Published online 16 February 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 11* ***4933–4941***
doi:10.1093/nar/gks153

Deciphering the Sox-Oct partner code by quantitative cooperativity measurements

Calista K. L. Ng[1,2], Noel X. Li[1], Sheena Chee[1], Shyam Prabhakar[3], Prasanna R. Kolatkar[1,4] and Ralf Jauch[1,*]

[1]Laboratory for Structural Biochemistry, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, [2]School of Biological Sciences, Nanyang Technological University, Singapore 637551, [3]Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672 and [4]Department of Biological Sciences, National University of Singapore, Singapore 117543

Received October 4, 2011; Revised and Accepted January 25, 2012

ABSTRACT

Several Sox-Oct transcription factor (TF) combinations have been shown to cooperate on diverse enhancers to determine cell fates. Here, we developed a method to quantify biochemically the Sox-Oct cooperation and assessed the pairing of the high-mobility group (HMG) domains of 11 Sox TFs with Oct4 on a series of composite DNA elements. This way, we clustered Sox proteins according to their dimerization preferences illustrating that Sox HMG domains evolved different propensities to cooperate with Oct4. Sox2, Sox14, Sox21 and Sox15 strongly cooperate on the canonical element but compete with Oct4 on a recently discovered compressed element. Sry also cooperates on the canonical element but binds additively to the compressed element. In contrast, Sox17 and Sox4 cooperate more strongly on the compressed than on the canonical element. Sox5 and Sox18 show some cooperation on both elements, whereas Sox8 and Sox9 compete on both elements. Testing rationally mutated Sox proteins combined with structural modeling highlights critical amino acids for differential Sox-Oct4 partnerships and demonstrates that the cooperativity correlates with the efficiency in producing induced pluripotent stem cells. Our results suggest selective Sox-Oct partnerships in genome regulation and provide a toolset to study protein cooperation on DNA.

INTRODUCTION

How regulatory information is genetically encoded is an overarching yet unresolved question in genome biology. This information is scanned and interpreted by sequence-specific transcription factor (TF) proteins. However, the biochemical basis for the selective recruitment of TFs to genomic enhancers that govern spatial and temporal gene expression remains elusive. Multiple studies have shown that TFs often bind to short and degenerate DNA-binding sites that have been discovered computationally in huge numbers throughout the genome (1–3). Yet, only 1–5% of these binding sites are actually occupied by the corresponding TF. How do TFs discriminate between functional and nonfunctional binding sites? It has been shown that TFs have a propensity to cluster and are more likely to target genomic regions that are co-bound by other factors (4,5). Potentially, enhancers of co-expressed genes could share their own distinctive 'fingerprint' or grammar of DNA motifs that recruit particular TF combinations. To predict gene expression patterns from DNA sequence and TF concentration alone, this motif grammar needs to be decoded. It is possible that enhancers of co-expressed genes are only loosely defined with an unconstrained arrangement of binding motifs over several 100 bps not necessitating the direct physical interactions of TFs (4,6). In contrast, the motif grammar may include binding sites with constrained spacing between them whose recognition is tied to specific protein–interaction surfaces of individual TF proteins. These protein interactions underlie their developmental specificities and selectively target TFs to genomic enhancers. However, while TF heterodimerization predominates among paralogous groups of TFs such as nuclear receptors (7), helix–loop–helix (8) and leucine zipper families (9), examples for the selective dimerization of structurally unrelated TFs are sparse. Nevertheless, several studies have highlighted the importance of a direct cooperation between unrelated TF pairs (10–13). Most prominently, the Sox and Oct families of TFs have been shown to cooperate to execute key developmental programs (14,15).

*To whom correspondence should be addressed. Tel: +65 6808 8102; Fax: +65 6478 9060; Email: jauchr@gis.a-star.edu.sg

On their own, all 20 mammalian Sox proteins bind to a CTTTGT-like sequence (2,16,17) while most Oct factors recognize an octamer related to a ATGCAAAT sequence (18). By combining these sequences, composite motifs can be constructed with different motif orientation and spacing (19). Several such composite motifs have been found to be functional targets for the synergistic regulation by Sox and Oct proteins. For example, (i) the Sox2-Oct4 pair drives stem-cell pluripotency genes on either a 0 or a 3 bp-spaced motif element (20–23); (ii) Sox17-Oct4 cooperate during endodermal differentiation (24) presumably on a compressed motif (19) and (iii) Sox2-Brn2 regulates brain development on a sox-oct motif with a 6-bp spacer (13). Notably, when the cooperative binding of Sox2 and Oct4 to DNA is perturbed by rational mutagenesis its ability to induce pluripotency is lost (19). Conversely, although wild-type Sox17 cannot induce pluripotency, a mutated version of Sox17 increases cooperative binding with Oct4 on pluripotency gene enhancers and thus has the potential to induce pluripotency. Such results suggested that there might be a Sox-Oct partner code that underlies cell fate decisions (14,15). To further investigate whether members of the Sox and Oct families evolved features to cooperatively target specific enhancer elements a global assessment of the Sox-Oct pairing profile is highly desirable. To this end, we have developed a method to measure heterodimer cooperativity factors revealing the mode TF heterodimerization on composite DNA elements. We used this method to study the heterodimerization propensities of representative members of all seven major Sox families with Oct4 on a range of composite DNA elements. As a result, we found that Sox families exhibit markedly different propensities to associate with Oct4 on distinctly configured binding motifs. By measuring cooperativity factors of rationally mutated Sox proteins, we found that the re-engineered Sox17EK behaves like an enhanced Sox2. This likely underlies its improved properties in producing induced pluripotent stem cells *in lieu* of the native pluripotency factor Sox2 (19). Together, we demonstrate that cooperativity measurements are critical to understand TF function and the *cis*-regulatory logic of developmental enhancers.

