Seminars in Cell & Developmental Biology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Molecular basis for the genome engagement by Sox proteins

Linlin Hou[a,b,c], Yogesh Srivastava[a,b,c], Ralf Jauch[a,b,c,*]

[a] Genome Regulation Laboratory, Drug Discovery Pipeline, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[b] Key Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[c] Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form 9 August 2016
Accepted 9 August 2016
Available online xxx

Keywords:
Sox
Transcription factors
Gene regulation
DNA binding
Cellular reprogramming

ABSTRACT

The Sox transcription factor family consists of 20 members in the human genome. Many of them are key determinants of cellular identities and possess the capacity to reprogram cell fates by pioneering the epigenetic remodeling of the genome. This activity is intimately tied to their ability to specifically bind and bend DNA alone or with other proteins. Here we discuss our current knowledge on how Sox transcription factors such as Sox2, Sox17, Sox18 and Sox9 'read' the genome to find and regulate their target genes and highlight the roles of partner factors including Pax6, Nanog, Oct4 and Brn2. We integrate insights from structural and biochemical studies as well as high-throughput assays to probe DNA specificity in vitro as well as in cells and tissues.

Contents

* Corresponding author at: Genome Regulation Laboratory, Drug Discovery Pipeline, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.
E-mail address: ralf@gibh.ac.cn (R. Jauch).

http://dx.doi.org/10.1016/j.semcdb.2016.08.005

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

1. Introduction

Transcription factor (TF) proteins determine cellular identities and direct embryonic development by selectively binding to genomic DNA to orchestrate gene expression programs. Amongst the ~21,000 human protein-coding genes about 1600 encode for sequence-specific DNA binding TF proteins [1,2]. Most TFs belong to gene families comprising a handful up to several hundred members in mammalian genomes. Members of such gene families are termed paralogs and evolved by the expansion of ancestral genes through gene or genome duplications [3]. The gene encoding for the TF Sry (sex-determining region Y) was discovered following an intense search for the testis-determining factor on the Y-chromosome [4,5]. The sequence conservation between mouse and human Sry genes is restricted to a region of 79 amino acids. This sequence motif encodes for a special version of the high-mobility group (HMG) box also found in a class of ubiquitous and highly abundant non-histone DNA binding proteins [6]. In the original Sry study, four more homologous genes were isolated from autosomal loci of mouse 8.5 d.p.c. (days post coitum) cDNA libraries corresponding to Sox1-4 [4,7]. The unifying feature of these genes is the Sry-like HMG box; hence this gene family was termed Sox. Additional members were subsequently detected and cloned in a wide array of tissues taking advantage of sequence signatures within the HMG box [7–10]. With the availability of whole genome sequences it became clear that the mouse and human genomes each encodes for 20 Sox genes [11]. Based on the sequence identity of the HMG box, the Sox genes are classified into 8 groups denoted SoxA to SoxH with 1–3 members each [12].

The Sox TFs were soon found to constitute essential molecules with key roles during virtually all phases of embryonic development and the fate determination of many cell types as summarized in a number of excellent reviews [13–23]. The prominence of the gene family received a further elevation when one of its members, Sox2, was found to be a core component of TF cocktails with the ability of converting mouse and human somatic cells to induced pluripotent stem cells (iPSCs) [24–26]. Most Sox TFs are highly pleiotropic as they bind and regulate different gene sets in different cellular contexts. Sox2, for example, acts in a staggeringly diverse array of cell and tissue types including pluripotent stem cells, neural lineages, lung tissue, the eye and the ear [27]. Yet, what endows Sox proteins with this versatility and developmental plasticity largely remains elusive. Moreover, Sox proteins are reported to function as 'pioneer' factors. That is, they are able to bind compact transcriptionally silent chromatin and to recruit non-pioneer TFs to drive cell fate conversions. In this review, we discuss recent progress in the understanding of the biochemical basis for DNA and chromatin recognition by Sox proteins, mechanisms for the partnership of Sox proteins with other TFs and mechanisms for their pioneering activity.

2. Principles of DNA recognition

2.1. Structural basis

Evidence that the Sox HMG box enables DNA binding was first provided for the SRY protein. Binding was found to be sequence specific with a preference for a $C_1T_2T_3T_4G_5T_6$ C_7-like motif [28,29]. This core-motif was later verified to be the preferred binding sequence for all 20 Sox proteins although there can be subtle variations especially in the flanks of the element and a substantial degeneracy is tolerated [30,31]. As several sex-reversing mutations of SRY profoundly reduced the affinity for DNA, it was immediately clear that DNA binding is of critical functional importance [28]. The first structural view on the DNA recognition by Sox TFs was provided by the group of Marius Clore with the nuclear magnetic resonance (NMR) structure of the HMG box of human SRY bound to a $G_1T_2T_3T_4G_5T_6G_7C$ dsDNA in 1995 [32]. In the same year, the NMR structure of the HMG box of the related Lef-1 protein bound to $CACCC_1T_2T_3T_4A_5A_6GCTC$ was reported [33]. More recently, crystal structures of Sox2 [34], Sox17 [35], Sox4 [36] and Sox18 [37] bound to cognate DNA elements were published and a Sox9/DNA complex was deposited to the protein data bank (PDB-id 4S2Q). Altogether, these studies provided valuable insights into the molecular basis for the DNA recognition by Sox proteins. The HMG box folds into an L-shaped 'boomerang' structure constructed of three alpha helices and extended N- and C-terminal tails with an irregular strand-like configuration (Fig. 1A, B). Two hydrophobic clusters stabilize the fold, which include a conserved set of aromatic amino acids such as Phe10, Trp13 and Trp41 and Phe52 (HMG numbering according to reference [12] used throughout this manuscript). The short and long arms of the 'L' have also been denoted as major and minor 'wings' [38–40]. The shorter major wing encompasses the bulk of the amino acids and is composed of helices 1, 2 and the N-terminal turn of helix 3. The minor wing consists of the remainder of helix 3 and the extended N-terminus, which packs against helix 3. Contrary to most other TFs that bind to the major groove of the DNA, the Sox HMG binds to the minor groove of the DNA and its binding induces an overall bend of 60–70° (Fig. 1C). All base pairs of the CATTGT core motif are directly contacted by amino acids via base-specific interactions (Fig. 1D, E). Several of the contact residues emanate from the $R_5PMNAF_{10}MVW$ Sox signature motif at the N-terminus of the HMG box. The $F_{10}M_{11}$ dipeptide constitutes a wedge that intercalates between the central $T_3A_3'T_4A_4'$ base pair forcing the kinking of the DNA. Notably, all residues engaged in base-specific DNA interactions are invariant amongst the 20 Sox TFs (Fig. 1E). Only some residues mediating non-specific interactions with the DNA backbone show conservative replacements such as residues 2 and 15. Overall, the highly positively charged DNA binding surface exhibits a strong evolutionary conservation, whereas interfaces pointing away from the DNA are variable amongst Sox TFs (Fig. 1F, G). Therefore, monomeric forms of all 20 Sox TFs are expected to bind DNA in an identical fashion. Nevertheless, with the availability of a growing number of structures some variations at the Sox/DNA interface have been observed. For example, Arg18 and Asn30 can undergo a concerted conformational switch [36]. Moreover, Arg5, His29 and Tyr72 can structurally re-orient to better accommodate changes in the sequence of the DNA binding element [37]. However, these changes are dictated by the chemical environment provided by the DNA sequence and do not reflect differences inherent to individual Sox TFs. Thus, other mechanisms must account for the multitude of non-redundant and cell-type specific functions of Sox TFs.

2.2. Differences between Sox TFs and other HMG box proteins

Sox TFs belong to the HMG box superfamily of proteins, which is evolutionarily ancient with members present in unicellular eukaryotes such as yeast species. The HMG superfamily can be broadly divided into two groups based on the mechanism of DNA interaction [41,42]. First, there are sequence specific HMG boxes (ssHMGs) including the Sox TFs, the Tcf/Lef TFs and the yeast mating protein MATA. Second, there are the non-sequence specific HMG boxes (nsHMG) including HMG1, HMG2, the SSRP1 subunit of FACT (facilitates chromatin transcription), the mitochondrial TFAM/mtTF1 and UBF1 [41,42]. It appears plausible that sequence specificity evolved after the divergence of ssHMG and nsHMG groups however this question is not ultimately resolved. In an alternative scenario, nsHMGs lost sequence specificity while sequence specificity was present in the ancestral protein.

Lef-1 of the Tcf/Lef family exhibits a similar fold and DNA binding mechanism as Sox proteins but has a number of features distin-

