

# Nucleosome-bound Sox2 and Sox11 structures elucidate pioneer factor function

Svetlana O. Dodonova<sup>1</sup>, Fangjie Zhu<sup>2</sup>, Christian Dienemann<sup>1</sup>, Jussi Taipale<sup>2</sup>, Patrick Cramer<sup>1\*</sup>

<sup>1</sup>*Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.* <sup>2</sup>*Department of Biochemistry, University of Cambridge, 8 Tennis Ct Rd, CB2 1QW Cambridge, UK.*

\*Email: [patrick.cramer@mpibpc.mpg.de](mailto:patrick.cramer@mpibpc.mpg.de)

**Pioneer transcription factors are required for stem cell pluripotency, cell differentiation, and cell reprogramming<sup>1,2</sup>. Pioneer factors can bind nucleosomal DNA to enable gene expression from regions of the genome with closed chromatin. Sox2 is a prominent pioneer factor that is essential for pluripotency and self-renewal of embryonic stem cells<sup>3</sup>. Here we report cryo-electron microscopy structures of the DNA-binding domains of Sox2 and its close homologue Sox11 bound to nucleosomes. These first structures of pioneer factors in complex with nucleosomes show that Sox factors can bind and locally distort DNA at superhelical location 2. The factors also facilitate detachment of terminal nucleosomal DNA from the histone octamer, and this increases DNA accessibility. Sox factor binding to the nucleosome can also lead to a repositioning of the N-terminal tail of histone H4, including residue lysine-16. This is incompatible with higher-order nucleosome stacking, which involves contacts of the H4 tail with a neighbouring nucleosome. These results indicate that pioneer transcription factors can use binding energy to contribute to initial chromatin opening and facilitate nucleosome remodelling and transcription.**

Transcription of the human genome is controlled by ~1,600 transcription factors (TFs)<sup>4</sup>. TFs recognize DNA motifs and recruit protein complexes that enable transcription initiation<sup>5</sup>. Binding of most TFs is restricted to regions of the genome that are not packaged into chromatin<sup>6</sup>. Some TFs can however bind to chromatin via contacts to its fundamental unit, the nucleosome<sup>7</sup>. These ‘pioneer’ TFs can initiate transcription in silent chromatin regions<sup>8</sup> and are required for embryo development, cell differentiation, and cell reprogramming<sup>9,10</sup>.

Sox2 and Oct4 are pioneer factors that are widely used for reprogramming of adult cells to induced pluripotent stem cells<sup>2,11,12</sup>. They can interact with nucleosomes *in vitro* and *in vivo*<sup>13,14</sup>. Sox2 alone can direct chromatin opening<sup>15</sup> and bind target DNA sites before Oct4<sup>11</sup> *in vivo*, indicating that Sox2 makes DNA accessible for binding of other factors. Most Sox family factors show pioneer factor function<sup>7</sup>, are essential for developmental processes<sup>16</sup>, and their mutation can lead to severe developmental defects and cancer<sup>17</sup>. How pioneer TFs such as Sox factors bind to the nucleosome and how they make DNA accessible is unknown.

Dodonova et al., Nucleosome-bound Sox2 and –Sox11 structures elucidate pioneer factor function

To investigate this, we determined the cryo-electron microscopy (cryo-EM) structure of human Sox2 in complex with a nucleosome (Methods). We used a 147-bp nucleosomal DNA sequence ('DNA-1', **Extended Data Figure 1**), that was selected for binding the closely related factor Sox11<sup>7</sup>. The DNA-binding domains (DBDs) of Sox2 and Sox11 share 83% sequence similarity (**Extended Data Figure 2**), and bind the same DNA motif, TTGT<sup>18</sup>. Sox2 and Sox11 also bind the same target sites in cells, but differ in regions that flank their DBDs and recruit different factors<sup>19,20</sup>. Nucleosomes containing DNA-1 indeed bound recombinant Sox2 or Sox11 DBDs (**Extended Data Figure 3**). We added purified Sox2 DBD in excess to reconstituted nucleosomes<sup>21</sup>, plunge-froze cryo-EM grids, and collected cryo-EM data (Methods). A subset of 32,301 particles resulted in a ~5.5 Å resolution map that showed extra density on the nucleosome surface (**Figure 1; Extended Data Figures 4-5, Extended Data Table 1**). We fitted the map with structures of the nucleosome<sup>22</sup> and the Sox2 DBD<sup>23</sup> (**Extended Data Figures 5-6**).

The obtained nucleosome–Sox2 structure revealed a single copy of the Sox2 DBD bound to DNA at superhelical location (SHL) +2 (**Figure 1**). The observation of a single Sox2 binding site was unexpected because the DNA-1 template contains multiple Sox motifs (**Extended Data Figure 1**), and because free DNA-1 can bind multiple copies of Sox2 (**Extended Data Figure 3**). However, nucleosomes containing DNA-1 prefer to bind only one Sox2 copy at SHL +2 (**Extended Data Figure 3**), consistent with our structure. Indeed, Sox2 binding to other motifs in DNA-1 is predicted to lead to pronounced clashes with DNA and histones (**Extended Data Figure 7**). The observed Sox2 binding involves specific interactions with the DNA motif, as confirmed by site-directed mutagenesis (**Extended Data Figure 8**) of involved residues<sup>24</sup>. Also in agreement with our finding, *in vivo* Sox factors preferentially occupy their target sites on nucleosomes<sup>25</sup> near their centre<sup>26</sup>.

Despite extensive efforts, the resolution of our nucleosome–Sox2 structure remained limited and prevented further mechanistic insights. We therefore determined the structure of a nucleosome bound to the DBD of Sox11 (Methods). The cryo-EM dataset contained 222,731 particles and resulted in a detailed reconstruction at 3.7 Å resolution (**Extended Data Figure 4, 5, 6**). To aid model building, we determined the crystal structure of the Sox11 DBD in complex with a DNA fragment at 2.5 Å resolution (**Extended Data Figure 9, Extended Data Table 2**).

The structure of the nucleosome–Sox11 complex was virtually identical to that of the nucleosome–Sox2 complex (**Extended Data Figures 5c, 6**). It is also a good model for nucleosome complexes of other Sox family members, which are highly conserved (**Extended Data Figure 2**). For comparisons we further determined the structure of the free nucleosome containing DNA-1 from 368,270 particles at 3.2 Å resolution (**Extended Data Figure 4, 5, 6**). This structure was highly similar to the canonical nucleosome structure (PDB 6FQ5<sup>27</sup>, RMSD (P)=1.0 Å).

Comparison of the nucleosome-Sox11 structure with the free nucleosome structure shows that Sox11 binding leads to strong local DNA distortions at SHL +2 (**Figure 2**) (local RMSD (P)=3.9 Å; calculated for 12 bp of DNA). Sox11 widens the DNA minor groove by 7 Å and pulls the DNA away from the histone octamer by up to 3.4 Å, which increases DNA bending (**Figure 2**). These DNA distortions are induced by Sox11 binding, and are also

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

observed in our Sox11-DNA crystal structure (RMSD (P)=1.4 Å, for 12 DNA bp). Thus the Sox factor uses binding energy to distort DNA locally despite competing histone-DNA interactions.

In both nucleosome-Sox factor structures the terminal ~2.5 turns of DNA near entry and exit sites are detached from the histone octamer and not visible in the EM densities (**Figure 3; Extended Data Figure 6**). This is consistent with the observation that several Sox factors facilitate DNA unwrapping from the nucleosome<sup>7</sup>, and with the known high dynamics of the terminal DNA<sup>28,29</sup>. A DNA cleavage assay supports the increase in accessibility of the terminal nucleosomal DNA in the presence of the Sox factor (**Extended Data Figure 10**). Comparison with the free nucleosome structure indicates that terminal DNA at SHL -7, -6, -5 is detached from the octamer because of a clash with helix h2 of the Sox factor (**Figure 3c, Supplementary Video 1**). Thus Sox factor binding to the nucleosome facilitates DNA detachment and increases accessibility of terminal DNA.

Our cryo-EM data reveals a possible route for how Sox factors invade the nucleosome. A set of particles from a separate dataset (151bp DNA-1, Methods) resulted in an alternative nucleosome-Sox11 structure where the terminal DNA near Sox11 was still associated with the histone octamer (nucleosome-Sox11\*) (**Figure 3; Extended Data Figure 4, 5, 6**). Thus Sox factors may initially bind to their target site without detaching the second DNA gyre. Movement of the DNA-bound Sox factor to the position observed in the nucleosome-Sox11 structure would then lead to terminal DNA detachment. This resulted in a model of nucleosome invasion and DNA unwrapping by Sox factor binding (**Figure 3; Supplementary Video 1**). This mechanism differs from that used by the yeast pioneer factor Reb1, which directly binds and traps terminal DNA<sup>30</sup>.

In our nucleosome-Sox factor structures, terminal DNA is detached on both sides of the nucleosome, suggesting an allosteric effect of the Sox factor, in addition to the steric clash described above. Detachment of terminal DNA on the other side of the nucleosome may be stabilized by binding of a second copy of the Sox factor at SHL -2, which we observed in another population of our cryo-EM particles (**Extended Data Figure 4, Extended Data Figure 6e**). In this nucleosome-Sox11<sub>2</sub> structure, Sox factor molecules are located at both, SHL +2 and SHL -2. In both locations, the orientation of Sox TF is dictated by the DNA motif, confirming interaction specificity. Sox11 binds to a TTCT motif at SHL -2 which likely has lower affinity than the canonical TTGT and is not clearly observed in EMSA.

The nucleosome-Sox11<sub>2</sub> structure shows that Sox11 binding at SHL -2 is incompatible with binding of terminal DNA at SHL +7, +6, +5, although the predicted clash at this location is with helix h3 and both termini of the Sox factor DBD (**Extended Data Figure 7**). Thus Sox factors can induce detachment of both DNA ends and can bind to both sides of the nucleosome (**Figure 3, Supplementary Video 1**). These observations agree with the recently described strong preference for Sox2 binding ~ ±25 bp around the nucleosome dyad *in vivo*<sup>26</sup>. However, it is not excluded that Sox factors may also bind additional nucleosomal positions in other contexts.

The nucleosome-Sox11 structure further shows that binding of Sox11 repositions the N-terminal tail of histone H4 (**Figure 4, Extended Data Figure 11**). In the free nucleosome structure, the H4 tail binds to its canonical site and follows a trajectory towards DNA at

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

SHL +2. However, in the nucleosome-Sox11 structure, the binding site of the H4 tail at SHL+2 is occupied by the Sox11 C-terminal tail (**Figure 4, Extended Data Figure 11**). The H4 tail is displaced, rotated by  $\sim 90^\circ$ , and extends towards SHL +1. The functionally important lysine residue K16 moves by  $\sim 33 \text{ \AA}$ . At SHL -2, Sox11 is however oriented differently and does not displace the H4 tail (**Extended Data Figure 11**).

Displacement of the H4 tail suggests that Sox factor binding is incompatible with the formation of canonical nucleosome-nucleosome contacts<sup>31</sup> (**Extended Data Figure 12**). Formation of nucleosome arrays depends on the H4 tail and is impaired by acetylation at position K16 or by tail truncation<sup>32-34</sup>. Nucleosome stacking is mediated by H4 tail residues K16-R19 that interact with the acidic patch of the H2A-H2B histone dimer of the neighbouring nucleosome<sup>22,35</sup>. Modelling the Sox DBD onto a nucleosome array structure<sup>35</sup> suggests that the pioneer factor could be accommodated. Thus Sox factors may bind nucleosomes within arrays and could displace the H4 tail and impair nucleosome stacking. To achieve efficient chromatin opening, Sox factors however also cooperate with other TFs such as Oct4, Klf4, Pax6, Nanog, Brn2, Prx1<sup>17</sup>, and with chromatin remodelers<sup>36</sup>.

### Acknowledgements

We thank members of the Cramer laboratory, in particular H. Hillen, D. Tegunov, and G. Kokic for advice. We thank the crystallization facility at the MPI BPC, in particular J. Wawrzinek and U. Steuerwald. Part of this work was performed at Beamline X10SA at the SLS at the PSI, Villigen, Switzerland. We thank W. Fischle for providing histone expression constructs. S.O.D. was supported by an EMBO long-term fellowship (ALTF-949-2016). P.C. was supported by the ERC Advanced Investigator Grant TRANSREGULON (grant agreement No 693023) and the Volkswagen Foundation.

### Author contributions

S.O.D. designed and carried out all experiments and data analysis. F.Z. supported by J.T. identified the original DNA template used in the study. C.D. assisted with cryo-EM data collection. P.C. designed and supervised research. S.O.D. and P.C. interpreted the data and wrote the manuscript, with input from all authors.

**Author information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declare that they have no competing financial interest. Correspondence and request of materials should be addressed to P.C. ([patrick.cramer@mpibpc.mpg.de](mailto:patrick.cramer@mpibpc.mpg.de)).

