552	47	
	Sox13	Mm		3.3 kb, 595aa	leucine zipper	AB006329, AJ000740	117, 118	
	Sox23	Om		4.5 kb, 585aa	leucine zipper	AB007906	32	
	E	SoxP1	Om		4.0 kb, 467aa		D83256	40
Sox9		Hs	Hs 17q24-25 Mm11 (D11Mit10)	4.3 kb, 509aa	C-term. TA	Z46629, S74504-6	35, 37	
		Mm		5.5 kb, 507aa		not deposited	43	
		Gg		4.1 kb, 494aa		U12533/AB012236	99	
		Om		4.5 kb, 488aa		AB006448	119	
Sox10 (SoxM, Sox21)		Hs	Hs 22q13 Mm15 (D15Mit71)	2.9 kb, 466aa	alternative splicing, weak	AJ001183	31, 36	
	Mm	3.1 kb, 466aa		C-term. TA	AF047043, AF017182 (U66141)	31, 45, 102		
	Rn		3.1 kb, 466aa		AJ001029	(51) 44		
F	xSox7	XI		2.0 kb, 362aa		D83649	41	
	Sox17	Mm	Mm1 (D1Mit1)	3.1 kb, 419aa	splice switch, C-term. TA	D49474/D49473	39	
	xSox17 α xSox17 β	XI			1.6 kb, 380aa		AJ001730	88
1.6 kb, 373aa						AJ001742		
Sox18	Mm	Mm 2 (D2Mit74) [Hs 20q]		1.6 kb, 378aa	central TA	L35032	42	
G	Sox15 ^b	Mm	Mm11	231aa		AB014474	-	
	Sox20	Hs	Hs 17p13	1.5 kb, 233aa		AB006867	48, 49	
	SoxD	XI		196aa		AB013896	90	

^aSox24 from rainbow trout is highly related to Sox11.

^bSox15 might be the mouse homologue of human Sox20. Species: Hs, man; Mm, mouse; Rn, rat; Gg, chicken; XI, *Xenopus*; Dr, zebrafish; Dm, *Drosophila*; Om, rainbow trout. Mapping: where possible, the name of the closest minisatellite marker in the mouse genome is mentioned. Ψ indicates the existence of a pseudogene. Size: length of transcripts is given in kilobases (kb), length of proteins in amino acids (aa). Remarks: TA, transactivation domain.

Sox binding sites (34). However, the leucine zipper does not only allow homodimerization or heterodimerization between group D Sox proteins, it also mediates interactions with other proteins (32). It is very tempting to speculate that not only the leucine zipper of group D Sox proteins, but also those regions found to be conserved in other classes of Sox proteins, exert their function by mediating interactions with other proteins.

Finally, members of the same group of Sox proteins tend to have similar genomic organizations in mammals. Mammalian genes for Sox proteins belonging to groups B and C are intronless, at least with respect to their coding region. As an aside, it deserves to be mentioned that this intronless character is not preserved in *C.elegans* Sox proteins belonging to these groups (Fig. 1). Open reading frames F40E10.2 and K08A8.2 for group B Sox proteins contain 2 and 4 introns, respectively, whereas 13 introns are found in the group C open reading frame C32E12.3.

In contrast to groups B and C, members of groups D–G possess exon–intron structure in mammals (Fig. 3). The exact genomic structure has only been determined for a handful of mammalian Sox proteins. Within the open reading frame of group E Sox genes, the position of introns is conserved (35–37; M.Wegner, unpublished). In group D gene Sox5 (38), group E genes Sox9 and Sox10 and group F gene Sox17 (39), the HMG domain is interrupted by an intron. Interestingly, even the *C.elegans* group D open reading frame W01C8.2 carries the intron within the HMG domain at exactly the same position as the mammalian group member Sox5, arguing for a conservation of intron positions within each subgroup. Furthermore, the group F gene Sox17 also carries its intron at the same position within the HMG domain as the group D gene Sox5. In group E genes, however, the intron has moved by four amino acids to the C-terminal side of the HMG domain. A look at the Sox open reading frames from *C.elegans* furthermore reveals that the ones belonging to groups B and C also contain introns within the HMG domain, sometimes up to three (see K08A8.2 in Fig. 1). These introns are found at positions other than the ones in the mammalian Sox proteins. Under the assumption that all Sox proteins are derived from a single ancestor, this finding indicates that at least some of the introns were not present in this common ancestor and were introduced later during the course of evolution. Sox proteins thus seem to be a good example in support of the ‘insertional theory of introns’.

Even those Sox proteins coded for by intron-containing genes are fairly small with sizes <20 kb. They are usually transcribed into messages of small to moderate sizes (Table 1). From the known chromosomal localizations of Sox genes it has to be concluded that they are interspersed throughout the mammalian genome and are not organized in clusters.

PARTIALLY CLONED SOX PROTEINS

As already mentioned, most Sox proteins were originally identified not as full-length proteins, but as PCR-derived partial sequences corresponding to their HMG domains. Because of the already mentioned high degree of sequence conservation within this part of the protein, PCR products sometimes differ from each other only by a limited number of bases. The deduced amino acid sequences differ even less.

Differences could indicate that PCR products are derived from highly related, but distinct, genes. Alternatively, they could be attributed to species-specific differences if different cDNA sources were used, or simply to PCR errors. Thus, it is generally

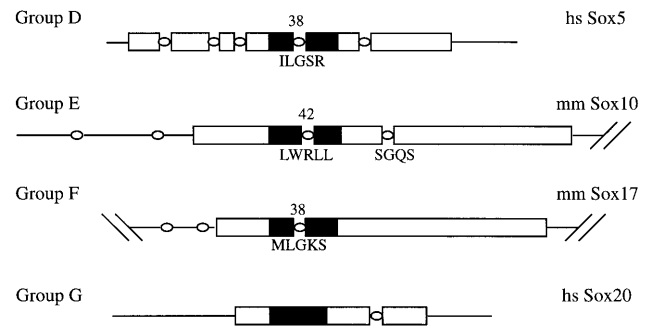


Figure 3. Gene structure for Sox proteins. Selected Sox genes are shown. Members of the Sox group E are known to share the same genomic organization in their open reading frame. The open reading frame is indicated as a box, the HMG domain is marked in black. Introns are depicted as ellipses. The exact position of introns within the HMG domain and the flanking amino acid sequences are indicated.

difficult to make statements about the nature of a particular Sox protein other than a preliminary classification, as long as it is only known from its HMG domain.

Some of the originally identified Sox proteins (2,22,23), which have not yet been determined outside the HMG domain and still await cloning, are: group B Sox14 (DDBJ/EMBL/GenBank accession no. Z18963), group C Sox12 and Sox19 (accession nos Z18961, U70442 and X98368), group E Sox8 (accession no. Z18957), group F Sox7 (accession no. X65660) and group G Sox16 (accession no. L29084). It should be noted, however, that proteins highly related to the group E factor Sox8 and to the group F factor Sox7 have already been identified in other species with SoxP1 from rainbow trout (40) and xSox7 from *Xenopus* (41). Whether these proteins are indeed true orthologs will only become clear after identification of full-length clones for the mammalian factor.

AMBIGUOUS NAMES

Another problem arising from the original cloning by PCR is that identification was so rapid that the same name was given to PCR fragments of different identity. This has been a source of confusion.

Sox8 was assigned originally to a group E PCR fragment from mouse (DDBJ/EMBL/GenBank accession no. Z18957) (23) and a group F fragment from man (accession no. X65664) (22). Given the fact that the mouse homolog of the group F fragment has already been fully cloned as Sox18 (42), I suggest that the term Sox8 should exclusively be used for the group E protein, that is still to be cloned.

Sox9 has been used to describe a group E PCR fragment from mouse (accession no. Z18958) (23) and a group B fragment from man (accession no. X65665) (22). The group E protein is fully cloned in several species and has been extensively characterized both with regards to its expression and function (35,37,43). The group B protein on the other hand is still only known as a PCR fragment. Therefore, I propose to use Sox9 for the well-characterized group E protein. Once fully cloned, the human group B protein could be renamed.