MATERIALS AND METHODS

Cloning, protein expression and purification

The POU domains of mouse Oct4 and high-mobility group (HMG) domains of mouse Sry, Sox2, 14, 21, 4, 5, 8, 9, 17, 18 and 15 were BP cloned from their respective Imagene clones (Oct4: IRAKp961K04111Q; Sry: IRAMp995I2211Q; Sox2: RPCIB731A06406Q; Sox14: IRAKp961K05125Q; Sox21: IRAKp961C14126Q; Sox4: IRAVp968F02163D; Sox5: IRAMp995N0310Q; Sox8: IRAVp968H01144D; Sox9: IRAVp968B0369D; Sox17: OCACo5052D058; Sox18: IRAVp968E0317D; Sox15: IRCKp5014B242Q) into a pENTRY vector, pDONR221 (see Supplementary Table S1C for primer sequences). The resulting pENTR-constructs were first verified by sequencing and recombined into either pETG20A or pHisMBP expression plasmids using the GATEWAY^{TM} technology (Invitrogen). Constructs were transformed into *Escherichia coli* BL21(DE3) cells, grown in 1 × Terrific Bertani broth supplemented with 0.1% glucose and 100 μg/ml ampicillin until OD600 nm ~0.6–0.8 before inducing with 0.5 mM isopropyl-β-thiogalactoside at 18°C for ~18 h. Fusion proteins were purified using previously published protocol (19,25,26). In short, fusion proteins were purified using an immobilized metal affinity chromatography step, tag cleavage using the TEV protease followed by ion-exchange chromatography and gel filtration.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) were carried out using forward strand 5′Cy5-labeled-dsDNA (Sigma Proligo, see Figures 2A and 3A). DNA probes were prepared by mixing equimolar amounts of complementary strands in 1 × annealing buffer (20 mM Tris–HCl, pH 8.0; 50 mM $MgCl_2$; 50 mM KCl), heated to 95°C for 5 min and subsequently with 1°C/min ramping down to 4°C in a PCR block. Typical binding reactions contain 100 nM dsDNA with varying concentrations of both Sox and Oct4 proteins in a 1 × EMSA buffer [20 mM Tris–HCl pH 8.0, 0.1 mg/ml bovine serum albumin (Biorad), 50 μM $ZnCl_2$, 100 mM KCl, 10% (v/v) glycerol, 0.1% (v/v) Igepal CA630 and 2 mM β-mercaptoethanol] and were incubated for 1 hr at 4°C in the dark to reach binding equilibrium. Reactions were loaded into a pre-run 12% native 1 × Tris–glycine (25 mM Tris pH 8.3; 192 mM glycine) polyacrylamide gel, and DNA complexes were separated at 4°C for 30 min at 200 V. The bands were detected using a Typhoon 9140 PhosphorImager (Amersham Biosciences) and quantified using the ImageQuant TL software (Amersham Biosciences).

Cooperativity factor measurement

As an extension of our homodimer model described previously (27), we defined four possible microstates for a heterodimer-binding model. The participating species are defined as D for DNA, P1 for protein 1 and P2 for protein 2. The equilibrium dissociation constant of each individual protein can be represented as in Equation 1 where [D] is the concentration of free DNA, [P1] and [P2] concentrations of free proteins and [DP1] and [DP2] solitary protein–DNA complexes.

$$K_{d1} = \frac{[D][P1]}{[DP1]} \quad \text{or} \quad K_{d2} = \frac{[D][P2]}{[DP2]} \tag{1}$$

If P1 and P2 are mixed with DNA in the same tube, the fourth state representing a ternary complex becomes feasible. Dissociation constants of secondary-binding events are described by Equation 2, where [DP1P2] is the equilibrium concentration of heterodimeric protein–DNA complexes.

$$K_{d12} = \frac{[DP1][P2]}{[DP1P2]} \quad \text{and} \quad K_{d21} = \frac{[DP2][P1]}{[DP1P2]} \tag{2}$$

With f_0, f_1, f_2 and f_3 defined as the fractional concentrations of the free DNA, monomer–DNA complexes 1 and 2 and the heterodimer–DNA complex, respectively ($f_0+f_1+f_2+f_3 = 1$), the heterodimer cooperativity factor ω can be straightforwardly calculated from the experimentally determined fractional concentrations f_0, f_1, f_2 and f_3 as follows:

$$\omega = \frac{K_{d1}}{K_{d21}} = \frac{K_{d2}}{K_{d12}} = \frac{[D][DP1P2]}{[DP1][DP2]} = \frac{f_0 f_3}{f_1 f_2} \quad (3)$$

As defined here, $\omega > 1$ implies positive cooperativity; $\omega = 1$ no cooperativity; $\omega < 1$ negative cooperativity. To reduce errors when calculating ω, we only included measurements, where each of the four fractional concentrations were at least 5%. Cooperativity factor heatmaps and graphs for the Sox-Oct4-DNA combinations were plotted using R (http://www.r-project.org/) with the Gplot package.

Structural modeling and analysis

Homology models for Sox HMG or the Oct4 POU domain proteins were generated using I-TASSER with the Sox17 HMG (pdb-id 3F27) and the Oct1 POU (pdb-id 1GT0) as templates (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (28). Sox HMGs were subsequently superimposed onto the Sox2 HMG/Oct1 POU complex bound to a canonical element (pdb-id 1O4X). Superpositions, visual inspections and figure generation were carried out using PyMol.