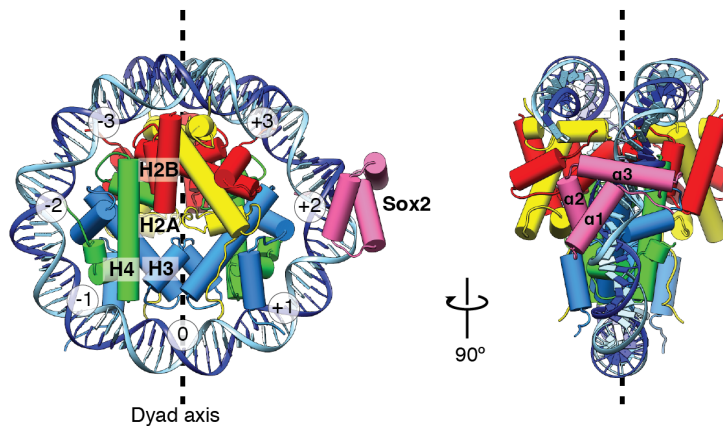
### Competing interests

The authors declare no competing interests.

**Data availability statement** The cryo-EM density reconstructions and final models were deposited with the EMDB (EMD-10390, EMD-10391, EMD-10392, EMD-10393, EMD-10394) and with the PDB (6T78, 6T79, 6T7A, 6T7B, 6T7C, 6T7D). All data is available in the main text or the supplementary materials.

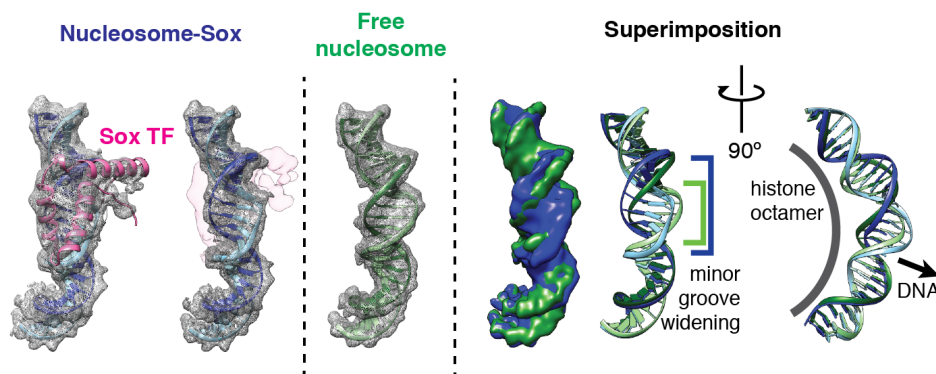


## FIGURE LEGENDS



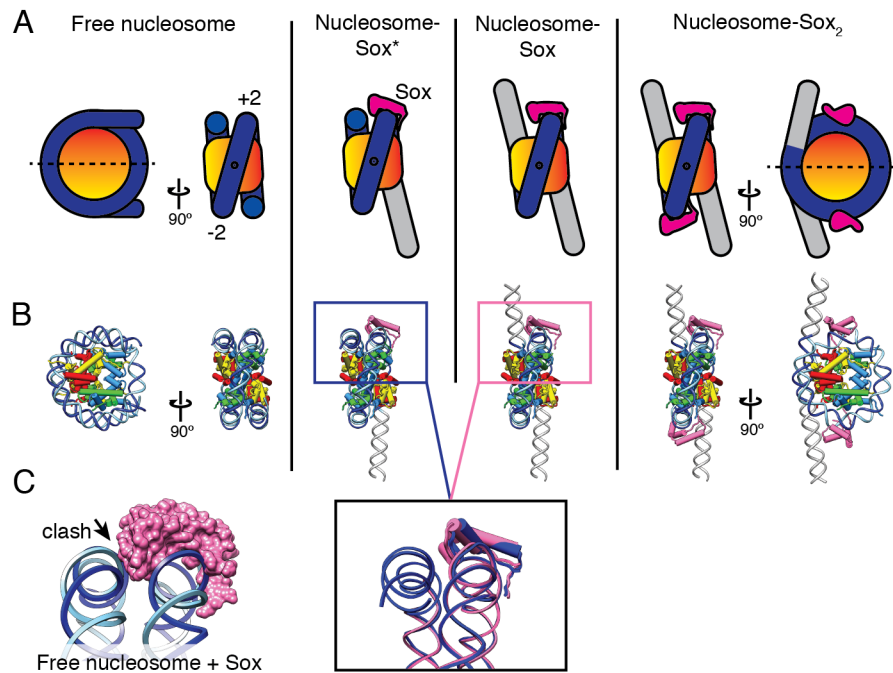
**Figure 1 | Structure of the nucleosome-Sox2 complex.**

Structure of the nucleosome-Sox2 complex reveals Sox2 binding at SHL +2. Top (left) and side (right) views are related by a 90-degree rotation around the dyad (dashed line). SHLs -3 to +3 are labeled. Sox2 DBD is in pink and its three  $\alpha$ -helices are labeled (right). Histones H2A, H2B, H3, and H4 are in yellow, red, blue, and green, respectively. DNA in dark and light blue.



**Figure 2 | Structure of the nucleosome-Sox11 complex and local DNA distortion.**

Sox11 is shown in pink, and DNA in dark and light blue. DNA in the free nucleosome structure is dark and light green. EM maps shown here were Gaussian-smoothened. For clarity, Sox density was segmented out on the right (blue).

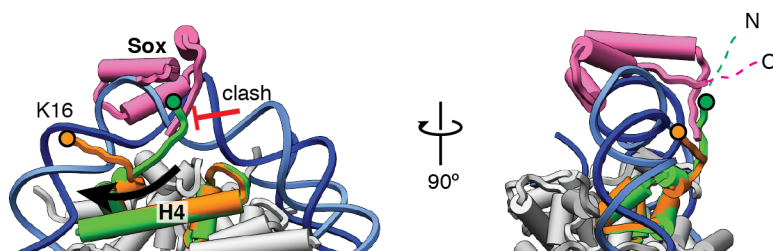


**Figure 3 | Model of nucleosome invasion by Sox factors.**

**a.** Nucleosome invasion by Sox factors and terminal DNA detachment. Schematic representation of four of the structures reported here. Left to right: free nucleosome, nucleosome-Sox11\*, nucleosome-Sox11, nucleosome-Sox11<sub>2</sub>. Histone octamer is in orange, Sox in pink, DNA in blue. Detached DNA is grey. The dyad is shown as a dashed line or as a dot.

**b.** Four structures colored as in Figure 1. Detached DNA was modelled as ideal B-DNA (grey). The black box shows a comparison of the nucleosome-Sox\* (dark blue) and nucleosome-Sox (pink) structures.

**c.** DNA superposition in the free nucleosome and the Sox factor (surface view) from the nucleosome-Sox structure illustrates the clash between Sox and the second DNA gyre.



**Figure 4 | Sox11 repositions the N-terminal tail of histone H4.**

Top and side views of SHL +2 with Sox TF (pink). Histones are grey, except for H4. H4 histone from the free nucleosome is shown in green, where the H4 N-terminal tail would clash with the C-terminal Sox tail. In the nucleosome-Sox structure the H4 tail is repositioned (orange). Residue K16 is marked with a coloured circle.

## METHODS

### Plasmids and strains

Full-length histone sequences from *H. sapiens* were incorporated into the following plasmids: pET22B-H2B, pET22b-H3.2, pET3a-H4 (kindly provided by the W. Fischle lab). H2A construct was cloned into a LIC 1B vector (MacroLabs) and contained an N-terminal 6xHis-tag followed by a tobacco etch virus (TEV) protease cleavage site [HHHHHHENLYFQS]. Sox2 DNA-binding domain (DBD) construct contained residues 36-121 of the full-length Sox2 (Uniprot ID P48431). The DBD sequence was codon-optimized and synthesized by IDT as a gBlock. The gBlock was inserted into a LIC-1B plasmid following an N-terminal 6xHis-tag and a TEV protease cleavage site sequences. The Sox11 DBD with short flanking sequences contained residues 33-138 of full-length Sox11 (Uniprot ID P35716). It was inserted into a LIC-1B plasmid. The construct was identical to the Sox11 construct used in <sup>7</sup>. Protein constructs are schematically shown in the Extended Data Figure 2.

### Protein purification

Histones were purified according to standard protocols <sup>21,37</sup>. Purified histones were flash-frozen and lyophilized. Histones were resuspended in unfolding buffer (6 M guanidine hydrochloride, 20 mM HEPES pH 7.5, 10 mM DTT). H2A, H2B, H3 and H4 were mixed in 1.2:1.2:1:1 ratios, and dialyzed against three changes of Refolding Buffer High (RB High: 20 mM Hepes pH=7.5, 1 mM EDTA, 2 M NaCl, 2 mM DTT). After dialysis the sample was concentrated and loaded onto a size-exclusion chromatography column (Superdex 200 10/300 GL, GE Healthcare). A peak corresponding to the complete octamer was collected and used for nucleosome reconstitution. The Sox2 DBD was expressed in BL21 (DE3) RIL *Escherichia coli* cells and purified over a sequence of columns: affinity His-Trap HP, cation exchange HiTrap SP-HP, and size-exclusion Superdex75 10/300 GL (GE Healthcare). The His-tag was cleaved off after the affinity purification step. Sox11 DBD was expressed and purified exactly as the Sox2 DBD. Purified proteins in the final buffer (20 mM Hepes pH=7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT) were flash-frozen and stored at -80°.

### DNA preparation

DNA-1 template sequence was:  
ATCTACACGACGCTCTCCGATCTAATTTATGTTTGTAGCGTTATACTATTCTAA  
TTCTTTGTTTCGGTGGTATTGTTTATTTGTTTCCTTTGTGCGTTCAGCTTAATGCCT  
AACGACACTCGGAGATCGGAAGAGCACACGTGAT. We note that this sequence was directly (no changes) adopted from the NCAP-SELEX experiment with nucleosomes and Sox TF performed in <sup>7</sup>.

DNA-1a sequence with all but one TTGT motifs replaced by a random sequence was:  
ATCTACACGACGCTCTCCGATCTAATTTATTCAGACTAGCGTTATACTATTCTAA  
TTTCAGACTTCGGTGGTCAGACTTATCAGACTCCTTTGTGCGTTCAGCTTAATGCC  
TAACGACACTCGGAGATCGGAAGAGCACACGTGAT.

Widom601 DNA template used as a control was:  
ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCG  
CTTAAACGCACGTACGCGCTGTCCCCCGCGTTTAAACCGCCAAGGGGATTACTCC

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

CTAGTCTCCAGGCACGTGTCAGATATATACATCCGAT. Three bases at each end were changed to accommodate an EcoRV restriction site.

The DNA\* template used for the nucleosome-Sox\* structure determination was 151 bp long and almost identical to the DNA-1: ATCCCTACACGACGCTCTTCCGATCTAATTTATGTTTGTAGCGTTATACTATTCT AATTCTTTGTTTCGGTGGTATTGTTTATTTTGTTCCTTTGTGCGTTCAGCTTAATGC CTAACGACACTCGGAGATCGGAAGAGCACACGTCTGAT. Two additional nucleotides on each side in the DNA\* template are highlighted in bold.

A plasmid pMK containing four consecutive copies of a DNA template of interest separated by EcoRV restriction sites was ordered from GeneArt (Thermo Fisher). The plasmid was produced in large quantities in *E. coli* XL1 blue cells, and purified with a NucleoBond PC 10000 kit (Macherey Nagel). The plasmid was digested with EcoRV enzyme (NEB) overnight, and produced 4 copies of the insert per plasmid. The plasmid was then precipitated with PEG-6000<sup>38</sup>. The insert was further purified by size-exclusion chromatography with a Superose 6 Increase column (GE Healthcare). Peak fractions were pooled and concentrated by ethanol precipitation.

### **Nucleosome reconstitution**

Nucleosomes were reconstituted from the histone octamer and DNA template with a salt gradient as described<sup>37</sup>. Briefly, octamer and DNA were mixed in 1.2:1 ratio in RB High, transferred into Slide-A-Lyzer MINI Dialysis Units 7,000 MWCO (Thermo Fisher), and dialyzed gradually over a course of 24 h from RB High into RB Low (20 mM Hepes pH=7.5, 1 mM EDTA, 20 mM NaCl, 2 mM DTT). Freshly reconstituted nucleosomes were concentrated in Amicon Ultra-0.5 centrifugal filters MWCO 10000 (Sigma Aldrich).