Similarly, both a group E fragment from mouse (accession no. Z18959) (23) and a group B fragment from man (accession no. X65666) (22) were originally designated as Sox10. The

group B fragment is highly similar, if not identical, to Sox3. The group E protein, on the other hand, has been recently identified in full by us and others and characterized as to its biological function (31,44,45). Thus, I suggest the term Sox10 to be used exclusively for the group E protein.

In the case of Sox12 the situation is even more complicated. The term has been used for a mouse group C (accession nos Z18961 and U70442) (23), a human group G (accession no. X73039) (46) and a *Xenopus* group D (accession no. D50552) (22) protein. The group D protein from *Xenopus* is the only one that has been fully cloned as xSox12 (47). The human group G fragment is identical to the recently cloned human Sox20 gene (48,49) and the mouse group C protein is still at large. Once cloned, it could either be given the name Sox12 and distinguished from xSox12 or, preferably, given a new name.

The same is true for Sox19 which was assigned to a group C PCR fragment from mouse (accession no. X98368) (49), but is also used to name the zebrafish group B Sox protein, zfSox19 (50). To avoid all complications, it would be most convenient to assign a new name to the group C factor once it has been cloned.

Finally, the term Sox21 has very recently been used to describe two fully cloned Sox proteins. One is identical to Sox10 (51), the other is a novel group B factor (52). Because one of the two is identical to Sox10, the term Sox21 should be exclusively used for the remaining group B factor.

SOX GENE FUNCTION

Between 15 and 20 different Sox proteins have already been identified in both mouse and man. If partially cloned Sox proteins and Sox proteins only known from other species are taken into account, it has to be assumed that the number of Sox proteins in any given vertebrate species will be >20. Given this high number, it cannot be surprising that most tissues and cell types express a Sox protein during at least one stage of their development (Table 2). In effect, a couple of studies have shown that a number of tissues (53–55) or cell types (27,34) express more than one Sox protein at certain times. Although it is conceivable that co-expressed Sox proteins perform different functions, especially when belonging to different subgroups and being distantly related, their recognition of similar, if not identical, DNA sequence motifs suggests that they might influence each other's activity or function redundantly. Functional redundancy is a recurring theme with Sox proteins. It also poses problems in the interpretation of gene deletion experiments or dominant-negative interference strategies. Nevertheless, these techniques combined with extensive analyses of developmental and adult expression patterns have helped in recent years to highlight some of the important biological roles of Sox proteins.

Sox proteins and sex determination

Sex determination is historically the classical domain of Sox protein function. Sry, the prototype of the Sox protein family, is a decisive factor for male sex determination in mammals located on the Y chromosome (1,2). Sry is expressed for a short period of time (10.5–12 days post-coitum in the mouse) in certain somatic cells of the genital ridge, triggering their differentiation into Sertoli cells and thereby initiating testis differentiation from the indifferent gonad (56–58). When expressed in mice as a transgene, mice develop as males even with an XX karyotype (59). When deleted from the Y chromosome, chromosomally male

Table 2. Summary of expression patterns and biological function of Sox proteins

Group	Factor	Expression pattern	Function
A	SRY	male-specific embryonic: genital ridge preceding differentiation adult: testis, hypothalamus, midbrain	testis determination <i>mutation in humans leads to XY sex reversal and gonadal dysgenesis</i>
B	Sox1	embryonic central nervous system (CNS), lens, urogenital ridge	lens development induction and maintenance of γ -crystallin gene expression <i>deletion in knockout mice leads to microphthalmia, cataracts, and spontaneous seizures</i>
	Sox2	inner cell mass, primitive ectoderm, developing CNS, lens	induction by chordin, repression by BMP-4, activation of FGF4, repression of osteopontin gene expression; induction of δ - and γ -crystallin gene expression modulation of responsiveness to neutralizing signals <i>deletion in knockout mice is embryonic lethal at implantation</i>
	Sox3	epiblast, embryonic CNS, urogenital ridge	candidate gene for human X-linked mental retardation syndromes, e.g. Borjeson-Forsman-Lehmann
	ZfSox19	transient; maternal, presumptive CNS in early gastrula, ventral region of embryonic diencephalon, midbrain and hindbrain; retina and lens	
	Sox21	primitive ectoderm, endoderm, developing gut, embryonic CNS: regional, longitudinal stripes in spinal cord	
	Sox70D (Dichaete, fishhook)	entire trunk of syncytial blastoderm, 7 irregular stripes in cellular blastoderm, ventral and cephalic neuroectoderm	<i>null mutation is lethal, exhibits segmentation defects, CNS defects regulation of pair-rule genes</i>
C	Sox4	embryonic heart (endocardial ridges and cushion) developing CNS, lung, tooth buds, mesonephros, gonads, thymus pre-B and T-cells	endocardial ridge development B-cell development <i>deletion in knockout mice leads to cardiac outflow tract malformation, lack of pro-B cell expansion, minor disturbances of thymocyte development</i>
	(IRE-ABP)		
	Sox11 (XLS13A, XLS13B)	widespread expression in developing CNS and peripheral nervous system (PNS), nascent somites, developing limbs, facial mesenchyme, somites, kidney, lung, specific postmitotic neuronal subpopulations	neural determination & differentiation events regulation of epithelio-mesenchymal interactions and inductive tissue remodelling
C	Sox22	throughout embryonic CNS and PNS; developing mesenchymal structures, fetal kidney, heart, pancreas, gonads	neural determination & differentiation events
	Sox24	oocytes	
D	Sox5	adult testis, highest in post-meiotic round spermatids	spermatogenesis
	L-Sox5	mesenchymal condensations, chondrocytes, embryonic CNS, notochord, otic vesicle	chondrogenesis
	Sox6 (SoxLZ)	rostrocaudal gradient in embryonic CNS, mesenchymal condensations, chondrocytes, notochord, otic vesicle	chondrogenesis
	xSox12	ovary	
	Sox13	embryonic: saccular component of inner ear, arterial walls, thymus adult: thymus, ovary, kidney	
Sox23	brain, ovary, weaker in heart, liver		
E	SoxP1	pituitary, gonads	
	Sox9	mesenchymal condensations chondrocytes, genital ridge and adult testis, notochord, otocysts, vibrissae, tubular heart structures, ventricular CNS cells	chondrogenesis, regulator of Col2a1 and aggrecan sex determination, Sertoli cell differentiation factor <i>mutation leads to campomelic dysplasia in humans</i>
	Sox10 (SoxM, Sox21)	neural crest, embryonic PNS and CNS, enteric nervous system, melanoblasts, Schwann cells, oligodendrocytes	determination of neural crest & glial cells, <i>mutation leads to megacolon, pigmentation defects and deafness in humans (Waardenburg-Hirschsprung syndrome) and mice (Dominant megacolon)</i>
F	xSox7	various adult tissues, e.g. gonads, kidney, lung, brain	
	Sox17	testis: full length form in premeiotic germ cells, truncated form in postmeiotic germ cells	
	xSox17 α,β	late blastula & gastrula, exclusively in presumptive endoderm	determination of endoderm; downstream of activin and Mixer; induces endoderm, IFABP, Xhlbox8
	Sox18	lung, cardiac and skeletal muscle	
G	Sox 20	fetal testis	
	SoxD	blastula and gastrula, prospective neuroectoderm, developing CNS	inducer of anterior neural tissues; enhanced expression by chordin, suppressed by BMP-4

mice adopt a female phenotype (2,60). Analogous observations have been made in human sex-reversed patients (see for example 61,62; for a full list of allelic variants see MIM entry 480000).

How Sry performs its function on the molecular level still remains a puzzle. Functional interaction with other proteins involved in sex determination such as Dax1 and SF1 is probably an important factor (63,64). Most of the SRY missense mutations identified in human patients localize to the HMG domain (65). Severe ones are usually *de novo*, those with moderate effects can be transmitted to male progeny. A significant number of these mutations interfere with the DNA-binding ability of the HMG domain or alter the bending characteristics (14). Other functions of the HMG domain might also be affected. Thus, it has been shown that the HMG domain of SRY and other Sox proteins contains the protein's nuclear localization signal (66,67). In analogy to other transcription factors, it might also be expected that the HMG domain engages in functionally important protein-protein interactions.