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

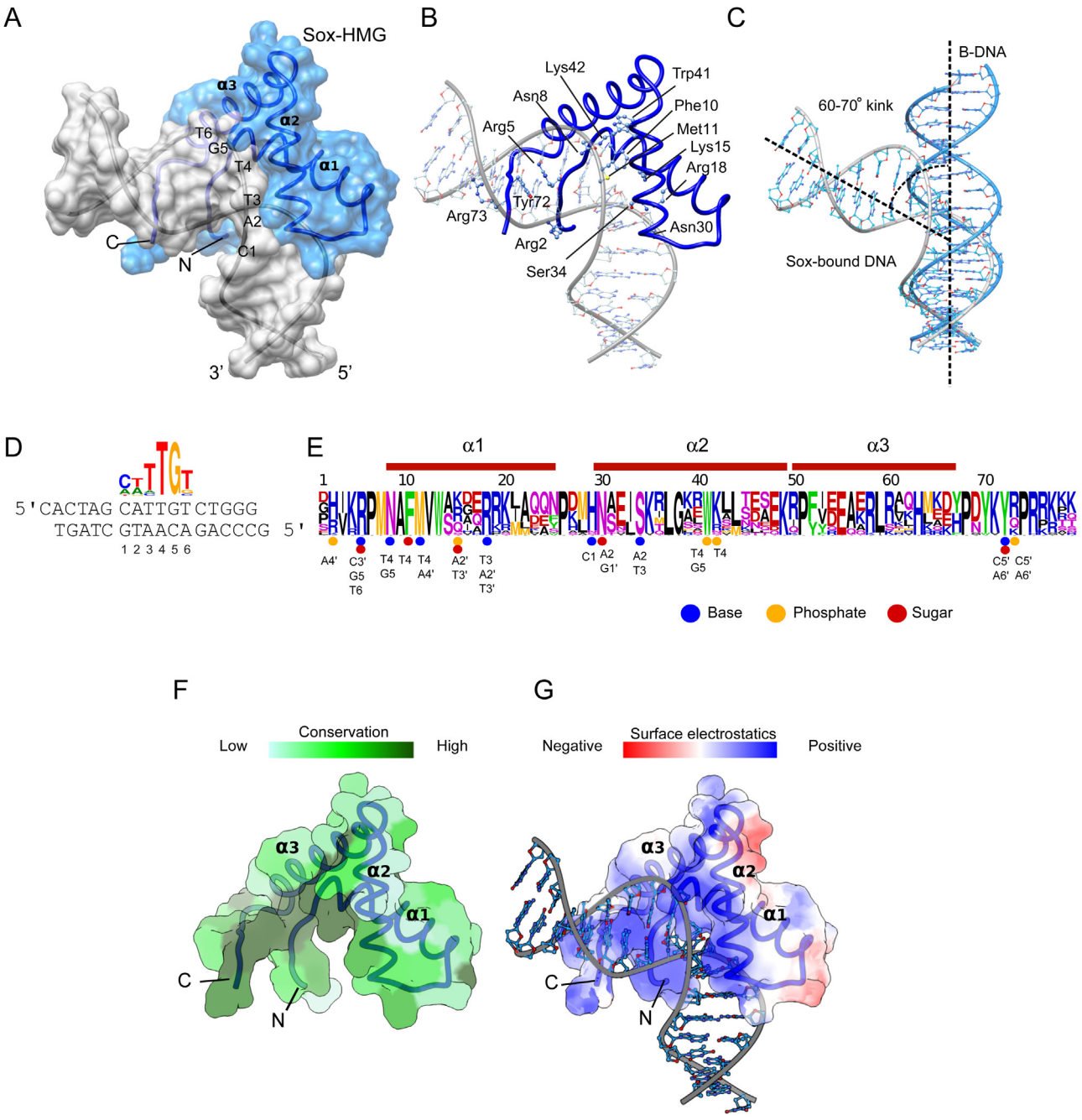

Fig. 1. DNA recognition by the Sox HMG box.
(A) Topological model of the Sox HMG (blue) bound to DNA (gray) shown as semi-transparent surface and ribbons. Helices a1–a3, N-and C-termini and the postion of the CATTGT core element are marked. (B) The DNA bases and the DNA contact amino acids are shown as ball-and-sticks.
(C) Ideal B-DNA is superimposed onto DNA when bound to the Sox HMG and the bending angle is highlighted. (D) A sequence logo of a Sox motif aligned with the cognate sequence of the *Prox1* gene. (E) Sequence logo constructed using the HMG box sequences of the 20 mouse Sox proteins. The residues involved in DNA binding are marked with spheres and the nucleotides they bind are indicated using the numbering scheme in (D). The apostrophe denotes nucleotides of the reverse strand. (F) Conservation scores of the 20 mouse Sox paralogs were mapped to the surface of the HMG box with light green indicating low conservation and dark green high conservation. (G) Electrostatic surface representation of the Sox18 HMG. Structural coordinates of SOX18 bound to *Prox1* DNA were used in all structural panels ([37]; PDB ID: 4Y60) and the Chimera software was used to generate structural cartoons (http://www.cgl.ucsf.edu/chimera/).

guishing it from Sox TFs [33,35]. First, a number of invariant Sox signature DNA contact residues are replaced in this family including Asn30, Ser34 and Tyr72 (Fig. 1E). These differences lead to variations in the consensus binding sequence in particular at position 6 ($CTTTGA_6A$ for Lef-1 versus $CTTTGT_6C$ for Sox) [30,43]. Next, the NMR structure of a Lef-1/DNA complex revealed an extended C-terminal tail that crosses over to make contacts with the major groove accompanied by a tilting of helix 3 in comparison to the Sry helix 3 [33,40]. The non-conservative replacement of Tyr72 in the Sox HMG by Ala72 in the Lef-1 HMG could be causative for this difference. In the case of the Sox HMG, Tyr72 interacts with the base pair at position 6 and positions the C-terminal tail firmly within the minor groove and likely prevents flipping over into the major groove [44].

nsHMG proteins HMG1 and HMG2 are highly abundant structural components of nuclear chromatin [6] and TFAM organizes the

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

mitochondrial genome [45,46]. Whereas all Sox genes encode only a single HMG box, certain nsHMGs such as HMG1 & HMG2 and TFAM possess tandem repeats of this domain [47,48]. A series of structures of free and DNA bound nsHMGs could clarify some key differences between sequence-specific and non-sequence specific DNA recognition by HMG boxes [39,45,46,49–53]. While the overall L-shaped and triple-helical fold is very similar for both nsHMG and ssHMGs, all nsHMGs contain two rather than one non-polar intercalation sites. The position corresponding to the secondary intercalation site corresponds to the polar Asn30 involved in base-specific contacts in the Sox HMG (Fig. 1B, E). A number of other Sox signature sequences mediating specific DNA binding are either absent or poorly conserved within the nsHMG group including Asn8, Ser34 and Tyr72 [51]. In sum, while the overall fold is similar for nsHMG and ssHMG boxes, the mechanism of DNA binding is profoundly different.

2.3. DNA bending

Circular permutations assays provided first evidence that Sox TFs, like nsHMG boxes, bend DNA [54,55]. Structural analysis revealed that the shape complementarity between the concave face of the L-shaped HMG box and intercalation of apolar residues facilitate DNA bending, minor groove widening and helical unwinding [32,33]. As the DNA bending activity of many nsHMGs is proposed to influence nuclear compartmentalization and chromatin remodeling, it is intriguing to speculate that Sox TFs are core determinants of the 3D-architectures of cell-type specific enhanceosomes. Indeed, the degree of DNA bending was reported to be critical for the transcriptional output of Sox2 [56]. Likewise, sex-reversing mutations of human SRY were reported to decrease the overall bending angle of SRY-bound DNA [44]. In agreement with this, FRET studies suggested that the bending angle depends on three critical factors: (i) the amino acid composition of the HMG box (i.e. Sry, Sox5, Lef-1), (ii) the sequences of the DNA element and (iii) at least for Lef-1 on the length of the C-terminal tail [57,58]. However, a detailed comparison of the structures of DNA elements bound to Sox2, Sox9, Sox17, Sox4 and Sox18 indicated very similar bending angles of the DNA duplex ranging from 60° to 70° [37]. Further, the DNA deformation at the level of base-pair parameters such as the twist angle describing the local unwinding and the roll angle describing the local opening of the base-pairs is also very similar for all five Sox HMG/DNA structures. Therefore, currently there is no evidence that the 20 Sox factors exhibit differential DNA bending activity and that specific bends explain the cell type specific activity of Sox TFs. Yet, the precise DNA sequence requirement and effects of mutations within the HMG box on bending should be more rigorously assessed.

2.4. Dynamics of the HMG box

A number of HMG box structures were analyzed as apo-proteins in the absence of DNA providing insights into the dynamics of the domain and structural adjustments accompanying DNA binding [38,39,50,53,59–61]. Both, DNA and the HMG box undergo mutually induced structural changes as the HMG is subjected to a pronounced disorder to order transitions (reviewed in reference [40]). Even at low temperatures HMG boxes are partially unfolded [33,38]. Major and minor wings are tied together by structurally independent hydrophobic cores leading to strong movements of the wings relative to each other. In the DNA-bound Sox17 HMG domain, helix 3 flips ~5° away from the DNA as compared to a superimposed apo-protein [35]. Further, the C-terminal tail is highly disordered in the apo-form but latches into the minor groove upon DNA recognition and adopts an ordered conformation. In sum, the HMG box is very dynamic and DNA binding leads to a reciprocal induced fit. Interestingly and counter-intuitively, deforming the DNA is energetically inexpensive and thought to be driven by favorable entropic contributions caused by the displacements of ordered waters from the minor groove [57,58].

3. Impact of partner factors on DNA recognition

The context dependent function of Sox TFs is driven by a 'Sox partner code', that is, the dynamic association of Sox TFs with partner proteins in a cell type specific fashion. The partner switches change genome engagement, gene expression programs and the developmental trajectories driven by Sox TFs. The partnering of Sox TFs has been covered in a number of excellent reviews [23,62,63]. Here we will focus on more recent insights on the molecular basis for these interactions and their functional consequences.

3.1. Homotypic interactions

While the majority of Sox TFs form monomers in solution, SoxD proteins (Sox5, Sox6 and Sox13) are an exception, as they possess a group-specific N-terminal coiled-coil domain that mediates DNA independent homo and SoxD heterodimerization [64]. Sox5 exists in short and long isoforms where the short isoform lacks the coiled-coil domain [65]. Dimerization of SoxD TFs facilitates binding to clustered Sox elements with little constraints on half-site spacing and a high tolerance for degeneracy in the binding site [65,66]. A Sox6 ChIPseq study reinforced this notion as single Sox motifs and a long A-rich sequence were identified rather than a distinctive dimer motif [67].

SoxE proteins (Sox8, Sox9, Sox10) encode a unique 40 amino acids dimerization (DIM) domain N-terminally preceding the HMG box [68–70]. SoxE TFs dimerize in a highly cooperative fashion but only in the presence of a GACAATGnnnnnC$_1$A$_2$T$_3$T$_4$G$_5$T$_6$C$_7$-like palindromic DNA binding sequence (Fig. 2A). Thus, in contrast to the constitutive SoxD dimers, SoxE dimerization is DNA dependent. Different from other dimerizing TFs that favor composite DNA elements with fixed spacing [71], SoxE factors are able to accommodate a range of flexibly spaced half-sites [72]. All three SoxE proteins can also effectively heterodimerize with each other but do not dimerize with non-SoxE proteins. Interestingly, truncated DIM-SoxE constructs can also effectively dimerize with isolated SoxE HMG boxes, suggesting that a single DIM is necessary and sufficient for dimer formation (Fig. 2A). Therefore, the dimerization must be driven by DIM-HMG rather than by DIM–DIM interactions and both, the SoxE DIM and the SoxE HMG contain structural elements facilitating it [72].