### **Cryo-EM grid preparation and data collection**

Nucleosomes at 1.6  $\mu$ M concentration were mixed with 20x molar excess of Sox TF at 4° C in the final buffer containing 20 mM HEPES pH=7.5, 1 mM EDTA, 30 mM NaCl, 2 mM DTT, and used for cryo-grid preparation. First, R 2/1 Cu 300 mesh grids (Quantifoil) were glow-discharged with PELCO easiGlow (Ted Pella) device for 120 s. Next, 3.5  $\mu$ l of sample were applied to the grid in the Vitrobot Mark IV (FEI) chamber at 100% humidity and 16° C. The excess of liquid was blotted away for 10 s, and the grid was vitrified by plunging into liquid ethane. Data collection was performed on a G2 Titan Krios microscope (FEI) equipped with a K2 Summit direct electron detector (Gatan). Data were collected with EPU software (Thermo Fisher), with defocus ranging from 0.9 to 3.4  $\mu$ m at a nominal magnification of 130,000x and a pixel size of 1.05 Å/pixel. Data were collected with an energy filter slit set to 30 eV. The total electron dose of 45 e<sup>-</sup>/Å<sup>2</sup> was distributed over 40 movie frames. For all imaged samples at least 50% of the data were collected at 25° stage tilt in order to partially compensate for preferred orientation of particles on the grid, and to improve angular distribution. The quality of the reconstructions was improved compared to the 0-tilt data. Data collection was monitored on-the-fly with Warp<sup>39</sup> and cryoSPARC 2D classification<sup>40</sup>.

### **Data processing and analysis**

## Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

Processing details are summarized in the Table S1. For every dataset, particles were picked with gAutomatch, CTF-determination was performed with Gctf<sup>41</sup>. The initial reference from the free-nucleosome set was obtained *ab initio* in cryoSPARC, low-pass filtered to 40 Å, and used as a starting point for the 3D classification of all datasets. For every dataset, in order to speed up the computation, binned particles with the pixel size of 4.2 Å were extracted and subjected to several rounds of 2D classification and 3D classification in Relion<sup>42</sup>. Classes showing high-resolution features were selected for further processing. Next, selected particles were reextracted with a pixel size of 2.1 Å, and were 3D classified and cleaned again. Finally, particles were reextracted at the final pixel size of 1.05 Å and box size of 400 pixels, and subjected to 3D refinement. For all datasets processing was performed without symmetry application (C1). Final FSC curves supplied with directional FSC curves and anisotropy estimates were calculated using 3DFSC server<sup>43</sup> (Extended Data Figure 5). Additionally, for each map local resolution was calculated in Relion (Extended Data Figure 4).

For the free-nucleosome dataset, after CTF-refinement and 3D refinement, final maps were sharpened using B-factor of -75. The final dataset contained 368,270 particles (see Table S1). When classified, this dataset showed typical levels of partial DNA unwrapping/bulging (~10 bp) at the nucleosome entry/exit sites in ~15% of the data similarly to<sup>29</sup>, however the overwhelming majority of particles contributed to a fully-wrapped nucleosome reconstruction.

In case of the nucleosome-Sox2 dataset, classes that showed additional densities were selected after 3D classification (with global soft mask applied). Next, selected subset was subjected to a round of focused classification with a small soft spherical mask centered at the additional density near SHL+2 of the nucleosome. A class showing strong additional density was selected and further refined. The final dataset was CTF-refined in order to compensate for local defocus variations. At a final step the dataset was subjected to non-uniform refinement in cryoSPARC<sup>40</sup>, which led to an improved local resolution distribution in the 3D reconstruction. The final map was sharpened using a B-factor of -100. The final dataset contained 32,301 particles.

In order to get an overview of the processing pipelines for both Nucleosome-Sox11 datasets, please see **Extended Data Figures 14 and 15**.

The nucleosome-Sox11 dataset was processed in a similar way. The final dataset after initial steps of coarse cleaning was classified into 4 classes, out of which two were of high quality. One of the classes (202,142 particles) showed a clear additional density at SHL+2 and detached terminal DNA. The corresponding final map was sharpened using a B-factor of -100. Another class with two additional densities (nucleosome-Sox11<sub>2</sub>) contained 114,104 particles. It was refined and sharpened using a B-factor of -120. In this nucleosome-Sox11<sub>2</sub> structure, the Sox factor molecules are located at SHL +2 and SHL -2, but are not related by the two-fold pseudo-symmetry of the nucleosome. This confirms that the density for the second Sox factor is not an artifact of particle misalignment during data processing. Lower occupancy of Sox11 at SHL -2 may be due to the presence of a weaker binding motif TTCT in that position. We also note that the local curvature induced by Sox binding at the SHL-2 not as pronounced as at SHL+2, possibly also due to a weaker binding motif.

The nucleosome-Sox11\* dataset resulted in two distinct classes. The first class (130,870 particles) resulted in a map virtually identical to the nucleosome-Sox structure (see above),

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

but with slightly lower resolution (~4.0 Å). The remaining 63,821 particles resulted in the nucleosome-Sox\* map. The final nucleosome-Sox\* map was sharpened using B-factor of -100.

### **Model fitting and refinement**

To model the free nucleosome structure, we started from a canonical nucleosome structure obtained by cryo-EM (PDB 6FQ5)<sup>27</sup> and altered the DNA sequence to correspond to the DNA-1 template using Chimera<sup>44</sup>. Several amino acid residues in the *X. laevis* histones were substituted with the ones corresponding to the *H. sapiens* histones in Coot<sup>45</sup>. Next, the model was fitted into the corresponding sharpened EM map of the free nucleosome and refined in real space using Phenix<sup>46</sup>.

The refined model of the free nucleosome was used to generate models for the nucleosome-Sox complex structures. In case of the nucleosome-Sox2 complex, both the nucleosome model and the Sox2 structure (PDB 1O4X)<sup>23</sup> were placed into the EM map, nucleosome DNA regions outside of the EM map were removed, and the model was refined in real space using Phenix. For the nucleosome-Sox11 models, the nucleosome and the X-ray structure of Sox11 (determined in this study, see below) were placed into the density and refined in real space using Phenix. In both cases extra reference model restraints ( $\sigma=1$ ) were imposed in order to keep the model close to the available higher resolution X-ray structure. Additionally base-pair and base-stacking restraints were used during refinements, excluding the region near the Sox TF binding site because strong local DNA distortion was evident in this region of the map. Equivalent procedure was used for modelling other structures described here.

### **Electrophoretic mobility shift assay (EMSA)**

Nucleosomes at a final concentration of 1.1 nM were mixed with purified proteins (Sox2 or Sox11 DBD). The final buffer contained 10 mM Hepes pH=7.5, 1 mM MgCl<sub>2</sub>, 0.01 mM ZnCl<sub>2</sub>, 1 mM DTT, 10 mM NaCl, 0.5 mg/ml BSA, 5% Glycerol as in<sup>13</sup> (Extended Data Figure 3). Samples were incubated at 10 min at RT, mixed with Novex Hi-Density TBE Sample Buffer (Thermo Fisher), and loaded onto a 6% TBE PAGE. Electrophoresis was performed at 4° at 100 V in 0.5x TBE buffer for 1.5-2 h. Gels were stained with SYBR Gold dye (Thermo Fisher), washed, and imaged with Typhoon 9500 FLA Imager (GE Healthcare Life Sciences).

EMSAs (Extended Data Figure 8) were performed identically to the procedure described above but with higher final glycerol concentration in order to better observe effects of point mutations of Sox11 on its nucleosome-binding properties in a wider range of apparent affinities. Control EMSA in the 12% glycerol buffer are shown in panels A and B of Extended Data Figure 8.

### **Digestion assay**

250 ng of nucleosome or DNA were mixed on ice with increasing amounts of Sox11 in digestion buffer (20 mM Hepes pH=7.5, 30 mM NaCl, 10 mM magnesium acetate, 0.1 mg/ml BSA). 0.125 units of restriction enzyme BfuCI (NEB) were added to each reaction. Samples were incubated at 37°C for 30 min, and the enzyme was inactivated by incubating at 65°C for 20 min. Samples were then incubated with proteinase K, and then were loaded onto a non-

Dodonova et al., Nucleosome-bound Sox2 and –Sox11 structures elucidate pioneer factor function

denaturing 4-20% TBE-gel. Electrophoresis was performed at 4°C at 180 V in 1x TBE buffer for 40 min. Gels were stained with SYBR Gold dye (Thermo Fisher), washed, and imaged with a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences). Two independent experiments were performed both for the DNA and Nucleosome digestion assays. Band intensities for the digestion product were measured in ImageJ according to standard routine<sup>47</sup>. Average values with standard deviations were plotted in Extended Data Figure 10A.

### **Crystallization and X-ray structure determination**

DNA oligonucleotides [TATTGTTTATTTTGTT and AACAAAATAACAATA] were synthesized and PAGE-purified by IDT. Complimentary oligonucleotides were annealed by heating to 95°C and stepwise cooling to 4°C (1°/90 s) at a concentration of 1.5 mM. Concentrated purified Sox11 DBD and 16-mer DNA were mixed in 1:1.2 ratio and incubated on ice for 30 min. Crystallization was achieved by the hanging drop vapor diffusion method at 20°C by mixing 1 µl of sample solution with 1 µl of reservoir solution containing 100 mM NaOAc pH 4.5, 200 mM CaOAc, 17% PEG400. Crystals were cryo-protected by 35% PEG400 (v/v) in the final storage solution and flash frozen in liquid nitrogen.

X-ray diffraction data were collected at beamline X10SA at the Swiss Light Source (Villigen, CH) using a Pilatus 6M detector. Data were indexed and integrated using XDS and scaled using XSCALE<sup>48</sup>. The structure was solved by molecular replacement with PHASER<sup>49</sup>, using the structure of the free Sox2 (PDB id 1GT0<sup>50</sup>) as the search model. The crystals belonged to space group P6<sub>1</sub> and diffracted to a resolution of 2.5 Å. The asymmetric unit contained two protein-DNA complexes (Extended Data Figure 6B). Density modification and model building was carried out with phenix.autobuild and manually completed in Coot. The model was iteratively refined with phenix.refine and outliers were fixed in Coot. The final R<sub>free</sub>-factor was 26%. The final model contained Sox11 residues 46-122 and DNA nucleotides 1-14. Diffraction data and refinement statistics are summarized in Table S2.

### **Estimation of the effect of Mg<sup>2+</sup> on the Nucleosome–Sox11 structure**

Since nucleosomes are known to be sensitive to Mg<sup>2+</sup> concentration, we wanted to test if Mg<sup>2+</sup> affects the Nucleosome-Sox11 structure. Nucleosomes at 1.6 µM concentration were mixed with 20x molar excess of Sox11 TF at 4° C in the final buffer containing 20 mM HEPES pH=7.5, 30 mM NaCl, 1 mM DTT supplied with additional 1 mM MgCl<sub>2</sub>, 0.01 mM ZnCl<sub>2</sub>, 0.5% Glycerol. Such sample buffer resembles the buffer used for our EMSAs except for glycerol and BSA, which should be avoided in cryo-EM samples. Next we used this sample for cryo-grid preparation. We collected a dataset on the Titan Krios equipped with a K3 Gatan detector, nominal pixel size 1.07 Å. Processing was done similarly to the other datasets described here. Final set contained 93493 particles. After light 3D classification (removing low resolution classes), the EM map (at 4Å resolution, with 0.73 sphericity) looked highly similar to our original Nucleosome–Sox11 structure. The comparison is shown in the Extended Data Figure 13. We concluded that Mg<sup>2+</sup> ions do not alter Nucleosome-Sox structure. We did not further process the dataset.

**EXTENDED DATA FIGURE LEGENDS**

**A**

DNA-1:

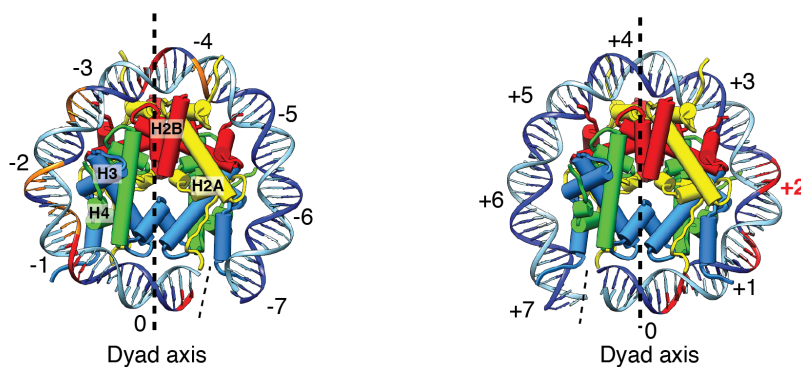


-43 -37 -28 -20 -14 -10  
 -73 ATCTACACGACGCTCTTCCGATCTAAT**TTATGTTTGT**TAGCGTTATACTAT**TTCTAA****TTCTTTGT**TCGGTGGT

A (SHL0)

**TTGTTTATTTTGTTCCTTTGT**GCGTTCAGCTTAATGCCTAACGACACTCGGAGATCGGAAGAGCACACGTGAT +73  
 +4 +8 +13 +21

**B**



**C**

DNA-1a: (One TTGT Sox site)

-73 ATCTACACGACGCTCTTCCGATCTAAT**TTAT**TCAGACTAGCG**TTATACTATTTCTAAT**TCAGACTTCGGTGGT

C (SHL0)