Outside the DNA-binding domain, Sry is highly divergent between species which is otherwise uncommon for Sox proteins. In fact, Sry has been evolving so rapidly that significant differences can even be detected between various mouse strains (68). A transactivation domain has been mapped to the C-terminus of mouse Sry, which is not conserved (either on a structural or functional level) in the human protein (69). Reversely, human SRY interacts with SIP-1, a novel PDZ domain protein via its C-terminal seven amino acids which are not conserved in the mouse protein (70). Human SRY has recently been shown to be phosphorylated by protein kinase A or a kinase of similar substrate specificity at a serine residue in the N-terminal part and as a consequence to be modulated in its DNA-binding activity (71). Again, this serine residue is absent from mouse Sry. This lack of conservation outside the HMG domain might indicate that all important functions of Sry are mediated by the HMG domain. Alternatively, sequences outside the HMG domain might have to perform different functions in different species and therefore might have been subject to directional selection with species-specific adaptive divergence.

Sry must be a relatively recent addition to the mechanism of sex determination as it does not seem to exist in vertebrates other than mammals. Given the fact that Sry most closely resembles the group B protein Sox3 which is localized on the X chromosome, it has been proposed that Sry evolved from Sox3 (72). Although the developing central nervous system is the major site of Sox3 expression in mammals (55), Sox 3 is also expressed in somatic cells of the early urogenital ridge, lending further support to a Sox3-derived origin of Sry. Unlike Sry, however, Sox3 is not only expressed in the male, but also in the female genital ridge.

Sry and its putative ancestor Sox3 are not the only Sox proteins present very early in the genital ridge. Another Sox protein is the group E protein Sox9. Like Sox3, Sox9 is not exclusively expressed during development in the genital ridge, but is also expressed in a number of other places, most prominently at sites of chondrogenesis (43; below).

Sox9 is initially expressed in both male and female genital ridges up until the onset of Sry expression. Following this event, Sox9 expression becomes restricted to the male gonad, where its expression follows differentiation of Sertoli cells so closely that it has been proposed to function as a critical Sertoli cell differentiation factor (73,74). Such a function is also supported by the fact that ~75% of karyotypically male patients with heterozygous

SOX9 mutations develop as intersexes or XY females (35,37). Thus, Sox9 is expressed in the same somatic cells that express Sry. However, whereas Sry expression is transient, Sox9 expression is maintained throughout further testis development. These and other observations have led to the assumption that, although originally turned on in the genital ridge in a Sry-independent manner, Sox9 must be under the control of Sry in mammals at the time of sex-determination. Contrary to Sry, Sox9 is an ancestral component of the vertebrate sex-determination pathway as is evident from its male-specific expression during the sex determination period in chicken genital ridges (73,74). In the chick embryonic gonad, Sox9 expression starts ~1 day after expression of the anti-Müllerian hormone (AMH) gene (75). Therefore, Sox9 does not seem to be essential for AMH gene induction. Nevertheless, Sox9 might be involved in the maintenance of AMH gene expression, as the AMH promoter contains Sox9 binding sites (75) and is activated in tissue culture by a combination of Sox9 and SF-1 (76).

Whereas Sry, Sox3 and Sox9 are all restricted to somatic cells of the gonad, there is a second set of Sox proteins expressed in germ cells. The group F protein Sox17, for instance, is found in pre-meiotic spermatogonia (39). In contrast, group D proteins Sox5 and probably Sox6 are both restricted in adult mice to post-meiotic germ cells, with highest levels in round spermatids (77,78). The occurrence of Sox proteins in the female germline is less well analyzed. Sox2 transcripts have been found in oocytes (55). Maternal transcripts were also detected for XLS13A and XLS13B, the *Xenopus* orthologues to mammalian group C protein Sox11 (79), rtSox24 (26) and the zebrafish group B protein zfSox19 (80).

Sox proteins and early embryogenesis

An important role for Sox proteins during the initial stages of ontogeny can so far only be inferred from maternal expression of some family members. Evidence exists, however, for an involvement of Sox proteins in the following early stages. The best studied case in mammals is the group B protein Sox2. This Sox protein is transiently expressed in the inner cell mass and the epiblast of mouse blastocysts, before it appears later again in the forming neuroepithelium (55,81). Mouse embryos homozygous for a targeted deletion of Sox2 die around implantation (82). Sox2 expression during early embryogenesis exhibits a pattern broadly overlapping with that of the POU transcription factor Oct-3/4 (83). Both factors have been found to cooperate in the transcriptional regulation of genes required during early embryogenesis, such as FGF4 and osteopontin (81,83,84). Early embryonic expression of the FGF4 gene is conferred by a strong enhancer in exon 3 with multiple transcription factor binding sites. One of these sites is a recognition element for POU proteins, one a binding site for Sox proteins. Both sites are separated by 3 bp. The function of this enhancer is critically dependent on the synergistic interaction between Sox-2 and Oct-3/4. Both factors have to be bound to their respective sites (81) and, additionally, engage in direct protein-protein interactions, thereby forming a ternary complex with exact stereospecific requirements (84). Because of these requirements, not any pair of Sox and POU proteins will function equally well together. Thus, it has been observed in tissue culture experiments that the group E protein Sox10 functions best with Tst-1/Oct6/SCIP and the group C protein Sox11 with Brn-1 or

Brn-2 (27,44). This argues for the existence of a specific partner code between members of both families.

Early embryonic expression of osteopontin is driven by an enhancer in the first intron of the gene. As in the FGF4 enhancer, binding sites for both Sox2 and Oct-3/4 are present, although the exact configuration of both sites is different (83). As a consequence, the effects of Oct-3/4 and Sox2 on osteopontin gene expression strongly differ from their effects on FGF4 transcription. Binding of Oct-3/4 to the enhancer is alone sufficient for activation, whereas binding of Sox2 to its site causes gene repression. Thus, the same set of transcription factors can be used simultaneously for gene activation and repression depending on the gene-specific context. The context-dependent, combinatorial function of Sox2 and Oct-3/4 is probably an important principle of transcription factor function during early development.

On a sequence level, Sox2 is closely related to the *Drosophila* Sox protein Sox70D/fish-hook/Dichaete with a 42% overall identity on the amino acid level (85,86). Sequence similarity is paralleled by a conservation of biochemical properties such that ectopic expression of the mouse Sox2 rescued phenotypic aspects of Dichaete mutations in transgenic flies (87). There are also resemblances between the expression pattern of Sox2 and Sox70D/fish-hook/Dichaete. Sox70D/fish-hook/Dichaete is expressed in the early *Drosophila* embryo, before taking on a second role later during neurogenesis (85,86). Expression initiates in the syncytial blastoderm in the form of a broad circumferential band corresponding to the entire trunk region. This band splits into two stripes and additional expression starts in the procephalic region. During cellularization of the blastoderm, Sox70D/fish-hook/Dichaete expression domains are further refined into seven irregular stripes. During and following gastrulation Sox70D/fish-hook/Dichaete is involved in maintaining the proper expression of several pair-rule and segment polarity genes, with the effect on *hairy*, *runt* and *even-skipped* expression being direct (15,86). Because of this role, functional inactivation of Sox70D/fish-hook/Dichaete leads to severe segmentation defects, including loss or fusion of abdominal denticle belts and to a defective organization of head structures. The observed segmentation defects are, however, extremely variable, indicating that Sox70D/fish-hook/Dichaete might not be absolutely required for pair-rule gene expression, but rather might have an accessory role, probably through alterations of chromatin structure and synergistic interactions with other transcription factors such as the POU proteins Pdm-1 and Pdm-2 in the case of *even-skipped* expression (15,86). Similarly, interaction with the POU protein Cf1a/drifter/ventral veinless has been postulated to explain Sox70D/fish-hook/Dichaete function during development of midline glia in later phases of development (87).

