Structural Basis for the SOX-Dependent Genomic Redistribution of OCT4 in Stem Cell Differentiation

Felipe Merino,[1,2] Calista Keow Leng Ng,[3] Veeramohan Veerapandian,[4] Hans Robert Schöler,[5,6] Ralf Jauch,[3,4,*] and Vlad Cojocaru[1,2,*]

[1]Computational Structural Biology Group, Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Röntgenstrasse 20, 48149 Münster, Germany
[2]Center for Multiscale Theory and Computation, Westfälische Wilhelms University, Corrensstrasse 40, 48149 Münster, Germany
[3]Laboratory for Structural Biochemistry, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore
[4]Genome Regulation Laboratory, Guangzhou Institutes for Biomedicine and Health, Chinese Academy of Sciences, 190 Kai Yuan Avenue, Science Park, 510530 Guangzhou, China
[5]Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Röntgenstrasse 20, 48149 Münster, Germany
[6]Medical Faculty, Westfälische Wilhelms University, Domagstrasse 3, 48149 Münster, Germany
*Correspondence: ralf@gibh.ac.cn (R.J.), vlad.cojocaru@mpi-muenster.mpg.de (V.C.)
http://dx.doi.org/10.1016/j.str.2014.06.014

SUMMARY

In pluripotent cells, OCT4 associates with SOX2 to maintain pluripotency or with SOX17 to induce primitive endoderm commitment. The OCT4-SOX2 and OCT4-SOX17 combinations bind mutually exclusive to two distinct composite DNA elements, known as the "canonical" and "compressed" motifs, respectively. The structural basis for the OCT4-SOX17 cooperativity is unknown. Whereas SOX17 has been engineered to replace SOX2 in the pluripotency circuitry, all generated SOX2 mutants have failed to act like SOX17. From molecular simulations, we revealed the OCT4-SOX17 interaction interface and elucidated the SOX-dependent motif preference of OCT4. Moreover, we designed a SOX2 mutant that we predicted and confirmed experimentally to bind cooperatively with OCT4 to the compressed motif. Ultimately, we found a strong correlation between the experimental and calculated relative cooperative-binding free energies of 12 OCT4-SOX-DNA complexes. Therefore, we validated the OCT4-SOX interfaces and demonstrated that in silico design of DNA-binding cooperativity is suitable for altering transcriptional circuitries.

INTRODUCTION

In eukaryotes, a limited number of transcription factors (Vaquerizas et al., 2009) achieve a fine spatio-temporal control of gene expression through a combinatorial approach in which only certain combinations will regulate specific sets of target genes (Biggin, 2011; Reményi et al., 2004). The biophysical basis for the recruitment of defined groups of transcription factors to specific locations in the genome is not completely understood.

Several families of transcription factors are known to control fundamental biological processes through direct interaction between their members (Reményi et al., 2004; Sarkar and Hochedlinger, 2013). For example, POU and SOX proteins coregulate gene transcription at different stages of development (Kondoh and Kamachi, 2010; Tantin, 2013). POU proteins have a bipartite DNA-binding domain with two helix-turn-helix subdomains, a POU-specific domain (POU$_S$), and a homeodomain (POU$_{HD}$), connected by a variable linker region (Phillips and Luisi, 2000). The globular regions of the POU$_S$ and POU$_{HD}$ bind to the major groove of the DNA, while the N-terminal tail of POU$_{HD}$ binds to the minor groove. The SOX proteins have a "high-mobility group" (HMG) domain, which strongly bends the DNA by binding to the minor groove (Werner et al., 1995).

From these two families, OCT4 (POU5F1) (Jerabek et al., 2014) and SOX2 lie in the core of the transcriptional network that controls stem cell pluripotency (Boyer et al., 2005) and are key to induce pluripotency in somatic cells (Takahashi and Yamanaka, 2006). OCT4 binds to an octamer site with the consensus sequence ATGC(A/T)AAT, whereas SOX2 recognizes a sequence related to C(T/A)TTGTT. Predominantly, they bind cooperatively to a composite motif formed by the juxtaposition of their individual binding sites (Chen et al., 2008), known as the "canonical" motif. This motif is found in the regulatory regions of key pluripotency genes, including *POU5F1*, *NANOG*, *UTF1*, and others (Aksoy et al., 2013a; Chen et al., 2008; Nishimoto et al., 1999; Rodda et al., 2005). Although OCT4 and SOX2 are able to bind to these sites separately, in many instances only their association activates transcription efficiently (Ambrosetti et al., 2000; Nishimoto et al., 1999). The direct interaction between OCT4 and SOX2 is DNA dependent (Lam et al., 2012) and involves the POU$_S$ helix α_1 and the HMG helix α_3.

In mouse embryos, OCT4 is required for the commitment of the pluripotent cells from the inner cell mass to the primitive endoderm lineage (Frum et al., 2013; Le Bin et al., 2014; Wu and Scholer, 2014). Based on experiments in cultured embryonic

stem cell (ESC) lines, it was proposed that this process involves a genomic redistribution of OCT4 to bind cooperatively with SOX17 to an alternative composite motif, known as the "compressed" motif, in which one base pair is lacking between the individual binding sites compared to the canonical motif (Aksoy et al., 2013a; Jauch et al., 2011). Importantly, OCT4 and SOX2 do not bind together to the compressed motif, whereas OCT4 and SOX17 bind only additively to the canonical motif (Ng et al., 2012). This discriminatory recognition is likely due to differences in the POU-HMG interaction interfaces, which depend on the motif configuration.

Whereas the POU-HMG interface was resolved for the complex between the OCT4 homolog OCT1 and SOX2 on the *HOXB1* enhancer, which bears a canonical motif (Williams et al., 2004), the structure of the OCT4-SOX17 complex on the compressed motif is not known. Based on the alignment of SOX sequences and the available structural data, the HMG residue 57 was proposed to be responsible for the motif preference of the OCT4-SOX2 and OCT4-SOX17 complexes. Indeed, the SOX17^{E57K} mutant is able to replace SOX2 in the pluripotency circuitry and in somatic cell reprogramming (Aksoy et al., 2013b; Jauch et al., 2011). However, although the overexpression of SOX2^{K57E} led to the upregulation of some primitive endoderm genes (Jauch et al., 2011), the genomic distribution of SOX2^{K57E} was different compared to SOX17 (Aksoy et al., 2013a). Moreover, SOX2^{K57E} does not cooperate with OCT4 to bind the compressed motif (Ng et al., 2012). Therefore, to understand the functions of OCT4-SOX complexes, it is necessary to reveal the structural basis for their motif preference.

Here we modeled six wild-type and mutant OCT4-SOX2 and OCT4-SOX17 complexes bound to the canonical and compressed motifs and performed molecular simulations combined with free energy calculations to explore the structural determinants of the SOX-dependent OCT4-DNA recognition. More generally, we aimed to probe for the suitability of this approach to study cooperative DNA binding by transcription factors in the delicate case in which this process involves small but biologically relevant free energy changes. To validate our predictions, we used quantitative electrophoretic mobility shift assays. From the simulations, we found that I21 from OCT4 forms a hydrophobic interaction either with the SOX residue 64 or 53 upon binding to the canonical or compressed motifs, respectively. We also found that E57 and R60 from SOX17 together with K17 from OCT4 form a network of interactions that stabilizes the OCT4-SOX17-compressed complex, which explains the crucial role of residue 57. Interestingly, L46 from SOX17 docks in a hydrophobic pocket of OCT4 upon binding to the compressed motif. In contrast, the acidic E46 in SOX2 is unable to support this interaction. Therefore, we designed the SOX2$^{E46L/K57E}$ mutant, which we predicted and confirmed experimentally to bind cooperatively with OCT4 to the compressed motif. Furthermore, we found a strong correlation between the calculated and experimentally determined relative cooperative binding free energies. Thus, we validated our proposed structural basis for the OCT4-SOX cooperativities. These findings shed light on the SOX-dependent genomic redistribution of OCT4 during primitive endoderm commitment of pluripotent cells and demonstrate that in silico design of cooperative DNA binding can be employed to rationally alter transcriptional circuitries.

RESULTS AND DISCUSSION

Cooperative DNA-Binding Properties Are Mirrored in the Genomic Distribution of SOX Proteins

To test whether the DNA-binding properties of the isolated proteins reflect their genomic distribution, we reanalyzed the previously determined genome-wide binding profiles of exogeneoulsy expressed SOX proteins in mouse ESC lines. To obtain an unbiased view of how SOX proteins select their genomic targets, we now avoided the previously applied filter of mandatory cobinding with OCT4. Rather, we probed for the DNA motif preferences within the 500 most strongly bound sites by SOX proteins independent of the presence of OCT4 (Figure 1). Remarkably, the canonical motif was found in 47% and 81% of the sites occupied by SOX2 and SOX17^{E57K}, respectively (Figures 1A and 1B), suggesting cobinding with OCT4 on these sites. The higher percentage in the case of SOX17^{E57K} reflects its ability to outperform SOX2 in maintaining and inducing pluripotency and is in agreement with its higher binding cooperativity with OCT4 measured with purified proteins (Ng et al., 2012). Furthermore, while 75% of the SOX17-bound sites bear the compressed motif, only 24.6% of the SOX2^{K57E} -bound sites bear at least one of the two composite motifs (Figures 1C and 1D). Hence, the ability of SOX2^{K57E} to cobind with OCT4 to these motifs is impaired. In agreement with previous findings, SOX17 and SOX2^{K57E} bind to different genomic loci (Figure S1 available online). This is consistent with the observation that purified SOX2^{K57E} is unable to cooperate with OCT4 on the compressed motif. In conclusion, the genomic distributions of SOX proteins reflect their in-vitro-measured DNA-binding cooperativities with OCT4, and the identity of residue 57 is not sufficient to explain the motif preference of OCT4-SOX complexes.