AGAC**TTAT**CAGACTCCT**TTGT**GCGTTCAGCTTAATGCCTAACGACACTCGGAGATCGGAAGAGCACACGTGAT +73

**Extended Data Figure 1 | DNA constructs. Related to Figure 1.**

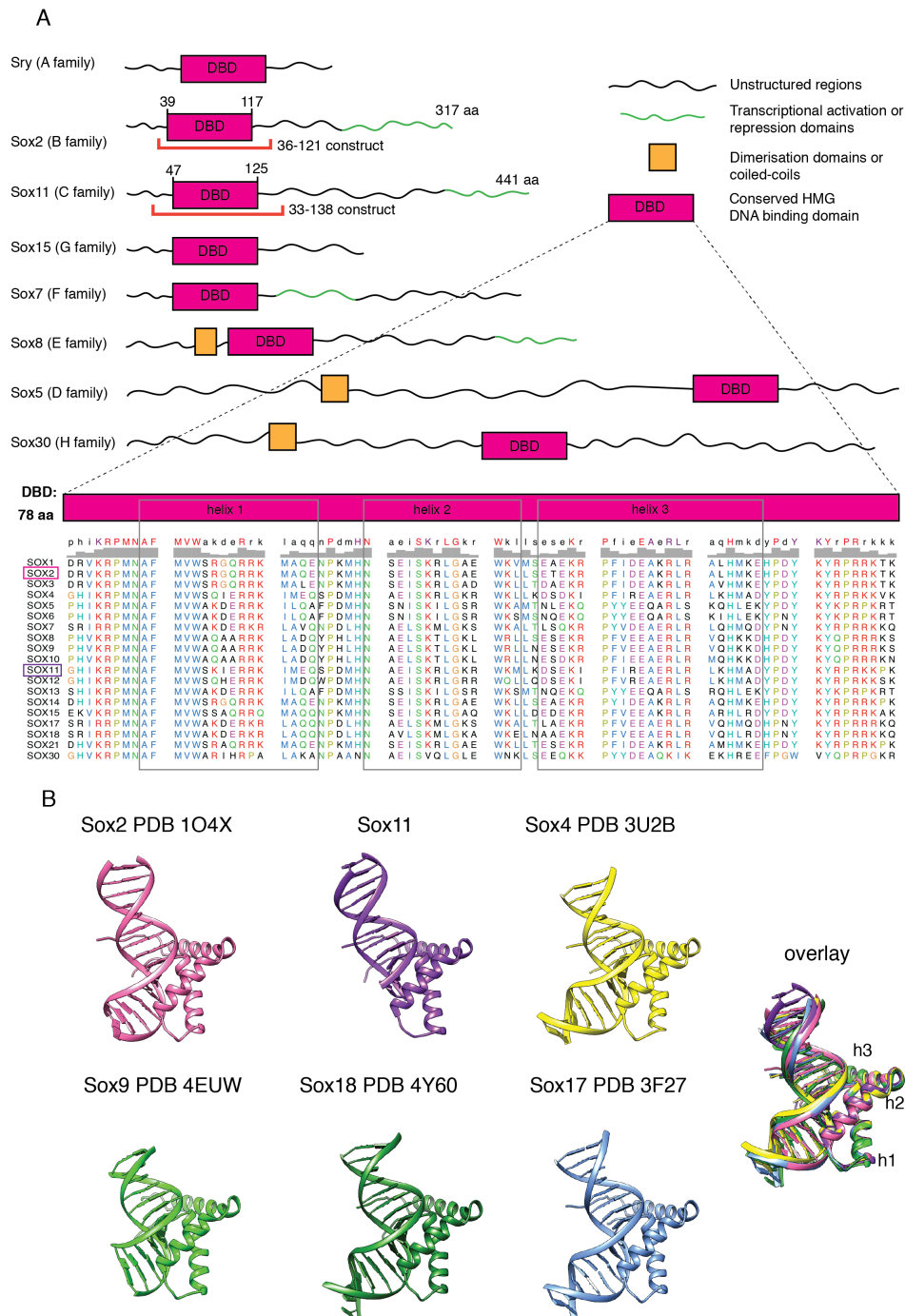
**a.** DNA-1 template sequence. Two DNA strands are colored in dark and light blue, canonical core motifs TTGT are colored in red, other TTXA motifs are orange. Only motifs that allow Sox binding to the DNA minor groove are considered. Position of the 3<sup>rd</sup> nucleotide of each motif in the complete DNA-1 sequence is indicated. Motifs at SHL+2 and SHL-2 are shown in bold font.

**b.** Top views of the nucleosome. Histones H2A, H2B, H3, and H4 are in yellow, red, blue, and green, respectively, DNA is in dark and light blue. Dyad axis is shown with a dashed line, SHLs are labeled.

**c.** DNA-1a template sequence. Only one TTGT motif is present (red).



Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

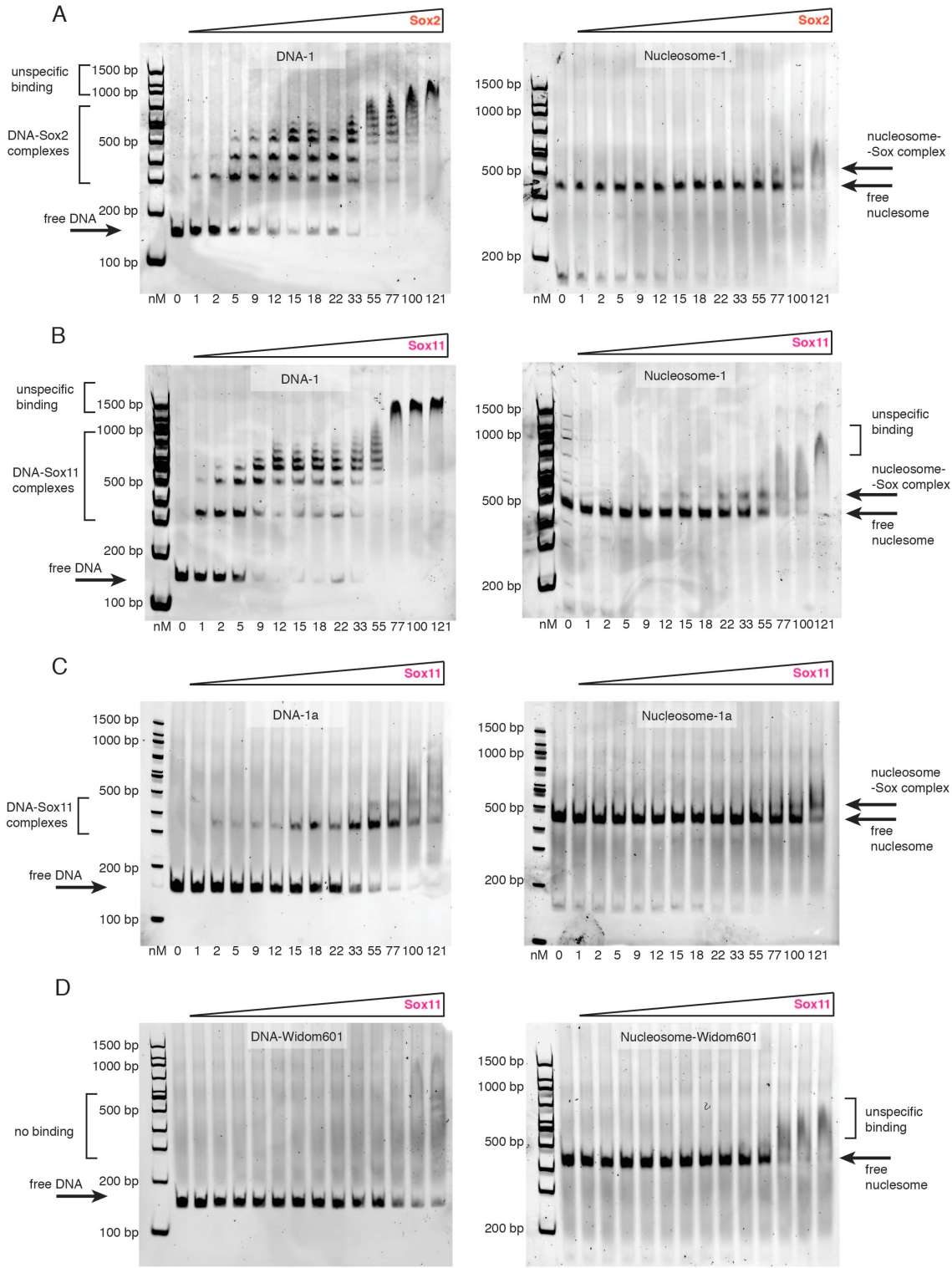


**Extended Data Figure 2 | Conservation of Sox family DBD sequence and structure. Related to Figure 1.**

**a.** Depiction of domain organization of the human Sox protein family. DBDs are depicted as pink rectangles, largely unstructured functionally diverse regions are shown as wavy lines. Protein constructs used in this study are marked. The alignment of Sox DBD sequences with ClustalO is shown below<sup>51</sup>. Helices 1-3 of the DBD are indicated with rectangles.

**b.** Structural conservation of Sox factors. Crystal and NMR structures of Sox family members: Sox2<sup>23</sup>, Sox11 (this study), Sox4<sup>52</sup>, Sox9, Sox18<sup>53</sup>, and also Sox17<sup>54</sup>. A superimposition of all structures is shown on the right and reveals that the structures are virtually identical.

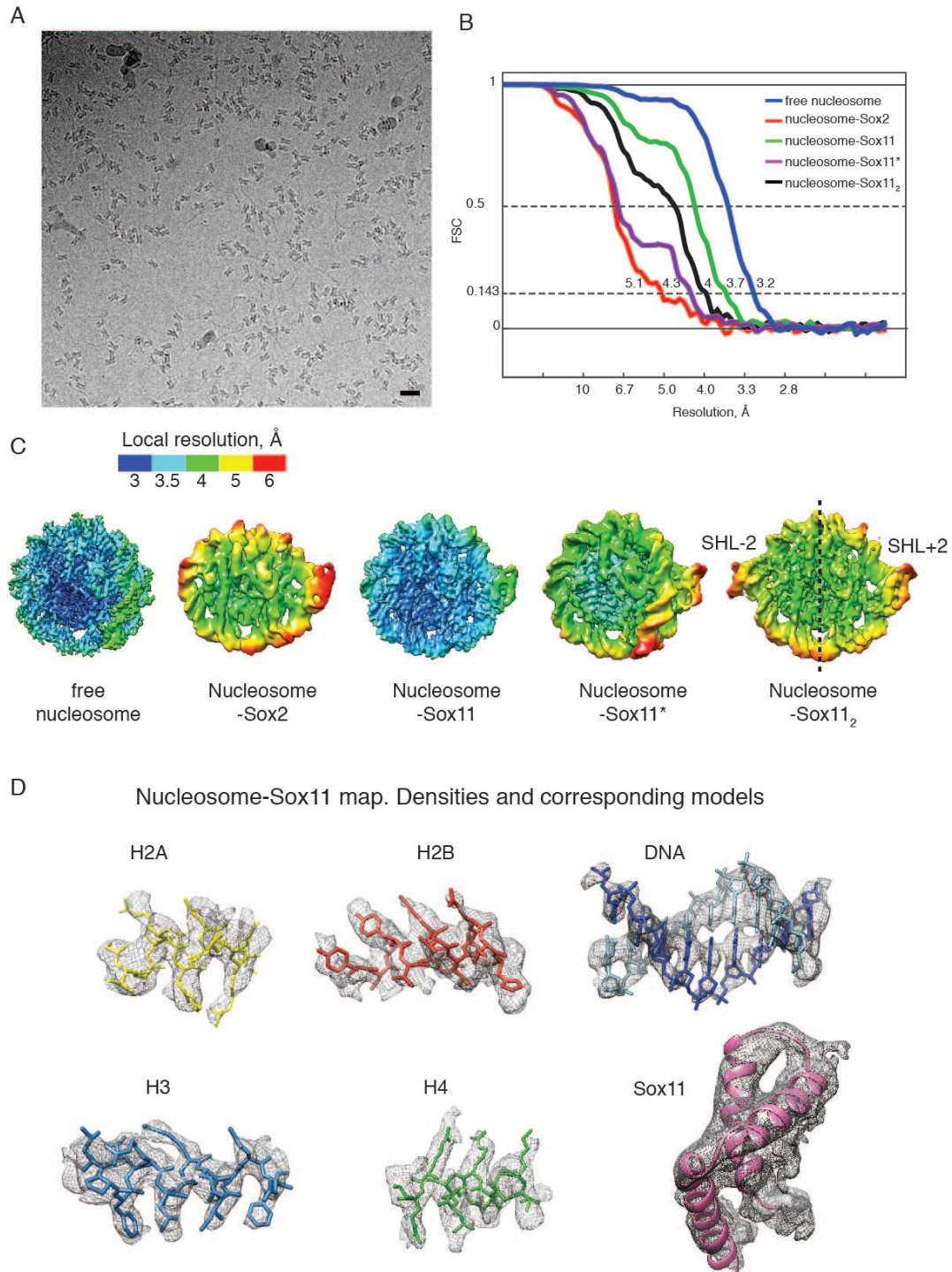
Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function



**Extended Data Figure 3 | Electrophoretic mobility shift assays. Related to Figures 1-3.**

Electrophoretic mobility shift assays (EMSAs) reveal formation of Sox factor complexes with DNA (left) or nucleosomes (right). DNA or nucleosome concentration is 1.1 nM. **(A)** EMSA of DNA1-Sox2 and Nucleosome1-Sox2 complexes. **(B)** EMSA of DNA1-Sox11 and Nucleosome1-Sox11 complexes. **(C)** EMSA of DNA1a-Sox11 and Nucleosome1a-Sox11 complexes. **(D)** EMSA of DNA Widom601-Sox11 and Nucleosome Widom601-Sox11 complexes. Relevant bands are labeled.

