Extended data figures and tables for

DNA Mechanical Strain Steers Transcription Factor Recognition

Yumi Minyi Yao[1], Michael P. O'Hagan[1], Karn Onoon[2], Lihee Givon[1], Shelly Hamer-Rogotner[3], Raul Salinas[4], Raj V. Nithun[5], Naama Kessler[1], Muhammad Jbara[5], Orly Dym[3], Tanadet Pipatpolkai[6], Maria A. Schumacher[5], Ariel Afek[1*]

[1]Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel;

[2]Department of Biology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand;

[3]Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot 7610001, Israel;

[4]School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore;

[5]Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710, USA

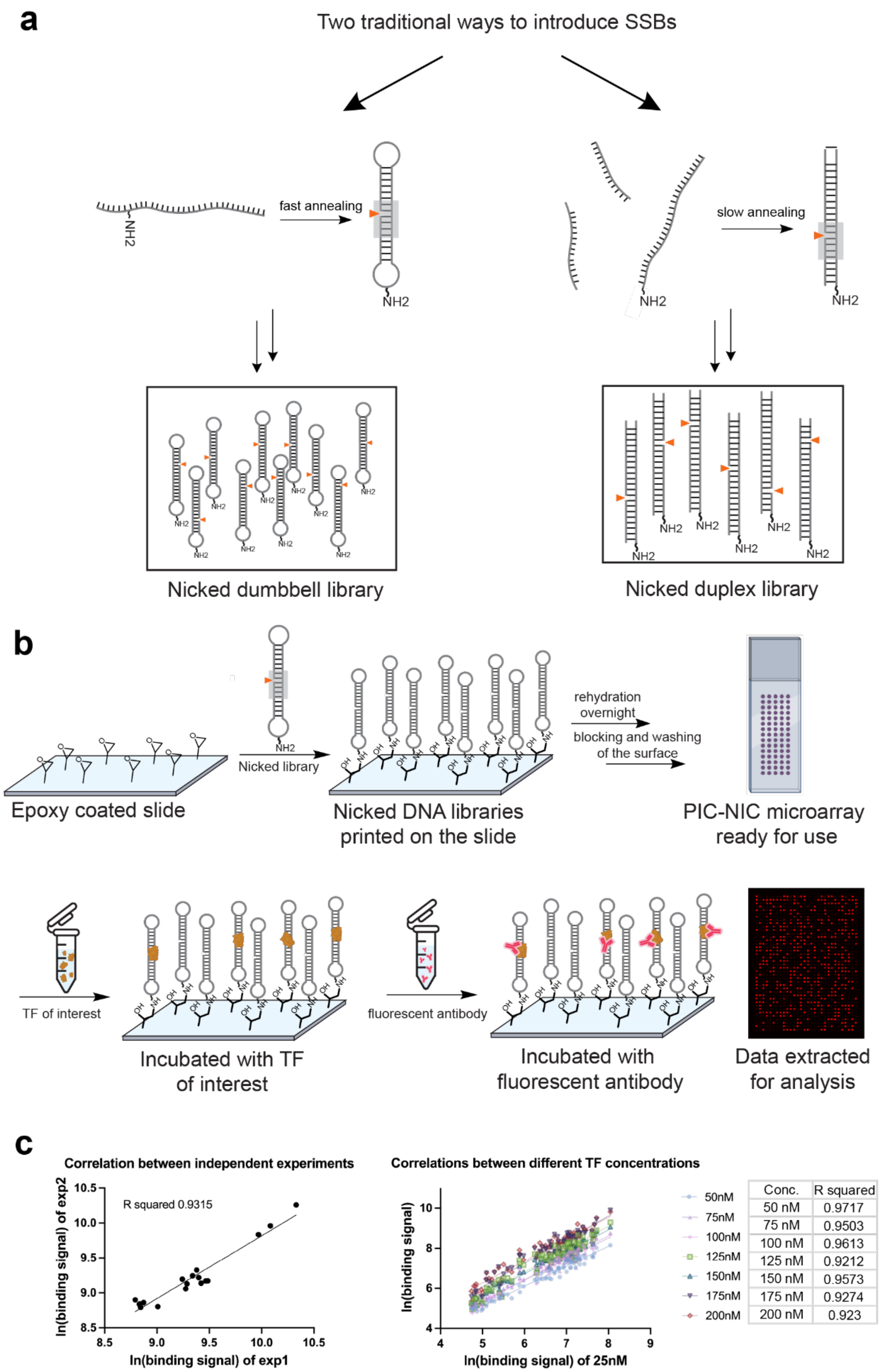
[6]School of Physical and Mathematical Sciences, Nanyang Technological University; 637371, Singapore

*Corresponding Author: ariel.afek@weizmann.ac.il

This PDF file includes:

Extended Data Figures 1-13
Extended Data Tables 1-3

Extended Data Fig. 1 PIC-NIC Pipeline and experimental reproducibility



Conc.	R squared
50 nM	0.9717
75 nM	0.9503
100 nM	0.9613
125 nM	0.9212
150 nM	0.9573
175 nM	0.9274
200 nM	0.923

