Mechanisms of OCT4-SOX2 motif readout on nucleosomes

Alicia K. Michael1, Ralph S. Grand1, Luke Isbel1, Simone Cavadini1, Zuzanna Kozicka1,2, Georg Kempf1, Richard D. Bunker1, Andreas D. Schenk1, Alexandra Graff-Meyer1, Ganesh R. Pathare1, Joscha Weiss1, Syota Matsumoto1, Lukas Burger1,3, Dirk Schübeler1,2,†, Nicolas H. Thomé1,‡

Transcription factors (TFs) regulate gene expression through chromatin where nucleosomes restrict DNA access. To study how TFs bind nucleosome-occupied motifs, we focused on the reprogramming factors OCT4 and SOX2 in mouse embryonic stem cells. We determined TF engagement throughout a nucleosome at DNA access. To study how TFs bind nucleosome-occupied motifs, we focused on the reprogramming factors OCT4 and SOX2, and they reveal how TFs distort nucleosomes to access chromatinized motifs.

Transcription factors (TFs) regulate gene expression and govern cell identity by interacting with specific sequence motifs. OCT4 and SOX2 serve as reprogramming factors (1) that cooperate as critical mediators of pluripotency (2–4). Although they are not obligate heterodimers in solution, the combined OCT4-SOX2 motif drives OCT4-SOX2 complex formation (5, 6).

Chromatin restricts DNA access (7, 8), but a specialized subset of TFs, termed pioneer factors, can engage chromatinized motifs to trigger cell-fate changes (9). Several TFs have been shown to bind motifs embedded in nucleosomes in vitro (1, 10, 11); however, the nucleosome architecture, with histones H2A, H2B, H3, and H4 and its two DNA gyres (12), restricts TF access to >95% of nucleosomal DNA (13). Two extreme scenarios for nucleosomal TF-engagement have been put forward: TF binding without changing the nucleosomal architecture (10, 14) or TF-mediated changes to the nucleosome by distorting the histone core, looping the DNA, or taking advantage of nucleosome unwrapping dynamics at the entry-exit sites (15, 16).

OCT4 binding is predicted to be incompatible with the nucleosome architecture on the basis of its engagement with free DNA (17–19), although partial motifs have been identified where OCT4 engages only a portion of its binding site to maintain nucleosome integrity (1). Despite being critical for genome regulation, the structural and mechanistic principles governing nucleosome engagement by single or multiple TFs have yet to be determined.

SeEN-seq reveals preferential binding of OCT4-SOX2 to nucleosomal entry-exit sites

Each rotational and translational setting of a motif around a nucleosome places the TF in a different histone-DNA environment. To assess all potential TF-binding registers on a nucleosome, we built on approaches that examine select loci (20, 21) and developed selected engagement on nucleosome sequencing (SeEN-seq), which measures the relative affinity of one or multiple TFs for each nucleosomal register in parallel (Fig. 1A). We focused on OCT4 and its dependence on SOX2 because both factors show strong co-occupancy at target genes in vivo (22, 23), and the OCT4-SOX2 protein-protein interface is required for pluripotency (24). The canonical OCT4-SOX2 motif (matrix ID MA0142.1) (25) was tiled at 1-base pair (bp) intervals through a 601-nucleosome positioning sequence (26).

The sequence library was then assembled into octameric nucleosome core particles (NCPs), incubated with the TF(s), and subjected to electrophoretic mobility shift assay (EMSA). TF-bound and unbound nucleosome complexes were excised and subjected to next-generation sequencing (NGS), which generated single molecule counts that approximate motif affinity as a function of position.