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function



**Extended Data Figure 4 | Global and local resolution of cryo-EM reconstructions. Related to Figures 1-3.**

**a.** Example micrograph (scale bar 20 nm) reveals a preferred orientation of nucleosomes.

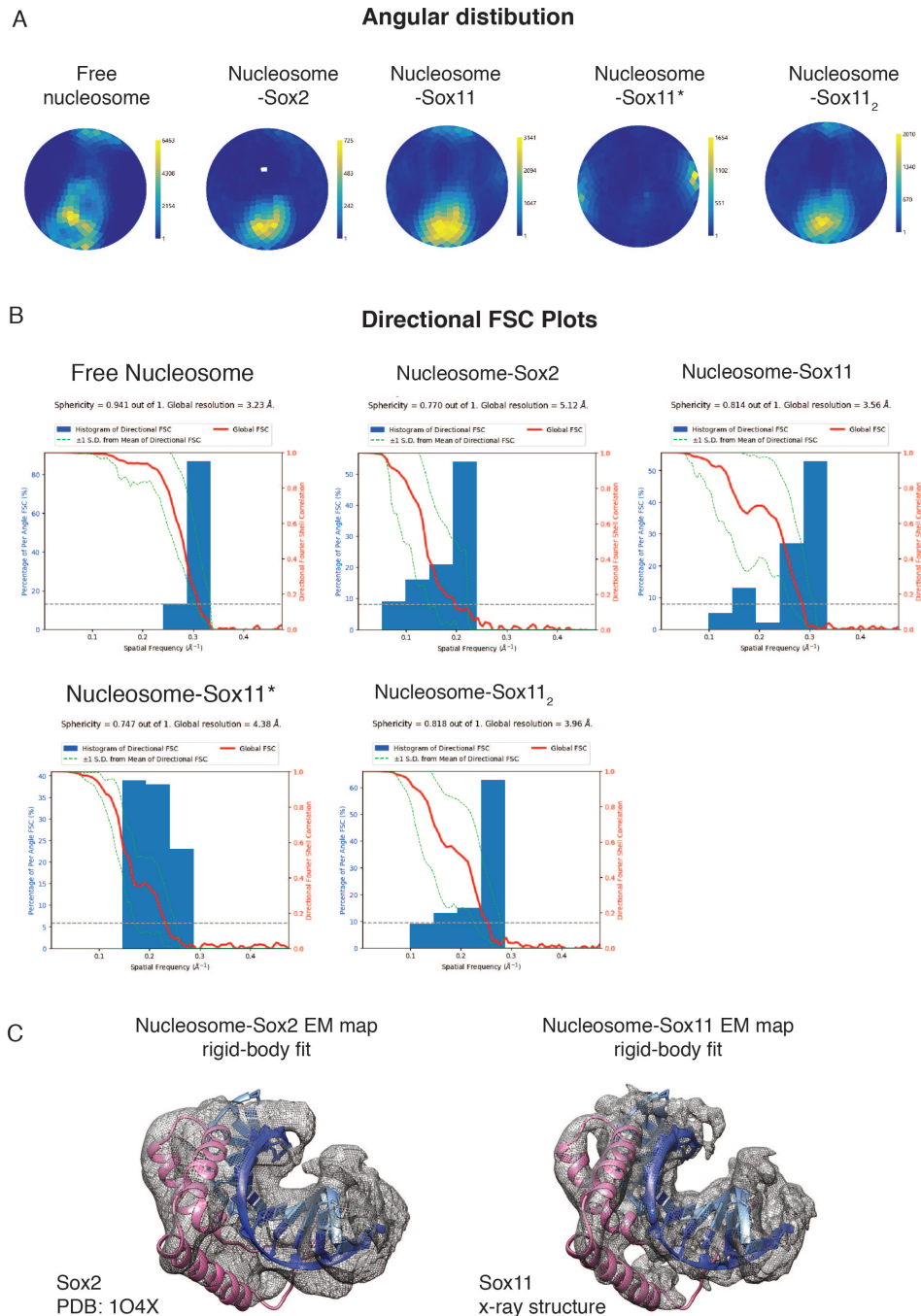
**b.** Fourier Shell Correlation (FSC) plots for all five reported reconstructions.

**c.** Local resolution distribution. In some maps the resolution varies from 3 Å (dark blue) through 4 Å (green) to 6 Å (red). The lower resolution of some DNA regions indicates flexibility.

**d.** Visualization of different regions of the nucleosome-Sox11 map. In the nucleosome core histone side chains are clearly visible, while Sox density has slightly lower resolution, but helical densities are clearly distinguishable.



Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

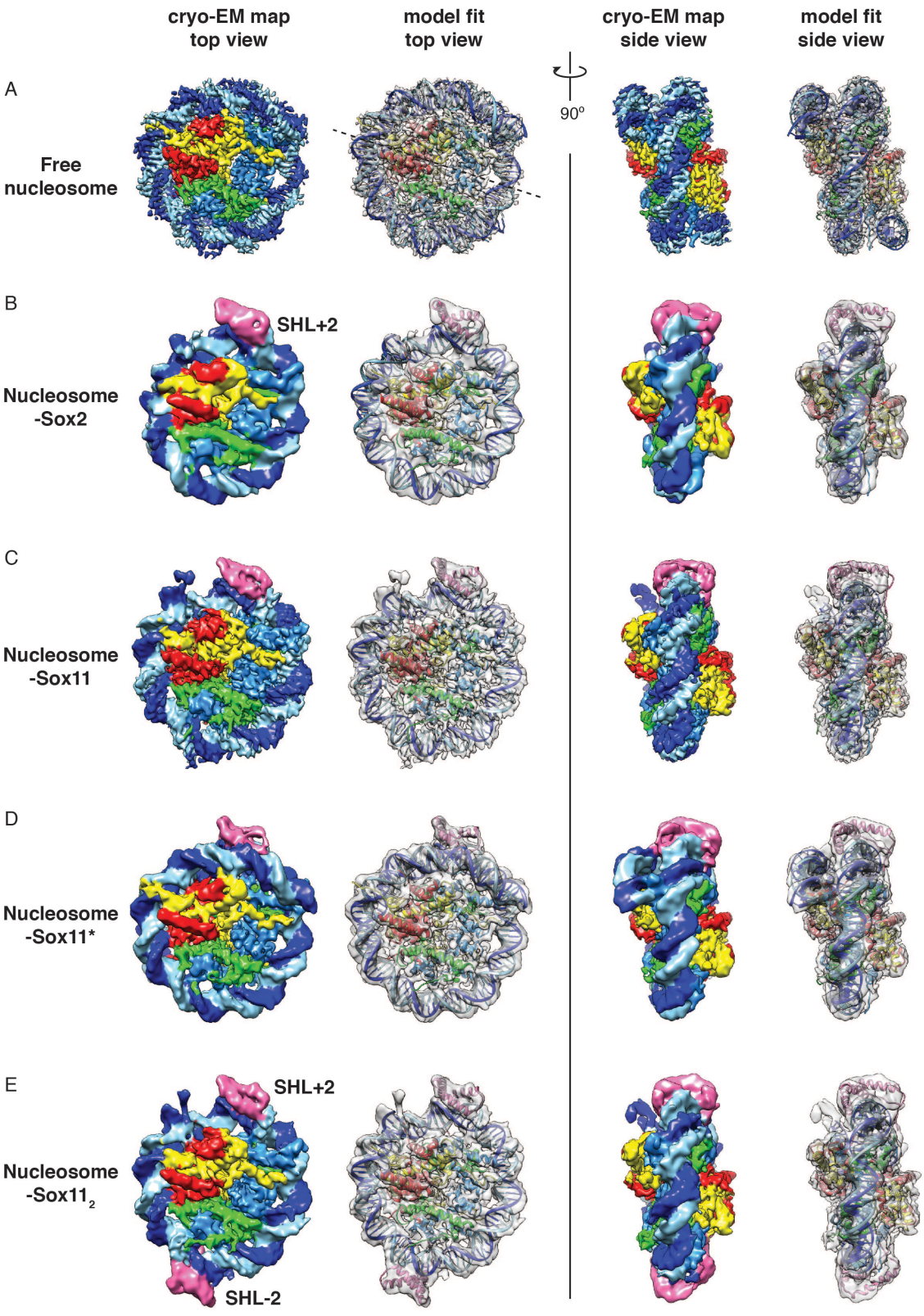


**Extended Data Figure 5 | Angular distribution and directional FSC plots. Related to Figures 1-3.**

**a.** Angular distribution plots for the five reconstructions. Scale shows number of particles assigned with a certain angle (blue – low number of particles, yellow - high).

**b.** Directional FSC plots for the five reconstructions calculated on the 3DFSC server<sup>43</sup>. Sphericity as a degree of anisotropy present in reconstructions is indicated above each plot. Histograms (blue bars) indicated portion of voxels with certain resolution in the EM maps.

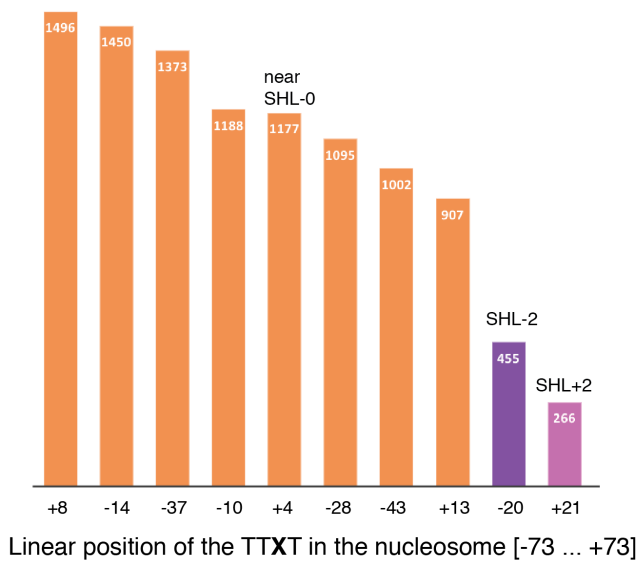
**c.** Rigid body fit of the Sox2-DNA structure (PDB 1O4X) into the nucleosome-Sox2 EM map (left). Rigid body fit of the Sox11-DNA X-ray structure into the nucleosome-Sox11 EM map (right). The region containing Sox and a short DNA stretch were isolated from the rest of the map for clarity.



**Extended Data Figure 6 | Gallery of cryo-EM structures. Related to Figures 1-4.**

EM maps and corresponding models of all reported structures. Top views (left) and side views (right) are related by a 90-degree rotation. EM maps are colored based on the fitted model (as in Figure 1) or transparent.