RESULTS AND DISCUSSION

Sox proteins exhibit diverse protein interaction surfaces

The 80 amino acid HMG domain of Sox proteins is highly conserved for all paralogs (Figure 1A). In accordance with similar DNA sequence preferences for all Sox proteins (2), amino acids that contact DNA bases are nearly invariant for all Sox proteins. However, protein contact interfaces as defined in structural studies on Sox2 and Oct1 show some disparity (30) (highlighted as blue empty circles). As an extension of earlier work on Sox2 and Sox17, we generated homology models for all Sox families and inspected the electrostatic charge distribution on the van der Waals' surface (Figure 1B). The protein surface of Sox proteins facing Oct4 when bound to canonical sox-oct motifs show pronounced differences distinguishing Sox families. The SoxC, E and F groups contain an acidic patch at this interface, the SoxB and SoxG groups are highly basic and the SoxD group is largely neutral. We have recently shown that residue 57 (numbering according to HMG conventions), which is causing the disparate electrostatic pattern of the Sox families, is critical for the effective dimerization of Sox2 with Oct4 on pluripotency enhancers (19). To understand how these structural differences affect Sox-Oct partnerships, we developed a quantitative method to study TF cooperation and analysed the interaction of 11 Sox proteins with Oct4.

A method for the determination of cooperativity factors

We first cloned HMG domains of mouse Sox proteins and screened for soluble protein expression in a 96-well format. Representative members of most Sox families were then purified to at least 95% purity. Next, we performed EMSAs to quantify the fractional binding of Sox proteins and Oct4 on DNA. At equilibrium conditions, the mixing of the two proteins with DNA results in macromolecular complexes corresponding to four microstates per EMSA lane: (1) free DNA; (2) Sox-bound DNA; (3) Oct4-bound DNA and (4) Sox-Oct4-DNA ternary complex (Figure 1C). The abundance of each microstate at equilibrium is directly proportional to its Boltzmann weight, which in turn is a function of the protein concentrations, the equilibrium dissociation constants and the cooperativity factor, ω (Figure 1D, see 'Materials and Methods' section). Substituting the fractional contribution of each microstate from equilibrium experiments into our heterodimerization cooperativity model allowed us to quantify the cooperativity factor (Figure 1D). The cooperativity factor is essentially the fold change in the equilibrium binding constants when a protein co-binds, relative to the equilibrium constant for solitary binding. Values greater than 1 represent positive cooperativity, where both proteins mutually lower their free energies of binding. That is, complex formation is favoured. Negative cooperativity ($\omega < 1$) represents a competitive-binding mode, where the protein has a preference for binding to unbound DNA rather than forming a ternary complex. Finally, values of about 1 correspond to additive binding with proteins having no specific preference to binding either DNA that was already bound by another protein or free DNA.

Sox proteins exhibit a unique dimerization preference with Oct4 on variant DNA configurations

We measured cooperativity factors in multiple replicates for 11 Sox proteins with Oct4 on nine differently configured composite DNA motifs including the 'canonical' Sox2-Oct4 site found in many embryonic stem cell enhancers (23,31), the plus3 site found in the *Fgf4* enhancer (32) and the newly discovered 'compressed' element (19). In particular, on the canonical and compressed elements, we observed differences in the cooperativity pattern for the Sox proteins, whereas all Sox proteins tested cooperated with Oct4 on the plus3 element (Figure 2B, C, E and Supplementary Figure S1). We combined the whole dataset of log2 transformed cooperativity factors and created a heat map using the hierarchical clustering method implemented in the heatmap.2 R package (Figure 2C). The clustering approach revealed that the Sox proteins can be categorized into five separate groups highlighting their cooperativity patterns (Figure 2C). Similarly, DNA motif configurations cluster into five groups. Cluster II contains only the plus3 element displaying cooperative recruitment of all Sox-Oct4 pairs. Widely spaced elements (plus 2–plus 10; cluster III) exhibit an essentially additive-binding mode for all proteins under study. The plus1 element, however, shows a strongly