A third example of how Sox proteins influence early developmental processes is given by Xsox17 α and Xsox17 β , two highly related *Xenopus* proteins with sequence similarity to the group F mouse protein Sox17 (88). Expression of both proteins is first detected in the late blastula of *Xenopus*, is restricted to the presumptive endoderm and remains endoderm-specific throughout gastrulation and neurulation. Xsox17 β becomes undetectable during tailbud stages, whereas Xsox17 α continues to be expressed in endoderm-derived tissues. Both proteins have been found to act downstream from the endoderm-inducing activin signal. Their expression is dependent on the homeodomain Mixer during gastrulation, not however during the initial phases (89). Xsox17 α and Xsox17 β themselves seem to be involved in the regulation of

such genes as *endodermin*, *HNF1 β* , *cerberus* and *Xhlbox8*, as their ectopic expression in animal caps activates these endodermal markers. Importantly, injection of Xsox17 proteins or dominant negative fusions between them and the *engrailed* repressor domain into whole embryos lead to significant disruptions of endoderm development. These data provide evidence that Xsox17 α and Xsox17 β function as early regulators of normal endoderm development.

Sox proteins and neural development

At late *Xenopus* blastula stages, not only Xsox17 proteins start to be expressed, but also SoxD. Contrary to Xsox17 proteins, however, SoxD is found in the prospective ectoderm (90). During gastrulation, expression levels increase, but at the same time become restricted to the dorsally located, prospective neuroectoderm. Later on, SoxD is widely expressed throughout neural plate and tube. SoxD expression can be induced by chordin and is suppressed by BMP-4 signaling. The importance of SoxD during neural development became obvious from the observation that microinjected SoxD mRNA induces ectopic formation of neural tissue in the *Xenopus* embryo, whereas microinjection of a dominant-negative form of SoxD interferes with proper formation of anterior neural structures. Thus, SoxD is an essential mediator of major neuralization pathways in *Xenopus*.

In addition to SoxD, *Xenopus* Sox2 is also induced by chordin and is widely expressed in the prospective neuroectoderm at the beginning of gastrulation (91). Contrary to SoxD, however, Sox2 did not exhibit any neuralizing activity upon microinjection of its mRNA into *Xenopus* embryos. Rather, Sox2 made the ectoderm responsive to extracellular signals such as bFGF. In agreement with this, experiments in the chick also showed that there is no strict correlation between Sox2 expression and the competence of a region for neural induction (92), thus assigning Sox2 a role in neural development downstream from SoxD.

Sox2 is widely expressed in the early neural plate and early neural tube of several species including *Xenopus*, chicken and mouse and thus is one of the earliest pan-neural markers (53,55,91,93). It is assumed that Sox2 strengthens the neural cell fate at this period and might help cells to acquire further specification. At later stages, Sox2 expression in the developing central nervous system becomes restricted to the neuroepithelial cells of the ventricular layer, which still divide and exhibit an immature phenotype. Cells that leave the ventricular layer lose Sox2 expression. Thus, Sox2 and markers of terminally differentiating cells are expressed in a mutually exclusive manner. In agreement with this, Sox2 expression is no longer detectable in the fully differentiated central nervous system of the adult.

As already mentioned, early neural development marks the second major period of Sox2 expression after the preceding presence in the preimplantation embryo. Whereas occurrence in the preimplantation embryo is unique among Sox proteins, Sox2 expression in the developing nervous system is highly similar to and strongly overlapping with the structurally related group B proteins Sox1 and Sox3 (55,82,93).

In the case of Sox1, it was possible to recapitulate this expression pattern in tissue culture (82). Upon retinoic acid induction of aggregated P19 cells to neuroectodermal derivatives, endogenous Sox1 is transiently induced and becomes down-regulated before markers for the differentiated neuronal or glial phenotype are turned on. Importantly, induction of ectopic Sox1

expression in aggregated P19 cells was able to substitute for retinoic acid, showing that Sox1 must be an early response to neural inducing signals.

From the overlapping expression pattern of group B Sox proteins, it has been concluded that these proteins might perform similar functions in the developing nervous system and might be functionally redundant. Such an assumption is supported by the recent targeted deletion of Sox1 in mice, which exhibit a surprisingly mild central nervous system phenotype characterized only by spontaneous seizures (94).

Sox1, Sox2 and Sox3 are, however, not the only examples of possible functional redundancy between Sox proteins. Group C proteins Sox4, Sox11 and Sox22 also show overlapping expression in the developing central and peripheral nervous systems (17,27,95–98).

Among the group C proteins, Sox11 has been best analyzed for its expression in the nervous system. In the central nervous system, it is co-expressed in the neuroepithelium with Sox1, Sox2 and Sox3 on a low level and transiently up-regulated in cells that leave the neuroepithelium. As these cells simultaneously down-regulate Sox2 and Sox3, a model has been postulated in which neural differentiation is characterized by an ordered switch from one group of Sox proteins to the next (53). Another interesting aspect of Sox11 concerns its expression in a subset of differentiating brain regions such as the cortical plate and the inferior colliculus during late embryonic development. This might indicate that Sox11 is not only important for the development of early neural precursors, but might also be involved in the differentiation of distinct neuronal subpopulations (27). Outside the nervous system, Sox11 seems to be primarily expressed at places of epithelio-mesenchymal interactions including somites, branchial arches, developing face and limbs (96).

Sox proteins and lens development

Sox1, Sox2 and Sox3 are also co-expressed during lens development in chicken. After initial induction of Sox2 and Sox3 in the prospective lens area of the surface ectoderm, Sox1 expression follows with a slight delay in the lens placode upon its invagination shortly after the onset of δ -crystallin gene expression (99). All three Sox proteins stimulate δ -crystallin gene expression. In the case of the δ 1-crystallin gene, activation is mediated by direct binding of Sox proteins to the DC5 enhancer which is localized in the third intron (100).

In the mouse, Sox expression during lens development exhibits several marked differences as compared with chicken. Sox3, for instance, is not expressed at all in the developing lens and Sox2 becomes strongly down-regulated the moment Sox1 is turned on around 12.5 days post-coitum. From this time onwards, Sox1 is by far the predominant Sox protein in the developing lens (99). As a consequence, deletion of both Sox1 alleles in mice leads to a failure of lens fibre cells to elongate, ultimately resulting in microphthalmia and cataract (94). Most of the γ -crystallin genes, which like the δ -crystallin genes of birds are under the control of Sox proteins, are not activated in Sox1-deficient mice. Those that are, probably get activated by Sox2, but are turned off at a later point, indicating that maintenance of their expression also requires Sox1 function.

Sox proteins and the neural crest

Another Sox protein with strong expression in the nervous system is the group E protein Sox10 (31,44,45,101). Sox10 is first very broadly expressed in cells of the neural crest at the time of their emergence. Sox10 expression continues in neural crest cells that contribute to the forming peripheral nervous system and can be detected in the sensory, sympathetic and enteric ganglia as well as along nerves in a manner typical for cells of the Schwann cell lineage (44). Additional Sox10 expression has been detected in melanoblasts (45). Whereas expression in melanoblasts and in the enteric nervous system is transient, expression in all other structures of the peripheral nervous system continues into adulthood and seems to be confined at later times to glial cells (44). Differing from the peripheral nervous system, Sox10 expression in the central nervous system starts late (around day 13 post-coitum in the mouse), but then increases in strength, reaching maximal levels during adulthood. In the adult central nervous system, Sox10 expression seems to be largely confined to glia of the oligodendrocyte type (44).

In the spontaneous mouse mutant *Dominant megacolon (Dom)*, Sox10 carries a frameshift mutation (45,102). This mutation leaves the HMG domain intact, but replaces the protein's normal C-terminal half with a structurally unrelated sequence. The resulting protein is functionally inactive (102). As a result of the mutation, *Dom* mice suffer from several neural crest defects which in the homozygous animal are characterized by a loss of neurons and glia in the peripheral nervous system and a complete lack of the enteric nervous system (45,102). In a large percentage of heterozygous animals, ganglia are absent from a variable part of the distal colon leading to the formation of an aganglionic megacolon, which in combination with pigmentation defects of the skin are the characteristic features of the *Dom* mouse. The phenotypic manifestation of the Sox10 mutation in the heterozygotes is variable and points to haploinsufficiency or to a dominant-negative mode of action.