Amino Acid Identities across the SOX Family Suggest Alternative Interfaces with OCT4

The transcriptional regulation involving the direct interaction of POU and SOX factors is evolutionary conserved (Dailey and Basilico, 2001; Leichsenring et al., 2013). To explore whether amino acids important for the OCT4-SOX cooperativity can be deduced from sequence information, we calculated the alignment covariance between the HMG residues of metazoan SOX factors and analyzed those whose identities correlate with residue 57. We consistently identified residues 39, 53, 64, and 66 within the top 100 scoring pairs (Figure 2; Table S1). Interestingly, the hydrophobicity of residues 53 and 64, which do not interact with DNA, is largely conserved. Notably, M64 is important for the OCT4-SOX2 interaction on the canonical motif (Reményi et al., 2003; Williams et al., 2004). Consistently, SOX8 and SOX9 have a lysine at position 64 and do not bind cooperatively with OCT4 to this motif (Ng et al., 2012). On the other hand, residue 53 has not yet been associated with any particular function of SOX proteins. Importantly, the properties of residues 53 and 64 vary significantly only when residue 57 is neither a lysine nor a glutamate. This suggests that they are involved in the OCT4-SOX interaction but not in the selectivity of OCT4-SOX complexes for a certain composite motif. Residue 39 bears a positive charge that interacts with the DNA backbone whenever residue 57 is negatively charged. As K57 from SOX2 also interacts with the DNA in the OCT1-SOX2-canonical complex
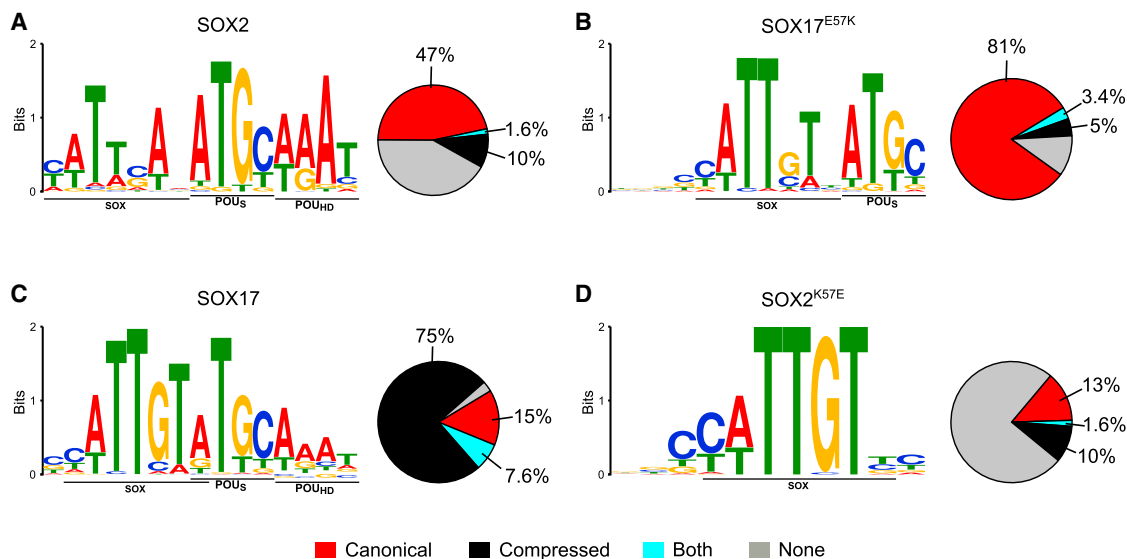
Figure 1. Motifs Discovered within 100 bp Regions Centered on the Top 500 ChiP-Seq Peak Summits from the Genomic Binding Profiles of SOX2, SOX17^E57K, SOX17, and SOX2^K57E
(A) Profile of SOX2.
(B) Profile of SOX17^E57K.
(C) Profile of SOX17.
(D) Profile of SOX2^K57E.
The pie charts illustrate the fraction of composite motifs within the top 500 SOX binding sites. "Both" refers to binding sites that contain canonical as well as compressed motifs. See also Figure S1.

(Williams et al., 2004), these residues may form a functional cluster, similar to those identified in other protein families (Halabi et al., 2009). The differences at position 66 are only related to side chain length. As this residue is oriented away from both the DNA and the POU-HMG interface, it is unlikely to contribute to the OCT4-SOX cooperativity.

Mutually Exclusive Interfaces Determine the OCT4-SOX Cooperative DNA Binding

To understand the structural basis for the OCT4-SOX cooperativity, we modeled the OCT4-SOX2 and OCT4-SOX17 complexes bound to the canonical and compressed motifs and then performed 150-ns-long molecular dynamics simulations (see representative structures in Figure 3 and the Supplemental Information). For detailed characterization (Table 1), we monitored the following: (1) the root-mean-square deviation (rmsd) from the initial model, 2) the DNA bending angle per SOX binding site (θ_{HMG}) and base pair (θ_{bp}), (3) the minor groove width per base pair (w_{bp}), (4) the angle between the POU helix α_1 and the HMG helix α_3 ($\varphi_{\alpha1-\alpha3}$), (5) the orientation of the HMG helix α_3 relative to the octamer site (φ) (Figure S2A), and (6) the OCT4-SOX interaction area (A_{OS}). To characterize the most relevant, long-lived interactions, we calculated contact maps including only the interactions stable for at least 50% of the simulated time (Figures 4A and 4B).

The canonical-bound OCT4-SOX2 and OCT4-SOX17 complexes (Figures 3B and 3C) remained stable during the simulations with an average rmsd of 2.5 ± 0.3 Å and 2.8 ± 0.5 Å, respectively (Figure S2B). Also, the DNA configuration, measured by θ_{HMG}, θ_{bp}, and w_{bp}, was stable and independent

of the SOX factor identity (Table 1; Figure S2B). The hydrogen bonds between the POU$_S$ and DNA bases were less stable in the OCT4-SOX17 complex, suggesting that the OCT4-SOX17 interaction has a slight detrimental effect on the OCT4-DNA interactions. Interestingly, the hydrogen bonds between the POU$_{HD}$ and DNA bases were unstable during the simulations (Figure S3). This agrees with the known ability of the POU$_{HD}$ from OCT4 to bind nonconsensus DNA sequences (Nishimoto et al., 1999), which is key for the biological function of OCT4 (Nishimoto et al., 2005). Moreover, DNA-bound conformations lacking the specific interactions with DNA bases have been observed in the crystal structures (Aishima and Wolberger, 2003) and simulations (Babin et al., 2013) of other homeodomains. The difference between the cooperative binding of OCT4 with SOX2 or SOX17 to the canonical motif can be explained almost entirely based on the differences in the residue-residue interactions. The orientation angle φ of the HMG site relative to the octamer site, the interaction area A_{OS} (Table 1; Figure S2B), and the number of stable contacts (Figures 4A and S4) are comparable between OCT4-SOX2 and OCT4-SOX17. However, in contrast to E57 from SOX17, K57 from SOX2 interacts favorably with the dipole of the POU$_S$ helix α_1 (Figures 3B, 3C, and S5A). This partly explains why the mutant SOX17^E57K gains the ability to bind cooperatively with OCT4 to the canonical motif (Ng et al., 2012). The POU$_S$-HMG orientation angle $\varphi_{\alpha1-\alpha3}$ is lower in the OCT4-SOX2 complex (Table 1), suggesting that the dynamics of the two helices defining the POU-HMG interface may correlate with the identity of residue 57. However, this small difference can also be caused by the incomplete sampling of the conformational space achieved during the simulations. The
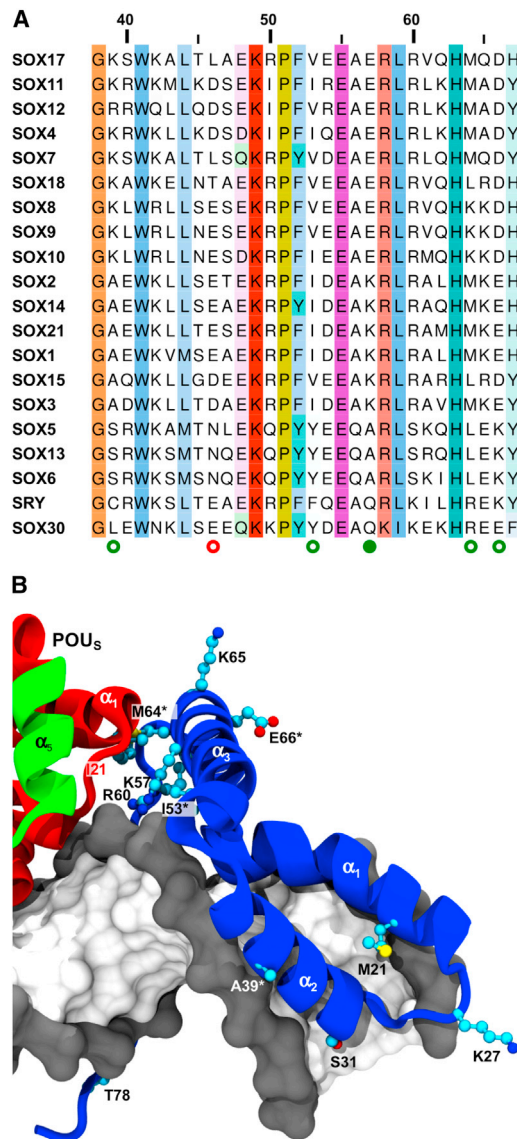
Figure 2. Alignment Covariance Analysis of SOX Factors

(A) Alignment of the α_3 helix sequences from *M. musculus*. The green circles mark the amino acids correlated with position 57 (filled circle). The red circle marks position 46.