Extended Data Fig. 1 PIC-NIC Pipeline and experimental reproducibility

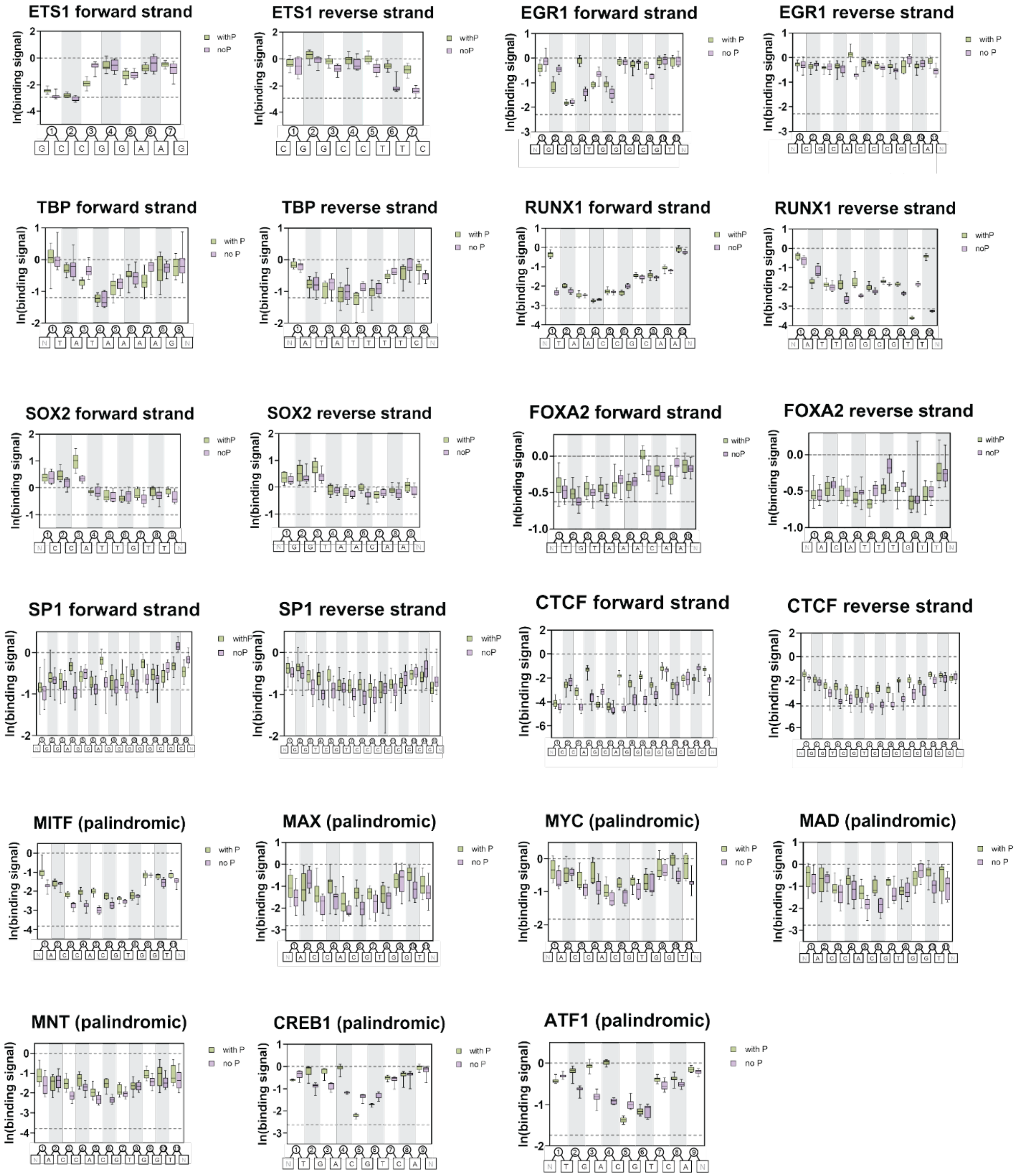
a. Two strategies to introduce SSBs into DNA TF binding sites using traditional approaches. *Left:* dumbbell-shaped DNA oligos fold on themselves upon fast annealing leaving a nick at the desired position; *Right:* slow annealing of three single-stranded DNA oligos leaves a nick at the desired position. Both methods were employed to yield oligo libraries of nicked TF binding sites. **b.** Schematic representation of the PIC-NIC experimental pipeline, for further details see Supplementary Methods. **c.** PIC-NIC experiments are highly reproducible and robust. *Left:* Binding measurements for the SOX2 protein show strong correlation between two independently synthesized replicate microarrays (Pearson's $R^2 = 0.9315$). *Right*: Data acquired at different TF concentrations remain highly consistent, demonstrating robustness across binding regimes.

Extended Data Fig. 2 Structural families and contact maps of the 15 TFs employed in this study

Extended Data Fig. 2 Structural families and contact maps of the 15 TFs employed in this study

Structural families, PDB ID and detailed protein-DNA H-bond contact maps for the 15 TFs showing backbone contacts (green) and base contacts (turquoise). Contact maps were produced in ChimeraX employing a general cutoff distance of 3.5 Å (tolerance 0.4 Å) and angle cutoff of 100° (tolerance 20°). The binding site shown in the figure is the same length as the binding site chosen for PIC-NIC experiments.

Extended Data Fig. 3 PIC-NIC binding profiles for the 15 TFs



Extended Data Fig. 3 PIC-NIC binding profiles for the 15 TFs

Box-and-whisker plots of binding signal of each TF to its respective binding site containing a nick at the indicated position, either with retention of 5’ phosphate (green) or removal of the phosphate (purple). The upper dashed line (y=0) represents the binding signal of the target TF to the intact DNA binding site. The lower dashed line represents the binding signal of the target TF to a non-specific site. Each box represents 10-20 spatially scattered replicates of the same DNA binding site on the array.

Extended Data Fig. 4 Within-family comparison of bHLH family

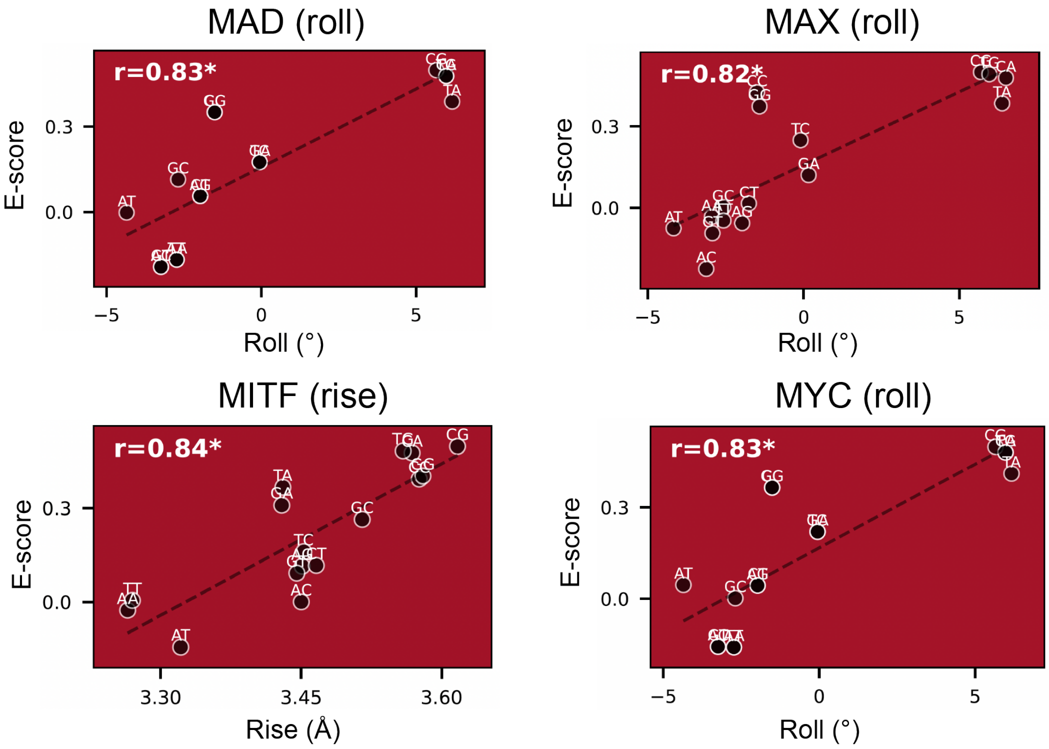
a

Structural family	Members in panel	Within-family profile similarity	Notes
bHLH	MITF, MAX, MYC, MAD, MNT	High	Consistent sensitivity at central palindrome position across all five members
bZIP	CREB1, ATF1	High	Consistent positional sensitivity pattern across both members
Zinc finger	EGR1, SP1, CTCF	Low	Divergent profiles, likely reflecting differences in finger number, binding site length, and contact geometry

b

Protein	Sequence	Feature	Feature type	NN _position	Predicted _position	Correlation
MAD	CCACGTGG	Roll	Base step	3	3	0.83
MAX	ACCACGTG	Roll	Base step	4	4	0.82
MITF	CACGTGA	Rise	Base step	2	2	0.84
MYC	CCACGTGG	Roll	Base step	3	3	0.83