Surprisingly, Sox2 from the SoxB1 group has recently been reported to exist in both monomeric and dimeric forms in human neutrophils [73]. Truncation analysis revealed that the group B homolog (GBH) domain C-terminally of the HMG box is required for the homodimerizatin of Sox2. Triggered by the presence of bacterial DNA, Sox2 homodimers, but not the monomers, activate the TAB2-TAK1 complex leading to the stimulation of the innate immune response [73]. Therefore, homotypic Sox interactions, dependent or independent of DNA, could be of more functional importance than previously assumed.

3.2. Heterotypic interactions

3.2.1. Sox/Pax

Sox2 has been identified to bind and regulate the expression of the δ-crystallin gene in the chicken eye cooperatively with Pax6 on a regulatory element termed *DC5* [74,75]. *DC5* expression requires precise and compact spacing between the Sox and Pax half-sites [75]. However, curiously, the Pax half-site in *DC5* is different from the high-affinity SELEX (systematic evolution of ligands

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

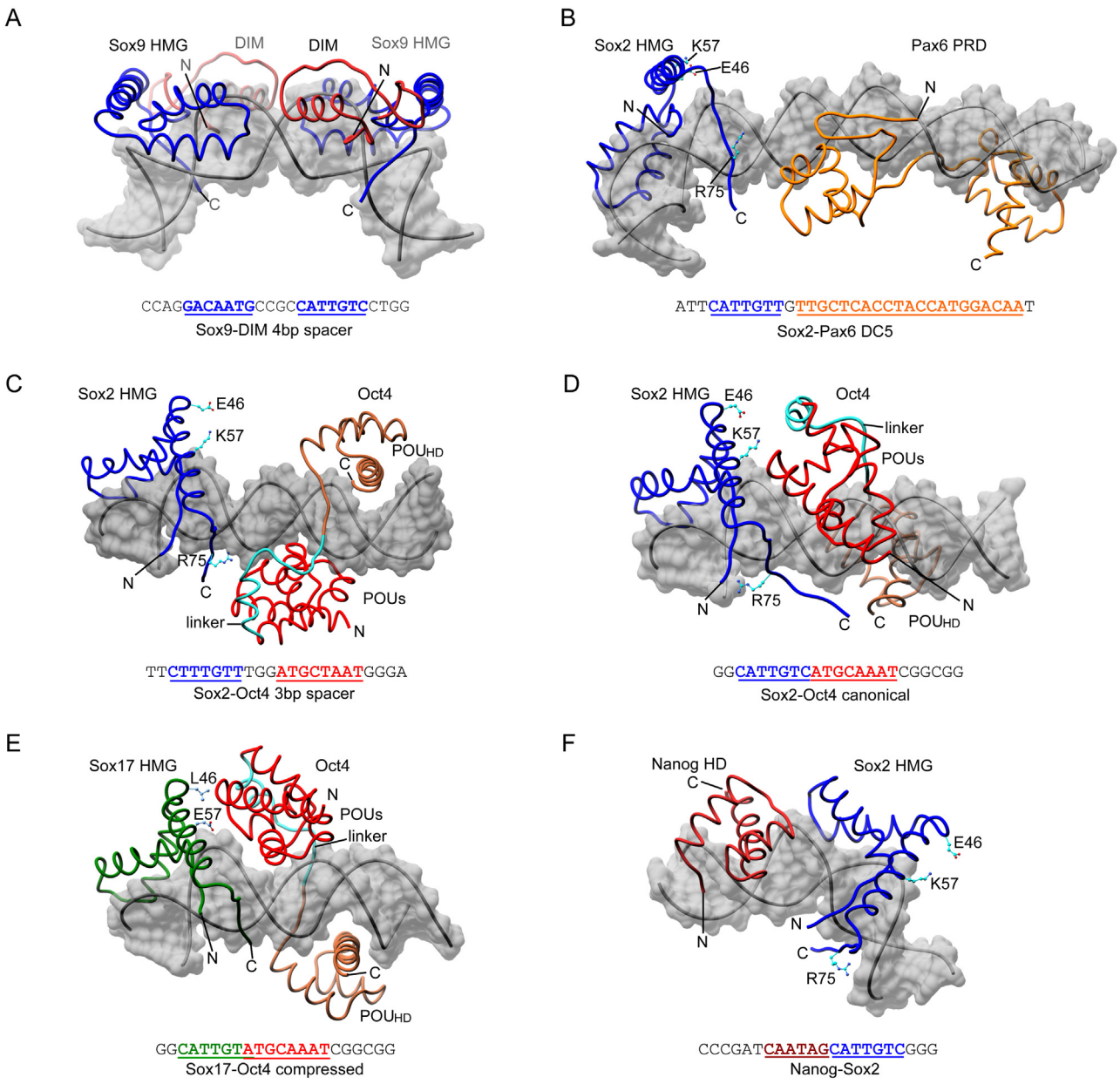

Fig. 2. Complexes of Sox HMGs with partner factors.
(A-F). Topological model of the Sox HMG (blue) bound. The DNA is shown as gray surface and the proteins as backbone ribbons with SoxE and Sox2 in HMG boxes in blue, Sox17 HMG in green, Pax6 PRD in orange, Nanog HD in red and bipartite Oct4 POU in red (POU specific domain, POU_S) or brown (POU homeodomain, POU_{HD}) with the POU linker in light blue. The sequence of the composite DNA element used to construct the models is shown underneath the structural cartoons. Models depict (A) a SoxE homodimer on palindromic Sox element with 4 bp spacer with blue HMG and red DIM [72], (B) Sox2/Pax6 heterodimer model including the Sox2 HMG and the Pax6 paired (PRD) domain on a *DC*5 element [77], (C) a Sox2/Oct4 heterodimers on an *Fgf4* and (D) a canonical *HoxB1*-like elements [34,92,95], (E) Sox17/Oct4 heterodimer on the compressed motif [93,94] and (F) a Sox2 HMG/Nanog homeodomain (HD) heterodimer on a SELEX-derived composite motif [107,147]. Amino acids discussed in the text as critical for the TF partnerships are marked.

by exponential enrichment) identified PAX consensus [76]. When the *DC5*-version of the Pax site is replaced with the SELEX consensus *(DC5con)* cooperative binding is abolished. This indicated that Pax6 binds DNA in two different modes: a high affinity mode not compatible with Sox2 cooperation and a high-cooperativity mode facilitating dimer formation and transactivation. Mutations of Arg75 of the Sox2 HMG box diminished complex formation on *DC5* suggesting that Arg75 directly mediates the Sox2/Pax6 complex formation [34]. However, when the Sox2 and Pax6 interaction on *DC5* DNA was probed by NMR, obvious protein interactions could not be mapped [77]. This suggests that the Sox2/Pax6 interaction on *DC5* is predominantly driven by a DNA-dependent allosteric mode rather than by direct protein–protein interactions (Fig. 2B). Yet, Arg75 could not be assigned in these experiments. A DNA element related to *DC5* termed *N3* has been found to autoregulate Sox2 [78]. However, there is no DNA sequence motif describing functional *DC5*-like sequences and genome-wide binding studies have only revealed DNA motifs resembling the SELEX consensus, albeit this element is present in only small fractions of the total binding sites [79]. Therefore, for the majority of Pax6 binding sites it remains unknown which sequence signatures exactly promote the functional recruitment of Pax6. When sites jointly targeted by SOX2 and PAX6 in human neural lineages where screened for *DC5*-like sites, a number of candidates elements could be identified [77].

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

Thus, the cooperation of SOX2 with PAX6 in the eye and neural lineages relies on a cryptic enhancers and indirect mechanisms such as DNA mediated allostery appear to contribute substantially to genomic target selection and gene regulation.