Mutations in one SOX10 allele have also been found in patients who suffer from congenital aganglionic megacolon (Hirschsprung disease) associated with a combination of pigmentation defects and deafness (Waardenburg syndrome) (30,36; for a full list of allelic variants see MIM 602229). This so called Waardenburg–Hirschsprung or Shah–Waardenburg syndrome closely resembles the phenotype of the *Dom* mouse both with respect to the extensive neural crest dysfunction and the phenotypic variability. One of the reasons for the observed phenotypic variability might be the role of Sox10 as a modulator of other transcription factors (44). While only possessing a relatively weak transactivation potential, Sox10 efficiently cooperates with co-expressed transcription factors such as Pax3 and Tst-1/Oct6/SCIP (30,44). Thus, it is conceivable that the Pax3 or Tst-1/Oct6/SCIP allelic variant present in a particular mouse strain or individual strongly influences the phenotypic manifestation of a given Sox10 mutation. Additionally, SOX10 mutations from patients exhibit differing degrees of functional inactivation in tissue culture experiments (30). Taken together, these results clearly prove a role for Sox10 at an early stage in neural crest development.

Sox proteins and chondrogenesis

Closely related to Sox10 is the group E protein Sox9. As already mentioned Sox9 is expressed very early in the genital ridge and is

likely to be involved in sex determination and testis development as a crucial Sertoli cell differentiation factor (73,74). Thus, mutation of one SOX9 allele in humans leads to autosomal sex reversal in the majority of affected chromosomally male individuals (35,37). Phenotypic manifestations are as variable as observed for SOX10.

Whereas sex reversal only affects karyotypically male patients with SOX9 mutations, all carriers are affected by severe skeletal malformations, including congenital bowing and angulation of the long bones, but also hypoplastic scapulae, dislocated hips and deformed pelvis. This autosomal dominant syndrome is known as campomelic dysplasia and affects ~1 in 20 000 births (for a detailed description see MIM entry 114290). Mutations found in patients affected by campomelic dysplasia often constitute missense mutation within the HMG domain or nonsense mutations. These either abrogate DNA binding or remove the protein's C-terminal transactivation domain (29,35,37,103). Alternatively, the SOX9 locus is affected by translocations with breakpoints being localized 50–600 kb upstream of the transcriptional start site (35,37,104). Transgene experiments prove for at least some of these translocations that they alter the expression rate of SOX9 (105).

In addition to the various skeletal defects, campomelic dysplasia patients sometimes exhibit absence of the olfactory bulbs, heart and renal malformations, defects of the tracheo-pulmonary system, deafness or mental retardation. This variety of symptoms in humans reflects the expression pattern of Sox9 in mouse development where it is found in brain, otic vesicle, urogenital system, lung and heart (28,43). Most importantly, Sox9 is predominantly expressed in the mesenchymal condensations from which the skeleton develops. There it is expressed before and during cartilage deposition, first in the prechondrogenic precursors and later on in the proliferating and maturing chondrocytes. From this expression pattern and the phenotype of campomelic dysplasia patients it can be concluded that Sox9 plays an important role in chondrocyte development.

One aspect of its function has recently been revealed by the observation that Sox9 is largely co-expressed with collagen II (*col2a1*), the major extracellular matrix component of cartilage, during development of skeletogenic and a number of other tissues in the mouse (28,34). Additionally, Sox9 has been shown to bind to sites within the chondrocyte-specific enhancer present in intron 1 of the *col2a1* gene and to activate this enhancer in tissue culture experiments (16,28,106). Mutation of these binding sites severely disturbs the chondrocyte-specific activity of this enhancer *in vitro* and *in vivo*. Additionally, expression of a *Sox9* transgene under the control of the *Hoxb2* promoter induces ectopic expression of both the endogenous *col2a1* gene and a *col2a1-lacZ* transgene, indicating that *col2a1* is probably a direct target of Sox9 (106). However, Sox9 activity alone is not sufficient to account for the full expression pattern of *col2a1* and there is evidence that the *col2a1* enhancer is targeted by other proteins, including L-Sox5 and Sox6 (34,107). These two Sox proteins are co-expressed with Sox9 at all chondrogenic sites of the mouse embryo and form with Sox9 a multiprotein complex on the *col2a1* enhancer (34). As a result the *col2a1* gene is activated by the combination of all three Sox proteins to a much higher extent than by Sox9 alone. A similar cooperative activation by L-Sox5, Sox6 and Sox9 has also been shown for a second chondrocyte marker, the *aggrecan* gene.

Sox proteins and haemopoiesis

Interaction with other proteins and functional redundancy also seem to be important determinants of Sox4 function. Sox4 is very widely expressed during embryogenesis in brain, gonads, lung, heart and thymus. It is a very prominent transcription factor in lymphocytes of the B- and T-cell lineages (17). Targeted deletion in the mouse led to embryonal lethality at day 14 post-coitum due to circulatory failure (97). This circulatory failure is caused by an impaired development of the Sox4-expressing endocardial ridges. As a consequence, the semilunar valves and the outlet portion of the muscular ventricular septum fail to develop. Using fetal liver cells from these knockout animals, it was also shown that there is a block in early B-cell development at the pro-B-cell stage, similar to that observed for Pax5 (108). This block was, however, not absolute as a small number of B-cells matured. Such leakiness is rather expected for proteins with accessory than with autonomous function.

Although T-cells also express Sox4, there was only a mild effect of Sox4 deletion on T-cell development. Only upon close inspection was it possible to detect slightly lower expansion and maturation rates which prevent cells of Sox4-deficient animals from efficiently competing in competitive intrathymic reconstitution assays with wild-type thymocytes (109). Although being highly expressed in both the T- and B-cell lineages, Sox4 is thus essential for B-cell development, but only of minor importance for T-cell development.

CONCLUDING REMARKS

Research on Sox proteins has made rapid advances over the last few years, with more and more family members being identified and shown to be intricately involved in regulation of development. The increasing number of cases in which mutations in Sox proteins are associated with human diseases further highlights the importance of this group of transcription factors. More will be learnt about the functions of Sox proteins from mouse models as many family members are currently subject to targeted deletion strategies. However, it will be equally important to understand their mode of action on a molecular level. Whether they function as transcriptional activators, modulators or architectural components will only become fully apparent with the identification of more target genes, interaction partners and the signals by which Sox proteins are activated. The coming years should prove to be exciting.

ACKNOWLEDGEMENTS

I thank Elisabeth Sock and Irm Hermans-Borgmeyer for critical reading of the manuscript. Research in my laboratory on Sox proteins is supported by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie and by grants from the Deutsche Forschungsgemeinschaft (We1326/5-2 and We1326/7-1).