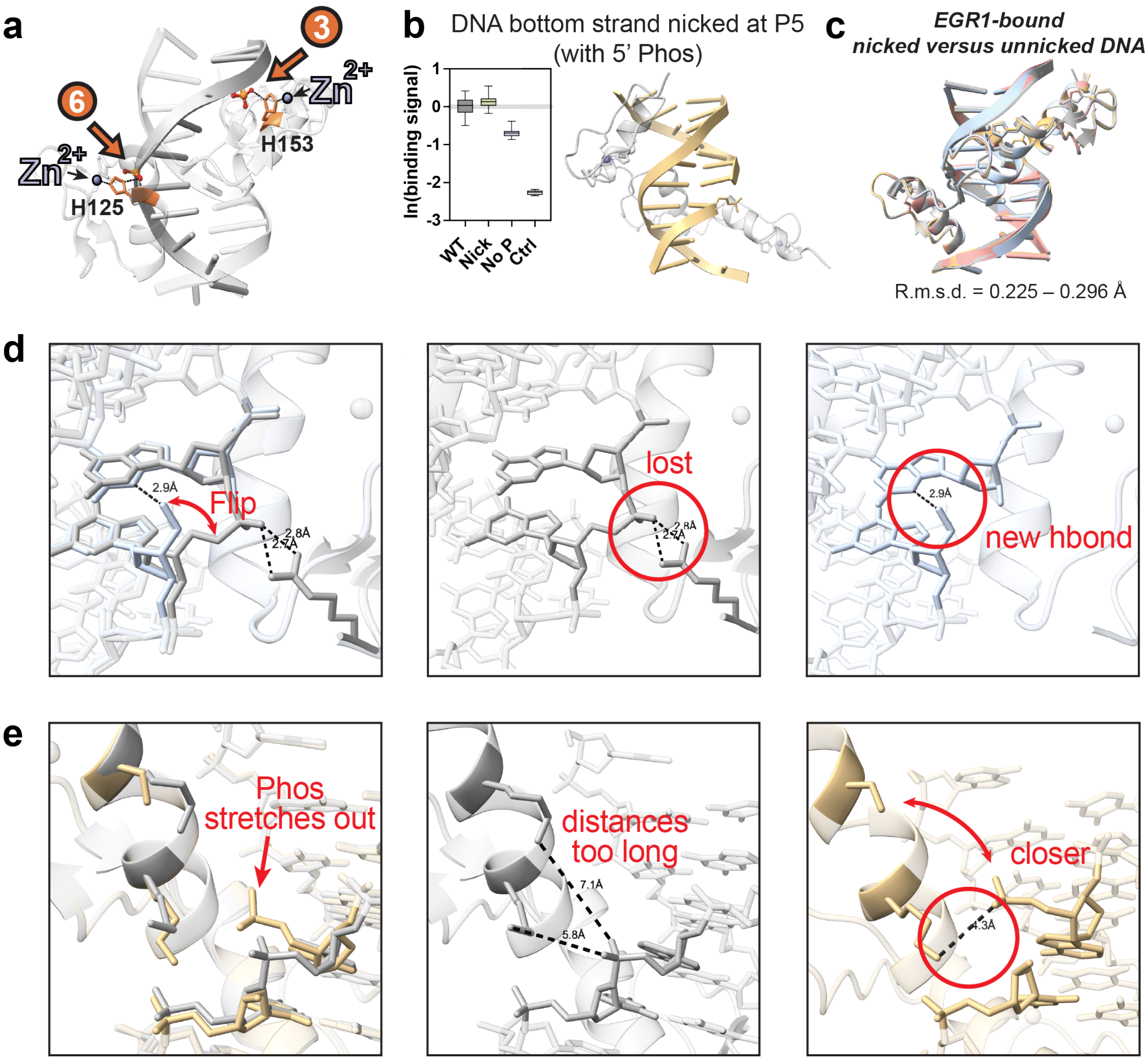
c



Extended Data Fig. 4 Within-family comparison of bHLH family

a. Within-family comparison of PIC-NIC nick sensitivity profiles across TFs sharing the same structural class. **b.** deepDNAshape correlations at the central CG palindrome position across bHLH family members. All correlations passed Bonferroni correction. **c.** deepDNAshape analysis reveals a conserved structural dependency at the central palindrome position across bHLH family members. Each point represents one of 16 possible dinucleotide substitutions at the central CG position (NN), with the x-axis showing the predicted roll (MAD, MAX, MYC) or rise (MITF) at the CG base step from deepDNAshape, and the y-axis showing the experimentally measured binding score (E-score) from universal PBM data. All four bHLH proteins show significant correlations ($r = 0.82–0.84$; *FDR and Bonferroni corrected).