3.2.2. Sox/Oct

In 1995, Sox2 was found to form a heterodimeric complex with Oct4 (also termed Oct3 and encoded by the *Pou5f1* gene) on an element regulating the expression of the *Fgf4* gene in the blastocyst [80]. In the *Fgf4* element, Sox and Oct half-sites are separated by a 3 bp spacer (CATTGTCnnnATGCAAAT, n denotes spacer nucleotides) and the precise spacing is important for *Fgf4* expression [81] (Fig. 2C). Subsequently, functional Sox2/Oct4 heterodimers were found to regulate a number of other genes including *UTF1* [82], *Sox2* [83], *Nanog* [84] and *Pou5f1* [85]. However, in all these genes the half-sites are closer together, and the 3 bp spacer seen in the *Fgf4* element is absent, resulting in a $C_1A_2T_3T_4G_5T_6C_7$ATGCAAAT-like sequence, which we henceforth term the 'canonical' element (Fig. 2D). Strikingly, a canonical Sox/Oct motif was also identified in genome wide ChIPseq (chromatin immunoprecipitation followed by deep sequencing) studies using *de novo* motif discovery tools even if only loci immunoprecipitated with Oct4, Nanog, Smad1 or p300 antibodies were taken into account [86–89]. Moreover, Sox2 and Oct4 were found to be core factors to induce pluripotency in mouse and human somatic cells in combination with other TFs [24–26] or alone [90,91]. This suggested that the Sox2/Oct4 heterodimer is of pivotal importance for the acquisition and maintenance of pluripotency. Based on the crystal structure of the Sox2 HMG and the Oct1 POU (Pit-Oct-Unc) domain on the *Fgf4* element the molecular basis for the Sox2/Oct4 binding to *Fgf4* and *Utf1* elements could be modeled and analyzed by rational mutagenesis [34]. The molecular picture was further clarified by an NMR structure of a Sox2 HMG/Oct1 POU dimer on *HoxB1* DNA [92]. Helix 3 and the ordered C-terminal tail of the Sox2 HMG box are the main contributors to direct protein–protein interactions during Sox2/Oct4/DNA ternary complex formation (Fig. 2C, D). However, the molecular interfaces and participating amino acids mediating the interactions switch on *Fgf4* and *HoxB1*-like DNA elements [34]. The POU interface is provided by the POU-specific (POU_S) part of the bi-partite POU but indirect allosteric interaction with the interface-distal POU homeodomain (POU_{HD}) do also contribute to complex formation [93]. Interestingly, some amino acids at the HMG interface facilitating POU binding on the canonical element including Glu46 and Lys57 are not conserved amongst Sox paralogs and could mediate Sox-specific dimerization profiles [94–96]. Indeed, when these amino acids are mutated, Sox2 is no longer able to cooperate with Oct4 on the canonical motif and looses the ability to induce pluripotency demonstrating that Sox2/Oct4 heterodimers are critical for this process [94,97,98]. The SoxF factor Sox17 contains an acidic glutamate at position 57 and only poorly cooperates with Oct4 on canonical motifs. However, when the glutamate is exchanged with the basic lysine of Sox2, the engineered Sox17EK protein now cooperates very effectively with Oct4 on all tested canonical motifs [94,96]. More surprisingly, Sox17EK can also replace wild-type (WT) Sox2 in iPSC generation experiments using somatic mouse or human cells and even outperforms Sox2 in terms of reprogramming speed and efficiency [94,97]. Consistently, ChIPseq analysis revealed that Sox17EK indeed occupies sites normally bound by Sox2, while WT Sox17 binds separate locations [99]. Nevertheless, Sox17 also frequently co-binds with Oct4. However, instead of the canonical sequence, the Sox17/Oct4 dimer is recruited to an alternative $C_1A_2T_3T_4G_5T_6$ATGCAAAT-like signature sequence (Fig. 2E). This element differs from the canonical element by the elimination of nucleotide 7 of the Sox element bordering with the Oct half-site and was therefore termed 'compressed' DNA motif [94]. The compressed motif is also enriched in Sox17/Oct4 co-bound sites in primitive endoderm-like cells [99]. The switch of Sox partners by Oct4 indicates an intimate link between DNA-sequence, TF–TF co-operation and developmental outcome. WT Sox2 is not normally able to dimerize with Oct4 on the compressed motif. However, replacing residues 46 and 57 with their counterparts in Sox17 producing a synthetic Sox2E46LK57E protein, installs strong cooperativity with Oct4 on the compressed motif [95]. Whether this Sox2 double mutant acquires the Sox17-like capacity of converting embryonic stem cells (ESCs) into bona fide extraembryonic endoderm (XEN) cells remains to be investigated [100]. These studies suggested a Sox/Oct partner code triggering early developmental switches. Namely, a Sox2/Oct4 dimer binds the canonical motif to induce and maintain pluripotency and a Sox17/Oct4 dimer binds the compressed motif to trigger the differentiation of endodermal lineages. A number of other synergistically acting Sox/Oct pairs have been described implying that Sox/Oct dimerization on specific composite motifs is a crucial mechanism for context specific functions of these proteins [22,63]. For example, in neural progenitor cells (NPCs) Sox2 is thought to switch partners from Oct4 to Brn2 [101]. Early work on *Nestin* regulation suggested that Sox2/Brn2 partnership is driven by novel configuration of Sox/Oct half-sites [102]. However, genome-wide studies in NPCs indicated that the Sox2/Brn2 preferentially forms on canonical motifs albeit on a different subset of canonical motif locations than bound by Sox2/Oct4 in ESCs [101]. Therefore, additional mechanisms must be at play directing the selection of canonical motifs in these two cell types.

3.2.3. Sox/Nanog

The Nanog TF constitutes one of the hallmark genes of pluripotency alongside Sox2 and Oct4 [103,104]. Nanog binds DNA with a variant Q50 homeodomain [105] and co-occupies many enhancers with Sox2 and Oct4 [86,88,89,106]. Co-immunoprecipitation assays indicated that Nanog and Sox2 robustly interact in ESCs [107]. As tyrosine residues of the Sox2 serine-rich domain and tryptophan residues of the Nanog WR repeat are necessary for the interaction, binding is likely mediated by aromatic stacking. A combined SELEX for Sox2 and Nanog co-expressed in *E.coli* revealed an AATnnCATT composited element with curtailed half-sites compared to the isolated TAATCG-like and CATTGTC-like consensus sequences for Nanog or Sox2, respectively [107] (Fig. 2F). This composite motif is present in a large number of ChIPseq peaks co-bound by Nanog and Sox2 and it was suggested that the Sox2/Nanog interaction is more critical for pluripotency than the Sox2/Oct4 dimer [108]. Possibly, the Nanog/Sox2 interaction modifies the binding specificity similarly to the Sox2/Pax6 interaction and relaxes the requirement for nucleotides flanking the core-binding site. Thus, mutually induced changes to the binding preferences appear to be a common mechanism for the target gene selection of Sox TFs and their dimerization partners.

4. High-throughput identification of Sox DNA target sites

In 1990, SELEX was developed to comprehensively assess the binding specificity of nucleic acid binding proteins [109]. This technique became soon widely used to construct binding models for DNA-protein interactions including Sox TFs [29]. Conventional SELEX is a highly stringent technique providing the highest affinity binding sites for a given protein. However, in a cellular context, not only the highest affinity consensus sequences but also the entirety of low and medium binding sites will determine the genomic binding profile of a TF population. Naturally, all exposed genomic elements compete for a TF and modify the search process, that is the time it takes for a given TF to arrive at a functional binding site. Therefore, to predict binding probability from

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

sequence alone it would be desirable to quantify binding energies for all kmer/TF interactions. To this end, a high-throughput method termed protein-binding microarrays (PBMs) was developed capable of probing the binding of large TF sets to 8–10 mer sequences [30,110]. PBM provides a semi-quantitative readout for the binding of tagged recombinant proteins to microarrays with dsDNA sequences, which is detected using fluorescent antibodies. Sox proteins analyzed by PBM were found to bind a $T_3T_4G_5T_6$ core sequence and which differs from the binding pattern of the Lef/Tcf family [30]. Nevertheless, subtle changes in the periphery of this core motif allowed the clustering of Sox proteins by their sequence preferences although some of the Sox kmer clusters appear to be distinct only because of changed definitions of the start site of the binding site [30]. Moreover, the PBM-derived binding landscape of many Sox TFs was characterized by a secondary motif. That is, a position weight matrix (PWM) that captures a subset of binding sequences that cannot be satisfactorily described by the primary PWM. For example, the primary motif of mouse Sox4 is C_1T_2TTGT and the secondary motif is A_1A_2TTGT ([36], uniprobe ID UP00062). Apparently, positions 1 and 2 are interdependent and need to be changed concomitantly to maintain a favorable binding energy. The switch of binding between the motifs appears to be facilitated by a coordinated conformational change of the side-chains of residues Arg18 and Asn30 [36]. Similar DNA induced changes to the TF conformation have been linked to global conformational switches of the regulatory complex that influences co-factor recruitment and the direction of the regulatory event (activation versus repression) [111]. Yet, it remains controversial whether primary and secondary motifs represent truly distinctive binding modes [112]. Nevertheless, it is conceivable that changes in the binding sequence of Sox TFs, such as the primary and secondary motif, modulate the regulatory circuitries controlled by these proteins.

High-throughput SELEX (HT-SELEX) combines SELEX with deep sequencing technologies [113]. It can probe longer sequences than PBMs (14–40 base pairs) and is therefore better suited to study TF complexes [114]. Modifications of HT-SELEX can in principle be used to derive highly accurate thermodynamic binding models [115]. When Sox TFs were studied by HT-SELEX, a head-to-head pseudopalindromic $ACAATANC_1A_2T_3T_4G_5T_6$ DNA sequences was obtained for many Sox family members [113]. This finding was interpreted as binding of Sox TFs to DNA hairpin structures. As HMG boxes are long known to recognize unusual DNA structures such as four-way junctions [55,56] it will be of interest to test if hairpin binding is of physiological relevance or if it is a technical limitation of HT-SELEX. Surprisingly, homodimeric binding of Sox HMGs to double-stranded $ACAATGNC_1A_2T_3T_4G_5T_6$-like sequences was also observed recently albeit it is less efficient than binding to more widely spaced half-sites [72].

5. Involvement of domains outside the HMG box in genome engagement

The regions outside of the HMG box of Sox TFs is largely devoid of conserved sequences and distinct structural or functional domains and has therefore received less attention with few exceptions such as the SoxE DIM and the coiled-coil of SoxD TFs. The regions outside the box have well-established roles in transactivation and repression but its is less clear how they influence DNA binding and target gene selection. Structural information of non-HMG regions is at present not available. Yet, regions flanking the HMG box have been proposed to play critical roles in partner factor interactions and selective recognition of genomic loci [69,97,107].

The non-HMG box regions of Sry are very poorly conserved and, curiously, several other Sox genes could replace Sry to induce XX sex reversal in mice [116]. In most mammals, Sry is composed of an N-terminal domain (~30–60 amino acids) and a C-terminal domain (~70–100 amino acids) flanking the central HMG box. In rat and mouse, however, the N-terminus of Sry is missing and the C-terminus consists of glutamine repeats interspersed by histidine-rich segregators. This gave rise to the idea that only the HMG box of Sry is required for sex determination. However, deficiencies in the glutamine-rich C-terminal region failed to confer XX male sex reversal, implying an important role in male sex determination [116]. In humans, amino acid substitutions occurring in the N-terminal domain were detected in patients with XY gonadal dysgenesis [117,118] some of which were demonstrated to modulate DNA binding [119]. Moreover, deleting the non-HMG portion of Sry reduced binding to the consensus DNA sequence *in vitro* [120]. Besides influencing DNA binding directly, non-HMG box regions can impact partner factor association thereby modify the co-targeting of genes by TF complexes analogously to HMG box mediated partnerships. For example, a naturally occurring Sox18 *ragged* allele leads to a C-terminal truncation of the protein but leaves the HMG box intact. This deletion compromises the interaction of Sox18 with the MADS-box TF MEF2C [121]. Perturbed dimer formation, target gene selection and activation could be causative for the phenotypic consequences of the truncation.