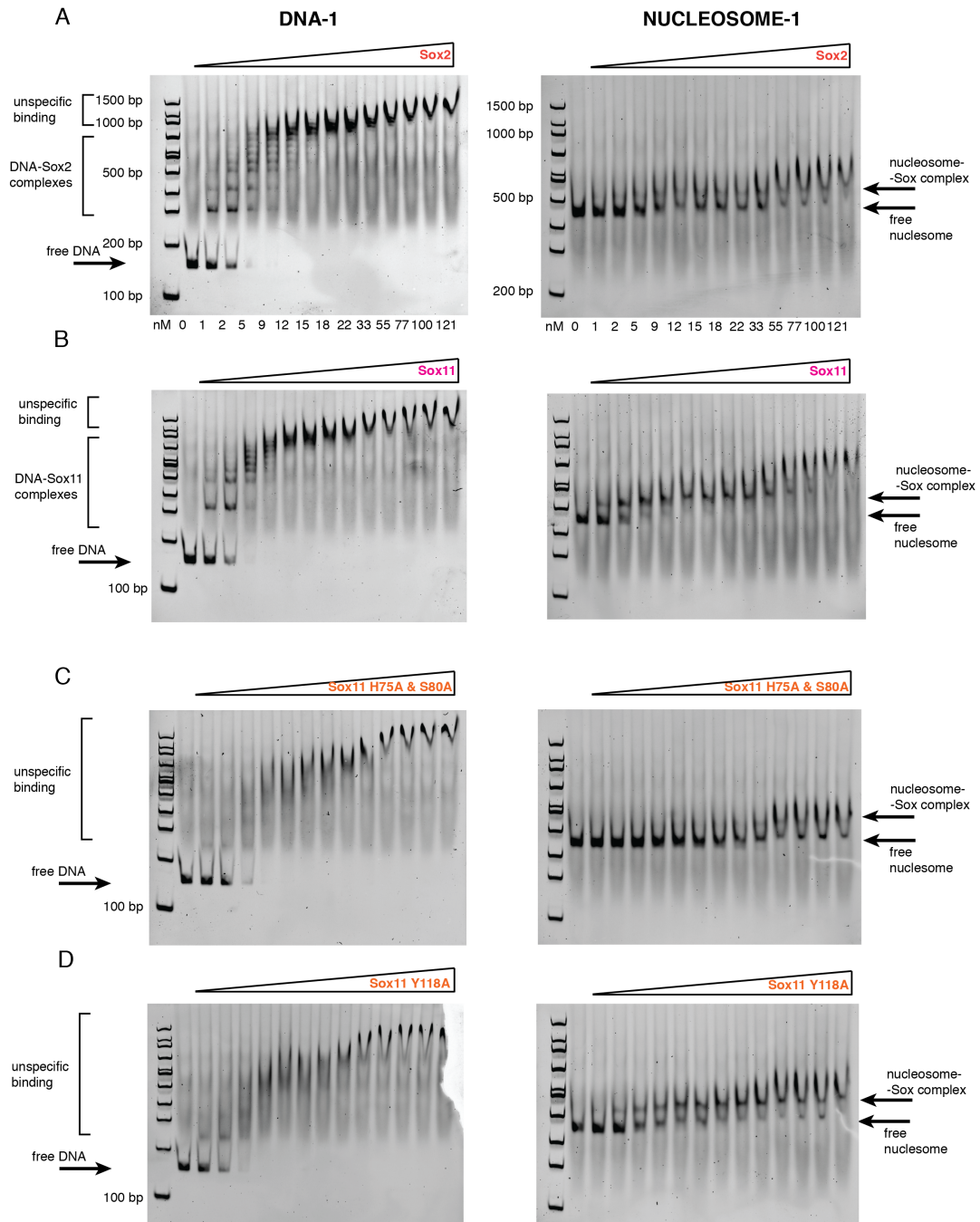
Clash score between Sox and canonical nucleosome at all positions with TTX motifs



**Extended Data Figure 7 | Clash score analysis between Sox TF and fully-wrapped nucleosome. Related to Figures 1-3.**

Structure of Sox2 bound to a short DNA fragment (PDB 1O4X) was aligned to each of the motifs present in the Nucleosome-1 sequence and allowing binding of Sox to the minor groove, and then number of clashes was calculated using *findclash* command in Chimera software. Sox2 binding to motifs located at SHL±2 is predicted to give rise to the least amount of clashes with DNA and histones compared to other motif locations. Exact locations of the 3<sup>rd</sup> motif nucleotide are indicated at the bottom of the plot.





**Extended Data Figure 8 | Electrophoretic mobility shift assays. Related to Figures 1-3.**

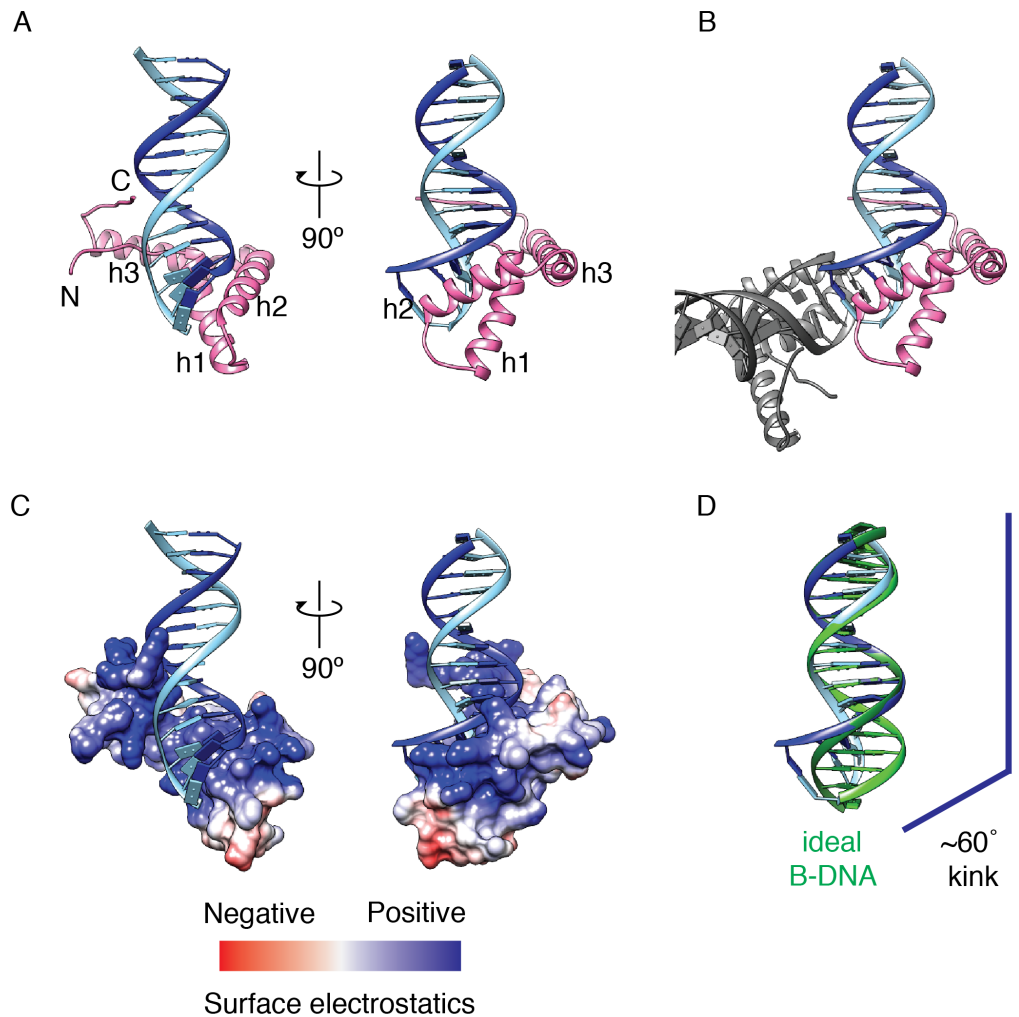
Electrophoretic mobility shift assays (EMSAs) reveal formation of Sox factor complexes with DNA (left) or nucleosomes (right). In order evaluate the mutants better and to observe a wider range of the binding curve, a higher amount of glycerol (12% final concentration) was used here as compared to the Extended Data Figure 3 (5% glycerol), thus the apparent affinity for is higher. DNA or nucleosome concentration is 1.1 nM

**a.** EMSA of DNA1 and Nucleosome1 Sox2 complexes.

**b.** EMSA of DNA1-Sox11 and Nucleosome-1 Sox11 wild type complexes.

**c.** EMSA of DNA-1 and Nucleosome-1 with Sox11 mutant H75A+S80A complexes.

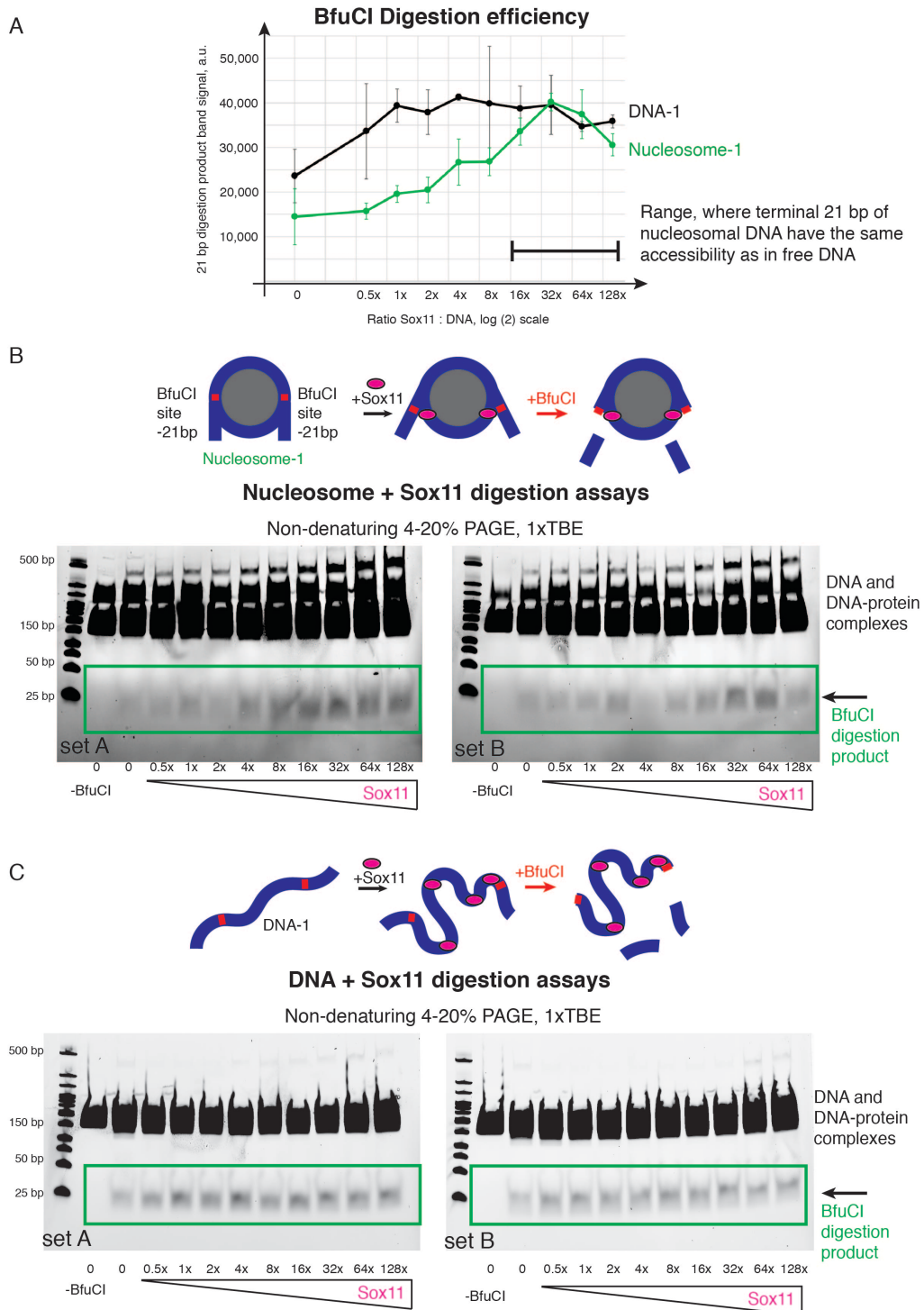
**d.** EMSA of DNA-1 and Nucleosome-1 with Sox11 Y118A mutant complexes. Relevant bands are labeled.



**Extended Data Figure 9 | X-ray structure of the Sox11-DNA complex. Related to Figures 1-3.**

- a.** Ribbon representation of the structure in bottom view (left) and side view (right). Helices 1-3 of the Sox11 DBD are labeled.
- b.** Two copies of Sox11-DNA in the asymmetric unit. The contact between the two is mediated by DNA-stacking.
- c.** Surface representation of the Sox11 molecule. DNA is engulfed by the strongly positively-charged inner surface of the Sox11 DBD domain.
- d.** Comparison of the observed DNA conformation with canonical B-DNA (green). Sox11 introduces a kink into DNA that is typical for HMG box proteins.