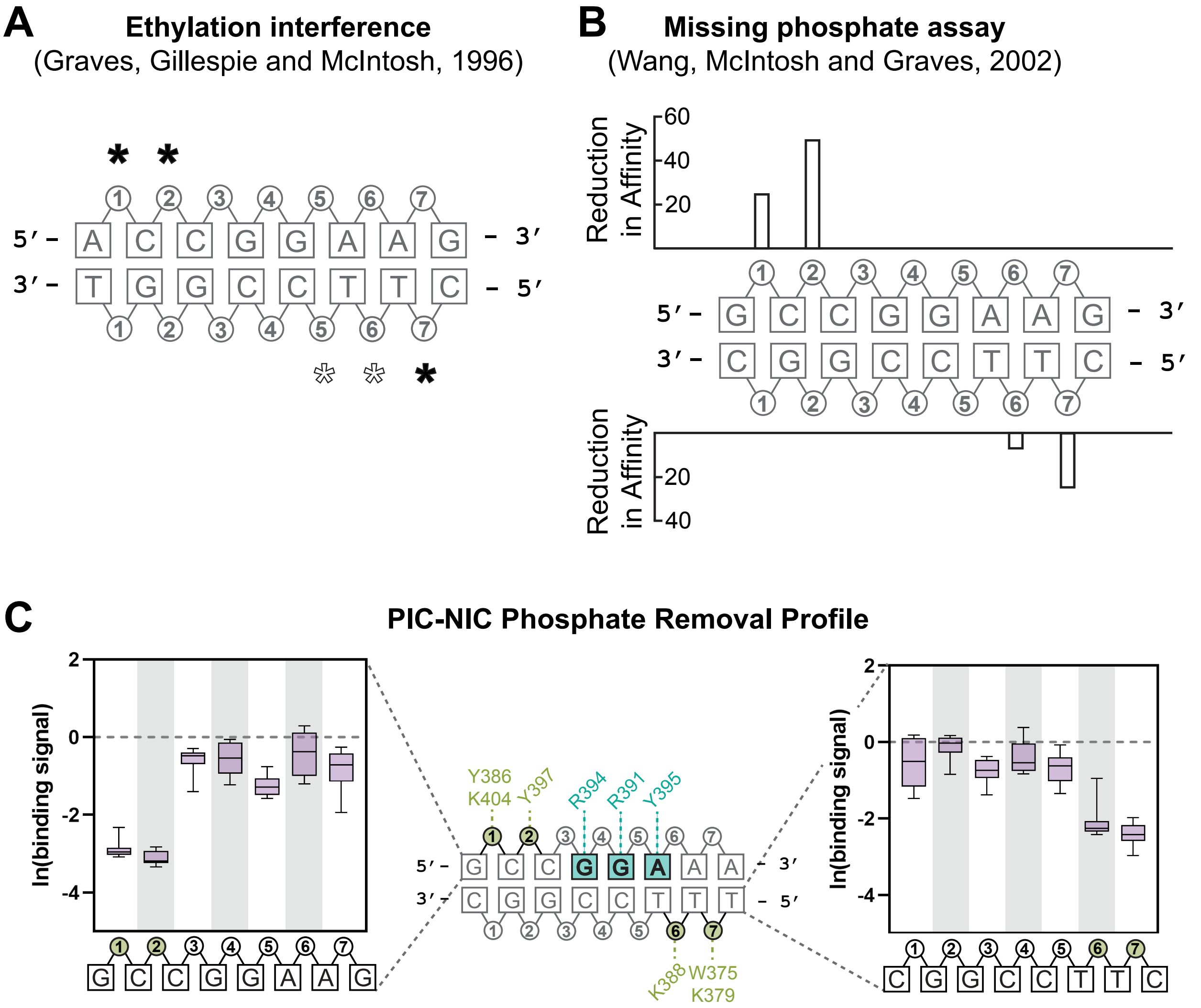
Extended Data Fig. 5 High-resolution crystal structure analysis of EGR1 in complex with nicked DNA



Extended Data Fig. 5 High-resolution crystal structure analysis of EGR1 in complex with nicked DNA

a. Structural zooms of two zinc coordination sites, highlighting the dual roles of H125 and H153 in DNA phosphate interaction and zinc finger coordination, rendering these positions highly sensitive to nicking. **b.** No disruption of EGR1 binding is observed when the DNA is nicked at position 5 on the bottom strand (5′ phosphate retained), while when the 5’ phosphate is removed, only minor disruption is observed. The high-resolution crystal structure of EGR1 in complex with DNA nicked at position 5 on the bottom strand (5' phosphate retained) is solved. **c.** Overlay of EGR1 bound to intact and three nicked DNA duplexes reveals high structural similarity (RMSD = 0.225 - 0.296 Å), indicating minimal global deformation. **d.** Structural changes when the DNA is nicked at position 7 on the top strand (phosphate removed). Gray: EGR1/intact DNA complex; blue: EGR1/nicked DNA complex. *Left panel*: Compared to the intact DNA, the backbone at the break site in the nicked DNA flips direction and induces a conformational change. *Middle panel*: The original H-bonds between Arg103 and the backbone phosphate are lost due to the phosphate removal at this position. *Right panel*: Due to the flipping of the backbone in the nicked structure, a new intramolecular H-bond is formed between the backbone OH and the neighboring guanine base. **e.** Structural changes when the DNA is nicked at position 5 on the bottom strand (phosphate retained). Gray: EGR1/intact DNA complex; gold: EGR1/nicked DNA complex. *Left panel*: When nicked, the backbone phosphate stretches away from its original position and the backbone geometry around the nicked phosphate is altered. *Middle panel*: In the intact DNA complex, the distances between the amino acid residues and DNA phosphate backbone at this position are too long to contribute to stabilizing interactions. *Right panel*: Due to the conformational stretch of the phosphate backbone in the nicked complex, the distances between nearby Arg and Lys residues and this backbone phosphate are shortened, and thus new favorable interactions emerge.

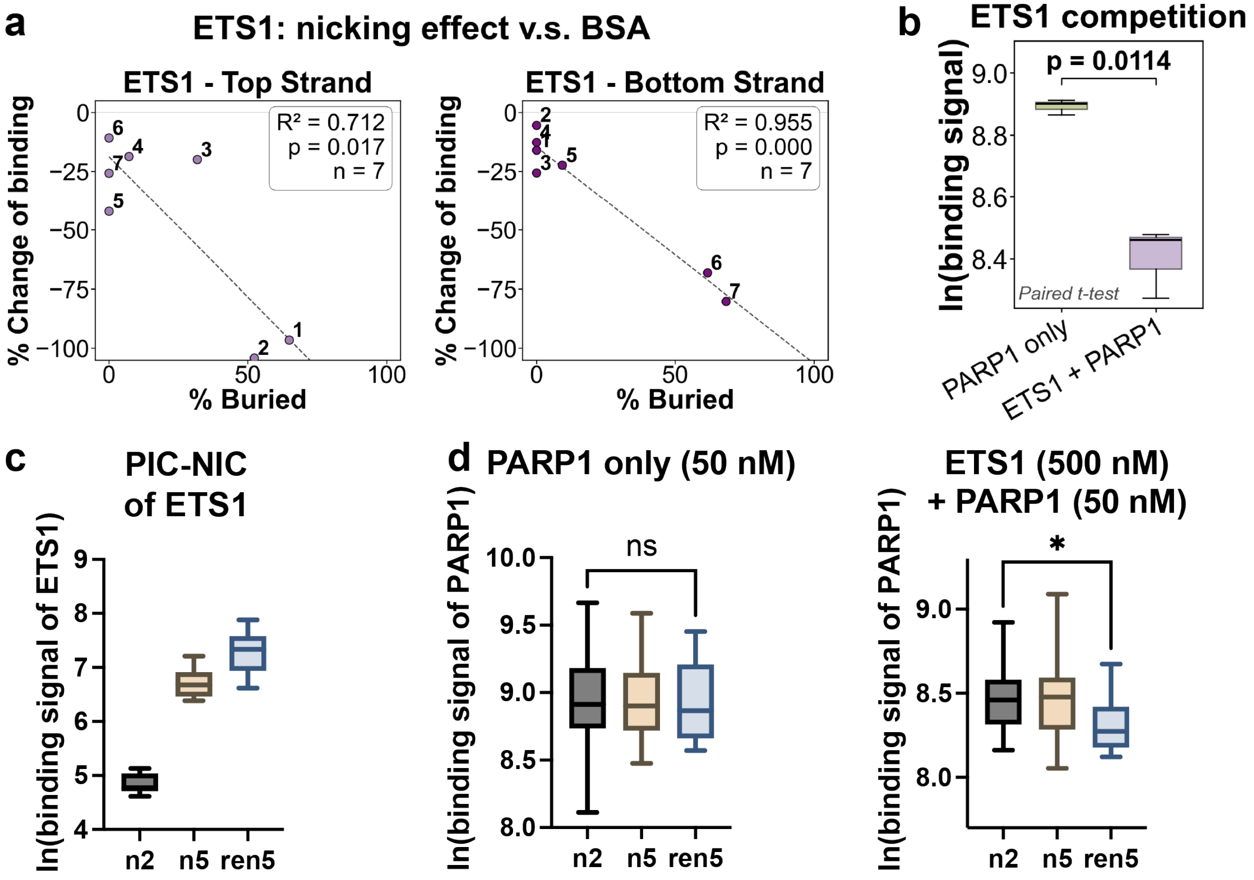
Extended Data Fig. 6 Orthogonal assays converge on key backbone phosphates in ETS1-DNA recognition