REFERENCES

- 1 Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischauf,A.M., Lovell-Badge,R. and Goodfellow,P.N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240–244.
- 2 Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Münsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, **346**, 245–250.
- 3 Laudet,V., Stehelin,D. and Clevers,H. (1993) Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res.*, **21**, 2493–2501.
- 4 Pevny,L.H. and Lovell-Badge,R. (1997) Sox genes find their feet. *Curr. Opin. Genet. Dev.*, **7**, 338–344.
- 5 Prior,H.M. and Walter,M.A. (1996) Sox genes: architects of development. *Mol. Med.*, **2**, 405–412.
- 6 Ferrari,S., Harley,V., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.*, **11**, 4497–4506.
- 7 Peters,R., King,C.-Y., Ukiyama,E., Falsafi,S., Donahue,P.K. and Weiss,M.A. (1995) An SRY mutation causing human sex reversal resolves a general mechanism of structure-specific DNA recognition: application to the four-way junction. *Biochemistry*, **34**, 4569–4576.
- 8 Vriz,S., Griffiths,B.L., Harley,V., Goodfellow,P. and Lovell-Badge,R. (1995) The SRY protein, like HMG1 recognizes (CA)_n sequences, an abundant repeat sequence in vertebrates. *Biochem. Mol. Biol. Int.*, **37**, 1137.
- 9 Trimmer,E.E., Zamble,D.B., Lippard,S.J. and Essigmann,J.M. (1998) Human testis-determining factor SRY binds to the major DNA adduct of cisplatin and a putative target sequence with comparable affinities. *Biochemistry*, **37**, 352–362.
- 10 Harley,V.R., Lovell-Badge,R. and Goodfellow,P.N. (1994) Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.*, **22**, 1500–1501.
- 11 van Houte,L.P.A., Chuprina,V.P., van der Wetering,M., Boelens,R., Kaptein,R. and Clevers,H. (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J. Biol. Chem.*, **270**, 30516–30524.
- 12 Werner,M.H., Huth,J.R., Gronenborn,A.M. and Clore,G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY–DNA complex. *Cell*, **81**, 705–714.
- 13 Connor,F., Cary,P.D., Read,C.M., Preston,N.S., Driscoll,P.C., Denny,P., Crane,R.C. and Ashworth,A. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nucleic Acids Res.*, **22**, 3339–46.
- 14 Pontiggia,A., Rimini,R., Harley,V.R., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1994) Sex-reversing mutations affect the architecture of SRY–DNA complexes. *EMBO J.*, **13**, 6115–6124.
- 15 Ma,Y., Niemitz,E.L., Nambu,P.A., Shan,X., Sackerson,C., Fujioka,M., Goto,T. and Nambu,J.R. (1998) Gene regulatory functions of *Drosophila* Fish-hook, a high mobility group domain Sox protein. *Mech. Dev.*, **73**, 169–182.
- 16 Lefebvre,V., Huang,W., Harley,V.R., Goodfellow,P.N. and DeCrombrugge,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.
- 17 van de Wetering,M., Oosterwegel,M., van Norren,K. and Clevers,H. (1993) Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.*, **12**, 3847–3854.
- 18 Werner,M.H. and Burley,S.K. (1997) Architectural transcription factors: proteins that remodel DNA. *Cell*, **88**, 733–736.
- 19 Wolffe,A.P. (1994) Architectural transcription factors. *Science*, **264**, 1100–1101.
- 20 Clevers,H. and van de Wetering,M. (1997) TCF/LEF factors earn their wings. *Trends Genet.*, **13**, 485–489.
- 21 Bienz,M. (1998) TCF: transcriptional activator or repressor? *Curr. Opin. Cell Biol.*, **10**, 366–372.
- 22 Denny,P., Swift,S., Brand,N., Dabhade,N., Barton,P. and Ashworth,A. (1992) A conserved family of genes related to the testis determining gene, SRY. *Nucleic Acids Res.*, **20**, 2887.
- 23 Wright,E.M., Snopak,B. and Koopman,P. (1993) Seven new members of the SOX gene family expressed during mouse development. *Nucleic Acids Res.*, **21**, 744.
- 24 Ryan,A.K. and Rosenfeld,M.G. (1997) POU domain family values: flexibility, partnerships and developmental codes. *Genes Dev.*, **11**, 1207–1225.
- 25 Wegner,M., Drolet,D.W. and Rosenfeld,M.G. (1993) POU-domain proteins: structure and function of developmental regulators. *Curr. Opin. Cell Biol.*, **5**, 488–498.
- 26 Kanda,H., Kojima,M., Miyamoto,N., Ito,M., Takamatsu,N., Yamashita,S. and Shiba,T. (1998) Rainbow trout Sox24, a novel member of the Sox family, is a transcriptional regulator during oogenesis. *Gene*, **211**, 251–257.
- 27 Kuhlbrodt,K., Herbarth,B., Sock,E., Enderich,J., Hermans-Borgmeyer,I. and Wegner,M. (1998) Cooperative function of POU proteins and Sox proteins in glial cells. *J. Biol. Chem.*, **273**, 16050–16057.
- 28 Ng,L.-J., Wheatley,S., Muscat,G.E.O., Conway-Campbell,J., Bowles,J., Wright,E., Bell,D.M., Tam,P.P.L., Cheah,K.S.E. and Koopman,P. (1997) SOX9 binds DNA, activates transcription and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev. Biol.*, **183**, 108–121.
- 29 Südbeck,P., Schmitz,M.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.
- 30 Kuhlbrodt,K., Schmidt,C., Sock,E., Pingault,V., Bondurand,N., Goossens,M. and Wegner,M. (1998) Functional analysis of Sox10 mutations found in human Waardenburg–Hirschsprung patients. *J. Biol. Chem.*, **273**, 23033–23038.
- 31 Pusch,C., Hustert,E., Pfeifer,D., Südbeck,P., Kist,R., Roe,B., Wang,Z., Balling,R., Blin,N. and Scherer,G. (1998) The SOX10/Sox10 gene from human and mouse: sequence, expression and transactivation by the encoded HMG domain transcription factor. *Hum. Genet.*, **103**, 115–123.
- 32 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, rtSox23 and its functional analysis. *Gene*, **209**, 193–200.
- 33 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.
- 34 Lefebvre,V., Li,P. and de Crombrugge,B. (1998) 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.*, **17**, 5718–5733.
- 35 Foster,J.W., Dominguez-Steglich,M.A., Guioli,S., Kwok,C., Weller,P.A., Stevanovic,M., Weissenbach,J., Mansour,S., Young,I.D., Goodfellow,P.N., Brook,J.D. and Schafer,A.J. (1995) A gene that is related to SRY and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, **372**, 525–530.
- 36 Pingault,V., Bondurand,N., Kuhlbrodt,K., Goerich,D.E., Prehu,M.-O., Puliti,A., Herbarth,B., Hermans-Borgmeyer,I., Legius,E., Matthijs,G., Amiel,J., Lyonnet,S., Ceccherini,I., Romeo,G., Smith,J.C., Read,A.P., Wegner,M. and Goossens,M. (1998) Sox10 mutations in patients with Waardenburg–Hirschsprung disease. *Nature Genet.*, **18**, 171–173.
- 37 Wagner,T., Wirth,J., Meyer,J., Zabel,B., Held,M., Zimmer,J., Pasantes,J., Bricarelli,F.D., Keutel,J., Hustert,E., Wolf,U., Tommerup,N., Schempp,W. and Scherer,G. (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene Sox9. *Cell*, **79**, 1111–1120.
- 38 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.
- 39 Kanai,Y., Kanaizuma,M., Noce,T., Saido,T.C., Shiroishi,T., Hayashi,Y. and Yazaki,K. (1996) Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region and their differential expression in mouse spermatogenesis. *J. Cell. Biol.*, **133**, 667–681.
- 40 Ito,M., Ishikawa,M., Suzuki,S., Takamatsu,N. and Shiba,T. (1995) A rainbow trout SRY-type gene expressed in pituitary glands. *FEBS Lett.*, **377**, 37–40.
- 41 Shiozawa,M., Hiraoka,Y., Komatsu,N., Ogawa,M., Sakai,Y. and Aiso,S. (1996) Cloning and characterization of *Xenopus laevis* xSox7 cDNA. *Biochim. Biophys. Acta*, **1309**, 73–76.
- 42 Dunn,T.L., Mynett,J.L., Wright,E.M., Hosking,B.M., Koopman,P.A. and Muscat,G.E. (1995) Sequence and expression of Sox-18 encoding a new HMG-box transcription factor. *Gene*, **161**, 223–225.
- 43 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.
- 44 Kuhlbrodt,K., Herbarth,B., Sock,E., Hermans-Borgmeyer,I. and Wegner,M. (1998) Sox10, a novel transcriptional modulator in glial cells. *J. Neurosci.*, **18**, 237–250.
- 45 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.