(B) Structure of the canonical-bound OCT4 (red)-SOX2 (blue) complex. The OCT4 linker is in green. I21 from OCT4 and the SOX residues with identities correlated to that of residue 57 found by at least three algorithms are shown as balls and sticks. Those found by all five algorithms are marked with a star. See also Table S1 and Document S2.

hydrophobic interactions of I21 from OCT4 with the SOX residues M64 (conserved between SOX2 and SOX17) and 61 (A in SOX2, V in SOX17) (Figures 3B, 3C, and S5A) are stable (Figures 4A and S4), thus contributing to the cooperative binding of OCT4 with both SOX factors to the canonical motif.

In the compressed-bound OCT4-SOX17 and OCT4-SOX2 complexes (Figures 3E and 3F), the DNA conformation is similar (Table 1; Figure S2B). As in the canonical-bound complexes, the interactions between the POU$_{HD}$ and the DNA were not stable

(Figure S3). However, the two complexes adopted substantially different arrangements with rmsd to the initial model of 2.4 ± 0.3 and 3.1 ± 0.3 Å for OCT4-SOX17 and OCT4-SOX2, respectively (Figure S2B). In the OCT4-SOX17 complex (Figure 3E), the interaction area A_{OS} (Table 1) and the number of stable OCT4-SOX contacts (Figures 4B and S4) are larger than in any other wild-type complex. In the OCT4-SOX2 complex (Figure 3F), the values for both measures were significantly reduced. Interestingly, the change in φ between the canonical- and compressed-bound complexes differs from the expected value of 36° (Table 1; Figure S2B) corresponding to the helical twist per one base pair of B-DNA. This suggests that the protein-protein interfaces affect the protein-DNA interactions. The POU$_S$-HMG orientation angle $\varphi_{\alpha 1-\alpha 3}$ is larger in the OCT4-SOX17 complex (Table 1), reflecting a tighter interface between the N terminus of the HMG helix α_3 and the C terminus of the POU helix α_1. R50 from SOX17 caps the POU$_S$ helix α_1, whereas E57 and R60 form a network of salt bridges with K17 from OCT4 (Figure S5B), thus explaining the crucial role of E57 for the OCT4-SOX17 cooperativity on the compressed motif (Jauch et al., 2011). In contrast, K57 from SOX2 contributes to the dismantling of this interface due to its positive charge (Figure 3F). Furthermore, I21 from OCT4 forms a stable hydrophobic interaction with V53 from SOX17 (Figures 3E, 4B, and S5B), which is mutually exclusive to the I21-M64 interaction observed in the canonical-bound complexes (Figure 3B, 4A, and S5A). This agrees with our findings from the amino acid correlation analysis (Figure 2).

To validate our qualitative characterization of the OCT4-SOX interaction interfaces, we estimated the contribution of each SOX residue to the binding free energy of OCT4 (the ligand) to the SOX-DNA complex (the receptor). We used only the simulations of the complexes, thus assuming that the dynamics of the free ligand and receptor are the same as in the complex. Although not valid for estimating absolute binding free energies, this approximation can be employed here as it does not affect the relative per-residue contributions. Importantly, the desolvation terms of the binding free energy (Supplemental Experimental Procedures) are not strictly decomposable (Miller et al., 2012). Thus, we only aim to semiquantitatively describe the differences in the contributions of individual SOX residues. On both motifs, the interaction between the HMG helix α_3 and the POU$_S$ contributes the most to the OCT4 binding affinity (Figures 4C and S5C). On the canonical motif, K57 from SOX2 has a substantially larger favorable contribution than E57 from SOX17. Other residues from SOX2 and SOX17 have similar contributions between both SOX proteins. Residues 61 and 64 contribute favorably, whereas R60 unfavorably due to its proximity to R20 from OCT4. In addition, residues in the loop following the helix α_3 contribute favorably mainly due to transient interactions involving the SOX residue K71, the OCT4 residue K17, and the main chains. In contrast, there is a clear difference in the contribution of SOX2 and SOX17 residues to the binding affinity of OCT4 to the compressed motif. While the SOX2 residues do not contribute significantly, L46, R50, V53, E54, R60, and E57 from SOX17 contribute favorably, whereas K57 from SOX2 contributes unfavorably. These results validate the qualitative characterization of the POU-HMG interfaces and are consistent with experiments (Aksoy et al., 2013a; Jauch et al., 2011).
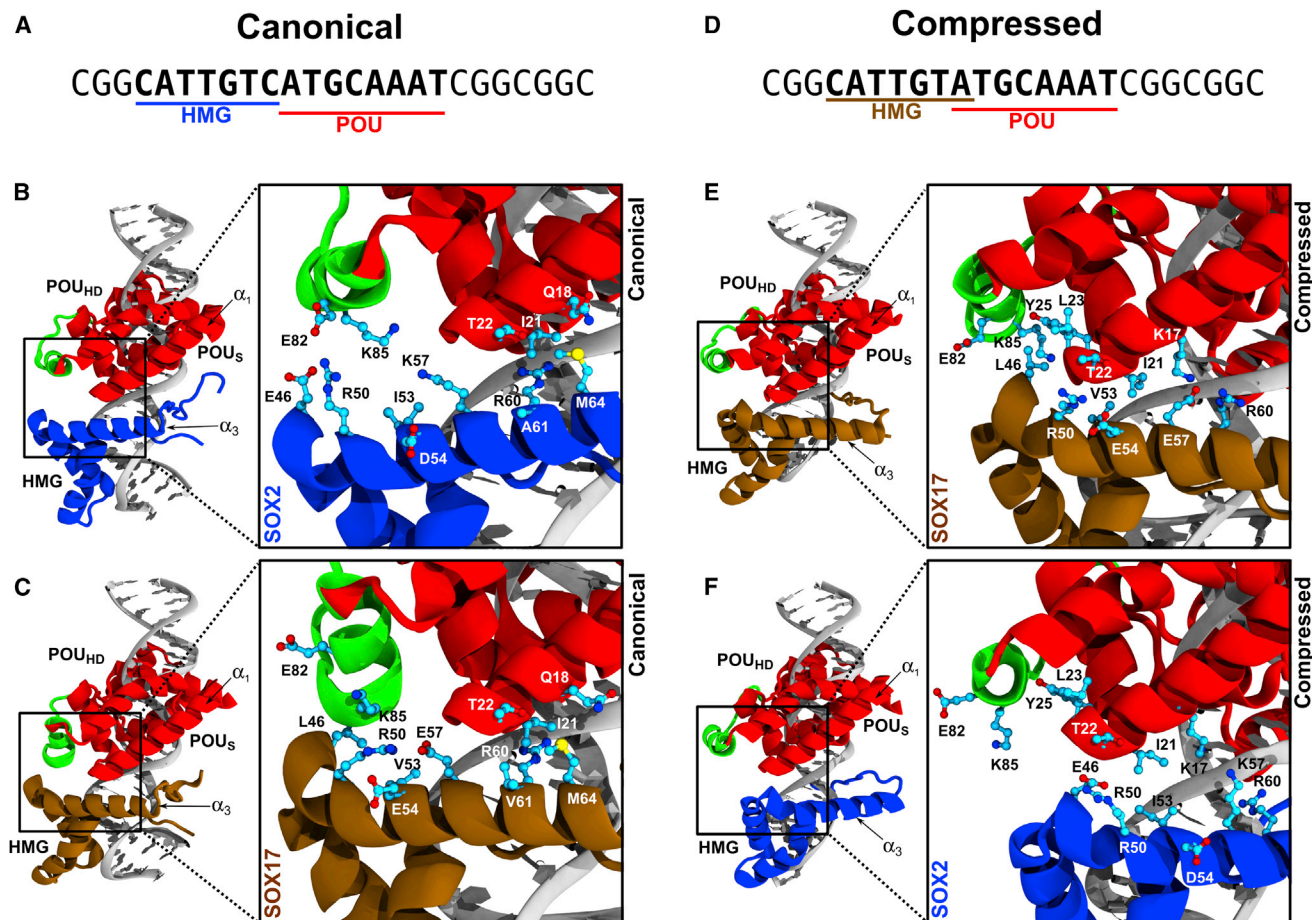
Figure 3. Representative Structures of OCT4-SOX-DNA Complexes
(A) Canonical motif sequence.
(B and C) Canonical-bound OCT4-SOX2 (B) and OCT4-SOX17 (C).
(D) Compressed motif sequence.
(E and F) Compressed-bound OCT4-SOX17 (E) and OCT4-SOX2 (F).
Overall topologies and zoomed-in illustrations of the POU-HMG interfaces are shown on the left and right sides of each panel, respectively. OCT4 and SOX2 are as in Figure 2, SOX17 is in ochre, and DNA is in gray. See also Figure S4 and 3D Molecular Models S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, and S12.

An Additional Hydrophobic Interface Is Formed in the OCT4-SOX17-Compressed Complex

The OCT4-SOX17-compressed complex is further stabilized by the docking of L46 from SOX17 into a hydrophobic pocket of OCT4 (Figure 3E), where it forms a stable interaction with Y25 from OCT4 (Figures 4B, S4, and S5B). L46 is part of the first turn of helix α_3, which docks almost perpendicular to the POU$_S$ helix α_1. The difference from the orientation of the rest of α_3 ($\varphi_{\alpha1-\alpha3}$ of ~45°; Table 1; Figure S2B) is due to a kink in α_3 induced by the highly conserved P51. Moreover, L46 contributes significantly to the OCT4-SOX17 cooperativity on the compressed motif (Figure 4C). In contrast, this interface is not supported by E46 from SOX2 (Figure 3F) and SOX2^{K57E}. Hence, we propose that due to E46, SOX2^{K57E} does not bind cooperatively with OCT4 to the compressed motif. Interestingly, the SOX proteins that cooperate with OCT4 on this motif (Ng et al., 2012) bear a neutral residue at this position (Figure 2).