Extended Data Fig. 6 Orthogonal assays converge on key backbone phosphates in ETS1-DNA recognition

a. Phosphate contacts between ETS domain and DNA as implicated by ethylation-interference analysis. Adapted from Graves, Gillespie and McIntosh, 1996. **b.** Summary of "missing phosphate" experiment data. Open and closed bars display the reduction in affinity of ΔN331 and ΔN280, respectively, for nicked duplexes relative to unnicked duplex: K_D(nicked)/K_D(unnicked). Adapted from Wang, McIntosh and Graves, 2002. **c.** Phosphate removal profile of ETS1 from PIC-NIC experiment. All three approaches agreed on the same phosphates that are important in ETS1/DNA interaction.

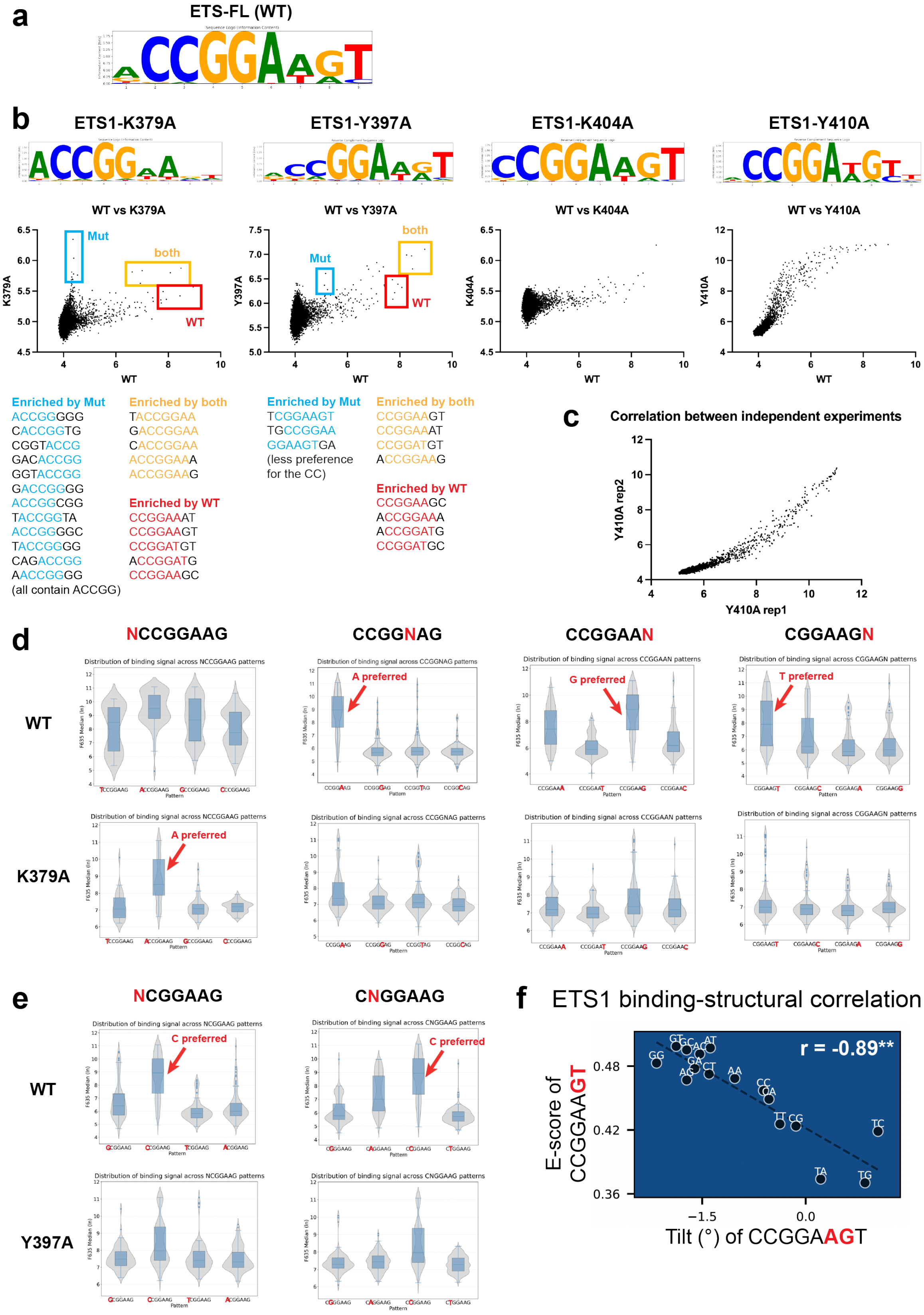
Extended Data Fig. 7 ETS1 BSA analysis and ETS1-PARP1 competition assay



Extended Data Fig. 7 ETS1 BSA analysis and ETS1-PARP1 competition assay

a. Correlation between phosphate burial and PIC-NIC nick sensitivity for ETS1. For each phosphate position along the top strand (left) and bottom strand (right) of the ETS1 binding site, the percentage of phosphate surface area buried upon protein binding (% Buried, x-axis; calculated as BSA/SASA×100) is plotted against the percentage change in binding signal upon phosphate removal in PIC-NIC (% Change of binding, y-axis). Numbers indicate binding site positions. Dashed line shows linear regression fit. Statistics shown are for Pearson correlation (n = 7 positions per strand). **b.** ETS1 and PARP1 compete for nicked sites. PARP1 binding signal (ln intensity values, median of 20 replicates per probe) across 3 nicked ETS1 binding site probes, measured in the absence (PARP1 only, green) or presence (ETS1 + PARP1, purple) of competitor ETS1. Addition of ETS1 significantly reduces PARP1 occupancy at nicked sites (p = 0.0114, paired t-test). **(c)** PIC-NIC binding profiles for three selected probes: n2 (largest reduction in ETS1 binding upon nicking), n5 (intermediate), and ren5 (minimal reduction), showing distinct nick-sensitivity profiles. **(d)** *Left*: In the absence of ETS1, all three probes show comparable PARP1 binding (50 nM PARP1; ns, not significant). *Right:* Pre-incubation with 500 nM ETS1 followed by 50 nM PARP1 reveals significantly lower PARP1 occupancy at ren5 compared to n2 (*p = 0.0112), consistent with stronger ETS1 competition at the less nick-sensitive position.