A range of protein concentrations of either OCT4, SOX2 (residues 37 to 118), or OCT4-SOX2 together were assessed, with a total of 4752

Fig. 1. SeEN-seq identifies preferred binding sites of OCT4 and OCT4-SOX2 across the nucleosome. (A) Principle of SeEN-seq. A library of TF motif–containing nucleosome positioning sequences is assembled into nucleosomes and incubated with TF(s). TF-bound and unbound nucleosome complexes are separated by EMSA and sequenced, revealing position-specific enrichments. Stars indicate a specific example sequence at each step of the assay. (B) SeEN-seq enrichment (n = 3 replicates) for nucleosome pool with or without TFs. Further details are provided in fig. S1. Motif position is indicated by superhelix location (SHL) that describes where the minor groove faces away (SHLs ±1, ±2, etc.) or toward (SHLs ±1.5, ±2.5, etc.) the histone octamer. (C) Autocorrelation analysis of OCT4 enrichment; dashed lines indicate 95% confidence interval. (D) DNA nucleosome unwrapping energy (25) versus OCT4 and OCT4-SOX2 SeEN-seq enrichment profile. Arrows indicate enriched positions in OCT4. DG, delta Gibbs free energy; kBT, 0.6 kcal/mol, where T = 300 K.
conditions measured with high reproducibility (Fig. 1B and fig. S1, A to C). The trends in OCT4 or SOX2 binding at selected positions were validated by fluorescence polarization measurements (fig. S2, A to E). Previously, two OCT4 motif locations were tested and found to provide similar OCT4 access to the nucleosome-embedded motif (27). Although this was recapitulated in SeEN-seq, our comprehensive examination of all motif locations reveals clear OCT4 preference for nucleosomal DNA entry-exit sites as well as discrete preferential binding sites −1 to 3 bp in width throughout the nucleosome (Fig. 1B and fig. S1C). SOX2 shows less differential enrichment in SeEN-seq, in line with published data (10), with some degree of preferred binding toward the entry-exit sites and near the dyad (Fig. 1B and fig. S1C). Given the small enrichment amplitudes for SOX2 alone, we focused on the robust and differential binding activity seen for OCT4 and how it is affected by SOX2. OCT4 and SOX2 cooperate strongly, which results in up to 650-fold increased binding compared with that caused by OCT4 alone (fig. S1C). This effect is most evident at entry-exit sites with weaker cooperativity observed at internal sites (fig. S1C).

Although OCT4-SOX2 binding appears roughly symmetrical across the dyad, for OCT4 alone, enrichment in the right half of the nucleosome [superhelix locations (SHLs) +4 to +6.5] is more pronounced compared with that in the left (SHLs −4 to −6.5) (Fig. 1B and fig. S1C). A notable difference is the motif orientation on either side of the dyad; on the left, the OCT4 portion of the motif is closest to the dyad, whereas on the right it is oriented toward the entry-exit site of the nucleosome. Positions enriched for OCT4 alone and OCT4-SOX2 exhibited stronger binding at discrete sites with 10-bp periodicity across the nucleosome (Fig. 1C and fig. S2, F and G). Both OCT4 and OCT4-SOX2 show a trend of stronger binding at the entry-exit site than at the dyad (Fig. 1D). This would be expected if OCT4 and OCT4-SOX2 binding was facilitated by nucleosomal breathing, which is more pronounced toward the nucleosomal entry-exit sites (Fig. 1C and fig. S2, F and G). OCT4 and OCT4-SOX2 together remove the DNA from core histones (Fig. 2, C and D). OCT4 has a bipartite POU-homeodomain (POUHD) separated by 17 residues, whereas SOX2 utilizes an HMG domain (18). When nucleosome-bound, the OCT4-POUS and SOX2-HMG DNA-binding domains engage major and minor grooves.

**Structure of OCT4-SOX2 bound at SHL −6** shows DNA release from the histone core

To dissect mechanisms of nucleosome engagement, we performed structural studies with OCT4-SOX2, which co-bind key target genes in vivo and have a cooperative effect in vitro (fig. S2, I to L) (22). For structural studies, we focused on sites that show discrete OCT4 SeEN-seq enrichment and cooperative binding in the presence of SOX2 (fig. S3A). A site 57 bp from the dyad (SHL −6) enabled structure determination from ~90,000 particles at an overall resolution of 3.1 Å (Fig. 2, A and B, and fig. S3). The nucleosome core and discrete domains of OCT4 [POU-specific (POUS)] and SOX2 [high-mobility group (HMG)] were well resolved (see fig. S3F for local resolution), which allowed conservative refinement with reference restraints to existing high-resolution OCT4 and SOX2 DNA-bound crystal structures (Fig. 2, B and C; fig. S4; and fig. S5, A to D) (18, 19).