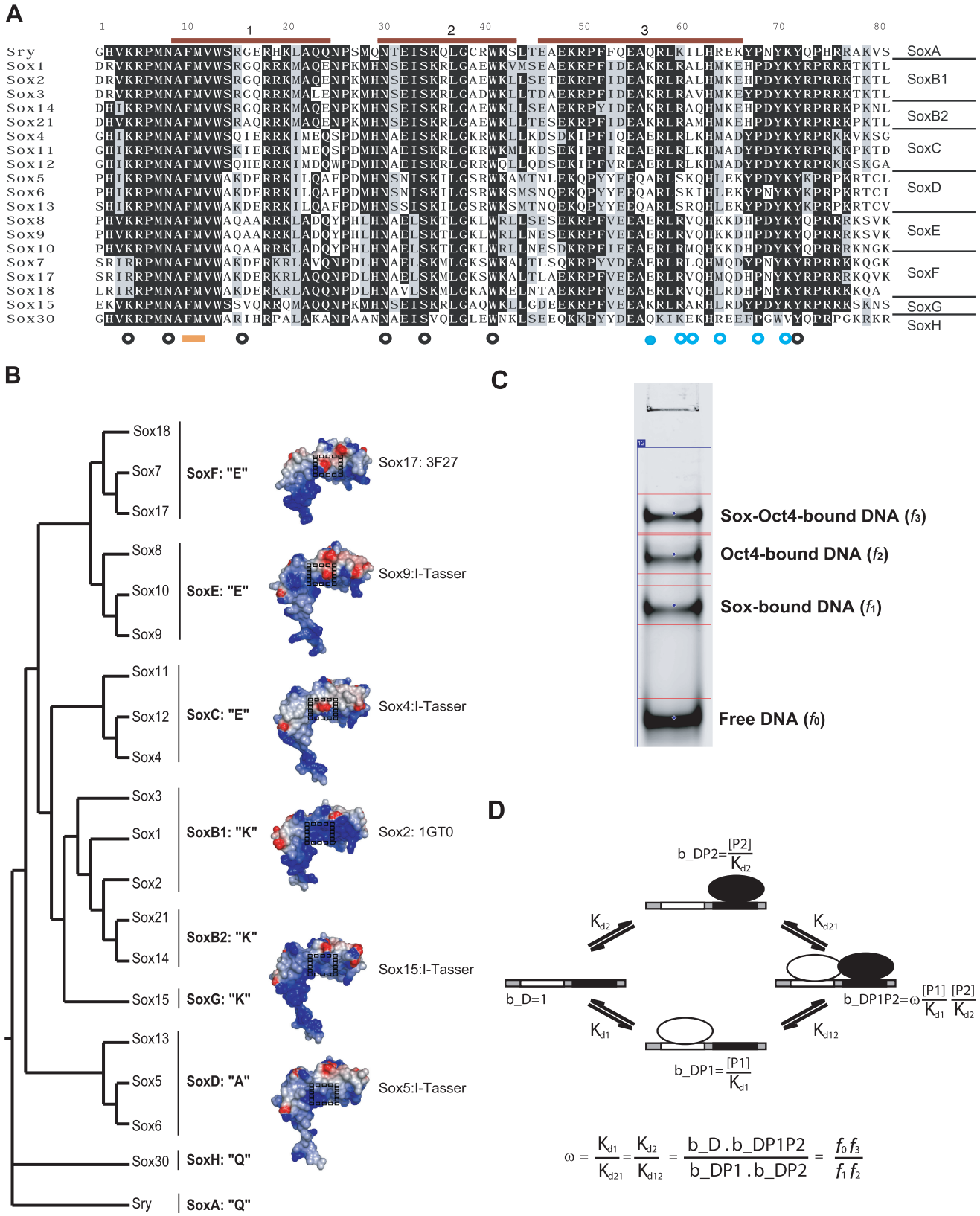


Figure 1. (**A**) Alignment of amino acid sequence of all mouse Sox-high-mobility group (HMG) domains shaded with BOXSHADE. The Sox subfamilies are indicated to the right. The numbering corresponds to the HMG convention (29). α-Helices are marked with a red bar. The Phe-Met wedge is indicated with an orange bar below the alignment. DNA interacting residues are marked by black empty circles while Sox-Oct interacting residues are marked by blue empty circles. Highly conserved and similar sequences are shaded in black or gray. (**B**) A phylogenetic tree calculated using PROML (http://caps.ncbs.res.in/iws/proml.html). This simplified tree largely corresponds to the more exhaustive phylogenetic analysis of Sox factors. Sox subgroups (29) and the amino acids found at position 57 of the HMG domains are indicated in single letter codes.

(continued)

competitive-binding mode for all Sox-Oct4. The canonical element and the compressed element (clusters I and IV) exhibit a strong disparity with regard to the Sox-Oct pairs they preferentially recruit and determine the clustering of Sox proteins. Sox8 and Sox9 (cluster B) are not capable of cooperating with Oct4 on either the compressed or the canonical element. By contrast, Sox5 and Sox18 (cluster E) bivalently cooperate on both elements. Clusters A and D, however, have inverse cooperativity profiles on those elements. While Sox2, Sox15, Sox14 and Sox21 (cluster A) strongly cooperate on the canonical element, they are not capable of co-binding with Oct4 on the compressed element. Similarly, Sry cooperates on the canonical element but retains an additive-binding mode on the compressed element. By contrast, Sox4 and Sox17 (cluster D) only weakly cooperate on the canonical element but strongly cooperate on the compressed element. Overall, while the HMG domains of the Sox proteins investigated here bind highly similar DNA sequences in isolation, they markedly differ in their potential to cooperate with Oct4.