Extended Data Fig. 8 ETS1 mutants at phosphate-contacting residues exhibit differential binding in comparison to the WT protein.

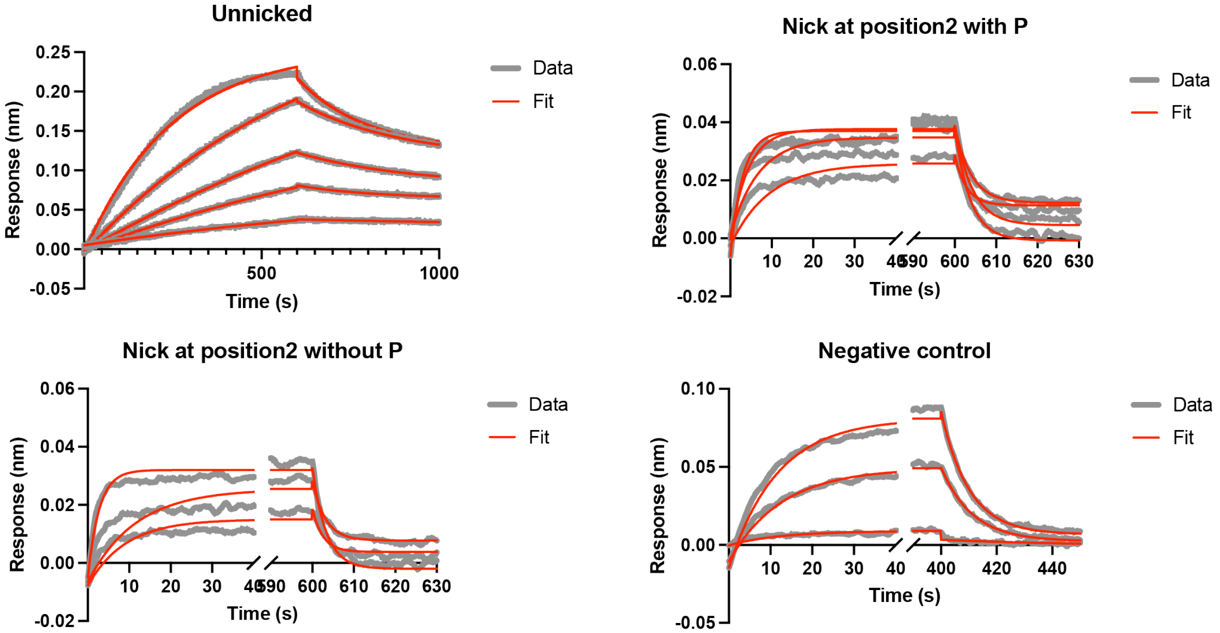

a
ETS-FL (WT)
b
ETS1-K379A
ETS1-Y397A
ETS1-K404A
ETS1-Y410A
WT vs K379A
WT vs Y397A
WT vs K404A
WT vs Y410A
Mut
both
WT
K379A
Y397A
K404A
Y410A
Enriched by Mut
ACCGGGGG
CACCGGTG
CGGTACCG
GACACCGG
GGTACCGG
GACCGGGG
ACCGGCGG
TACCGGTA
ACCGGGGC
TACCGGGG
CAGACCGG
AACCGGGG
(all contain ACCGG)
Enriched by both
TACCGGAA
GACCGGAA
CACCGGAA
ACCGGAAA
ACCGGAAG
Enriched by WT
CCGGAAAT
CCGGAAGT
CCGGATGT
ACCGGATG
CCGGAAGC
Enriched by Mut
TCGGAAGT
TGCCGGAA
GGAAGTGA
(less preference for the CC)
Enriched by both
CCGGAAGT
CCGGAAAT
CCGGATGT
ACCGGAAG
Enriched by WT
CCGGAAGC
ACCGGAAA
ACCGGATG
CCGGATGC
c
Correlation between independent experiments
Y410A rep2
Y410A rep1
d
NCCGGAAG
CCGGNAG
CCGGAAN
CGGAAGN
WT
K379A
Distribution of binding signal across NCCGGAAG patterns
Distribution of binding signal across CCGGNAG patterns
Distribution of binding signal across CCGGAAN patterns
Distribution of binding signal across CGGAAGN patterns
F635 Median (ln)
Pattern
TCCGGAAG
ACCGGAAG
GCCGGAAG
CCCGGAAG
CCGGAAG
CCGGGAG
CCGGTAG
CCGGCAG
CCGGAAA
CCGGAAT
CCGGAAG
CCGGAAC
CGGAAGT
CGGAAGC
CGGAAGA
CGGAAGG
A preferred
G preferred
T preferred
A preferred
e
NCGGAAG
CNGGAAG
WT
Y397A
Distribution of binding signal across NCGGAAG patterns
Distribution of binding signal across CNGGAAG patterns
GCGGAAG
CCGGAAG
TCGGAAG
ACGGAAG
CGGGAAG
CAGGAAG
CCGGAAG
CTGGAAG
C preferred
C preferred
f
ETS1 binding-structural correlation
r = -0.89**
E-score of CCGGAAGT
Tilt (°) of CCGGAAGT
0.48
0.42
0.36
−1.5
0.0
GG
GT
GC
AC
AT
AG
GA
CT
AA
CC
CA
TT
CG
TC
TA
TG

Extended Data Fig. 8 ETS1 mutants at phosphate-contacting residues exhibit differential binding in comparison to the WT protein.

a. PWM logo of full-length wild-type ETS1 protein, generated from experimental data of the universal protein binding microarray containing all possible DNA 10-mer sequences. **b.** *Top panels*: PWM logos of four ETS1 mutants (K379A, Y397A, K404A and Y410A), generated using an identical universal 10-mer microarray design. *Middle panels*: Binding signal correlations of the four ETS1 mutants versus the ETS1 WT. Each point represents a DNA 8-mer sequence. Blue box highlights the 8-mer sequences preferred by the respective ETS1 mutant; red box highlights the 8-mer sequences preferred by the WT ETS1; yellow box highlights the 8-mer sequences highly bound by both the mutant and WT ETS1. *Bottom panels*: 8-mer sequences corresponding to the highlighted boxes in the middle panels. **c.** Correlation of replicate experiments (from independent protein expression) confirms high reproducibility. **d-e.** Comparison of nucleotide preference at this indicated position (N) for WT ETS1 and **d** K379A and **e** Y397A**. f.** deepDNAshape analysis for ETS1. Predicted tilt at the AG base step correlates with experimentally measured binding scores (E-score) when the flanking GT dinucleotide is systematically varied ($r = -0.89$, **Bonferroni corrected). Each point represents one of 16 possible dinucleotide substitutions at the GT position; the varied dinucleotide is highlighted in red in the sequence labels. Dashed line shows linear regression fit.