Post-translational modifications of non-HMG domains present another mechanism for the regulation of the DNA-binding activity, transactivation, protein stability and nuclear localization of Sox proteins [15,17,122,123]. Several reports indicate that phosphorylation, sumoylation, glycosylation, acetylation, *O*-linked-*N*-acetylglucosamination (*O*-GlcNAcylation) and methylation directly influence DNA recognition by Sox TFs. For example, phosphorylation of Sox2 primarily at the C-terminal serine-rich region is a prerequisite for the subsequent sumoylation [124]. Sumoylation of a nearby lysine, in turn, decreases affinity for *Fgf4* DNA and inhibits transcriptional activation [125]. By contrast, phosphorylation of Sry and Sox9 in the flanks of the HMG box has the opposite effect as it increases affinity for DNA [126,127]. *O*-GlcNAcylation in the C-terminal domain of Sox2 was found to be a hallmark of pluripotency but the modification was readily removed upon differentiation [128]. Possibly, *O*-GlcNAcylation contributes to a discrimination of regulatory programs driven by Sox2 in pluripotent versus somatic cells. Lastly, methylation of Sox2 by CARM1 at Arg113 enhances the self-association of Sox2 and elevates Sox2-mediated transactivation [129]. Overall, post-translational modifications are clearly influence whether and how Sox TFs homo and heterodimerize and bind DNA to regulate gene expression. However, the role of non-HMG box regions on genome engagement and gene regulation has not yet been systematically explored so the mechanism of how these modifications affect regulatory programs remains largely elusive. As the non-HMG box regions are the most diverse parts of Sox TFs amongst both orthologs and paralogs, they are expected to critically impact their cell type and species-specific functions.

6. Pioneering activity

Eukaryotic genomes are organized into compacted and open chromatin regions that stabilize gene expression programs, cellular identities and deter unauthorized cell fate conversions. An elite subset of TFs such as FoxA1 was suggested to have a unique ability to directly bind the nucleosomal DNA, to loosen compacted chromatin and to activate otherwise silenced genes [130–132]. The paradigm pioneer TF FoxA1 has structural similarity to linker histones and its pioneering activity has been attributed to this similarity [133]. However, the TFs Sox2, Oct4 and Klf4 do not resemble linker histones but were recently shown to effectively bind nucleosome wrapped DNA *in vitro* and to associate with closed chromatin

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

at early stages of iPSC generation [134,135]. The authors performed *de novo* motif analysis for Sox2, Oct4 and Klf4 binding sites in nucleosome free versus nucleosome bound loci as distinguished using MNase-seq data (micrococcal nuclease digestion followed by deep sequencing). All three TFs bound their canonical motifs in nucleosome-free loci but bound to degenerate or partial motifs in nucleosome wrapped regions. In the case of Sox2, the 5th position of the $CTTTG_5T$ degenerated to a $CTTTN_5T$ version. Notably, mutations of the G at the 5th position were shown to profoundly reduce DNA bending upon Sox2 binding using circular permutations assays [56]. Given the smooth curvature of nucleosome bound DNA, the lack of a conserved G at position 5 has been interpreted to better accommodate Sox2 binding to nucleosomal DNA as a strong bend is neither possible nor necessary in the context of the nucleosome [135]. Thus, the ability of Sox2 to bind and only moderately bend DNA with a degenerate Sox motif on nucleosome surfaces has been suggested to be a key driver for its pioneering activity.

Interestingly, a single molecule tracking study in live ESCs demonstrated that Sox2 reduced the search time of Oct4 to cognate binding sites and increased the residence time of Oct4 [136]. Conversely, Oct4 did not shorten the search of Sox2 but moderately increased residence time. This suggested that Sox2 binds first to cognate Sox/Oct DNA elements and subsequently recruits Oct4. The resulting Sox2/Oct4/DNA ternary complex is overall more stable than the individual binary complexes [136]. The single molecule assays did not distinguish between binding to free DNA or nucleosomal DNA leaving open whether the hierarchy of binding is similar in either context.

In a different study, Sox as wells as Lef/Tcf TFs were classified as 'migrant TFs' in an effort to probe the ability of TFs to open up chromatin as measured by dynamic changes of DNase I hypersensitivity footprint sites during the differentiation of ESCs to endoderm [137]. By this definition, migrants TFs, in contrast to pioneer TFs, cannot open up chromatin by themselves and rely on additional stimuli such as co-factors to select their cognate target sites in open chromatin. Collectively, the pioneering function of Sox TFs leaves several open questions and further work is needed to unify the various experimental approaches into a common picture. In principle, the shared features of Sox TFs with nsHMGs that include architectural components of chromatin and subunits of the chromatin loosening FACT complex appear to qualify them as potent pioneers. Possibly, Sox TFs concomitantly function to direct transcriptional regulation as well as to reorganize chromatin, which is necessary for cellular reprogramming.

7. Perspectives

Since their discovery more than a quarter of a century ago, Sox TFs captivate an ever increasing community of researchers for their peculiar mode of DNA binding, their ability to pioneer cell fate conversions and the staggeringly vast array of cellular identities they determine. In addition to the topics discussed here, unexpected novel themes are emerging such as the interplay of Sox proteins and RNAs. In 2001, it was found that Sry and Sox6 co-localize with splicing factors in nucleus and modulate intron selection *in vitro* via their HMG box [138]. More recently, a direct cooperation between the long non-coding RNA *RMST* and Sox2 was found to influence target gene selection by Sox2 and to regulate neurogenesis [139]. Well-established partner factors directing cell-type specific functions of TFs such as Pax and Oct TFs may just represent the tip of the iceberg. Otx2 and Esrrb have also been reported to team up with Sox2 in the eye or in epiblast stem cells [140–142]. In addition, an extension of HT-SELEX to obtain dimer elements recovered a number of co-motifs of Sox factors with, for example, Tead4 and Meis1 [114]. Finally, it was suggested that the engineering of synthetic TFs has a great potential to more effectively drive cell fate conversions to produce cells for regenerative biomedicine [143–146]. It is thus of interest to explore the utility of synthetic Sox TFs, such as Sox17EK, for mammalian synthetic biology and the production of functional 'transplantation ready' cells.

Acknowledgements

We thank all members of the Jauch Lab, Andrew Hutchins and Kamesh Narasimhan for comments on the manuscript. This work was funded by the National Science Foundation for Distinguished Young Scholars of China (Grant No. 31501058 to L.H), a 2013 MOST China-EU Science and Technology Cooperation Program [2013DFE33080 to R.J.]; National Natural Science Foundation of China [31471238 to R.J.]; Chinese Academy of Sciences 100 talent award (to R.J.); Science and Technology Planning Project of Guangdong Province, China [2014B030301058 to R.J.]; Chinese Government Scholarship (CGS) (to Y.S.); University of the Chinese Academy of Sciences (UCAS) (to Y.S.).

References

[1] J.M. Vaquerizas, S.K. Kummerfeld, S.A. Teichmann, N.M. Luscombe, A census of human transcription factors: function, expression and evolution, Nature reviews, Genetics 10 (4) (2009) 252–263.
[2] H.M. Zhang, T. Liu, C.J. Liu, S. Song, X. Zhang, W. Liu, H. Jia, Y. Xue, A.Y. Guo, AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors, Nucleic Acids Res. 43 (2015) D76–D81, Database issue.
[3] E.V. Koonin, Orthologs, paralogs, and evolutionary genomics, Annu. Rev. Genet. 39 (2005) 309–338.
[4] J. Gubbay, J. Collignon, P. Koopman, B. Capel, A. Economou, A. Munsterberg, N. Vivian, P. Goodfellow, R. Lovell-Badge, 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 (6281) (1990) 245–250.
[5] A.H. Sinclair, P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A.M. Frischauf, R. Lovell-Badge, P.N. Goodfellow, A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif, Nature 346 (6281) (1990) 240–244.
[6] M. Bustin, R. Reeves, High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function, Prog. Nucleic Acid Res. Mol. Biol. 54 (1996) 35–100.
[7] P. Denny, S. Swift, N. Brand, N. Dabhade, P. Barton, A. Ashworth, A conserved family of genes related to the testis determining gene, SRY, Nucleic Acids Res. 20 (11) (1992) 2887.
[8] T.L. Dunn, L. Mynett-Johnson, E.M. Wright, B.M. Hosking, P.A. Koopman, G.E. Muscat, Sequence and expression of Sox-18 encoding a new HMG-box transcription factor, Gene 161 (2) (1995) 223–225.
[9] M. van de Wetering, M. Oosterwegel, K. van Norren, H. Clevers, Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes, EMBO J. 12 (10) (1993) 3847–3854.
[10] E.M. Wright, B. Snopek, P. Koopman, Seven new members of the Sox gene family expressed during mouse development, Nucleic Acids Res. 21 (3) (1993) 744.
[11] G.E. Schepers, R.D. Teasdale, P. Koopman, Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families, Dev. Cell 3 (2) (2002) 167–170.
[12] J. Bowles, G. Schepers, P. Koopman, Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators, Dev. Biol. 227 (2) (2000) 239–255.
[13] E.M. Abdelalim, M.M. Emara, P.R. Kolatkar, The SOX transcription factors as key players in pluripotent stem cells, Stem Cells Dev. 23 (22) (2014) 2687–2699.
[14] L. Dailey, C. Basilico, Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes, J. Cell. Physiol. 186 (3) (2001) 315–328.
[15] Y. Kamachi, H. Kondoh, Sox proteins: regulators of cell fate specification and differentiation, Development 140 (20) (2013) 4129–4144.
[16] J.C. Kiefer, Back to basics: Sox genes, Dev. Dyn. 236 (8) (2007) 2356–2366.
[17] V. Lefebvre, B. Dumitriu, A. Penzo-Mendez, Y. Han, B. Pallavi, Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors, Int. J. Biochem. Cell Biol. 39 (12) (2007) 2195–2214.
[18] R. Lovell-Badge, The early history of the Sox genes, Int. J. Biochem. Cell Biol. 42 (3) (2010) 378–380.
[19] L.H. Pevny, R. Lovell-Badge, Sox genes find their feet, Curr. Opin. Genet. Dev. 7 (3) (1997) 338–344.