Cooperativity patterns of rationally mutated Sox proteins

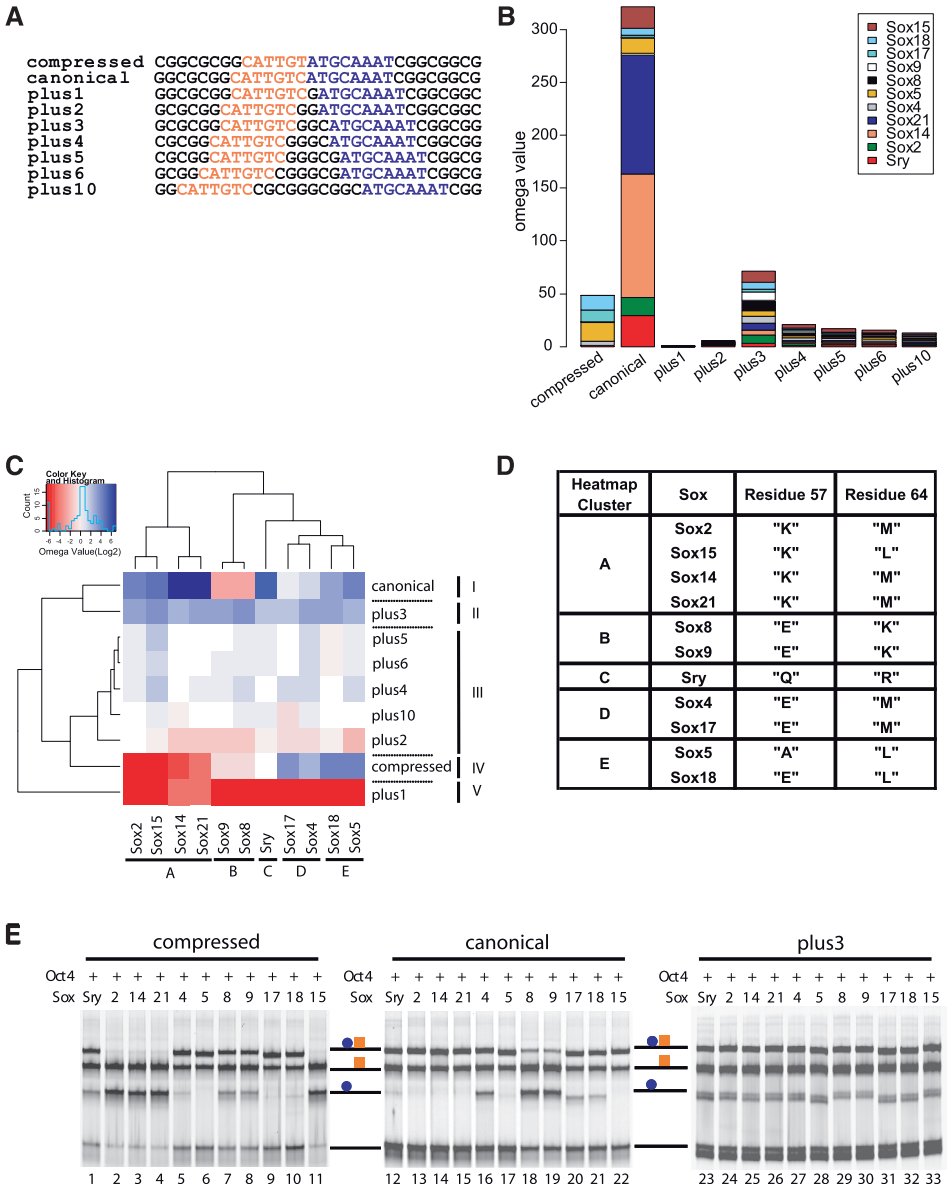
We noticed that the cooperativity-based classification of Sox proteins shows some relationship with their evolutionary classification (Figure 1B). Residue 57, which was predicted to affect the cooperativity of Sox2 and Sox17 with Oct4, provides a partial mechanistic explanation for this result. A lysine (Lys, K) residue at position 57 appears to favour cooperativity on canonical and plus3 configuration but is not compatible with binding to the compressed motif. We have previously shown that residue 57 is a critical determinant of the developmental function of Sox proteins (19). When this residue is swapped between Sox2 and Sox17, that is the *K* is replaced by a glutamate (Glu, E) and vice versa, their biological functions are interchanged and Sox17EK turns into an inducer of pluripotency and Sox2KE into a trigger of endodermal differentiation. To quantify the effect these mutations have on Oct4 cooperation, we compared the cooperativity of the mutant HMG domains with their wild-type counterparts. For these experiments, we used an element derived from the enhancer of the *Nanog* gene that behaves similarly as the idealized sox-oct element in cooperativity measurements (21). We found that the cooperativity of the Sox17EK protein with Oct4 is roughly 30 times stronger than that of wild-type Sox17 and even three times stronger than of wild-type Sox2 (Figure 3B and C).

Sox5 contains an alanine (Ala, A) at position 57 and strongly cooperates on both the canonical and compressed motifs. Given the pronounced effect of residue 57 on Sox-Oct cooperativity, we asked whether the presence of an A residue at position 57 in Sox5 might perhaps explain its ability to bind Oct4 cooperatively on both the canonical and compressed motifs. Indeed, we found that the Sox17EA mutation raises the cooperativity factor of Sox17 on the canonical element 10 times more and brings it up on par with Sox2 and Sox5 in binding the canonical sequence (Figure 3B and C).

Next, we asked whether these amino acid swap mutations also interchange the dimerization propensities on the compressed motif. As expected, the Sox17EK mutation causes a 30-fold drop of cooperation compared to wild-type Sox17 and now behaves like wild-type Sox2. However, Sox2KE cooperates only marginally better than wild-type Sox2 on this element indicating that further modifications in Sox2 are required to engineer a Sox17-like dimerization propensity on the compressed element. Further, introducing the Sox5-like alanine into Sox17 to generate the Sox17EA protein results in a 20-fold drop in the cooperativity although Sox5 cooperates strongly. We noted that Sox5 contains a glutamine (Gln, Q) at position 56, which is unique for the SoxD group (Figure 1A). All other Sox proteins contain an alanine at this position. It is conceivable that Gln56 impacts the cooperation of Sox5 on the compressed element by compensating for Glu57. The lack of both, Gln56 and Glu57, could explain why Sox17EA and Sox2KA cannot cooperate on the compressed element.

We were intrigued by the observation that Sox17EK cooperates more strongly with Oct4 on the canonical element than wild-type Sox2 (Figure 3B and C). We therefore decided to further explore this by allowing two different Sox proteins to compete for co-binding with Oct4 in the same reaction tube (Figure 3D). Consistently, Sox2-Oct4 complexes predominated when mixed with Sox17, whereas the situation was inversed in the presence of Sox17EK (Figure 3D). Interestingly, these results correlated with our earlier observation that Sox17EK produces induced pluripotent stem cells more efficiently than Sox2 in reprogramming cocktails (19). The enhanced ability of Sox17EK to cooperate with Oct4 on pluripotency enhancers may thus be the basis for this observation.