In the OCT4-SOX2-NCP<sub>S</sub>−6 structure, OCT4 and SOX2 together remove the DNA from core histones (Fig. 2, C and D). OCT4 has a bipartite DNA-binding domain composed of a POUS and POU-homeodomain (POUHD) separated by 17 residues, whereas SOX2 utilizes an HMG domain (18). When nucleosome-bound, the OCT4-POUS and SOX2-HMG DNA-binding domains engage major and minor grooves,
respectively, with protein-DNA interactions consistent with those previously seen for the individual OCT4-POUS domain and SOX2 on free DNA (fig. S5E). The DNA remains attached and straightened around the OCT4 site but is detached around the SOX2 motif (Fig. 2 and fig. S5F). SOX2 kinks the DNA and, together with OCT4, synergistically releases DNA from the core histones (movie S1). OCT4-SOX2 forms no discernible histone contacts, with the DNA separating them from the nucleosome core (Fig. 2, C and D). Density for the OCT4 transactivation domains was not observed, which is consistent with similar nucleosome-affinity measurements for full-length OCT4 and OCT4 DNA-binding domain only (fig. S5G). OCT4 recognizes a partial motif, engaging DNA with its POUS domain, whereas the
POUHD is not engaged. On free DNA, both POU domains engage the major groove spanning 8 bp on opposite sides of the DNA and would clash with either the histones or DNA gyres at all locations on the nucleosome (19, 28) (Fig. 2E and fig. S6A). To stabilize the OCT4-SOX2-NCPSHL+6 complex for imaging, GraFix cross-linking was necessary and a cross-link with SOX2-NCPSHL+6 was evident between H2A and H2B (fig. S6B) (29). To test if POUHD access was blocked by the cross-linking of histones, we solved an ~4 Å map of the non–cross-linked OCT4-SOX2-NCPSHL−6 nucleosome (fig. S7, A to E), which resulted in a largely indistinguishable model (root mean square deviation, 1.3 Å; fig. S7, F and G). As the OCT4 POUHD motif is occluded by H2A-H2B, the binding mechanism involving only the OCT4 POUS is consistent with partial motif engagement (Fig. 2F) (7). Partial motif recognition, however, does not necessarily render TF binding compatible with the nuclease-occupied DNA structure (fig. S5F).

In the context of the nucleosome, SOX2 competes with histones for DNA binding and kinks DNA by ~90° at SHL −6.5 away from the histones, similar to HMG domains on free DNA (Fig. 3, A and B, and fig. S8A) (18). This is accomplished by intercalation of the SOX2 Phe48 and Met49 wedge at the TT base step (18). Variant SOX2 motifs that lessen distortion of the DNA induced by SOX2 have been described and may facilitate nucleosome-compatible dyad binding (1, 10, 30). However, with the canonical SOX2 motif used here, SOX2 facing the entry-exit site is not compatible with the canonical nucleosome architecture and triggers DNA release (3f).

Despite disruption of histone-DNA contacts at SHL −6.5, no histone rearrangements were observed after OCT4-SOX2 binding and DNA release across 14 bp (fig. S8, B and C). To verify OCT4-SOX2 binding and DNA release using an orthogonal approach, we performed deoxyribonuclease I (DNaseI) footprinting in the presence and absence of OCT4-SOX2. This revealed increased digestion at the nucleosomal entry-exit site (SHL −7 to −5.5) in the presence of OCT4-SOX2. OCT4 alone triggers DNA release, and OCT4 and OCT4-SOX2 also destabilize the opposite nucleosomal entry-exit site (SHL +5.5 to +7), which is also evident in the cryo–electron microscopy (cryo-EM) map (Fig. 2 and fig. S3). The structures, footprinting profiles (Fig. 3), and thermal stability assays (fig. S9C) together support the idea that OCT4 and OCT4-SOX2 release DNA from the histones and have a global effect on the DNA structure of the nucleosome.