Notably, L46 from SOX17 also increases the binding affinity of OCT4 for the canonical motif (Figures 4C and S5C). In contrast, E46 from SOX2 has only a marginal favorable effect mainly because it electrostatically clashes with E78 and E82 from the short helix of the OCT4 linker (Esch et al., 2013), which destabilizes the interaction network in which K85 and E82 from the OCT4 linker form salt bridges with D29 from the POU$_S$ and R50 from the HMG, respectively (Figures 3C and S5A). In the OCT4-SOX17-canonical complex, this network is destabilized by the presence of E57 in SOX17 (Figure 3D). Thus, we propose that the OCT4-SOX interaction interface on the canonical motif can be further optimized by the presence of a hydrophobic side chain at position 46. This proposal is supported by the observation that SOX17^{E57K}, which bears a leucine at position 46, cooperates with OCT4 stronger than SOX2 to bind the canonical motif and is capable of maintaining and inducing pluripotency more efficiently than SOX2 (Aksoy et al., 2013b; Jauch et al., 2011).

Table 1. Summary for the Simulations of the Wild-Type OCT4-SOX-DNA Complexes

	Canonical Motif[a]		Compressed Motif[a]	
SOX factor	SOX2	SOX17	SOX2	SOX17
Simulation time (ns)	150	150	150	150
Residue 46	Glu	Leu	Glu	Leu
Residue 57	Lys	Glu	Lys	Glu
θ_{HMG} (°)[b]	41.8 ± 8.3	36.9 ± 5.2	35.7 ± 6.5	40.0 ± 4.9
$\varphi_{\alpha1-\alpha3}$ (°)[c]	37.8 ± 4.3	46.0 ± 4.8	39.2 ± 4.9	43.9 ± 4.0
φ (°)[d]	133.9 ± 3.3	133.7 ± 3.6	92.3 ± 3.5	101.4 ± 2.2
A_{OS} (Å²)[e]	369.9 ± 61.3	413.8 ± 61.2	164.0 ± 40.9	436.2 ± 62.1

See also Figures S2 and S3.
[a]Structural properties are shown as averages ± standard deviation.
[b]θ_{HMG} is the DNA bending per HMG binding site.
[c]$\varphi_{\alpha1-\alpha3}$ is the angle between helix α_1 from OCT4 and helix α_3 from the SOX factor.
[d]φ is the orientation of helix α_3 of the HMG relative to the helical axis of the POU binding site.
[e]A_{OS} is the OCT4-SOX interaction area.

The Double Mutant SOX2[E46L/K57E] and OCT4 Bind Cooperatively to the Compressed Motif

The results from our simulations suggest that the combined identities of residues 46 and 57 are sufficient to explain and predict the SOX-dependent composite motif preference of OCT4. Therefore, SOX2[E46L] and SOX2[E46L/K57E] should show similar binding cooperativities with OCT4 as SOX17[E57K] and SOX17, respectively. To test this hypothesis, we measured the cooperativity factors using a quantitative electrophoretic mobility shift assay (Figure 5). Briefly, we estimated the cooperativity factor $\omega = (f_D \times f_{DOS})/(f_{DO} \times f_{DS})$ from the fractional amount of free DNA (f_D), OCT4-bound DNA (f_{DO}), SOX-bound DNA (f_{DS}), and OCT4-SOX-bound DNA (f_{DOS}). Here ω represents the ratio between the association constant of the OCT4-DNA complex in the presence and absence of the SOX protein (Ng et al., 2012). Then, ω bigger, equal, or smaller than one represent positive cooperative binding, additive binding, or negative cooperative binding, respectively. To avoid errors due to the background noise, we only determined ω if each band contributed at least 1% of the total fluorescence per lane. From these experiments, we could classify the SOX factors in three categories according to their motif preference (Figures 5 and S6): (1) SOX2, SOX2[E46L], and SOX17[E57K] prefer the canonical over the compressed motif, (2) SOX2[K57E] binds additively with OCT4 to both motifs with a slight preference for the canonical motif, and (3) SOX2[E46L/K57E] and SOX17 prefer the compressed over the canonical motif. As we predicted, SOX2[E46L] cooperates stronger than SOX2 with OCT4 on the canonical motif. Moreover, SOX2[E46L/K57E] is the first SOX2 mutant that binds cooperatively with OCT4 to the compressed motif. For further validation, we also designed the triple mutant SOX2[E46L/K57E/K65Q]. From the simulations, we found that Q65 does not contribute to the OCT4-SOX17 cooperativity on the compressed motif (Figure 4C). Therefore, we predicted and confirmed experimentally that this mutant acts like SOX2[E46L/K57E] (Figure S6).

Prediction and Design of the OCT4-SOX Cooperative DNA Binding from Molecular Simulations

To investigate whether the OCT4-SOX cooperative DNA binding can be predicted quantitatively, we simulated four mutant complexes, OCT4-SOX17[E57K], OCT4-SOX2[K57E], OCT4-SOX2[E46L], and OCT4-SOX2[E46L/K57E], bound to the canonical and compressed motifs (Table 2) and estimated the cooperative binding free energy relative to the wild-type SOX2 and SOX17, respectively. A similar approach has been applied to zinc-finger-containing transcription factors to estimate absolute binding cooperativities (Lee et al., 2010). However, OCT4-SOX complexes have significantly more degrees of freedom and pose a particular challenge due to their small cooperative binding free energies. The calculations are challenging as they depend on numerous parameters, and results in good agreement with experiments may be a consequence of fortuitous error cancellation. Therefore, we investigated 12 wild-type and mutant complexes not only for direct comparison with experiments but also to minimize the risk of false positive results.

During the simulations, all mutant complexes adopted stable conformations with similar DNA structure (Table 2; Figures S2 and S3). As in the simulations of the wild-type complexes, the interactions of the POU$_{HD}$ with the DNA bases were not stable (Figure S3). In few simulations these interactions reformed, suggesting that this is a reversible process on longer timescales.

On the canonical motif, the interactions between the POU$_S$ and DNA bases are destabilized in the OCT4-SOX2[E46L/K57E] complex as also observed in the OCT4-SOX17 complex (Figure S3). This further supports the hypothesis that the lower cooperativity of these proteins with OCT4 is partly due to a less optimal arrangement of the POU$_S$-HMG interface, which affects the POU$_S$-DNA interaction. The interaction area A_{OS} is very similar among all the canonical-bound complexes (Table 2) and does not correlate with the experimental cooperativity factors. The larger A_{OS} in the wild-type OCT4-SOX17 complex (Table 1), despite the low cooperative factor, is likely due to incomplete sampling of the OCT4 linker region in the simulations. Configurations with larger $\varphi_{\alpha1-\alpha3}$ similar to those observed in the simulations of the OCT4-SOX17 complex are sampled by the highly cooperative OCT4-SOX2[E46L] complex, whereas the lowly cooperative OCT4-SOX2[K57E] also adopts configurations with lower $\varphi_{\alpha1-\alpha3}$. This suggests that the identity of residue 57 does not define a certain configuration of the interface but may affect the ratio between different configurations on longer timescales. In the canonical-bound OCT4-SOX2[E46L] and OCT4-SOX17[E57K] complexes, E82 and K85 from the OCT4 linker form a stable network of salt bridges with R50 from the HMG and D29 from the POU$_S$ (Figures 6A and 6B). This network is not present in the OCT4-SOX2 complex (Figures 6C and 6D), thus explaining the higher cooperativity with OCT4 of these two mutants. Consistently, SOX17[E57K] outperforms SOX2 in inducing pluripotency in somatic cells and driving transcription together with OCT4 from a reporter gene bearing a canonical motif (Aksoy et al., 2013a; Jauch et al., 2011). Although these observations may be partly attributed to the transactivation potential of different SOX factors (Aksoy et al., 2013b), they suggest a close relationship between the OCT4-SOX cooperative binding and the biological function of OCT4.
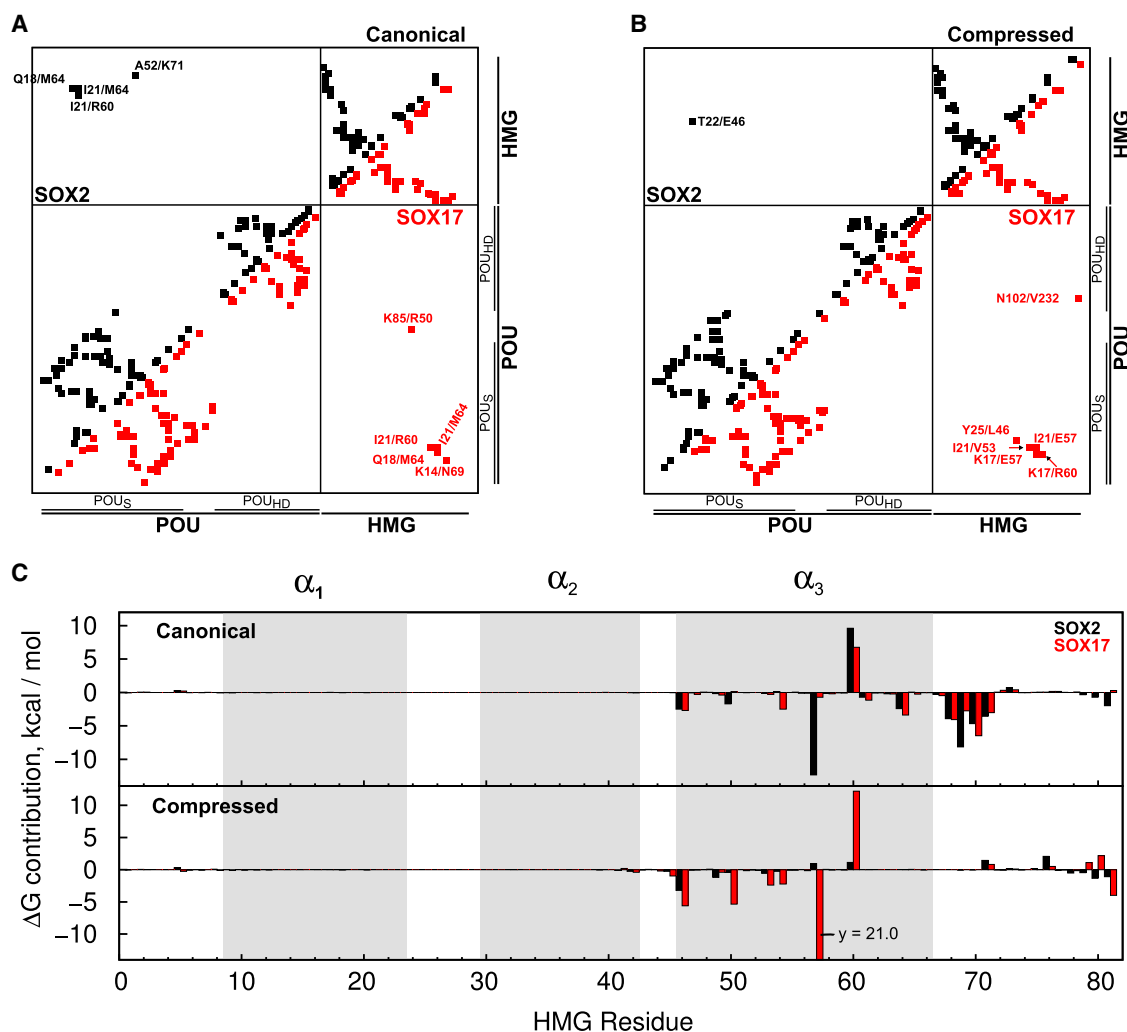
Figure 4. Residues Defining the POU-HMG Interfaces
(A and B) Maps of stable protein-protein interactions from the simulations of the OCT4-SOX2 and OCT4-SOX17 complexes bound to the canonical (A) and compressed (B) motifs.
(C) Per residue contribution to the binding free energy of OCT4 to the SOX-DNA complex obtained from the MMPBSA calculation with the dielectric boundary modeled using the solvent excluded surface. The shaded regions highlight the position of the HMG helices.
See also Figure S5.