Figure 1. Continued
Electrostatic surface maps of representing Sox members were calculated as described (26). Positively and negatively charged regions were represented in red and blue patches, respectively. Homology models for Sox HMGs were generated using I-TASSER (28) and surface patches that differ for Sox groups are boxed. (**C**) Illustration of how the microstates of the DNA complexes were quantified using the ImageQuant TL software. The cy5-labeled dsDNA migrated differently on native gel depending on how the proteins and DNA associate. Thus, the fractional contribution of the microstates of the free DNA (f_0), Sox-DNA (f_1), Oct4-DNA (f_2) and ternary complex (f_3) can be quantified. (**D**) Schematic diagram highlighting the approach to calculate the cooperativity of TF pairs on composite DNA elements. Boltzmann weights of the respective complexes are denoted as b_D, b_DP1, b_DP2 and b_DP1P2 and scaled so that the b_D = 1. [P1] and [P2] are the concentrations of the free proteins. The cooperativity factor omega does not depend on the concentration of the reactants but solely on the relative ratios of the four microstates represented by their fractional contributions measured in (C) (see main text and alternate derivation of the equation in the 'Materials and Methods' section).



Heatmap Cluster	Sox	Residue 57	Residue 64
A	Sox2	"K"	"M"
	Sox15	"K"	"L"
	Sox14	"K"	"M"
	Sox21	"K"	"M"
B	Sox8	"E"	"K"
	Sox9	"E"	"K"
C	Sry	"Q"	"R"
D	Sox4	"E"	"M"
	Sox17	"E"	"M"
E	Sox5	"A"	"L"
	Sox18	"E"	"L"

Figure 2. (**A**) Sequences of the idealized composite Sox-Oct-labeled probes used. The Sox-binding sites are indicated in orange while the Oct-binding sites are indicated in blue; (**B**) Bar plots showing cumulative mean cooperativity factors for 11 Sox HMG domains for elements shown in (A). Raw values and individual bar plots per element are shown in Supplementary Table S1 and Figure S1. To derive reliable omega values and to minimize errors in band quantification, the concentration of Sox HMG and the Oct4 POU was adjusted, such that the fractional contribution of each of the four microstates was at least 5%. If such conditions could not be established, that is, for maximally competitive binding excluding ternary complexes as seen on the plus1 element for most Sox HMGs or Sox2-Oct4 pairing on the compressed element, omega values were set to 0.01. Constitutive cooperativity was not observed in this study. (**C**) Heat map of cooperativity factors representing the different Sox-Oct4 dimers on the various DNA motifs. Log2-transformed mean cooperativity factors are expressed in a three-color gradient: red (competitive), white (additive binding) and blue (positive cooperativity). The matrix was hierarchically clustered using the heatmap.2 function in R with default parameters. Different categorizations were labeled as Clusters A–E and I–V. Each cooperativity factor was derived from at least 3 and maximally 30 replicates (see Supplementary Table S1). (**D**) Summary of the differential assembly dataset grouping Sox HMG domains exhibiting similar Oct4 cooperativity profiles. Candidate amino acids that likely explain the disparate Oct4 interactions at positions 57 and 64 are shown. (**E**) Differential assemblies of different Sox HMG members (50 nM) with the Oct4 POU protein (150 nM) were performed on compressed (left), canonical (center) and plus3 (right) element DNA. The cartoon to the left symbolizes free DNA (black line), Sox (blue circles) and Oct (orange squares).

Structural determinants for Sox -Oct cooperation

Our findings show that residue 57 is critically important for the discrimination between the canonical and compressed motifs. To study the structural basis for the differential assembly of Sox HMGs, we generated homology models of several Sox HMG/Oct4 POU complexes on the canonical element (Supplementary Figure S2). We observed that K57 of Sox2 interacts with a backbone carbonyl of the POU specific domain (Supplementary Figure S2). When K57 is replaced by