**OCT4-SOX2 bound at SHL +6 induces minimal distortion to nucleosomal DNA**

The SeEN-seq profile suggests that OCT4-SOX2 engagement depends on motif orientation (Fig. 1B). To examine this structurally, we utilized the same position but with the motif inverted across the dyad axis, i.e., SHL +6 (Fig. 4A). Doing so places the ~90° kink-inducing SOX2 motif in a dyad-proximal orientation and OCT4 closer to the entry-exit site. The SHL +6 site was enriched for OCT4-SOX2 binding in SeEN-seq, and the use of this position enabled cryo-EM structure determination of an OCT4-SOX2-NCPSHL+6 complex at an overall resolution of 3.5 Å (Fig. 4B and fig. S10; see fig. S10E for local resolution). The map allowed unambiguous rigid body docking of the nucleosome and of SOX2 (figs. S11 and S12). OCT4 (POUS) density was less continuous but sufficient to dock a Cα model (fig. S11D). The resulting OCT4-SOX2 interface was consistent with previously determined structures (Fig. 2, fig. S11C, and materials and methods).

In the structure, OCT4 engages its motif with the POUS domain only (Fig. 4B), and the POUHD is unable to access its motif in the observed DNA configuration (fig. S13). OCT4-
SOX2 together give rise to an extended DNA-binding surface across 11 bp that further bends the nucleosomal DNA by ~90° at the SOX2 site (SHL +5) and straightens the DNA near the OCT4 site (SHL +6), producing an L-shaped DNA arrangement (Fig. 4B). This conformation locally lifts the duplex away from the histones and DNA gyre, but, in contrast to the reversed orientation, does not fully release the DNA from histones at the entry-exit site. At the SOX2 motif with the DNA locally detached, SOX2 helices 1 and 2 widen the minor groove, and the C terminus (residues 110 to 114) wedges between the DNA and histones (SHL +5) (Fig. 4C). Despite partial engagement of an internal-nucleosomal motif, the SOX2 and OCT4 (POUS) DNA interactions and induced DNA distortions are again similar to those previously seen on free DNA (fig. S11C). Within one helical turn, on either side of OCT4-SOX2, these DNA distortions are largely absorbed into the canonical DNA trajectory of the nucleosome (Fig. 4D). The histone core architecture again shows no substantial distortion. The DNA at the opposite nucleosomal entry-exit site (SHLs −7 to −5.5) appears to be disordered in the cryo-EM map (fig. S10E and fig. S14). In both structures (Figs. 2 and 4), OCT4 binds a partial DNA motif through its POU domain and, along with SOX2, affects the entire nucleosomal DNA structure to varying extents.

The HMG-POUS partial motif is sufficient for TF engagement and the opening of chromatin in vivo

Previous work has identified that either Oct4 or Sox2 alone engage reduced-complexity motifs on nucleosomes during reprogramming (1). A recent study has also identified weaker-scoring motifs for SOX proteins on nucleosome-sized fragments (32). The structures now reveal that OCT4-SOX2 partial motif engagement is utilized in both orientations on the nucleosome. This led us to test whether OCT4-POUS and SOX2-HMG domains are sufficient to engage chromatin in vivo (Fig. 5A and fig. S15A). Through in-depth analysis of existing chromatin immunoprecipitation sequencing (ChIP-seq) datasets (33, 34), we found that the partial HMG-POUS motif is sufficient to drive genomic binding, although the full motif was bound more frequently (Fig. 5A). To test this experimentally, we introduced full and partial motifs at a defined genomic position in mouse embryonic stem cells (mESCs) by recombination-mediated cassette exchange (35) (Fig. 5B). Motifs were introduced in the SHL −6 position of the 601 sequence (see materials and methods), and Oct4 binding was determined. This revealed significant Oct4 enrichment at both the full and partial motifs but not in the control (Fig. 5B and fig. S15B). Thus, single motifs recapitulate genome-wide Oct4 binding to partial motifs.