On the compressed motif, the interaction area A_{OS} varies between 164.0 ± 40.9 Å^2 in the OCT4-SOX2 complex to 491.6 ± 90.5 Å^2 in the OCT4-SOX2^{K57E} complex (Table 2; Figure S2B). This shows that some of the differences in the OCT4-SOX cooperativities on this motif are due to large conformational rearrangements of the protein-protein interface. Interestingly, in the OCT4-SOX2^{E46L} and OCT4-SOX17^{E57K} complexes the OCT4-SOX interface is similar to the OCT4-SOX17 complex, as judged by φ, $\varphi_{\alpha1-\alpha3}$, and A_{OS} values (Table 2; Figure S3). This is due to the availability of L46 for the hydrophobic interaction with Y25 from OCT4, which leads to an increase in $\varphi_{\alpha1-\alpha3}$. In the OCT4-SOX2^{K57E} complex, $\varphi_{\alpha1-\alpha3}$ has a similar value as in OCT4-SOX2 reflecting a less optimal POU-HMG interface. As in the case of the canonical-bound OCT4-SOX17 complex, the larger A_{OS} in OCT4-SOX2^{K57E} is probably due to incomplete sampling of the OCT4 linker region.

Notably, it is possible to predict the motif preference of some of the simulated SOX factors just by monitoring A_{OS}. For instance, A_{OS} is significantly larger in the OCT4-SOX2-canonical than in the OCT4-SOX2-compressed complex (Table 2; Figure S3). Likewise, A_{OS} in the OCT4-SOX2$^{E46L/K57E}$ bound to the compressed motif is larger than in the corresponding complexes bound to the canonical motif. SOX17 also shows this behavior, although in this case the differences are considerably smaller. This prediction may be optimized by improving the sampling of the OCT4 linker region.

From the simulations, we estimated the relative cooperative binding free energies $\Delta\Delta\Delta G_{OCT4/SOX_I}^{OCT4/SOX_R}$ (SOX_R, reference wild-type SOX; SOX_I, measured SOX factor) using two alternative methods based on the Molecular Mechanics Poisson-Boltzman Surface Area approach (MMPBSA) (Miller et al., 2012), omitting the contribution of the conformational entropy (see the
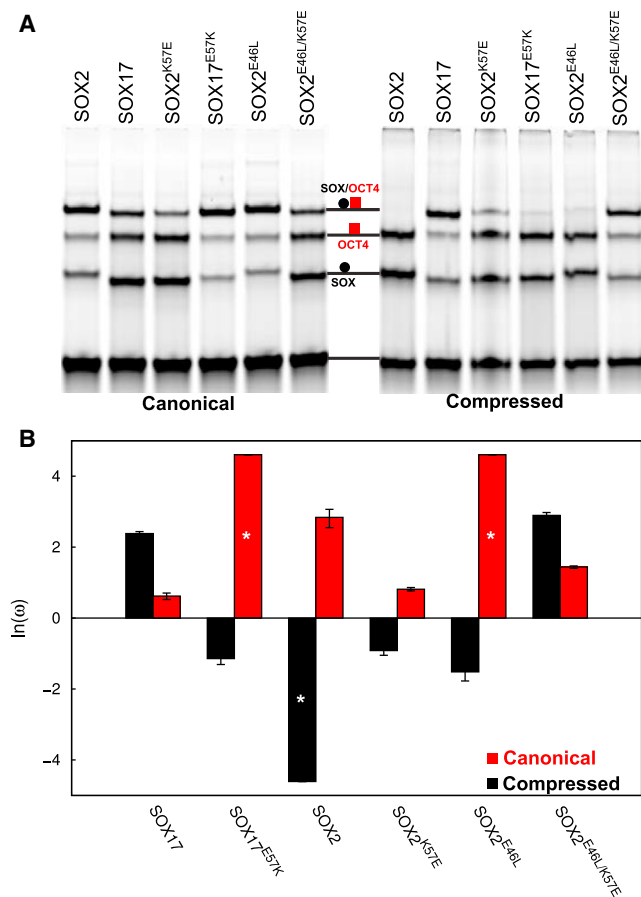
Figure 5. Measured OCT4-SOX Cooperativity Factors

(A) Gel from the electrophoretic mobility shift assay.

(B) Cooperativity factors. Binding to the canonical (A, left and B, red) and the compressed (A, right and B, black) motifs was measured. A star marks the systems in which one of the species is less than 1% of the total fluorescence; therefore, ω is estimated to be smaller than 0.01 or bigger than 100. See also Figure S6.

Supplemental Experimental Procedures). The measured cooperativity factor ω can be related to the cooperative binding free energy $\Delta\Delta G^{OCT4}_{OCT4/SOX_I}$ by $\Delta\Delta G^{OCT4}_{OCT4/SOX_I} = -RT \ln(K^{OCT4}_{SOX_I}/K^{OCT4}) = -RT \ln(\omega_I)$, where R is the gas constant, T is the tem-

perature, and $K^{OCT4}_{SOX_I}$ and K^{OCT4} are the association constant for the binding of OCT4 to the DNA in the presence and absence of the SOX factor, respectively. Therefore, the relative co-operativities can be estimated from ω as $\Delta\Delta\Delta G^{OCT4/SOX_R}_{OCT4/SOX_I} = \Delta\Delta G^{OCT4}_{OCT4/SOX_I} - \Delta\Delta G^{OCT4}_{OCT4/SOX_R} = -RT \ln(\omega_I/\omega_R)$.

We calculated only the binding affinities of OCT4 (ligand) for the SOX-DNA complexes (receptor). The relative cooperativities should be independent of whether OCT4 or the SOX factor is considered as a ligand. However, in the latter case, the receptor (OCT4-DNA) adopts significantly different conformations in the presence and absence of the ligand due to the large DNA bending by SOX factors. Therefore, the assumption that the entropic contribution cancels out between different complexes becomes invalid. This leads to a significant increase in the number and length of the simulations required for the same type of calculation.

The calculated OCT4 affinities were highly dependent on the molecular surface definition of the solute-solvent dielectric boundary (Table 3). The values obtained with the smoothed surface were more reasonable, in agreement with previous reports (Harris et al., 2013). The calculations converged slowly (Figures 7A, 7B, S7A, and S7B). For most systems 90–120 ns of simulation were necessary, whereas for some systems, even after 150 ns, convergence was not completely reached. This suggests that longer or additional simulations are required for a more thorough sampling of the conformational space. However, the convergence achieved was sufficient to study the relative cooperative binding free energies. Irrespective of the method, we found a strong correlation between the estimated and measured values ranging from 0.93 to 0.94 for the canonical to 0.83 to 0.89 for the compressed motif (Figures 7C, 7D, S7C, and S7D). Interestingly, the correlation coefficients converged significantly faster than the binding affinities (Figures 7E, 7F, S7E, and S7F), and the high correlation seen was not due to any particular complex (Figures 7G, 7H, S7G, and S7H).