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

[20] A. Sarkar, K. Hochedlinger, The sox family of transcription factors: versatile regulators of stem and progenitor cell fate, Cell stem Cell 12 (1) (2013) 15–30.
[21] M. Wegner, From head to toes: the multiple facets of Sox proteins, Nucleic Acids Res. 27 (6) (1999) 1409–1420.
[22] M. Wegner, All purpose Sox: the many roles of Sox proteins in gene expression, Int. J. Biochem. Cell Biol. 42 (3) (2010) 381–390.
[23] M. Wilson, P. Koopman, Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators, Curr. Opin. Genet. Dev. 12 (4) (2002) 441–446.
[24] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell 131 (5) (2007) 861–872.
[25] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (4) (2006) 663–676.
[26] J. Yu, M.A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J.L. Frane, S. Tian, J. Nie, G.A. Jonsdottir, V. Ruotti, R. Stewart, I.I. Slukvin, J.A. Thomson, Induced pluripotent stem cell lines derived from human somatic cells, Science 318 (5858) (2007) 1917–1920.
[27] H. Kondoh, R. Lovell-Badge, Sox2: Biology and Role in Development and Disease, Academic Press, 2015.
[28] V.R. Harley, D.I. Jackson, P.J. Hextall, J.R. Hawkins, G.D. Berkovitz, S. Sockanathan, R. Lovell-Badge, P.N. Goodfellow, DNA binding activity of recombinant SRY from normal males and XY females, Science 255 (5043) (1992) 453–456.
[29] V.R. Harley, R. Lovell-Badge, P.N. Goodfellow, Definition of a consensus DNA binding site for SRY, Nucleic Acids Res. 22 (8) (1994) 1500–1501.
[30] G. Badis, M.F. Berger, A.A. Philippakis, S. Talukder, A.R. Gehrke, S.A. Jaeger, E.T. Chan, G. Metzler, A. Vedenko, X. Chen, H. Kuznetsov, C.F. Wang, D. Coburn, D.E. Newburger, Q. Morris, T.R. Hughes, M.L. Bulyk, Diversity and complexity in DNA recognition by transcription factors, Science 324 (5935) (2009) 1720–1723.
[31] S. Mertin, S.G. McDowall, V.R. Harley, The DNA-binding specificity of SOX9 and other SOX proteins, Nucleic Acids Res. 27 (5) (1999) 1359–1364.
[32] M.H. Werner, J.R. Huth, A.M. Gronenborn, G.M. Clore, Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex, Cell 81 (5) (1995) 705–714.
[33] J.J. Love, X. Li, D.A. Case, K. Giese, R. Grosschedl, P.E. Wright, Structural basis for DNA bending by the architectural transcription factor LEF-1, Nature 376 (6543) (1995) 791–795.
[34] A. Remenyi, K. Lins, L.J. Nissen, R. Reinbold, H.R. Scholer, M. Wilmanns, Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers, Genes Dev. 17 (16) (2003) 2048–2059.
[35] P. Palasingam, R. Jauch, C.K. Ng, P.R. Kolatkar, The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms, J. Mol. Biol. 388 (3) (2009) 619–630.
[36] R. Jauch, C.K. Ng, K. Narasimhan, P.R. Kolatkar, Crystal structure of the Sox4 HMG/DNA complex suggests a mechanism for the positional interdependence in DNA recognition, Biochem. J 443 (1) (2012) 39–47.
[37] M. Klaus, N. Prokoph, M. Girbig, X. Wang, Y.H. Huang, Y. Srivastava, L. Hou, K. Narasimhan, P.R. Kolatkar, M. Francois, R. Jauch, Structure and decoy-mediated inhibition of the SOX18/Prox1-DNA interaction, Nucleic Acids Res. 44 (8) (2016) 3922–3935.
[38] P.D. Cary, C.M. Read, B. Davis, P.C. Driscoll, C. Crane-Robinson, Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5, Protein Sci. 10 (1) (2001) 83–98.
[39] C.M. Read, P.D. Cary, C. Crane-Robinson, P.C. Driscoll, D.G. Norman, Solution structure of a DNA-binding domain from HMG1, Nucleic Acids Res. 21 (15) (1993) 3427–3436.
[40] M.A. Weiss, Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition, Mol. Endocrinol. 15 (3) (2001) 353–362.
[41] V. Laudet, D. Stehelin, H. Clevers, Ancestry and diversity of the HMG box superfamily, Nucleic Acids Res. 21 (10) (1993) 2493–2501.
[42] S. Soullier, P. Jay, F. Poulat, J.M. Vanacker, P. Berta, V. Laudet, Diversification pattern of the HMG and SOX family members during evolution, J. Mol. Evol. 48 (5) (1999) 517–527.
[43] K. Giese, A. Amsterdam, R. Grosschedl, DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1, Genes Dev. 5 (12B) (1991) 2567–2578.
[44] E.C. Murphy, V.B. Zhurkin, J.M. Louis, G. Cornilescu, G.M. Clore, Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation, J. Mol. Biol. 312 (3) (2001) 481–499.
[45] H.B. Ngo, J.T. Kaiser, D.C. Chan, The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA, Nat. Struct. Mol. Biol. 18 (11) (2011) 1290–1296.
[46] A. Rubio-Cosials, J.F. Sidow, N. Jimenez-Menendez, P. Fernandez-Millan, J. Montoya, H.T. Jacobs, M. Coll, P. Bernado, M. Sola, Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter, Nat. Struct. Mol. Biol. 18 (11) (2011) 1281–1289.
[47] C.S. Malarkey, M.E. Churchill, The high mobility group box: the ultimate utility player of a cell, Trends Biochem. Sci. 37 (12) (2012) 553–562.
[48] A. Rubio-Cosials, M. Sola, U-turn DNA bending by human mitochondrial transcription factor A, Curr. Opin. Struct. Biol. 23 (1) (2013) 116–124.
[49] F.H. Allain, Y.M. Yen, J.E. Masse, P. Schultze, T. Dieckmann, R.C. Johnson, J. Feigon, Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding, EMBO J. 18 (9) (1999) 2563–2579.
[50] D.N. Jones, M.A. Searles, G.L. Shaw, M.E. Churchill, S.S. Ner, J. Keeler, A.A. Travers, D. Neuhaus, The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*, Structure 2 (7) (1994) 609–627.
[51] F.V.t. Murphy, R.M. Sweet, M.E. Churchill, The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition, EMBO J. 18 (23) (1999) 6610–6618.
[52] R. Sanchez-Giraldo, F.J. Acosta-Reyes, C.S. Malarkey, N. Saperas, M.E. Churchill, J.L. Campos, Two high-mobility group box domains act together to underwind and kink DNA, Acta Crystallogr. D Biol. Crystallogr. 71 (Pt. 7) (2015) 1423–1432.
[53] H.M. Weir, P.J. Kraulis, C.S. Hill, A.R. Raine, E.D. Laue, J.O. Thomas, Structure of the HMG box motif in the B-domain of HMG1, EMBO J. 12 (4) (1993) 1311–1319.
[54] S. Ferrari, V.R. Harley, A. Pontiggia, P.N. Goodfellow, R. Lovell-Badge, M.E. Bianchi, SRY, like HMG1, recognizes sharp angles in DNA, EMBO J. 11 (12) (1992) 4497–4506.
[55] A. Pontiggia, R. Rimini, V.R. Harley, P.N. Goodfellow, R. Lovell-Badge, M.E. Bianchi, Sex-reversing mutations affect the architecture of SRY-DNA complexes, EMBO J. 13 (24) (1994) 6115–6124.
[56] P. Scaffidi, M.E. Bianchi, Spatially precise DNA bending is an essential activity of the sox2 transcription factor, J. Biol. Chem. 276 (50) (2001) 47296–47302.
[57] C. Crane-Robinson, A.I. Dragan, C.M. Read, Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry, Methods Mol. Biol. 543 (2009) 625–651.
[58] A.I. Dragan, C.M. Read, E.N. Makeyeva, E.I. Milgotina, M.E. Churchill, C. Crane-Robinson, P.L. Privalov, DNA binding and bending by HMG boxes: energetic determinants of specificity, J. Mol. Biol. 343 (2) (2004) 371–393.
[59] N. Gao, W. Jiang, H. Gao, Z. Cheng, H. Qian, S. Si, Y. Xie, Structural basis of human transcription factor sry-related box 1 box 17 binding to DNA, Protein Pept. Lett. 20 (4) (2013) 481–488.
[60] J.J. Love, X. Li, J. Chung, H.J. Dyson, P.E. Wright, The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA, Biochemistry 43 (27) (2004) 8725–8734.
[61] L.P. van Houte, V.P. Chuprina, M. van der Wetering, R. Boelens, R. Kaptein, H. Clevers, Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4, J. Biol. Chem. 270 (51) (1995) 30516–30524.
[62] Y. Kamachi, M. Uchikawa, H. Kondoh, Pairing SOX off: with partners in the regulation of embryonic development, Trends Genet. 16 (4) (2000) 182–187.
[63] H. Kondoh, Y. Kamachi, SOX-partner code for cell specification: regulatory target selection and underlying molecular mechanisms, Int. J. Biochem. Cell Biol. 42 (3) (2010) 391–399.
[64] V. Lefebvre, The SoxD transcription factors-Sox5, Sox6, and Sox13-are key cell fate modulators, Int. J. Biochem. Cell Biol. 42 (3) (2010) 429–432.
[65] V. Lefebvre, P. Li, B. de Crombrugghe, 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 (19) (1998) 5718–5733.
[66] N. Takamatsu, H. Kanda, I. Tsuchiya, S. Yamada, M. Ito, S. Kabeno, T. Shiba, S. Yamashita, A gene that is related to SRY and is expressed in the testes encodes a leucine zipper-containing protein, Mol. Cell. Biol. 15 (7) (1995) 3759–3766.
[67] C.F. Liu, V. Lefebvre, The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis, Nucleic Acids Res. 43 (17) (2015) 8183–8203.
[68] R.I. Peirano, D.E. Goerich, D. Riethmacher, M. Wegner, Protein zero gene expression is regulated by the glial transcription factor Sox10, Mol. Cell. Biol. 20 (9) (2000) 3198–3209.
[69] R.I. Peirano, M. Wegner, The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences, Nucleic Acids Res. 28 (16) (2000) 3047–3055.
[70] B. Schlierf, A. Ludwig, K. Klenovsek, M. Wegner, Cooperative binding of Sox10 to DNA: requirements and consequences, Nucleic Acids Res. 30 (24) (2002) 5509–5516.
[71] A. Jankowski, E. Szczurek, R. Jauch, J. Tiuryn, S. Prabhakar, Comprehensive prediction in 78 human cell lines reveals rigidity and compactness of transcription factor dimers, Genome Res. 23 (8) (2013) 1307–1318.
[72] Y.H. Huang, A. Jankowski, K.S. Cheah, S. Prabhakar, R. Jauch, SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains, Sci. Rep. 5 (2015) 10398.
[73] P. Xia, S. Wang, B. Ye, Y. Du, G. Huang, P. Zhu, Z. Fan, Sox2 functions as a sequence-specific DNA sensor in neutrophils to initiate innate immunity against microbial infection, Nat. Immunol. 16 (4) (2015) 366–375.
[74] Y. Kamachi, S. Sockanathan, Q. Liu, M. Breitman, R. Lovell-Badge, H. Kondoh, Involvement of SOX proteins in lens-specific activation of crystallin genes, EMBO J. 14 (14) (1995) 3510–3519.
[75] Y. Kamachi, M. Uchikawa, A. Tanouchi, R. Sekido, H. Kondoh, Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development, Genes Dev. 15 (10) (2001) 1272–1286.