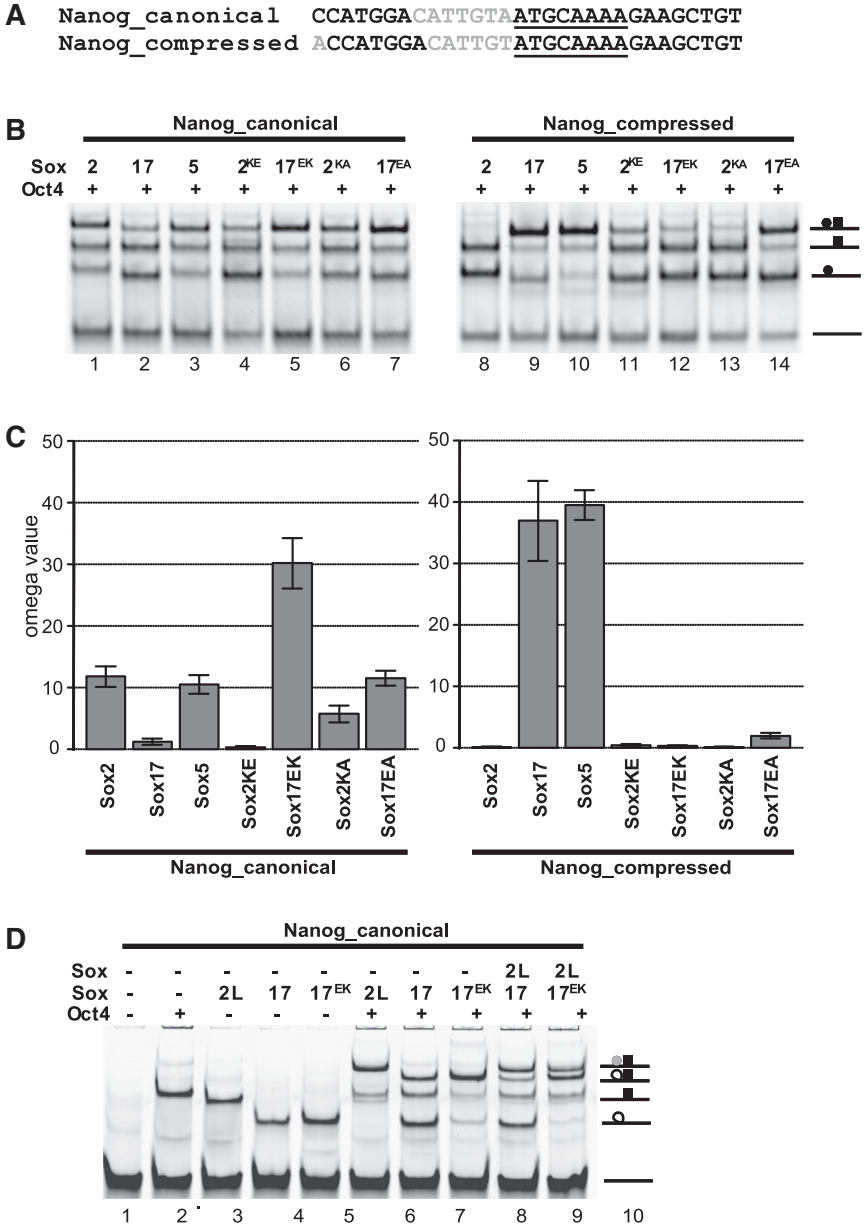


Figure 3. (**A**) Sequences of the labeled Nanog element probes used (21). The Sox-binding sites are indicated in grey while the Oct-binding sites are indicated as underlined; (**B**) Representative EMSAs of different Sox proteins with Oct4 on canonical and compressed motif. The indicate mutants refer to amino acid position 57 of the HMG domain. (**C**) Cooperativity factors for various Sox mutants compared to their wild-type counterparts expressed as mean ± standard deviation. (**D**) Competitive EMSA analysis of showing that the Sox17EK-Oct4 complexes predominate Sox2-Oct4 complexes, whereas the Sox2-Oct4 complex clearly outcompetes Sox17-Oct4 (lanes 9 and 10). A N-and C-terminally extended Sox2 HMG domain (2L) comprising 109 amino acids (residues 33–141 of full length Sox2 protein) was used to distinguish the various complexes. The cartoon to the right symbolizes free DNA (black line), Sox2L (grey-filled circles), Sox17 and Sox17EK (grey empty circles) and Oct4 (black squares).

E57, as in Sox17, Sox8 and Sox9, the negatively charged carboxyl group of Glu likely causes unfavourable charge–charge repulsions, leading to a drop in cooperativity. In contrast, an A57, as found in Sox5 or in Sox17EA, is compatible with binding.

However, residue 57 alone cannot explain the cooperativity profiles of the whole Sox family and there must be a combination of contributing elements. For example, while the E57 proteins Sox17, Sox18 and Sox4 cooperate strongly on the compressed element, Sox8 and Sox9 cannot, and Sox18 retains some cooperativity on the canonical element. The structural modeling suggested residue 64 as an additional candidate underlying the differential behaviour of Sox factors in our cooperativity assays (Supplementary Figure S2). Residue 64 is placed at the interaction interface and shows a strong divergence within the Sox family (Figure 2D, Supplementary Figure S2). Importantly, a M64E mutation has been demonstrated to abrogate Sox2–Oct4 interaction (30). Likewise, the charged K64 in Sox8 and Sox9 could

underlie their inability to interact with Oct4 on canonical and compressed elements.

While exhibiting overall similar cooperativity patterns, the degree of cooperativity still differs for Sox HMG domains with identical residues at positions 57 and 64 (i.e. Sox14/Sox21 cooperate ~5 times more strongly than Sox2 on the canonical element). We hypothesize that those residual difference are due to variations within the flexible and only poorly conserved C-terminal tail of the HMG domain that was shown to contribute to Sox2–Oct4 interactions (30) (Figure 1A). Experimental structures of Sox-HMG-Oct4 combinations on canonical and compressed elements combined with mutagenesis experiments are desirable to put those hypotheses to a test.

CONCLUSION

In this study, we have developed a biochemical system that enables quantitative measurements for protein heterodimerization on DNA and illustrated how such measurements can dissect protein partnerships of a whole protein family. This approach delivers a thermodynamic constant, the cooperativity factor, which allows discriminating between competitive, additive and cooperative interaction modes of TF proteins. Our results suggest that the protein interaction region of individual Sox HMG domains encodes features enabling them to target specific enhancers by teaming up with partner factors. By contrast, the actual DNA recognition interface is very similar and seems unable to explain the functional uniqueness of individual family members (2,26,30,33). A limitation of the approach presented here is its limited throughput. However, high throughput methods such as protein-binding microarrays (34) and HT-SELEX (3) have not yet been adapted to identify composite DNA motifs of heterodimers. Even if those methods can be adjusted to multi-component systems, the development of computational tools that can accurately model cooperativity will pose a significant challenge. Thus, we expect that our method will complement high throughput efforts by validating composite motifs and by providing quantitative estimates of the physical cooperativity in TF-DNA binding.

For Sox-Oct partnerships and probably many other TF pairs, direct cooperativity is likely a major determinant for the recruitment to functional-binding sites, and therefore a major determinant of cell-type-specific biological function (14,35). Thus, interrogating TF heterodimerization will allow inferring coding principles for developmental enhancers and molecular mechanisms for selective and combinatorial enhancer recognition by TF proteins. We have previously used such insights to re-engineer the endoderm differentiation factor Sox17 into an inducer of pluripotency that speeds up stem-cell production (19). The proof-of-concept that TFs can be optimized by tweaking their heterodimerization and that their function can be rationally altered has broad application in stem cell biology and tissue engineering for regenerative medicine.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1A–1C, Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

The authors are grateful to Andrew Hutchins for critical reading of the manuscript and suggestions and to Siew Hua Choo for technical support. Author contributions: C.K.L.N. and R.J.: conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; S.P.: derivation of cooperativity formula, data analysis and interpretation, manuscript writing; N.X.L., S.C.: collection of data; P.K.: financial support, administrative support, final approval of manuscript.