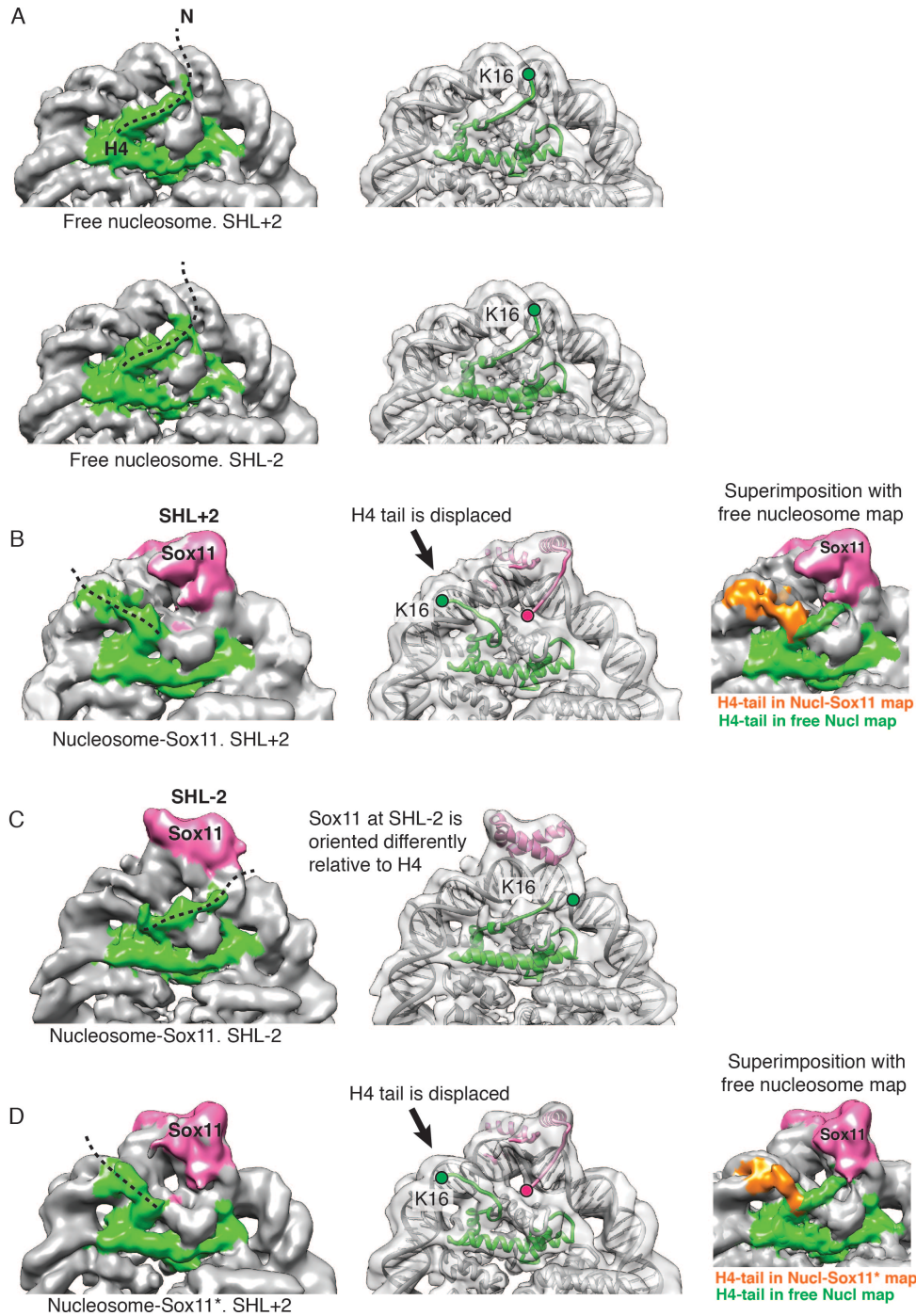


**Extended Data Figure 10 | Nucleosome DNA end unwrapping. Related to Figures 1, 3.**

**a.** BfuCI digestion assay for free DNA-1 (black) and nucleosome-1 (green) in the presence of increasing amounts of Sox11. Each experiment was repeated twice: standard deviations are shown. Band intensity was calculated using standard routine in ImageJ<sup>47</sup>.

**b.** BfuCI digestion assay (two experiments) of the Nucleosome-1 in the presence of increasing amounts of Sox11. The restriction site (21 nucleotides away from the entry and exit sites of the nucleosome) becomes more accessible with higher concentrations of Sox11.

**c.** BfuCI digestion assay (two experiments) of the free DNA-1 in the presence of increasing amounts of Sox11. The amount of digestion product (21 bp) increases slightly in the low Sox11 concentration range and then stays constant over a broad concentration range.



**Extended Data Figure 11 | Repositioning of the N-terminal tail of histone H4. Related to Figure 4.**

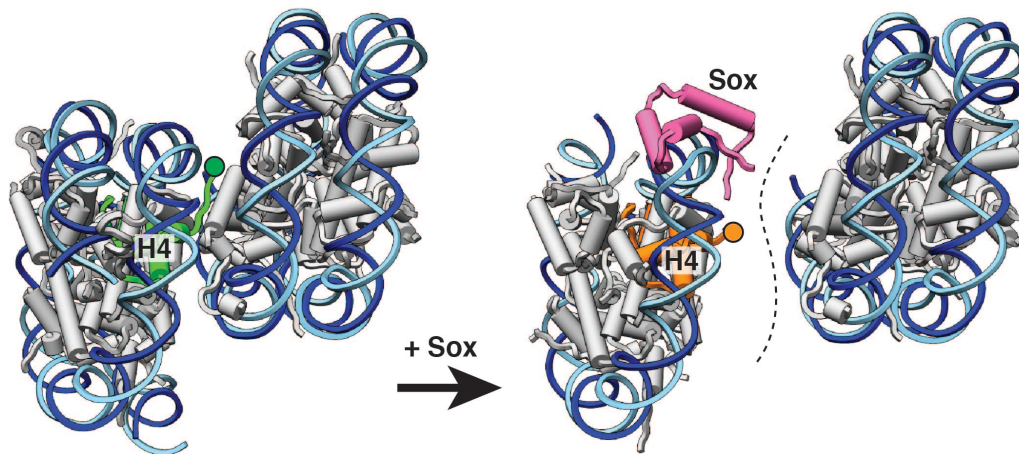
**a.** EM map of the free nucleosome with Gaussian smoothening filter applied for clarity in Chimera<sup>44</sup>. Views of SHL+2 and SHL-2 are shown to illustrate the position of the H4 tail (marked with a dashed line on the left). H4 histone and corresponding density are colored in green. Residue K16 is marked with a circle.

**b.** EM map of the nucleosome-Sox11 complex (Gaussian smoothened) with Sox11 (pink) located at the SHL+2. On the right a superimposition with the free nucleosome EM map is shown to highlight different H4 tail orientations. H4 tail orientation in the nucleosome-Sox11\* map is shown in orange.

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function

**c.** EM map of the nucleosome-Sox11 complex with Sox11 (pink) located at the SHL-2. Note that in this location Sox is oriented differently and does not clash with and does not reposition the H4 tail.

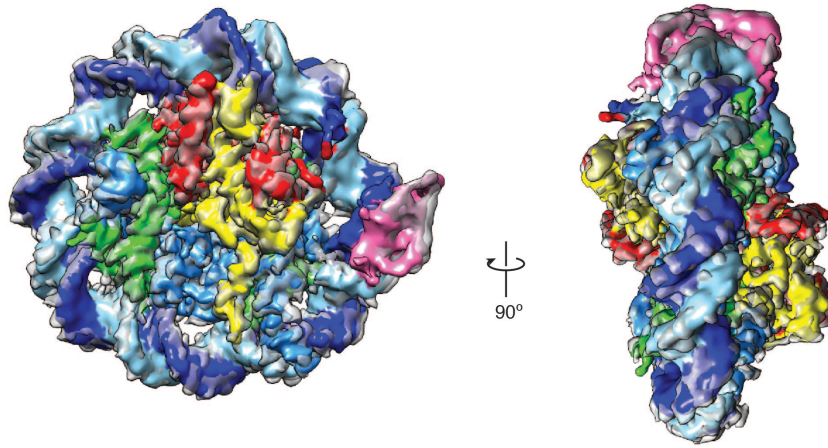
**d.** EM map of the nucleosome-Sox11\* complex (Gaussian smoothed). The H4 N-terminal tail is repositioned compared with the free nucleosome structure. Note that repositioning of the H4 tail was reported in case of strong distortions in the nucleosome structure: the H4 tail moved into the neighboring major DNA groove<sup>27</sup>. On the right a superimposition with the free nucleosome EM map is shown to highlight different H4 tail orientations. H4 tail orientation in the nucleosome-Sox11\* map is shown in orange.



**Extended Data Figure 12 | Sox binding might impair nucleosome stacking. Related to Figure 4.**

Sox binding repositions the H4 N-terminal tail and may impair nucleosome stacking. Side view of two stacking nucleosomes from PDB 1AOI<sup>22</sup>. H4 (green) interacts with the acidic patch on the neighboring H2A-H2B histone dimer. H4 tail repositioning is incompatible with nucleosome stacking.

Dodonova et al., Nucleosome-bound Sox2 and –Sox11 structures elucidate pioneer factor function



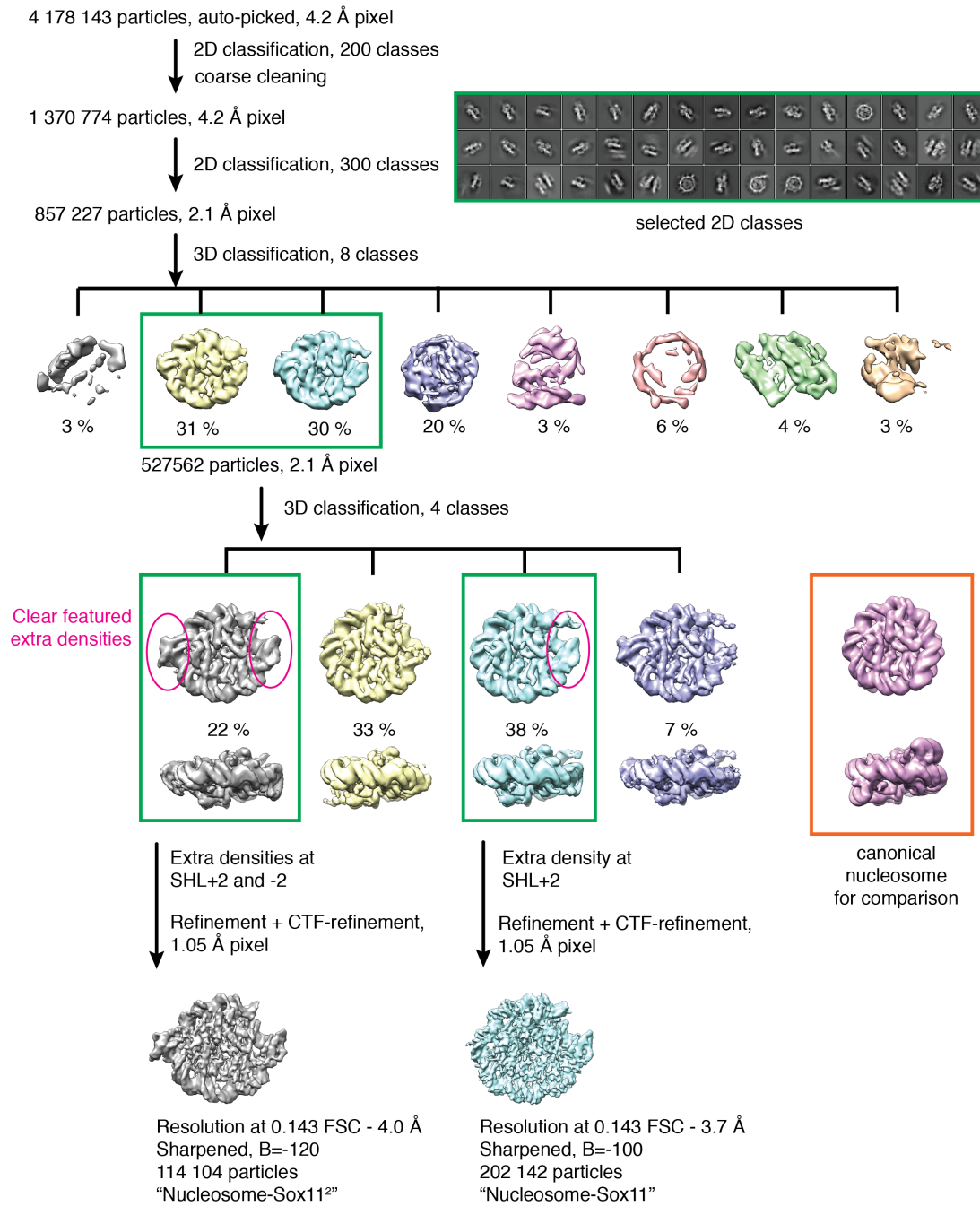
Superimposition of a Nucleosome-Sox11 in the presence (colored) or absence (grey) of Mg<sup>2+</sup> ions.  
Cross-correlation=0.94

**Extended Data Figure 13 | Sox binding to nucleosome is independent of Mg<sup>2+</sup>. Related to Figures 1-3.**

Superimposition of two nucleosome-Sox11 EM maps obtained in the presence of 1 mM MgCl<sub>2</sub> (sample was in “EMSA”-buffer, see Methods, shown as grey density) or in the absence of MgCl<sub>2</sub> (colored). Magnesium does not influence the structure of the Sox-nucleosome complex. Cross-correlation between two unmasked maps is 0.94, as measured in Chimera software.