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

[76] J. Epstein, J. Cai, T. Glaser, L. Jepeal, R. Maas, Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes, J. Biol. Chem. 269 (11) (1994) 8355–8361.
[77] K. Narasimhan, S. Pillay, Y.H. Huang, S. Jayabal, B. Udayasuryan, V. Veerapandian, P. Kolatkar, V. Cojocaru, K. Pervushin, R. Jauch, DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors, Nucleic Acids Res. 43 (3) (2015) 1513–1528.
[78] M. Inoue, Y. Kamachi, H. Matsunami, K. Imada, M. Uchikawa, H. Kondoh, PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development, Genes Cells 12 (9) (2007) 1049–1061.
[79] A. Bhinge, J. Poschmann, S.C. Namboori, X. Tian, S. Jia Hui Loh, A. Traczyk, S. Prabhakar, L.W. Stanton, MiR-135b is a direct PAX6 target and specifies human neuroectoderm by inhibiting TGF-beta/BMP signaling, EMBO J. 33 (11) (2014) 1271–1283.
[80] H. Yuan, N. Corbi, C. Basilico, L. Dailey, Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3, Genes Dev. 9 (21) (1995) 2635–2645.
[81] D.C. Ambrosetti, C. Basilico, L. Dailey, 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 (11) (1997) 6321–6329.
[82] M. Nishimoto, A. Fukushima, A. Okuda, M. Muramatsu, The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2, Mol. Cell. Biol. 19 (8) (1999) 5453–5465.
[83] M. Tomioka, M. Nishimoto, S. Miyagi, T. Katayanagi, N. Fukui, H. Niwa, M. Muramatsu, A. Okuda, Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex, Nucleic Acids Res. 30 (14) (2002) 3202–3213.
[84] D.J. Rodda, J.L. Chew, L.H. Lim, Y.H. Loh, B. Wang, H.H. Ng, P. Robson, Transcriptional regulation of nanog by OCT4 and SOX2, J. Biol. Chem. 280 (26) (2005) 24731–24737.
[85] J.L. Chew, Y.H. Loh, W. Zhang, X. Chen, W.L. Tam, L.S. Yeap, P. Li, Y.S. Ang, B. Lim, P. Robson, H.H. Ng, Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells, Mol. Cell. Biol. 25 (14) (2005) 6031–6046.
[86] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, H.H. Ng, Integration of external signaling pathways with the core transcriptional network in embryonic stem cells, Cell 133 (6) (2008) 1106–1117.
[87] J. Kim, J. Chu, X. Shen, J. Wang, S.H. Orkin, An extended transcriptional network for pluripotency of embryonic stem cells, Cell 132 (6) (2008) 1049–1061.
[88] G. Kunarso, N.Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y.S. Chan, H.H. Ng, G. Bourque, Transposable elements have rewired the core regulatory network of human embryonic stem cells, Nat. Genet. 42 (7) (2010) 631–634.
[89] Y.H. Loh, Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu, L. Lipovich, V.A. Kuznetsov, P. Robson, L.W. Stanton, C.L. Wei, Y. Ruan, B. Lim, H.H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, Nat. Genet. 38 (4) (2006) 431–440.
[90] A. Giorgetti, N. Montserrat, I. Rodriguez-Piza, C. Azqueta, A. Veiga, J.C. Izpisua Belmonte, Generation of induced pluripotent stem cells from human cord blood cells with only two factors: oct4 and Sox2, Nat. Protoc. 5 (4) (2010) 811–820.
[91] D. Huangfu, K. Osafune, R. Maehr, W. Guo, A. Eijkelenboom, S. Chen, W. Muhlestein, D.A. Melton, Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2, Nat. Biotechnol. 26 (11) (2008) 1269–1275.
[92] D.C. Williams Jr., M. Cai, G.M. Clore, Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1. Sox2. Hoxb1-DNA ternary transcription factor complex, J. Biol. Chem. 2 (2004) 1449–1457.
[93] F. Merino, B. Bouvier, V. Cojocaru, Cooperative DNA recognition modulated by an interplay between protein-Protein interactions and DNA-mediated allostery, PLoS Comput. Biol. 11 (6) (2015) e1004287.
[94] R. Jauch, I. Aksoy, A.P. Hutchins, C.K. Ng, X.F. Tian, J. Chen, P. Palasingam, P. Robson, L.W. Stanton, P.R. Kolatkar, Conversion of sox17 into a pluripotency reprogramming factor by reengineering its association with oct4 on DNA, Stem Cells 29 (6) (2011) 940–951.
[95] F. Merino, C.K. Ng, V. Veerapandian, H.R. Scholer, R. Jauch, V. Cojocaru, Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation, Structure 22 (9) (2014) 1274–1286.
[96] C.K. Ng, N.X. Li, S. Chee, S. Prabhakar, P.R. Kolatkar, R. Jauch, Deciphering the Sox-Oct partner code by quantitative cooperativity measurements, Nucleic Acids Res. 40 (11) (2012) 4933–4941.
[97] I. Aksoy, R. Jauch, V. Eras, W.B. Chng, J. Chen, U. Divakar, C.K. Ng, P.R. Kolatkar, L.W. Stanton, Sox transcription factors require selective interactions with oct4 and specific transactivation functions to mediate reprogramming, Stem Cells 31 (12) (2013) 2632–2646.
[98] N. Tapia, C. MacCarthy, D. Esch, A. Gabriele Marthaler, U. Tiemann, M.J. Arauzo-Bravo, R. Jauch, V. Cojocaru, H.R. Scholer, Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency, Sci. Rep. 5 (2015) 13533.
[99] I. Aksoy, R. Jauch, J. Chen, M. Dyla, U. Divakar, G.K. Bogu, R. Teo, C.K. Leng Ng, W. Herath, S. Lili, A.P. Hutchins, P. Robson, P.R. Kolatkar, L.W. Stanton, Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm, EMBO J. 32 (7) (2013) 938–953.
[100] A.C. McDonald, S. Biechele, J. Rossant, W.L. Stanford, Sox17-mediated XEN cell conversion identifies dynamic networks controlling cell-fate decisions in embryo-derived stem cells, Cell Rep. 9 (2) (2014) 780–793.
[101] M.A. Lodato, C.W. Ng, J.A. Wamstad, A.W. Cheng, K.K. Thai, E. Fraenkel, R. Jaenisch, L.A. Boyer, SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state, PLoS Genet. 9 (2) (2013) e1003288.
[102] S. Tanaka, Y. Kamachi, A. Tanouchi, H. Hamada, N. Jing, H. Kondoh, Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells, Mol. Cell. Biol. 24 (20) (2004) 8834–8846.
[103] I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells, Cell 113 (5) (2003) 643–655.
[104] K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, S. Yamanaka, The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells, Cell 113 (5) (2003) 631–642.
[105] R. Jauch, C.K. Ng, K.S. Saikatendu, R.C. Stevens, P.R. Kolatkar, Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog, J. Mol. Biol. 376 (3) (2008) 758–770.
[106] L.A. Boyer, T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, R.A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, Cell 122 (6) (2005) 947–956.
[107] A. Gagliardi, N.P. Mullin, Z. Ying Tan, D. Colby, A.I. Kousa, F. Halbritter, J.T. Weiss, A. Felker, K. Bezstarosti, R. Favaro, J. Demmers, S.K. Nicolis, S.R. Tomlinson, R.A. Poot, I. Chambers, A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal, EMBO J. 32 (16) (2013) 2231–2247.
[108] F.C.K. Wong, I. Chambers, N.P. Mullin, SOX2-dependent regulation of pluripotent stem cells, Sox2: biology and role in development and disease, Academic Press, 2015, pp. 163.
[109] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, Science 249 (4968) (1990) 505–510.
[110] L.A. Barrera, A. Vedenko, J.V. Kurland, J.M. Rogers, S.S. Gisselbrecht, E.J. Rossin, J. Woodard, L. Mariani, K.H. Kock, S. Inukai, T. Siggers, L. Shokri, R. Gordan, N. Sahni, C. Cotsapas, T. Hao, S. Yi, M. Kellis, M.J. Daly, M. Vidal, D.E. Hill, M.L. Bulyk, Survey of variation in human transcription factors reveals prevalent DNA binding changes, Science 351 (6280) (2016) 1450–1454.
[111] S.H. Meijsing, M.A. Pufall, A.Y. So, D.L. Bates, L. Chen, K.R. Yamamoto, DNA binding site sequence directs glucocorticoid receptor structure and activity, Science 324 (5925) (2009) 407–410.
[112] Y. Zhao, G.D. Stormo, Quantitative analysis demonstrates most transcription factors require only simple models of specificity, Nat. Biotechnol. 29 (6) (2011) 480–483.
[113] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors, Cell 152 (1-2) (2013) 327–339.
[114] A. Jolma, Y. Yin, K.R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, J. Taipale, DNA-dependent formation of transcription factor pairs alters their binding specificity, Nature 527 (7578) (2015) 384–388.
[115] Y. Zhao, D. Granas, G.D. Stormo, Inferring binding energies from selected binding sites, PLoS Comput. Biol. 5 (12) (2009) e1000590.
[116] L. Zhao, P. Koopman, SRY protein function in sex determination: thinking outside the box, Chromosome Res. 20 (1) (2012) 153–162.
[117] P. Canto, E. de la Chesnaye, M. Lopez, A. Cervantes, B. Chavez, F. Vilchis, E. Reyes, A. Ulloa-Aguirre, S. Kofman-Alfaro, J.P. Mendez, A mutation in the 5' non-high mobility group box region of the SRY gene in patients with Turner syndrome and Y mosaicism, J. Clin. Endocrinol. Metab. 85 (5) (2000) 1908–1911.
[118] G. Gimelli, S. Gimelli, N. Dimasi, R. Bocciardi, E. Di Battista, T. Pramparo, O. Zuffardi, Identification and molecular modelling of a novel familial mutation in the SRY gene implicated in the pure gonadal dysgenesis, Eur. J. Hum. Genet. 15 (1) (2007) 76–80.
[119] J.G. Assumpcao, C.E. Benedetti, A.T. Maciel-Guerra, G. Guerra Jr., M.T. Baptista, M.R. Scolfaro, M.P. de Mello, Novel mutations affecting SRY DNA-binding activity: the HMG box N65H associated with 46,XY pure gonadal dysgenesis and the familial non-HMG box R30I associated with variable phenotypes, J. Mol. Med. 80 (12) (2002) 782–790.
[120] I. Sanchez-Moreno, R. Coral-Vazquez, J.P. Mendez, P. Canto, Full-length SRY protein is essential for DNA binding, Mol. Hum. Reprod. 14 (6) (2008) 325–330.
[121] B.M. Hosking, S.C. Wang, S.L. Chen, S. Penning, P. Koopman, G.E. Muscat, SOX18 directly interacts with MEF2C in endothelial cells, Biochem. Biophys. Res. Commun. 287 (2) (2001) 493–500.
[122] P. Bernard, V.R. Harley, Acquisition of SOX transcription factor specificity through protein–protein interaction, modulation of Wnt signalling and