In agreement with the qualitative results from the simulations and the experiments, SOX17[E57K] and SOX2[E46L] cooperate strongest with OCT4 to bind the canonical motif, while SOX17 cooperates the weakest (Figures 7A and 7C). On the other hand, SOX17 and SOX2[E46L/K57E] have the strongest predicted cooperativities on the compressed motif, while SOX2 has the weakest (Figures 7B and 7D).

Table 2. Summary for the Simulations of the Mutant OCT4-SOX-DNA Complexes

	Canonical Motif[a]				Compressed Motif[a]			
SOX mutant	SOX17[E57K]	SOX2[K57E]	SOX2[E46L]	SOX2[E46L/K57E]	SOX17[E57K]	SOX2[K57E]	SOX2[E46L]	SOX2[E46L/K5E]
Simulation time (ns)	150	150	150	150	150	150	150	150
Residue 46	Leu*	Glu*	Leu	Leu	Leu*	Glu*	Leu	Leu
Residue 57	Lys	Glu	Lys*	Glu	Lys	Glu	Lys*	Glu
θ_{HMG} (°)[a]	44.4 ± 5.7	39.3 ± 6.6	37.7 ± 5.6	38.1 ± 9.1	36.0 ± 9.5	40.9 ± 5.4	44.2 ± 7.0	29.3 ± 8.9
$\varphi_{\alpha1-\alpha3}$ (°)[a]	37.8 ± 3.5	38.1 ± 4.3	47.1 ± 5.7	35.4 ± 5.0	45.5 ± 4.7	39.4 ± 5.0	47.4 ± 4.2	43.9 ± 3.8
φ (°)[a]	129.3 ± 3.6	132.8 ± 3.0	135.9 ± 3.4	130.1 ± 3.6	101.6 ± 3.1	101.3 ± 2.7	103.6 ± 2.3	102.3 ± 2.2
A_{OS} (Å^2) [a]	368.7 ± 50.9	389.6 ± 72.4	386.9 ± 71.6	292.2 ± 63.5	404.8 ± 52.6	491.6 ± 90.5	433.6 ± 88.8	416.9 ± 92.0

See also Figures S2 and S3.

[a]See footnotes of Table 1 for explanations.

*Residue identical in the mutant and the corresponding wild-type factor.
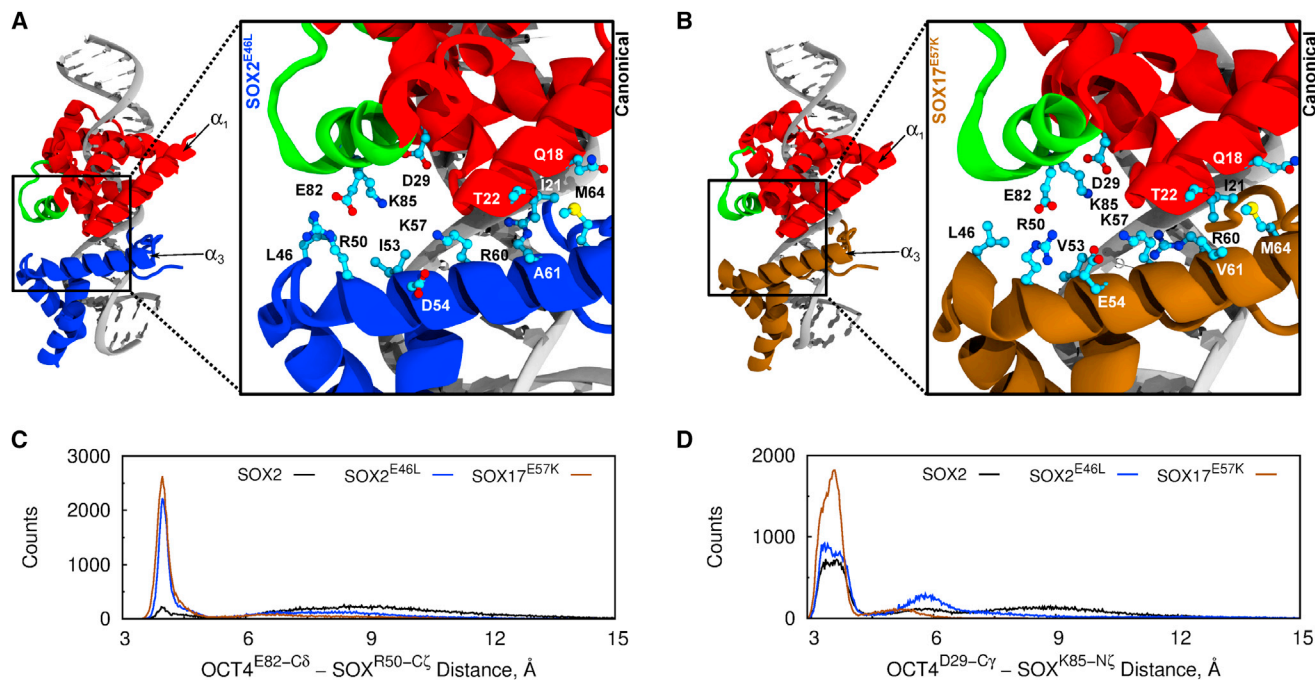
Figure 6. Representative Structures of the Mutant OCT4-SOX Complexes Bound with High Cooperativity to the Canonical Motif
(A and B) Mutant complexes OCT4-SOX2^{E46L} (A) and OCT4-SOX17^{E57K} (B). Illustrations are as in Figure 3.
(C and D) Histograms representing the stability of the E82^{OCT4}-R50^{SOX} (C) and D29^{OCT4}-K85^{OCT4} (D) interactions during the simulations.

Testing the reproducibility of the results, we found a slightly out-of-trend trajectory of the OCT4-SOX2$^{E46L/K57E}$-compressed complex. This was due to the poor convergence of this simulation mainly due to incomplete sampling of the POU$_{HD}$ dynamics (Figure S7I). Therefore, the slow dynamics leading to slow convergence of the estimated OCT4 affinities involve mostly the POU$_{HD}$. This is consistent with our observation of the unstable POU$_{HD}$-DNA interactions (Figure S4). However, since the POU$_{HD}$ does not contribute to the POU-HMG interface, our analysis was not affected.

Conclusions

In summary, we used molecular simulations and free energy calculations to study the structural basis for the genomic redistribution of OCT4 during primitive endoderm commitment of pluripotent cells. An important role in this process was attributed to the mutually exclusive cooperative binding of OCT4 with SOX2 or SOX17 to the canonical and compressed composite motifs, respectively (Aksoy et al., 2013a). We provided further evidence that the interactions with OCT4 are reflected in the genome-wide DNA binding by SOX factors. Also, we revealed the OCT4-SOX17 interface on the compressed motif and found that the identities of two amino acids are sufficient to explain the motif preference of OCT4-SOX complexes. We further validated our structural models by designing a SOX2 mutant that mimics SOX17 properties. Ultimately, we found a strong correlation between the estimated relative cooperative binding free energies and the measured cooperativity factors for 12 OCT4-SOX complexes. Therefore, we demonstrated that molecular simulations may be employed to design cooperative DNA recognition in a

rational manner. In addition, further research on cell fate transitions may benefit from the details we provide about the OCT4-SOX interaction interfaces. More generally, we further underscored the versatility with which POU proteins such as OCT4 select their target genes. In addition to their SOX-dependent binding to composite motifs, POU proteins were also found to bind as homodimers to palindromic DNA sequences (Reményi et al., 2001; Tantin et al., 2008). Whereas such motifs were not enriched among the OCT4 binding sites during endodermal differentiation (Aksoy et al., 2013a), they may play important roles in other processes. Therefore, we anticipate that the ability of transcription factors such as OCT4 to induce different environment-dependent biological outcomes depends on switches in their preference for alternative DNA motifs controlled by the availability of selective interaction partners.

EXPERIMENTAL PROCEDURES

Analysis of Genomic Distribution of SOX Factors
De novo motif analysis was performed with MEME (Bailey et al., 2009) for the SOX2, SOX17^{E57K}, SOX17, and SOX2^{K57E} proteins (Supplemental Experimental Procedures) based on the previously published data from chromatin immunoprecipitation followed by high-throughput sequencing (Aksoy et al., 2013a) (Gene Expression Omnibus GSE43275).

Alignment Covariance Analysis
The sequences for the HMG domains of eukaryotic SOX factors were collected from UniProt (http://www.uniprot.org) and aligned with Clustal (http://www.clustal.org) after removing identical HMGs (Document S2). The amino acid substitution covariance was calculated with five algorithms using Fodor's java code (Fodor and Aldrich, 2004) (Supplemental Experimental Procedures).

Table 3. Calculated Affinities for the Binding of OCT4 to the SOX-DNA Complexes

	Canonical[a]		Compressed[a]	
	Solvent Excluded[b]	Smooth[b]	Solvent Excluded[b]	Smooth[b]
SOX17	14.3 ± 0.69	−6.26 ± 0.83	−3.86 ± 0.74	−31.4 ± 0.84
SOX17^{E57K}	−1.44 ± 0.63	−26.3 ± 0.74	−2.23 ± 0.69	−28.6 ± 0.82
SOX2	8.09 ± 0.73	−15.0 ± 0.90	9.30 ± 0.64	−11.7 ± 0.77
SOX2^{K57E}	10.7 ± 0.67	−13.0 ± 0.80	0.58 ± 0.60	−30.0 ± 0.78
SOX2^{E46L}	3.30 ± 0.72	−22.0 ± 0.86	−0.35 ± 0.66	−27.8 ± 0.76
SOX2$^{E46L/K57E}$	10.2 ± 0.67	−13.7 ± 0.84	−3.69 ± 0.62	−30.9 ± 0.77

[a]The affinity values and standard errors are in kcal/mol.
[b]For the MMPBSA calculation, the dielectric boundary was defined either as the solvent excluded or the smoothed van der Waals surface.