FUNDING

Funding for open access charge: Agency for Science, Technology and Research (A*STAR) Singapore.

Conflict of interest statement. None declared.

REFERENCES

1. Bryne,J.C., Valen,E., Tang,M.H., Marstrand,T., Winther,O., da Piedade,I., Krogh,A., Lenhard,B. and Sandelin,A. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.*, **36**, D102–D106.
2. Badis,G., Berger,M.F., Philippakis,A.A., Talukder,S., Gehrke,A.R., Jaeger,S.A., Chan,E.T., Metzler,G., Vedenko,A., Chen,X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
3. Jolma,A., Kivioja,T., Toivonen,J., Cheng,L., Wei,G., Enge,M., Taipale,M., Vaquerizas,J.M., Yan,J., Sillanpää,M.J. *et al.* (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.*, **20**, 861–873.
4. Biggin,M.D. (2011) Animal transcription networks as highly connected, quantitative continua. *Dev. Cell*, **21**, 611–626.
5. Kadonaga,J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, **116**, 247–257.
6. Mirny,L.A. (2011) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. USA*, **107**, 22534–22539.
7. Glass,C.K. (1994) Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr. Rev.*, **15**, 391–407.
8. Grove,C.A., De Masi,F., Barrasa,M.I., Newburger,D.E., Alkema,M.J., Bulyk,M.L. and Walhout,A.J. (2009) A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. *Cell*, **138**, 314–327.
9. Hai,T.W., Liu,F., Coukos,W.J. and Green,M.R. (1989) Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. *Genes Dev.*, **3**, 2083–2090.
10. Garvie,C.W., Hagman,J. and Wolberger,C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol. Cell*, **8**, 1267–1276.
11. Hollenhorst,P.C., Chandler,K.J., Poulsen,R.L., Johnson,W.E., Speck,N.A. and Graves,B.J. (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet.*, **5**, e1000778.

12. Kamachi,Y., Uchikawa,M., Tanouchi,A., Sekido,R. and Kondoh,H. (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev.*, **15**, 1272–1286.
13. Tanaka,S., Kamachi,Y., Tanouchi,A., Hamada,H., Jing,N. and Kondoh,H. (2004) Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Mol. Cell. Biol.*, **24**, 8834–8846.
14. Wilson,M. and Koopman,P. (2002) Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr. Opin. Genet. Dev.*, **12**, 441–446.
15. Kondoh,H. and Kamachi,Y. (2010) SOX-partner code for cell specification: regulatory target selection and underlying molecular mechanisms. *Int. J. Biochem. Cell Biol.*, **42**, 391–399.
16. Nasrin,N., Buggs,C., Kong,X.F., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1991) DNA-binding properties of the product of the testis-determining gene and a related protein. *Nature*, **354**, 317–320.
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. Ryan,A.K. and Rosenfeld,M.G. (1997) POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.*, **11**, 1207–1225.
19. Jauch,R., Aksoy,I., Hutchins,A.P., Ng,C.K., Tian,X.F., Chen,J., Palasingam,P., Robson,P., Stanton,L.W. and Kolatkar,P.R. (2011) Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells*, **29**, 940–951.
20. 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.
21. Rodda,D.J., Chew,J.L., Lim,L.H., Loh,Y.H., Wang,B., Ng,H.H. and Robson,P. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.*, **280**, 24731–24737.
22. Nishimoto,M., Fukushima,A., Okuda,A. and Muramatsu,M. (1999) 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**, 5453–5465.
23. Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
24. Stefanovic,S., Abboud,N., Désilets,S., Nury,D., Cowan,C. and Puceat,M. (2009) Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J. Cell Biol.*, **186**, 665–673.
25. Ng,C.K., Palasingam,P., Venkatachalam,R., Baburajendran,N., Cheng,J., Jauch,R. and Kolatkar,P.R. (2008) Purification, crystallization and preliminary X-ray diffraction analysis of the HMG domain of Sox17 in complex with DNA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **64**, 1184–1187.
26. Palasingam,P., Jauch,R., Ng,C.K. and Kolatkar,P.R. (2009) The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J. Mol. Biol.*, **388**, 619–630.
27. BabuRajendran,N., Palasingam,P., Narasimhan,K., Sun,W., Prabhakar,S., Jauch,R. and Kolatkar,P.R. (2010) Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-{beta} effectors. *Nucleic Acids Res.*, **38**, 3477–3488.
28. Zhang,Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.
29. Bowles,J., Schepers,G. and Koopman,P. (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.*, **227**, 239–255.
30. Reményi,A., Lins,K., Nissen,L.J., Reinbold,R., Schöler,H.R. and Wilmanns,M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.*, **17**, 2048–2059.
31. Chen,X., Xu,H., Yuan,P., Fang,F., Huss,M., Vega,V.B., Wong,E., Orlov,Y.L., Zhang,W., Jiang,J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
32. Yuan,H., 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.
33. Jauch,R., Narasimhan,K., Ng,C.K. and Kolatkar,P.R. (2011) Crystal structure of the Sox4 HMG/DNA complex suggests a mechanism for the positional interdependence in DNA recognition. *Biochem. J.*, December 19 (doi:10.1042/BJ20111768; epub ahead of print).
34. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Peña-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
35. Kamachi,Y., Uchikawa,M. and Kondoh,H. (2000) Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet.*, **16**, 182–187.