Extended Data Fig. 9 Bio-Layer Interferometry (BLI) data for ETS1 and SOX2

a



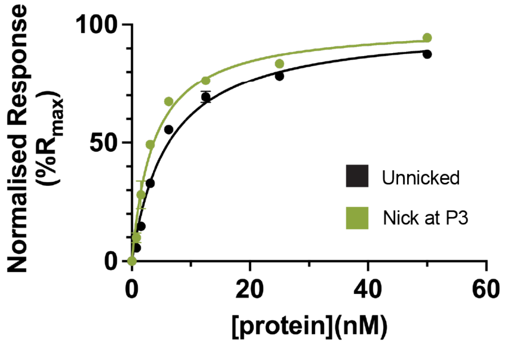
	Kd (× 10^{-7} M)	Kon (× 10^{5} M^{-1} s^{-1})	Koff (× 10^{-3} s^{-1})
Unnicked	0.3 ± 0.1	2.5 ± 1.0	6.2 ± 0.2
n2P	9 ± 4	4 ± 2	251 ± 40
n2	9 ± 7	4 ± 2	220 ± 82
Neg Ctrl	10 ± 3	2.7 ± 0.8	267 ± 7

~40 fold faster dissociation compared to unnicked

b

	Kd (× 10^{-9} M)	Kon (× 10^{5} M^{-1} s^{-1})	Koff (× 10^{-3} s^{-1})
Unnicked	5.0 ± 0.6	3.0 ± 0.3	1.49 ± 0.05
nick at P3	2.9 ± 0.4	5.8 ± 0.9	1.68 ± 0.04

c



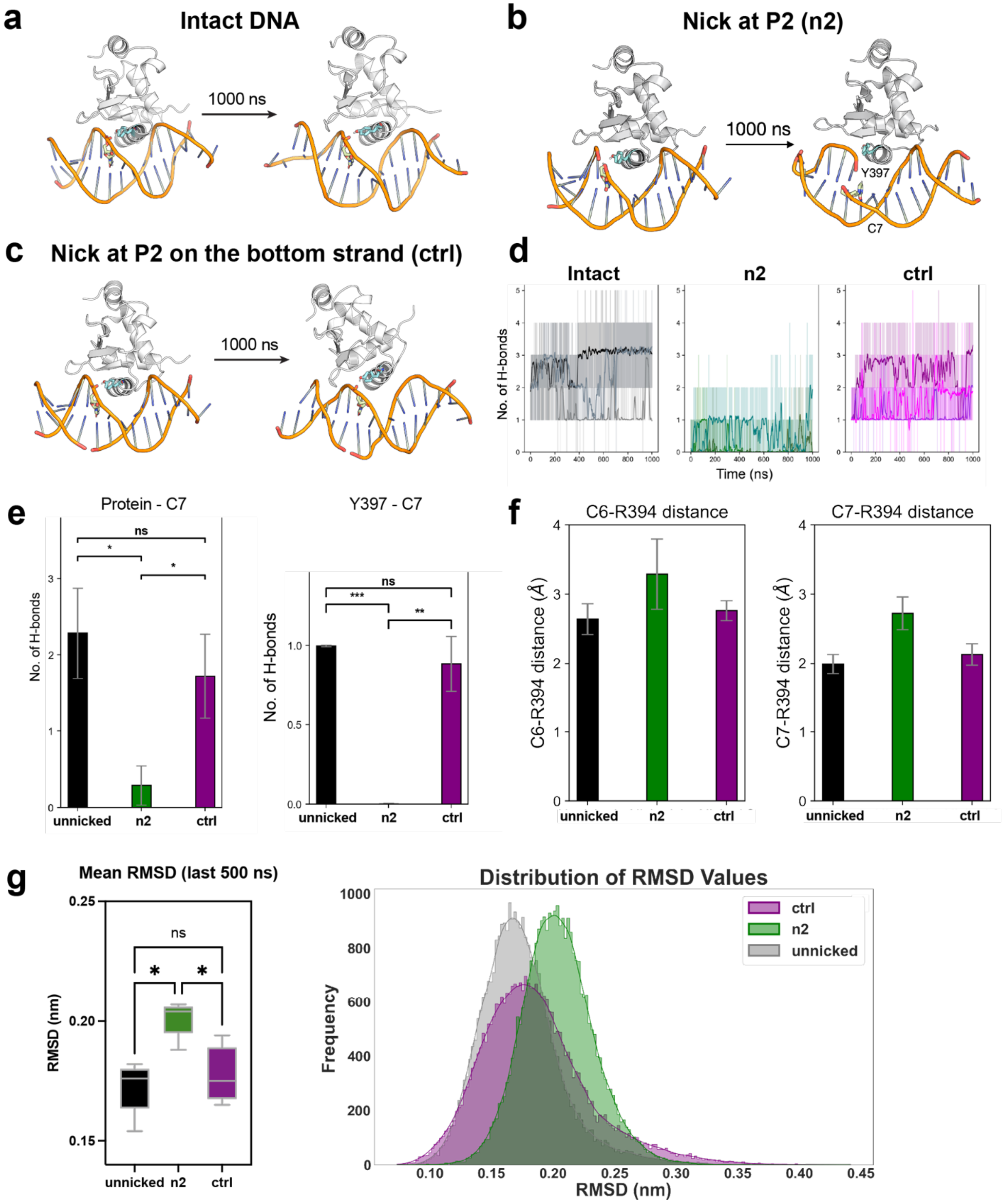
Steady State Kinetics

	Thermodynamic Kd (× 10^{-9} M)
Unnicked	6 ± 1
nick at P3	3.7 ± 0.9

Extended Data Fig. 9 Bio-Layer Interferometry (BLI) data for ETS1 and SOX2

a. Bio-Layer Interferometry (BLI) sensorgrams (grey) of WT ETS1 binding to the intact (unnicked) DNA probe, the DNA probe nicked at position 2 (n2P), the DNA probe nicked at position 2 with the removal of 5’ phosphate (n2), and a probe containing a non-specific site (neg ctrl). Sensorgrams at each concentration were fit locally to a 1:1 binding model (red) at each concentration to determine k_{on} and k_{off} values. K_{D} values were calculated for each curve, and the means and standard deviations derived from these values are presented. Compared to the intact DNA probe, the binding to n2P, n2 and neg ctrl exhibited ~40-fold faster dissociation. **b.** Kinetic K_{D}, k_{on} and k_{off} values derived from sensorgrams of SOX2 binding to the intact DNA probe and DNA probe nicked at position 3 (Fig. 4c). k_{on} and k_{off} values were obtained by global fitting of sensorgrams to a 1:1 binding model. **c.** Thermodynamic K_{D} of SOX2 binding to the intact DNA probe and DNA probe nicked at position 3 were calculated from 7 concentrations using a one-site binding model at equilibrium. The observed ~1.6-fold change in the thermodynamic Kd aligns with the ~1.7-fold change in the obtained kinetic K_{D} above.