Modeling the OCT4-SOX-DNA Complexes

We modeled the canonical- and compressed-bound complexes between OCT4 and SOX2, SOX17, SOX2^{K57E}, SOX17^{E57K}, SOX2^{E46L}, and SOX2$^{E46L/K57E}$ using MODELER (Sali and Blundell, 1993). To avoid noise from random sequence variability in natural enhancers, we used idealized canonical, 5′-CGG**CATTGT CATGCAAAT**CGGCGGC-3′ and compressed, 5′-CGG**CATTGTATGCAAAT** CGGCGGC-3′ motif sequences (composite binding sites are in bold).

The structural templates for OCT4 were its crystal structure (Protein Data Bank [PDB] ID 3L1P) and our previous models (Esch et al., 2013) built based on the NMR structure of the OCT1-SOX2-*HOXB1* complex (PDB ID 1O4X) (Williams et al., 2004). The models were further optimized with a simulated annealing procedure to improve side chain positions (Supplemental Experimental Procedures). The templates for the SOX factors were the crystal structure of SOX17 (PDB ID 3F27) (Palasingam et al., 2009) and the structure of SOX2 from our optimized models.

To model the murine OCT4-SOX-canonical complexes, we aligned the OCT4 and SOX17 crystal structures with our optimized model of human OCT4-SOX2-*HOXB1* complex. We then mutated the DNA with SPDBV (Guex and Peitsch, 1997) to match our chosen sequence, generated 15 models, and selected that with the lowest Discrete Optimized Protein Energy (DOPE) score (Shen and Sali, 2006) for simulations.

To model the OCT4-SOX-compressed complexes we employed the following stepwise procedure: (1) the structure of the DNA was generated from the canonical DNA by mutating the bases in order to move the POU binding site one base pair toward the HMG binding site with SPDBV; given that the POU site adopts a B-DNA like structure, the resulting composite site was not distorted and the bending of the SOX site was conserved. (2) The OCT4-DNA and SOX-DNA structures were aligned on the newly generated DNA by superimposing the POU and HMG binding sites. (3) Fifteen models were generated, and the one with the lowest DOPE score was selected for the simulations.

Molecular Dynamics Simulations

Ionizable residues were assigned their standard protonation state at neutral pH. The 5′ and 3′ ends of the DNA and the N termini of the proteins were methylated, whereas the C termini of the proteins were acetylated to avoid potential truncation artifacts. We optimized side chain rotamers, calculated structural water positions, and determined the orientation of asparagines, glutamines, and histidines with FoldX (Guerois et al., 2002; Schymkowitz et al., 2005). We then curated the results manually. This optimization step is only suitable for models that do not depart significantly from local energy minima. Therefore, we omitted it for the compressed-bound complexes because an experimentally derived configuration corresponding to an energy minimum for the POU-HMG interface on this motif was not available.

Then, the systems were (1) solvated in a truncated octahedral box of TIP3P water extending at least 12 Å from any protein or DNA atom, (2) neutralized with 100 mM KCl using the Smith-Dang ions (Dang, 1995) to mimic the experimental ionic strength, and (3) equilibrated for 3.525 ns using a stepwise protocol including restraints and accelerated molecular dynamics (Hamelberg et al., 2004) to improve the poorly modeled regions (Supplemental Experimental Procedures). Afterward, we used NAMD (Phillips et al., 2005) to perform 150 ns of molecular dynamics simulations per system in the isobaric-isothermic (NPT, p = 1 atm, T = 300 K) ensemble. We used a standard protocol for explicit solvent under periodic boundary conditions (Supplemental Experimental Procedures). Coordinates were saved every 3 ps.

We used the Amber-ff99SB force field (Hornak et al., 2006) including the bsc0 (Pérez et al., 2007), the NMR-based (Li and Brüschweiler, 2010), and the ILDN (Lindorff-Larsen et al., 2010) corrections for the DNA, protein backbone, and side chains, respectively.

Structural Analysis

The rmsd, distances, A_{OS}, φ, and $\varphi_{\alpha1-\alpha3}$ were calculated in VMD (Humphrey et al., 1996). For rmsd, each trajectory frame was superimposed on the initial model using all the protein C_α atoms and the P atoms of the composite motif. For the orientation of the SOX factor relative to the octamer (POU) site (φ), we defined the following coordinate system: v_x is the vector connecting the center of mass of the first and last base pairs of the POU site; v_t is the vector connecting the backbone of the bases from the first base pair of the POU site; v_z is the cross product of v_x and v_t; and v_y is the cross product of v_x and v_z. φ is the angle between the helical axis of helix α_3 of the HMG projected on the v_z/v_y plane and v_y (Figure S2). In essence, it represents the orientation of the helix around the helical axis of the octamer site, neglecting minor adjustments within the binding groove. The structural properties of the DNA were calculated with Curves+ (Lavery et al., 2009). The contact maps were calculated with Wordom (Seeber et al., 2007). For this, the number of atom-atom contacts (N) (contact threshold is 4.5 Å) was calculated for every pair of protein residues. Then, a relative strength of the interaction (I) was calculated as described in Brinda and Vishveshwara (2005). Briefly, $I = N/\sqrt{(N_i \times N_j)}$, where N_i and N_j are normalization factors that relate to the maximum number of contacts that residues i and j can make. The interactions with $I \geq 2.5\%$ present in more than 50% of the simulations were considered.

Continuum Solvation Free Energy Calculations

We defined the relative cooperative binding free energy as $\Delta\Delta G^{OCT4/SOX_R}_{OCT4/SOX_I} = \Delta\Delta G^{OCT4}_{OCT4/SOX_I} - \Delta\Delta G^{OCT4}_{OCT4/SOX_R}$, where $\Delta\Delta G^{OCT4}_{OCT4/SOX_I} = \Delta G^{OCT4}_{SOX_I} - \Delta G^{OCT4}$ represents the absolute cooperativity (change in OCT4 DNA-binding affinity in the presence of the SOX factor). Therefore, $\Delta\Delta\Delta G^{OCT4/SOX_R}_{OCT4/SOX_I} = \Delta G^{OCT4}_{SOX_I} - \Delta G^{OCT4}_{SOX_R}$. We used the MMPBSA.py program from AmberTools 13 (Miller et al., 2012) to estimate the binding free energies of OCT4 for the SOX-DNA complex (ΔG^{OCT4}_{SOX}) with MMPBSA. In principle the contribution of OCT4 (ligand), SOX-DNA (receptor), and the complex to the binding free energy can be approximated as $G = E_{MM} + G_{Solv} - TS$, where E_{MM} is the gas phase energy of the macromolecule estimated from the force field, S is the entropy, T is the temperature, and G_{Solv} is the solvation free energy. The latter is further divided into polar solvation, calculated with the Poisson-Boltzmann equation and an apolar solvation term. We omitted the conformational entropy calculation because we assumed that this term is similar among all complexes. Moreover, the quasi-harmonic analysis usually used to calculate it converges very slowly (Furini et al., 2013). We calculated the polar solvation by solving the nonlinear Poisson-Boltzmann equation at 100 mM ionic strength. The ratio between the longest solute dimension and the grid was set to 4, and the grid spacing was set to 0.5 Å. The solvent and solute dielectric constants were set to 80 and 4, respectively, as appropriate for highly charged systems
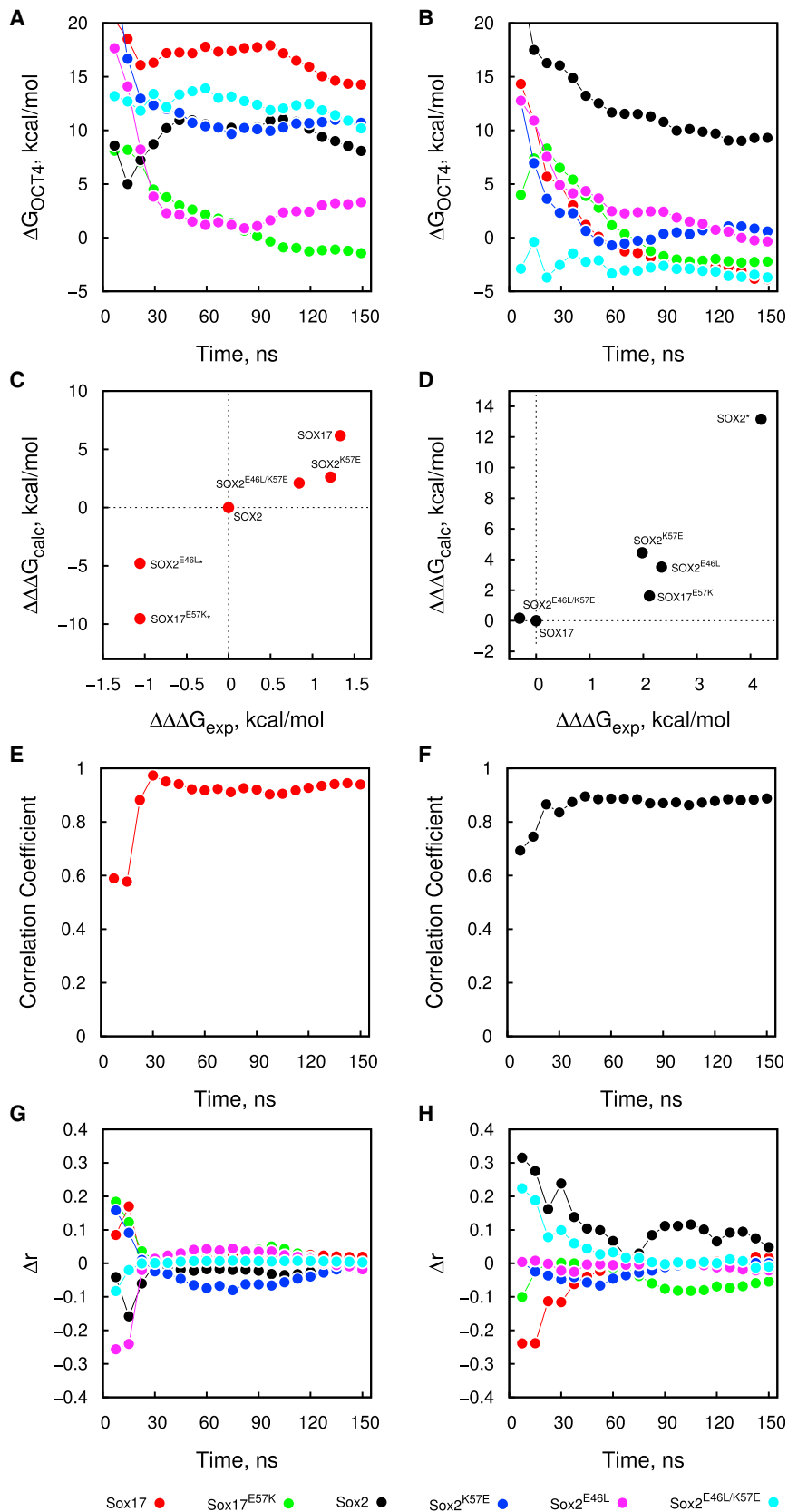
Figure 7. Correlation between the Estimated and Measured Relative Cooperative Binding Free Energies