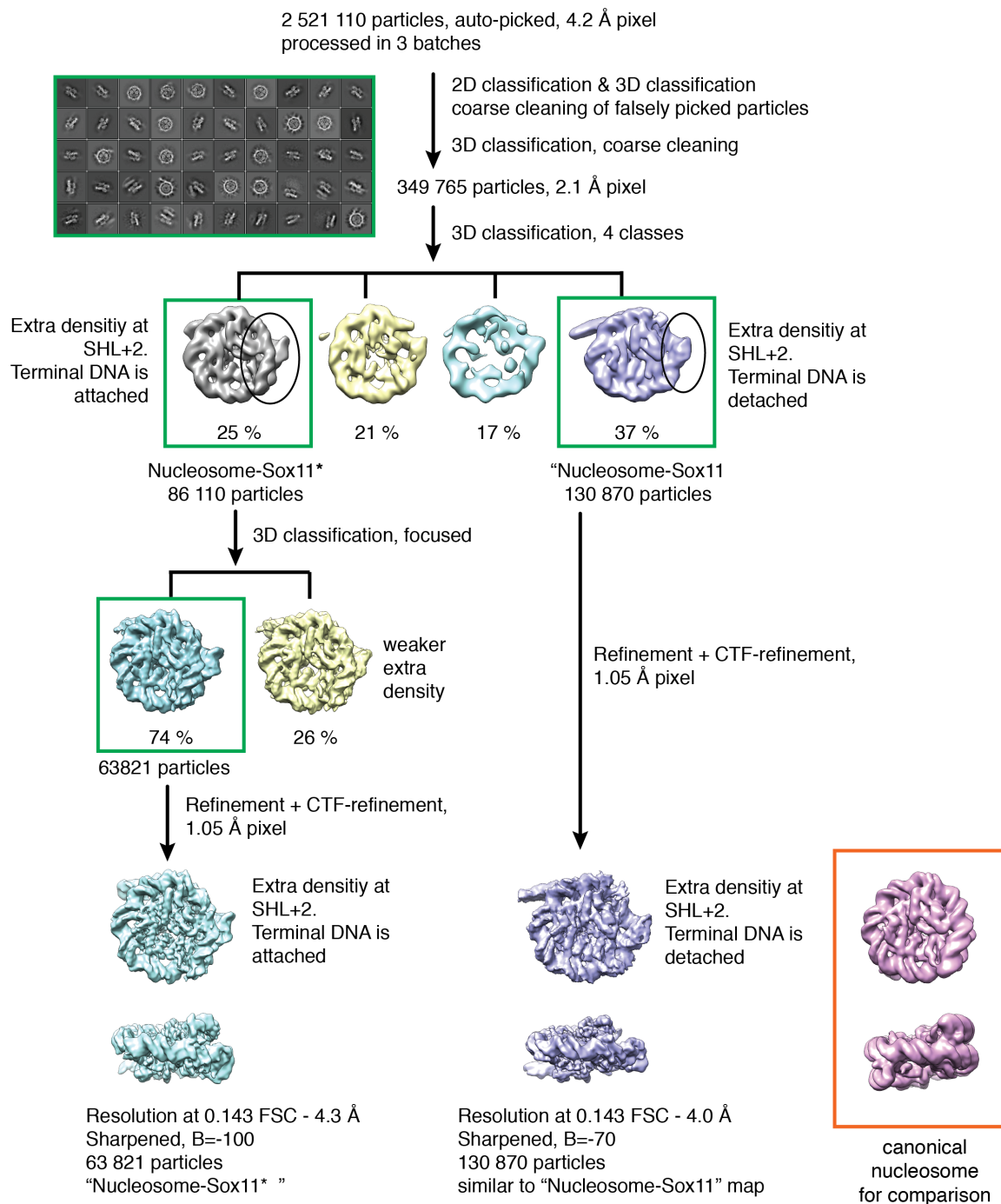
Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function



**Extended Data Figure 14 | Flow chart for cryo-EM structure determination of Nucleosome-Sox11<sub>2</sub> and Nucleosome-Sox11 complexes with 147bp DNA-1. Related to Figures 1-3.**

The processing chart for the 147bp DNA-1 nucleosome and Sox11 is depicted. The two resulting structures are "Nucleosome-Sox11<sub>2</sub>" and Nucleosome-Sox11.

Dodonova et al., Nucleosome-bound Sox2 and -Sox11 structures elucidate pioneer factor function



**Extended Data Figure 15 | Flow chart for cryo-EM structure determination of the Nucleosome-Sox11\* complex with 151bp DNA-1. Related to Figures 1-3.**

The processing chart for the 151bp DNA-1 nucleosome and Sox11 is depicted. The two resulting structures are “Nucleosome-Sox11\*” and a map virtually identical to the Nucleosome-Sox11 from the 147bp sample.

**EXTENDED DATA TABLES**

Sample	Free nucleosome	Nucleosome-Sox2	Sox11 + 151bp DNA-1 nucleosome	Sox11 + 147bp DNA-1 nucleosome	
Subset			Nucleosome-Sox11*	Nucleosome-Sox11	Nucleosome-Sox11 <sup>2</sup>
<b>Sample preparation and data acquisition</b>					
Buffer composition	20 mM HEPES pH=7.5, 1 mM EDTA, 30 mM NaCl, 2mM DTT				
Cryo-grid preparation	R 2/1 Cu 300 mesh grids, glow-discharged for 120 s, Vitrobot chamber set to 100% humidity and 16° C				
Voltage	300 kV				
Magnification	130,000 x				
Pixel size	1.05 Å				
Electron dose	45 e <sup>-</sup> /Å <sup>2</sup>				
Frame number	40				
Defocus range	0.9 μm – 3.4 μm				
<b>Processing</b>					
Final number of particles	368,270	32,301	63,821	202,142	114,104
Map resolution at 0.143 FSC	3.2 Å	5.1 Å	4.3 Å	3.7 Å	4.0 Å
Local resolution range	3.0 - 3.9 Å	4.5 - 8.3 Å	4.1 - 6.6 Å	3.5 - 5.5 Å	3.7 - 5.3 Å
Map sharpening B factor	-75	-100	-100	-100	-120

**Extended Data Table 1 | Cryo-EM data collection and processing information.**



<b>Data collection</b>	
Wavelength (nm)	1
Resolution range (Å)	46 - 2.5 (2.6 - 2.5)
Space group	P 6 <sub>1</sub>
Unit cell dimensions, a b c (Å)	106.1 106.1 76.9
Unit cell dimensions, $\alpha$ $\beta$ $\gamma$ (°)	90 90 120
Total reflections	350790 (34371)
Unique reflections	17071 (1694)
Multiplicity	20.5 (20.3)
Completeness (%)	99.7 (98.7)
Mean I/sigma (I)	17.3 (0.8)
Wilson B-factor (Å <sup>2</sup> )	74
R-merge	0.13 (2.8)
R-meas	0.13 (2.9)
R-pim	0.03 (0.63)
CC1/2	1 (0.67)
CC*	1 (0.90)
<b>Refinement</b>	
Reflections used in refinement	17033 (1682)
Reflections used for R-free	853 (85)
R-work	0.23 (0.42)
R-free	0.26 (0.46)
CC(work)	0.94 (0.68)
CC(free)	0.97 (0.65)
Number of non-hydrogen atoms	2496
macromolecules	2490
solvent	6
Protein residues	151
RMS(bonds)	0.006
RMS(angles)	0.78
Ramachandran favored (%)	98.6
Ramachandran allowed (%)	1.4
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	7.75
Average B-factor (Å <sup>2</sup> )	104
DNA	128
Protein	83
Solvent	65
Number of TLS groups	1

**Extended Data Table 2 | X-ray data collection, refinement, and validation statistics for the Sox11-DNA complex.**

**Supplementary Video 1 | Overview of Nucleosome-Sox complexes.**

Pioneer factor binding remodels the nucleosome. Binding of Sox11 to the nucleosome leads to local DNA distortion and unwrapping of terminal DNA. The five cryo-EM structures reported here have been used to generate this animation. For details compare text.

## REFERENCES

- 1 Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143**, 1833-1837, doi:10.1242/dev.133900 (2016).
- 2 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).
- 3 Adachi, K., Suemori, H., Yasuda, S. Y., Nakatsuji, N. & Kawase, E. Role of SOX2 in maintaining pluripotency of human embryonic stem cells. *Genes Cells* **15**, 455-470, doi:10.1111/j.1365-2443.2010.01400.x (2010).
- 4 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:10.1016/j.cell.2018.01.029 (2018).
- 5 Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-192, doi:10.1038/nature08449 (2009).
- 6 Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**, 1798-1812, doi:10.1101/gr.139105.112 (2012).
- 7 Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76-81, doi:10.1038/s41586-018-0549-5 (2018).
- 8 Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-289 (2002).
- 9 Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* **37**, 76-81, doi:10.1016/j.gde.2015.12.003 (2016).
- 10 Boller, S., Li, R. & Grosschedl, R. Defining B Cell Chromatin: Lessons from EBF1. *Trends Genet* **34**, 257-269, doi:10.1016/j.tig.2017.12.014 (2018).
- 11 Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285, doi:10.1016/j.cell.2014.01.062 (2014).
- 12 Velychko, S. *et al.* Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs. *Cell Stem Cell*, doi:10.1016/j.stem.2019.10.002 (2019).
- 13 Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).
- 14 Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Mol Cell* **75**, 562-575 e565, doi:10.1016/j.molcel.2019.05.025 (2019).
- 15 Malik, V. *et al.* Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Nat Commun* **10**, 3477, doi:10.1038/s41467-019-11054-7 (2019).
- 16 Julian, L. M., McDonald, A. C. & Stanford, W. L. Direct reprogramming with SOX factors: masters of cell fate. *Curr Opin Genet Dev* **46**, 24-36, doi:10.1016/j.gde.2017.06.005 (2017).
- 17 Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-4144, doi:10.1242/dev.091793 (2013).
- 18 Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720-1723, doi:10.1126/science.1162327 (2009).
- 19 Bergsland, M. *et al.* Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* **25**, 2453-2464, doi:10.1101/gad.176008.111 (2011).

- 20 Wiebe, M. S., Nowling, T. K. & Rizzino, A. Identification of novel domains within Sox-2 and Sox-11 involved in autoinhibition of DNA binding and partnership specificity. *J Biol Chem* **278**, 17901-17911, doi:10.1074/jbc.M212211200 (2003).
- 21 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16, doi:10.1385/1-59259-681-9:1 (1999).
- 22 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).
- 23 Williams, D. C., Jr., Cai, M. & Clore, G. M. 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, doi:10.1074/jbc.M309790200 (2004).
- 24 Hou, L., Srivastava, Y. & Jauch, R. Molecular basis for the genome engagement by Sox proteins. *Semin Cell Dev Biol* **63**, 2-12, doi:10.1016/j.semcd.2016.08.005 (2017).
- 25 Voong, L. N. *et al.* Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* **167**, 1555-1570 e1515, doi:10.1016/j.cell.2016.10.049 (2016).
- 26 Li, S., Zheng, E. B., Zhao, L. & Liu, S. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep* **28**, 2689-2703 e2684, doi:10.1016/j.celrep.2019.07.103 (2019).
- 27 Bilokapic, S., Strauss, M. & Halic, M. Structural rearrangements of the histone octamer translocate DNA. *Nat Commun* **9**, 1330, doi:10.1038/s41467-018-03677-z (2018).
- 28 Hall, M. A. *et al.* High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat Struct Mol Biol* **16**, 124-129, doi:10.1038/nsmb.1526 (2009).
- 29 Bilokapic, S., Strauss, M. & Halic, M. Histone octamer rearranges to adapt to DNA unwrapping. *Nat Struct Mol Biol* **25**, 101-108, doi:10.1038/s41594-017-0005-5 (2018).
- 30 Donovan, B. T., Chen, H., Jipa, C., Bai, L. & Poirier, M. G. Dissociation rate compensation mechanism for budding yeast pioneer transcription factors. *Elife* **8**, doi:10.7554/eLife.43008 (2019).
- 31 Pepenella, S., Murphy, K. J. & Hayes, J. J. Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. *Chromosoma* **123**, 3-13, doi:10.1007/s00412-013-0435-8 (2014).
- 32 Gordon, F., Luger, K. & Hansen, J. C. The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. *J Biol Chem* **280**, 33701-33706, doi:10.1074/jbc.M507048200 (2005).
- 33 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847, doi:10.1126/science.1124000 (2006).
- 34 Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol* **327**, 85-96 (2003).
- 35 Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380, doi:10.1126/science.1251413 (2014).
- 36 Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet* **43**, 607-611, doi:10.1038/ng.825 (2011).

- 37 Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375**, 23-44 (2004).
- 38 Lis, J. T. & Schleif, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res* **2**, 383-389 (1975).
- 39 Tegunov, D. & Cramer, P. Real-time cryo-EM data pre-processing with Warp. *bioRxiv*, doi:<https://doi.org/10.1101/338558> (2018).
- 40 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
- 41 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).
- 42 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).
- 43 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).
- 44 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 45 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
- 46 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 47 Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529, doi:10.1186/s12859-017-1934-z (2017).
- 48 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
- 49 McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41, doi:10.1107/S0907444906045975 (2007).
- 50 Remenyi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-2059, doi:10.1101/gad.269303 (2003).
- 51 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).
- 52 Jauch, R., Ng, C. K., Narasimhan, K. & Kolatkar, P. R. The crystal structure of the Sox4 HMG domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. *Biochem J* **443**, 39-47, doi:10.1042/BJ20111768 (2012).
- 53 Klaus, M. *et al.* Structure and decoy-mediated inhibition of the SOX18/Prox1-DNA interaction. *Nucleic Acids Res* **44**, 3922-3935, doi:10.1093/nar/gkw130 (2016).
- 54 Palasingam, P., Jauch, R., Ng, C. K. & Kolatkar, P. R. The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J Mol Biol* **388**, 619-630, doi:10.1016/j.jmb.2009.03.055 (2009).