Next, we asked if binding creates open chromatin. Comparing accessibility in mESCs (36) revealed that full and partial motifs can both generate accessible chromatin in an Oct4-dependent manner (Fig. 5C and fig. S15, C and D). Consistently, the same effect is evident upon knockdown of Oct4 or Sox2, which shows that both TFs are required for full accessibility at these loci (37) (fig. S15, E to H). This confirms the cooperativity observed in SeEN-seq; however, the local effect is expected to be highly context dependent, as additional proteins contribute to accessibility in vivo (36).

We interpret our structures to depict an initial encounter complex between OCT4-SOX2 and the nucleosome en route to open chromatin. Upon nucleosome removal, OCT4 is expected to engage a full motif with its two POU domains, thereby accounting for the stronger enrichment of the full versus the partial motif. Together, genome-wide binding, single-locus insertion, and genome-wide accessibility data demonstrate that the OCT4-POUS domain is sufficient to engage and open chromatin in conjunction with Sox2. This reveals the potential for such nucleosome-compatible motifs to function as bona fide binding sites beyond the ability to initially engage closed chromatin.

Discussion

The structures illustrate binding mechanisms of OCT4-SOX2 at two positions on the nucleosome. At SHL ±6, the structures depict OCT4-SOX2 near the entry-exit sites, where both TFs cooperate to access DNA. At the SHL −6 site, where SOX2 faces the entry-exit site, OCT4-SOX2 releases the DNA duplex from the histones. In the OCT4-SOX2-NCPSHL−6 structure, where SOX2 faces the dyad, the nucleosomal DNA assumes an L-shaped trajectory and is not fully released from the histones (Fig. 5D). The SHL −6 structure demonstrates that partial motif recognition and DNA release are not mutually exclusive (Figs. 2 and 3), whereas the SHL +6 structure depicts how more-internally bound sites can be accommodated without full removal of nucleosomal DNA ends (Fig. 4). We consistently find only the OCT4 POUS engaged, with the POUHD motif occluded by the nucleosome architecture. The OCT4 POUS and POUHD domains could, in principal, engage the full OCT4 motif if the DNA was further unwrapped from the histones, which we do not observe in our structures (Figs. 2 and 4) or DNaseI experiments (Fig. 3C). Thus, partial motif recognition allows TFs to minimize DNA unwrapping when engaging nucleosomal sites, although partial motifs do not fully preempt nucleosome distortions. We show that partial OCT4 motifs, in conjunction with SOX2, are recognized in vivo and create open chromatin (Fig. 5).

SeEN-seq binding profiles, combined with the structural data, allow us to rationalize OCT4-SOX2 engagement throughout the nucleosome. OCT4 and OCT4-SOX2 bind best at the nucleosomal entry-exit sites, where DNA breathing is expected to facilitate access. We observe distinct exceptions from such end-binding behavior for OCT4 in particular, where narrow regions of pronounced binding are juxtaposed to nonengaged regions. To correlate these accessibility profiles to the structure, we computationally translated isolated OCT4 POUS or POUHD domains along the DNA of the unbound nucleosome and calculated predicted atomic clash scores at each position (fig. S16). A comparison of OCT4 SeEN-seq data to the POUS domain nucleosome-clash scores revealed that the solvent accessibility of the POUS—but not of the POUHD—correlates with OCT4 binding (fig. S16, A and B). The solvent accessibility for OCT4 POUS is also a good predictor for OCT4-SOX2 engagement (fig. S16, C and D). The presence of SOX2 enables tight OCT4-SOX2 binding at the nucleosome ends but has a limited effect on more-internal sites (Fig. 1). At nucleosome ends, SOX2 can also drive binding at motifs where the OCT4 POUS is inward-facing and OCT4 alone does not bind (fig. S2H). Our structural and functional findings are consistent with a model where cooperative binding between OCT4 and SOX2 not only strengthens DNA binding affinity but also triggers additional DNA distortions that must be accommodated. These distortions are better tolerated at the entry-exit sites, where nucleosomal DNA is more loosely bound (38). Whereas partial motifs delimit the TF footprint and DNA unwrapping, the available structures (19) show that protein domains bound to nucleosomal DNA retain their free DNA-binding mode (Figs. 2 and 4). TF-induced DNA distortions intrinsically destabilize the nucleosome (10), which likely facilitates the binding of additional factors and disrupts the internucleosomal interactions of higher-order chromatin (15, 39) (fig. S17).