(A and B) Convergence of the OCT4 binding affinities for the SOX-DNA complexes.

(C and D) Correlation coefficients.

(E and F) Convergence of the correlation coefficients.

(G and H) Contribution of each SOX protein ($\Delta r = r - r_{no\text{-}SOX}$) to the correlation.

The canonical-bound and compressed-bound complexes are shown in the panels on the left (A, C, E, and G) and right (B, D, F, and H), respectively. For the MMPBSA calculation, the dielectric boundary was modeled using the solvent excluded surface. The stars are as in Figure 5. See also Figure S7.

(Hou et al., 2011). We tested two definitions of the dielectric boundary: (1) the solvent excluded surface and (2) the smoothed van der Waals surface (Ye et al., 2010).

The apolar solvation energy was modeled as the sum of a repulsive term related to the formation of a cavity in the solvent and an attractive term related to the solute-solvent interaction approximated by the van der Waals attraction energy. The repulsive term was $\Delta G_{rep} = \gamma(SASA) + b$, where $\gamma = 0.0378$ kcal/mol·A^2, and $b = -0.5692$ kcal/mol (Tan et al., 2007). For the calculation, we used 200 frames spaced by 0.75 ns per simulation.

Quantitative Cooperativity Measurements with Electrophoretic Mobility Shift Assay

The procedure was described in Ng et al. (2012) and summarized in the Supplemental Experimental Procedures. Briefly, pure, recombinant OCT4 and SOX DNA-binding domains and DNA probes labeled with Cy5 at the 5′ were prepared. The 100 nM dsDNA probe mixed with 100–500 nM OCT4 and 20–100 nM SOX protein was incubated for 1 hr and loaded on a native polyacrylamide gel for separation. The bands corresponding to free DNA, OCT4-DNA, SOX-DNA, and OCT4-SOX-DNA were quantified, and the cooperativity factors ω were determined (Figure S6). At least three replicates were performed to obtain the average ω and standard deviations.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and Supplemental Alignment, and can be found with this article online at http://dx.doi.org/10.1016/j.str.2014.06.014.

REFERENCES

Aishima, J., and Wolberger, C. (2003). Insights into nonspecific binding of homeodomains from a structure of MATalpha2 bound to DNA. Proteins 51, 544–551.

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013a). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. 32, 938–953.

Aksoy, I., Jauch, R., Eras, V., Chng, W.B., Chen, J., Divakar, U., Ng, C.K., Kolatkar, P.R., and Stanton, L.W. (2013b). Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming. Stem Cells 31, 2632–2646.

Ambrosetti, D.C., Schöler, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J. Biol. Chem. 275, 23387–23397.

Babin, V., Wang, D., Rose, R.B., and Sagui, C. (2013). Binding polymorphism in the DNA bound state of the Pdx1 homeodomain. PLoS Comput. Biol. 9, e1003160.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev. Cell 21, 611–626.

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.

Brinda, K.V., and Vishveshwara, S. (2005). A network representation of protein structures: implications for protein stability. Biophys. J. 89, 4159–4170.

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.

Dailey, L., and Basilico, C. (2001). Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J. Cell. Physiol. 186, 315–328.

Dang, L.X. (1995). Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. J. Am. Chem. Soc. 117, 6954–6960.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C.A., Araúzo-Bravo, M.J., Ng, C.K.L., et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. 15, 295–301.

Fodor, A.A., and Aldrich, R.W. (2004). Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. Proteins 56, 211–221.

Frum, T., Halbisen, M.A., Wang, C., Amiri, H., Robson, P., and Ralston, A. (2013). Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. Dev. Cell 25, 610–622.

Furini, S., Barbini, P., and Domene, C. (2013). DNA-recognition process described by MD simulations of the lactose repressor protein on a specific and a non-specific DNA sequence. Nucleic Acids Res. 41, 3963–3972.

Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. 320, 369–387.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. Cell 138, 774–786.

Hamelberg, D., Mongan, J., and McCammon, J.A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J. Chem. Phys. 120, 11919–11929.

Harris, R.C., Boschitsch, A.H., and Fenley, M.O. (2013). Influence of grid spacing in Poisson-Boltzmann equation binding energy estimation. J. Chem. Theory Comput. 9, 3677–3685.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65, 712–725.

Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. J. Comput. Chem. 32, 866–877.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Jauch, R., Aksoy, I., Hutchins, A.P., Ng, C.K.L., 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.

Jerabek, S., Merino, F., Schöler, H.R., and Cojocaru, V. (2014). OCT4: dynamic DNA binding pioneers stem cell pluripotency. Biochim. Biophys. Acta *1839*, 138–154.

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.

Lam, C.S., Mistri, T.K., Foo, Y.H., Sudhaharan, T., Gan, H.T., Rodda, D., Lim, L.H., Chou, C., Robson, P., Wohland, T., and Ahmed, S. (2012). DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. Biochem. J. *448*, 21–33.

Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D., and Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res. *37*, 5917–5929.

Le Bin, G.C., Muñoz-Descalzo, S., Kurowski, A., Leitch, H., Lou, X., Mansfield, W., Etienne-Dumeau, C., Grabole, N., Mulas, C., Niwa, H., et al. (2014). Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. Development *141*, 1001–1010.

Lee, J., Kim, J.S., and Seok, C. (2010). Cooperativity and specificity of Cys2His2 zinc finger protein-DNA interactions: a molecular dynamics simulation study. J. Phys. Chem. B *114*, 7662–7671.

Leichsenring, M., Maes, J., Mössner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science *341*, 1005–1009.

Li, D.W., and Brüschweiler, R. (2010). NMR-based protein potentials. Angew. Chem. Int. Ed. Engl. *49*, 6778–6780.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–1958.

Miller, B.R., III, McGee, T.D., Jr., Swails, J.M., Homeyer, N., Gohlke, H., and Roitberg, A.E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. J. Chem. Theory Comput. *8*, 3314–3321.

Ng, C.K.L., Li, N.X., Chee, S., Prabhakar, S., Kolatkar, P.R., and Jauch, R. (2012). Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. Nucleic Acids Res. *40*, 4933–4941.

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.

Nishimoto, M., Miyagi, S., Yamagishi, T., Sakaguchi, T., Niwa, H., Muramatsu, M., and Okuda, A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol. Cell. Biol. *25*, 5084–5094.

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.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A., and Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J. *92*, 3817–3829.

Phillips, K., and Luisi, B. (2000). The virtuoso of versatility: POU proteins that flex to fit. J. Mol. Biol. *302*, 1023–1039.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Reményi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Schöler, H.R., and Wilmanns, M. (2001). Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. Mol. Cell *8*, 569–580.

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.

Reményi, A., Schöler, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. Nat. Struct. Mol. Biol. *11*, 812–815.

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.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell *12*, 15–30.

Schymkowitz, J.W., Rousseau, F., Martins, I.C., Ferkinghoff-Borg, J., Stricher, F., and Serrano, L. (2005). Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc. Natl. Acad. Sci. USA *102*, 10147–10152.

Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics *23*, 2625–2627.

Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci. *15*, 2507–2524.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tan, C., Tan, Y.H., and Luo, R. (2007). Implicit nonpolar solvent models. J. Phys. Chem. B *111*, 12263–12274.

Tantin, D. (2013). Oct transcription factors in development and stem cells: insights and mechanisms. Development *140*, 2857–2866.

Tantin, D., Gemberling, M., Callister, C., and Fairbrother, W.G. (2008). High-throughput biochemical analysis of in vivo location data reveals novel distinct classes of POU5F1(Oct4)/DNA complexes. Genome Res. *18*, 631–639.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Werner, M.H., Huth, J.R., Gronenborn, A.M., and Clore, G.M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell *81*, 705–714.

Williams, D.C., Jr., Cai, M., and Clore, G.M. (2004). 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. *279*, 1449–1457.

Wu, G., and Schöler, H. (2014). Role of Oct4 in the early embryo development. Cell Regener. *3*, 7.

Ye, X., Wang, J., and Luo, R. (2010). A revised density function for molecular surface definition in continuum solvent models. J. Chem. Theory Comput. *6*, 1157–1169.