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005

post-translational modification, Int. J. Biochem. Cell Biol. 42 (3) (2010) 400–410.
[123] S. Malki, B. Boizet-Bonhoure, F. Poulat, Shuttling of SOX proteins, Int. J. Biochem. Cell Biol. 42 (3) (2010) 411–416.
[124] D. Van Hoof, J. Munoz, S.R. Braam, M.W. Pinkse, R. Linding, A.J. Heck, C.L. Mummery, J. Krijgsveld, Phosphorylation dynamics during early differentiation of human embryonic stem cells, Cell Stem Cell 5 (2) (2009) 214–226.
[125] S. Tsuruzoe, K. Ishihara, Y. Uchimura, S. Watanabe, Y. Sekita, T. Aoto, H. Saitoh, Y. Yuasa, H. Niwa, M. Kawasuji, H. Baba, M. Nakao, Inhibition of DNA binding of Sox2 by the SUMO conjugation, Biochem. Biophys. Res. Commun. 351 (4) (2006) 920–926.
[126] M. Desclozeaux, F. Poulat, P. de Santa Barbara, J.P. Capony, P. Turowski, P. Jay, C. Mejean, B. Moniot, B. Boizet, P. Berta, Phosphorylation of an N-terminal motif enhances DNA-binding activity of the human SRY protein, J. Biol. Chem. 273 (14) (1998) 7988–7995.
[127] W. Huang, X. Zhou, V. Lefebvre, B. de Crombrugghe, Phosphorylation of SOX9 by cyclic AMP-dependent protein kinase A enhances SOX9's ability to transactivate a Col2a1 chondrocyte-specific enhancer, Mol. Cell. Biol. 20 (11) (2000) 4149–4158.
[128] H. Jang, T.W. Kim, S. Yoon, S.Y. Choi, T.W. Kang, S.Y. Kim, Y.W. Kwon, E.J. Cho, H.D. Youn, *O*-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network, Cell Stem Cell 11 (1) (2012) 62–74.
[129] H.Y. Zhao, Y.J. Zhang, H. Dai, Y. Zhang, Y.F. Shen, CARM1 mediates modulation of Sox2, PLoS One 6 (10) (2011) e27026.
[130] L.A. Cirillo, F.R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K.S. Zaret, Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4, Mol. Cell 9 (2) (2002) 279–289.
[131] M. Iwafuchi-Doi, K.S. Zaret, Pioneer transcription factors in cell reprogramming, Genes Dev. 28 (24) (2014) 2679–2692.
[132] K.S. Zaret, J.S. Carroll, Pioneer transcription factors: establishing competence for gene expression, Genes Dev. 25 (21) (2011) 2227–2241.
[133] L.A. Cirillo, C.E. McPherson, P. Bossard, K. Stevens, S. Cherian, E.Y. Shim, K.L. Clark, S.K. Burley, K.S. Zaret, Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome, EMBO J. 17 (1) (1998) 244–254.
[134] A. Soufi, G. Donahue, K.S. Zaret, Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome, Cell 151 (5) (2012) 994–1004.
[135] A. Soufi, M.F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, K.S. Zaret, Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming, Cell 161 (3) (2015) 555–568.
[136] J. Chen, Z. Zhang, L. Li, B.C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, Z. Liu, Single-molecule dynamics of enhanceosome assembly in embryonic stem cells, Cell 156 (6) (2014) 1274–1285.
[137] R.I. Sherwood, T. Hashimoto, C.W. O'Donnell, S. Lewis, A.A. Barkal, J.P. van Hoff, V. Karun, T. Jaakkola, D.K. Gifford, Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape, Nat. Biotechnol. 32 (2) (2014) 171–178.
[138] K. Ohe, E. Lalli, P. Sassone-Corsi, A direct role of SRY and SOX proteins in pre-mRNA splicing, Proc. Natl. Acad. Sci. U. S. A. 99 (3) (2002) 1146–1151.
[139] S.Y. Ng, G.K. Bogu, B.S. Soh, L.W. Stanton, The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis, Mol. Cell 51 (3) (2013) 349–359.
[140] C. Buecker, R. Srinivasan, Z. Wu, E. Calo, D. Acampora, T. Faial, A. Simeone, M. Tan, T. Swigut, J. Wysocka, Reorganization of enhancer patterns in transition from naive to primed pluripotency, Cell Stem Cell 14 (6) (2014) 838–853.
[141] H. Danno, T. Michiue, K. Hitachi, A. Yukita, S. Ishiura, M. Asashima, Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression, Proc. Natl. Acad. Sci. U. S. A. 105 (14) (2008) 5408–5413.
[142] A.P. Hutchins, S.H. Choo, T.K. Mistri, M. Rahmani, C.T. Woon, C.K. Ng, R. Jauch, P. Robson, Co-motif discovery identifies an Esrrb-Sox2-DNA ternary complex as a mediator of transcriptional differences between mouse embryonic and epiblast stem cells, Stem Cells 31 (2) (2013) 269–281.
[143] R. Jauch, P.R. Kolatkar, What makes a pluripotency reprogramming factor? Curr. Mol. Med. 13 (5) (2013) 806–814.
[144] A. Soufi, Mechanisms for enhancing cellular reprogramming, Curr. Opin. Genet. Dev. 25 (2014) 101–109.
[145] J.C. Tsang, X. Gao, L. Lu, P. Liu, Cellular reprogramming by transcription factor engineering, Curr. Opin. Genet. Dev. 28 (2014) 1–9.
[146] X. Yang, V. Malik, R. Jauch, Reprogramming cells with synthetic proteins, Asian J. Androl. 17 (3) (2015) 394–402.
[147] Y. Hayashi, L. Caboni, D. Das, F. Yumoto, T. Clayton, M.C. Deller, P. Nguyen, C.L. Farr, H.J. Chiu, M.D. Miller, M.A. Elsliger, A.M. Deacon, A. Godzik, S.A. Lesley, K. Tomoda, B.R. Conklin, I.A. Wilson, S. Yamanaka, R.J. Fletterick, Structure-based discovery of NANOG variant with enhanced properties to promote self-renewal and reprogramming of pluripotent stem cells, Proc. Natl. Acad. Sci. U. S. A. 112 (15) (2015) 4666–4671.

Please cite this article in press as: L. Hou, et al., Molecular basis for the genome engagement by Sox proteins, Semin Cell Dev Biol (2016), http://dx.doi.org/10.1016/j.semcdb.2016.08.005