The OCT4-SOX2 structures and the accompanying in vitro and in vivo evidence provide a framework by which TFs use nucleosomal DNA distortion and not histone rearrangement to access parts of their motif. The degree of DNA distortion imposed on the nucleosome architecture depends on the position of the motif. Our structures suggest principal recognition mechanisms for nucleosome-incompatible TFs as well as for those TFs accommodated on the nucleosome without DNA release, illustrating how TFs can read out chromatinized binding sites.

REFERENCES AND NOTES

1. A. South et al., Cell 161, 555–566 (2015).
2. D. J. Rodda et al., J. Biol. Chem. 280, 24731–24737 (2005).
3. K. Takahashi, S. Yamanaka, Cell 126, 663–676 (2006).
4. V. Malik et al., Nat. Commun. 10, 3477 (2019).
30. P. Scaffidi, M. E. Bianchi, J. Biol. Chem. 276, 47306–47302 (2001).

31. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Nature 389, 251–260 (1997).

32. M. P. Meers, D. H. Janssens, S. Henikoff, Mol. Cell 75, 562–575, e5 (2019).

33. C. Choromos et al., Cell 168, 442–459.e20 (2017).

34. W. Liu, L. Kraus, Mol. Cell 65, 589–603.e9 (2017).

35. F. Lienert et al., Nat. Genet. 43, 1090–1097 (2011).

36. H. W. King, R. J. Kloos, eLife 6, 22631 (2017).

37. E. T. Friman et al., eLife 8, e50087 (2019).

38. M. A. Hall et al., Nat. Struct. Mol. Biol. 16, 124–129 (2009).

39. T. Schalch, S. Duda, D. F. Sargent, T. J. Richmond, Nature 436, 138–141 (2005).

ACKNOWLEDGMENTS

Funding: N.H.T. and D.S. acknowledge support from the Novartis Research Foundation, the European Research Council under the European Union’s (EU) Horizon 2020 research and innovation program grant agreement (N.H.T., 666068; D.S., 667951), and the Swiss National Science Foundation (N.H.T., Sinergia CRSII3_160734/1 and SNF 31003A_179541; D.S., 310030B_176394). A.K.M. acknowledges the Human Frontier Science Program, R.S.G. and L.I. acknowledge the EU Horizon 2020 research and innovation program grant agreement (R.S.G., 765445). Author contributions: A.K.M., R.S.G., and L.I. developed SeEN-seq. L.I. and R.S.G. performed biochemical assays. L.I. and L.B. performed analysis of periodicity mediated cassette exchange (RMCE) insertions. A.K.M., R.S.G., and L.I. developed SeEN-seq. L.I. and R.S.G. performed fluorescence polarization assays. G.K. and R.D.B. prepared atomic models with A.K.M. and S.C. R.S.G., performed DNase footprinting assay and analysis. A.D.S. and A.G.-M. provided technical support for cryo-EM. G.R.P., J.W., and S.M. contributed nucleosome preparations and reagents. A.K.M., R.S.G., L.I., D.S., and N.H.T. wrote the manuscript. The research was directed by D.S. and N.H.T. Competing interests: The authors declare no competing interests. Data and materials availability: Plasmids and cell lines generated by this study are available from the Protein Data Bank (accession codes: 6T90, 6T93, and 6T95). All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

References (40–83)

MBR Reproducibility Checklist

Movie S1

View/request a protocol for this paper from Bio-protocol.

22 January 2020; accepted 16 April 2020

Published online 23 April 2020

10.1126/science.abb0074