- 46 Goze,C., Poulat,F. and Berta,P. (1993) Partial cloning of SOX-11 and SOX-12, two new human SOX genes. *Nucleic Acids Res.*, **21**, 2943.
- 47 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.
- 48 Meyer,J., Wirth,J., Held,M., Schempp,W. and Scherer,G. (1996) SOX20, a new member of the SOX gene family, is located on chromosome 17p13. *Cytogenet. Cell Genet.*, **72**, 246–249.
- 49 Hiraoka,Y., Ogawa,M., Sakai,Y., Taniguchi,K., Fujii,T., Umezawa,A., Hata,J.-I. and Aiso,S. (1998) Isolation and expression of a human SRY-related cDNA, hSOX20. *Biochim. Biophys. Acta*, **1396**, 132–137.
- 50 Vriza,S. and Lovell-Badge,R. (1995) The zebrafish Zf-Sox 19 protein: a novel member of the Sox family which reveals highly conserved motifs outside of the DNA-binding domain. *Gene*, **153**, 275–276.
- 51 Tani,M., ShindoOkada,N., Hashimoto,Y., Shiroishi,T., Takenoshita,S., Nagamachi,Y. and Yokota,J. (1997) Isolation of a novel Sry-related gene that is expressed in high-metastatic K-1735 murine melanoma cells. *Genomics*, **39**, 30–37.
- 52 Rex,M., Uwanogho,D.A., Orme,A., Scotting,P.J. and Sharpe,P.T. (1997) cSox21 exhibits a complex and dynamic pattern of transcription during embryonic development of the chick central nervous system. *Mech. Dev.*, **66**, 39–53.
- 53 Uwanogho,D., Rex,M., Cartwright,E.J., Pearl,G., Healy,C., Scotting,P.J. and Sharpe,P.T. (1995) Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. *Mech. Dev.*, **49**, 23–36.
- 54 Stock,D.W., Buchanan,A.V., Zhao,Z.Y. and Weiss,K.M. (1996) Numerous members of the Sox family of HMG box-containing genes are expressed in developing mouse teeth. *Genomics*, **37**, 234–237.
- 55 Collignon,J., Sockanathan,S., Hacker,A., Cohentannoudji,M., Norris,D., Rastan,S., Stevanovic,M., Goodfellow,P.N. and Lovell-Badge,R. (1996) A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development*, **122**, 509–520.
- 56 Hacker,A., Capel,B., Goodfellow,P. and Lovell-Badge,R. (1995) Expression of Sry, the mouse sex determining gene. *Development*, **121**, 1603–1614.
- 57 Jeske,Y.W., Bowles,J., Greenfield,A. and Koopman,P. (1995) Expression of a linear Sry transcript in the mouse genital ridge. *Nature Genet.*, **10**, 480–482.
- 58 Koopman,P., Münsterberg,A., Chapel,B., Vivian,N. and Lovell-Badge,R. (1990) Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature*, **348**, 450–452.
- 59 Koopman,P., Gubbay,J., Vivian,N., Goodfellow,P. and Lovell-Badge,P. (1991) Male development of chromosomally female mice transgenic for Sry. *Nature*, **351**, 117–121.
- 60 Lovell-Badge,R. and Robertson,R. (1990) XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. *Development*, **109**, 635–646.
- 61 Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448–450.
- 62 Jäger,R.J., Anvret,M., Hall,K. and Scherer,G. (1990) A human XY female with a frameshift mutation in the candidate testis-determining gene SRY. *Nature*, **348**, 452–454.
- 63 Luo,X., Ikeda,Y. and Parker,K.L. (1994) A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*, **77**, 481–490.
- 64 Swain,A., Narvaez,V., Burgoyne,P., Camerino,G. and Lovell-Badge,R. (1998) Dax1 antagonizes Sry action in mammalian sex determination. *Nature*, **391**, 761–767.
- 65 Hawkins,J.R. (1993) Mutational analysis of Sry in XY females. *Hum. Mutat.*, **2**, 347–350.
- 66 Poulat,F., Girard,F., Chevrone,M.-P., Goze,C., Rebillard,X., Calas,B., Lamb,N. and Berta,P. (1995) Nuclear localization of the testis determining gene product SRY. *J. Cell Biol.*, **128**, 737–748.
- 67 Südbek,P. and Scherer,G. (1997) Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and Sox9. *J. Biol. Chem.*, **272**, 27848–27852.
- 68 Kunieda,T. and Toyoda,Y. (1992) Nucleotide sequence of mouse Sry gene is different between Y chromosome originating from *Mus musculus musculus* and *Mus musculus domesticus*. *Genomics*, **13**, 236–237.
- 69 Dubin,R.A. and Ostrer,H. (1994) Sry is a transcriptional activator. *Mol. Endocrinol.*, **8**, 1182–1192.
- 70 Poulat,F., de Santa Barbara,P., Desclozeaux,M., Soullier,S., Moniot,B., Bonneaud,N., Boizet,B. and Berta,P. (1997) The human testis determining factor SRY binds a nuclear factor containing PDZ protein interaction domains. *J. Biol. Chem.*, **272**, 7167–7172.
- 71 Desclozeaux,M., Poulat,F., de Santa Barbara,P., Capony,J.-P., Turowski,P., Jay,P., Mejean,C., Moniot,B., Boizet,B. and Berta,P. (1998) Phosphorylation of an N-terminal motif enhances DNA-binding activity of the human SRY protein. *J. Biol. Chem.*, **273**, 7988–7995.
- 72 Foster,J.W. and Graves,J.A. (1994) An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc. Natl Acad. Sci. USA*, **91**, 1927–1931.
- 73 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.
- 74 daSilva,S.M., 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.*, **14**, 62–68.
- 75 Oreal,E., Pieau,M.-G., Mattei,M.-G., Josso,N., Picard,J.-Y., Carre-Eusebe,D. and Magre,S. (1998) Early expression of AMH in chicken embryonic gonads precedes testicular Sox9 expression. *Dev. Dyn.*, **212**, 522–532.
- 76 De Santa Barbara,P., Bonneaud,N., Boizet,B., Desclozeaux,M., Moniot,B., Südbek,P., Scherer,G., Poulat,F. and Berta,P. (1998) Direct interaction of SRY-related protein SOX9 and Steroidogenic Factor 1 regulates transcription of the human Anti-Müllerian Hormone gene. *Mol. Cell Biol.*, **18**, 6653–6665.
- 77 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.
- 78 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.
- 79 Hiraoka,Y., Komatsu,N., Sakai,Y., Ogawa,M., Shiozawa,M. and Aiso,M. (1997) XLS13A and XLS13B: SRY-related genes of *Xenopus laevis*. *Gene*, **197**, 65–71.
- 80 Vriza,S., Joly,C., Boulekbache,H. and Condamine,H. (1996) Zygotic expression of the zebrafish Sox-19, an HMG box-containing gene, suggests an involvement in central nervous system development. *Mol. Brain Res.*, **40**, 221–228.
- 81 Yuan,H.B., Corbi,N., Basilico,C. and Dailey,L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.*, **9**, 2635–2645.
- 82 Pevny,L.H., Sockanathan,S., Placzek,M. and Lovell-Badge,R. (1998) A role for SOX1 in neural determination. *Development*, **125**, 1967–1978.
- 83 Botquin,V., Hess,H., Fuhrmann,G., Anastassiadis,C., Gross,M.K., Friend,G. and Schöler,H.R. (1998) New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev.*, **12**, 2073–2090.
- 84 Ambrosetti,D.C., Basilico,C. and Dailey,L. (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein–protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol. Cell Biol.*, **17**, 6321–6329.
- 85 Nambu,P.A. and Nambu,J.R. (1996) The *Drosophila* fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development*, **122**, 3467–3475.
- 86 Russell,S.R.H., Sanchez Soriano,N., Wright,C.R. and Ashburner,M. (1996) The Dichaete gene of *Drosophila melanogaster* encodes a SOX-domain protein required for embryonic segmentation. *Development*, **122**, 3669–3676.
- 87 Sanchez Soriano,N. and Russell,S. (1998) The *Drosophila* Sox-domain protein Dichaete is required for the development of the central nervous system. *Development*, **125**, 3989–3996.
- 88 Hudson,C., Clements,D., Friday,R.V., Stott,D. and Woodland,H.R. (1997) Xsox17 α and - β mediate endoderm formation in *Xenopus*. *Cell*, **91**, 397–405.
- 89 Henry,G.L. and Melton,D.A. (1998) Mixer, a homeobox gene required for endoderm development. *Science*, **281**, 91–96.
- 90 Mizuseki,K., Kishi,M., Shiota,K., Nakanishi,S. and Sasai,Y. (1998) SoxD: an essential mediator of induction of anterior neural tissues in *Xenopus* embryos. *Neuron*, **21**, 77–85.

- 91 Mizuseki, K., Kishi, M., Matsui, M., Nakanishi, S. and Sasai, Y. (1998) *Xenopus* Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. *Development*, **125**, 579–587.
- 92 Streit, A., Sockanathan, S., Perez, L., Rex, M., Scotting, P.J., Sharpe, P.T., Lovell-Badge, R. and Stern, C.D. (1997) Preventing the loss of competence for neural induction: HGF/SF, L5 and Sox-2. *Development*, **124**, 1191–1202.
- 93 Rex, M., Orme, A., Uwanogho, D., Tointon, K., Wigmore, P.M., Sharpe, P.T. and Scotting, P.J. (1997) Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Dev. Dyn.*, **209**, 323–332.
- 94 Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. and Episkopou, V. (1998) Sox1 directly regulates the γ -crystallin genes and is essential for lens development in mice. *Genes Dev.*, **12**, 776–781.
- 95 Jay, P., Goze, C., Marsollier, C., Taviaux, S., Hardelin, J.-P., Koopman, P. and Berta, P. (1995) The human Sox11 gene: cloning, chromosomal assignment and tissue expression. *Genomics*, **29**, 541–545.
- 96 Hargrave, M., Wright, E., Kun, J., Emery, J., Cooper, L. and Koopman, P. (1997) Expression of the Sox11 gene in mouse embryos suggests roles in neuronal maturation and epithelio-mesenchymal induction. *Dev. Dyn.*, **210**, 79–86.
- 97 Schilham, M.W., Oosterwegel, M.A., Moerer, P., Ya, J., Deboer, P.A.J., Vandewetering, M., Verbeek, S., Lamers, W.H., Kruisbeek, A.M., Cumano, A. and Clevers, H. (1996) Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature*, **380**, 711–714.
- 98 Jay, P., Sahly, I., Goze, C., Taviaux, S., Poulat, F., Couly, G., Abitbol, M. and Berta, P. (1997) Sox22 is a new member of the Sox gene family, mainly expressed in human nervous tissue. *Hum. Mol. Genet.*, **6**, 1069–1077.
- 99 Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R. and Kondoh, H. (1998) Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development*, **125**, 2521–2532.
- 100 Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R. and Kondoh, H. (1995) Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO J.*, **14**, 3510–3519.
- 101 Bondurand, N., Kobetz, A., Pingault, V., Lemort, N., Encha-Razavi, F., Couly, G., Goerich, D.E., Wegner, M., Abitbol, M. and Goossens, M. (1998) Expression of the SOX10 gene during human development. *FEBS Lett.*, **432**, 168–172.
- 102 Herbarth, B., Pingault, V., Bondurand, N., Kuhlbrodt, K., Hermans-Borgmeyer, I., Puliti, A., Lemort, N., Goossens, M. and Wegner, M. (1998) Mutation of the Sry-related Sox10 gene in *Dominant megacolon*, a mouse model for human Hirschsprung disease. *Proc. Natl Acad. Sci. USA*, **95**, 5161–5165.
- 103 Meyer, J., Südbeck, P., Held, M., Wagner, T., Schmitz, M.L., Bricarelli, B.D., Eggermont, E., Friedrich, U., Haas, O.A., Kobelt, A., Leroy, J.G., Van Maldergem, L., Michel, E., Mitulla, B., Pfeiffer, R.A., Schinzel, A., Schmidt, H. and Scherer, G. (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.
- 104 Wirth, J., Wagner, T., Meyer, J., Pfeiffer, R.A., Tietze, H.-U., Schempp, W. and Scherer, G. (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. *Hum. Genet.*, **97**, 186–193.
- 105 Wunderle, V.M., Critcher, R., Hastie, N., Goodfellow, P.N. and Schedl, A. (1998) Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc. Natl Acad. Sci. USA*, **95**, 10646–10654.
- 106 Bell, D.M., Leung, K.K.H., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P.L. and Cheah, K.S.E. (1997) Sox9 directly regulates the type-II collagen gene. *Nature Genet.*, **16**, 174–178.
- 107 Zhou, G., Lefebvre, V., Zhang, Z., Eberspaecher, H. and de Crombrughe, B. (1998) Three high-mobility group-like sequences within a 48-base pair enhancer of the Col2a1 gene are required for cartilage-specific expression *in vivo*. *J. Biol. Chem.*, **273**, 14989–14997.
- 108 Urbanek, P., Wang, Z.-Q., Fetka, I., Wagner, E.F. and Busslinger, M. (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell*, **79**, 901–912.
- 109 Schilham, M.W., Moerer, P., Cumano, A. and Clevers, H.C. (1997) Sox-4 facilitates thymocyte differentiation. *Eur. J. Immunol.*, **27**, 1292–1295.
- 110 Malas, S., Duthie, S.M., Mohri, F., Lovell-Badge, R. and Episkopou, V. (1997) Cloning and mapping of the human Sox1: a highly conserved gene expressed in the developing brain. *Mamm. Genome*, **8**, 866–868.
- 111 Stevanovic, M., Zuffardi, O., Collignon, J., Lovell-Badge, R. and Goodfellow, P. (1994) The cDNA sequence and chromosomal location of the human SOX2 gene. *Mamm. Genome*, **5**, 640–642.
- 112 Stevanovic, M., Lovell, B.R., Collignon, J. and Goodfellow, P.N. (1993) SOX3 is an X-linked gene related to SRY. *Hum. Mol. Genet.*, **2**, 2013–2018.
- 113 Penzel, R., Oswald, R., Chen, Y., Tacke, L. and Grunz, H. (1997) Characterization and early embryonic expression of a neural specific transcription factor xSox3 in *Xenopus laevis*. *Int. J. Dev. Biol.*, **41**, 667–677.
- 114 Nasrin, N., Buggs, C., Kong, X.F., Carnazza, J., Goebel, M. and Alexander, B.M. (1991) DNA-binding properties of the product of the testis-determining gene and a related protein. *Nature*, **354**, 317–320.
- 115 Miyata, S., Miyashita, K. and Hosoyama, Y. (1996) Sry-related genes in *Xenopus* oocytes. *Biochim. Biophys. Acta*, **1308**, 23–27.
- 116 Hiraoka, Y., Ogawa, M., Sakai, Y., Kido, S. and Aiso, S. (1998) The mouse Sox5 gene encodes a protein containing the leucine zipper and the Q box. *Biochim. Biophys. Acta*, **1399**, 40–46.
- 117 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.
- 118 Roose, J., Korver, W., Oving, E., Wilson, A., Wagenaar, G., Markman, M., Lamers, W. and Clevers, H. (1998) High expression of the HMG box factor Sox-13 in arterial walls during embryonic development. *Nucleic Acids Res.*, **26**, 469–476.
- 119 Takamatsu, N., Kanda, H., Ito, M., Yamashita, A., Yamashita, S. and Shiba, T. (1997) Rainbow trout Sox9: cDNA cloning, gene structure and expression. *Gene*, **202**, 167–170.

NOTE ADDED IN PROOF

The putative translation product W01C8.2 described in Figure 1 has recently been identified as a *C.elegans* Sox protein necessary for establishing a functional vulval–uterine connection and has been named COG-2 [Hanna-Rose, W. and Han, M. (1999) COG-2, a Sox domain protein necessary for establishing a functional vulval–uterine connection in *Caenorhabditis elegans*. *Development*, **126**, 169–179].