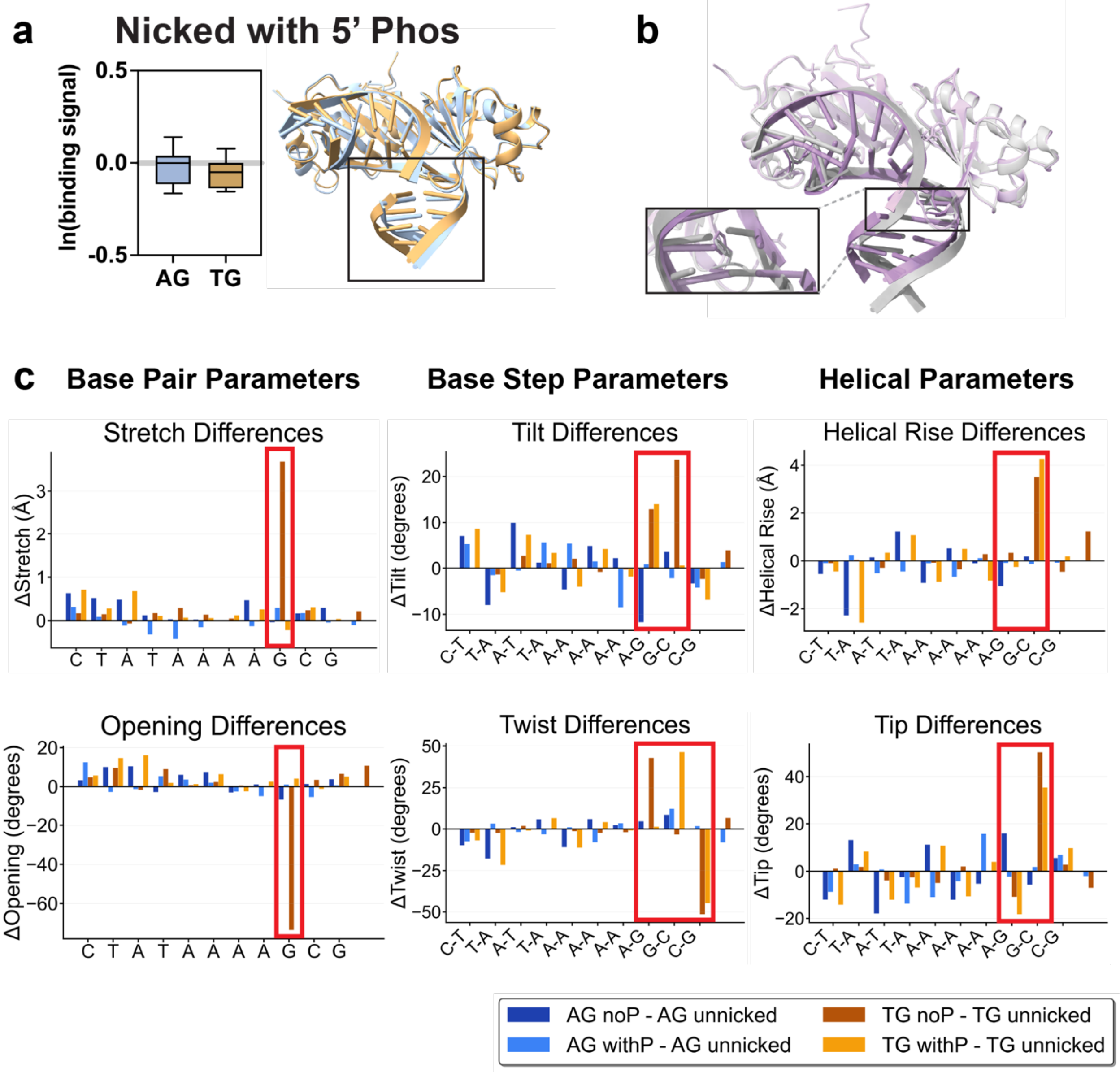
Extended Data Fig. 10 All atom Molecular Dynamic (MD) simulations for ETS1



Extended Data Fig. 10 All atom Molecular Dynamic (MD) simulations for ETS1

a. MD simulation snapshots of wild-type ETS1 bound to intact DNA over a 1000 ns trajectory, serving as a structural reference for nick-induced perturbation analysis. **b.** MD simulation snapshots of wild-type ETS1 bound to DNA nicked at position 2 (with the removal of 5' phosphate), where the nick abolishes binding, over a 1000 ns trajectory. Distinctive local distortion of the DNA backbone geometry is observed at the break site. **c.** MD simulation snapshots of wild-type ETS1 bound to DNA nicked at position 2 at the bottom strand (with the removal of 5' phosphate), where the nick does not induce a major effect in binding of the TF, over a 1000 ns trajectory, serving as negative control. **d.** Time-dependent H-bond analysis across the simulation trajectory for the three constructs. In n2, the removal of the phosphate in contact with Y397 leads to the overall loss of hydrogen bonds over the 1000 ns trajectory. **e.** Hydrogen bonds formed between the protein and DNA at C7 (left), and specifically between Y397 and C7 (right), across the simulation trajectories. The n2 nick significantly reduces hydrogen bonding, unlike the control nick, and this effect is caused by the removal of C7 backbone phosphate at the break site. **f.** Average distances from C6 and C7 to R394 reveal that the n2 nick increases these distances relative to the wild-type and control, weakening interactions and reflecting geometric disruption of the protein–DNA interface. **g.** *Left panel*: Root-mean-square deviation (RMSD) of backbone atoms in the last 500 ns of simulation (assuming equilibrium has been reached) shows that the n2 nick increases structural fluctuations compared to the wild-type and control. *Right panel*: Distribution of RMSD values for the last 500 ns trajectory shows a rightward shift in the n2 nicked complex (green), indicating elevated structural flexibility compared to the wild-type (gray) and control nick (magenta).

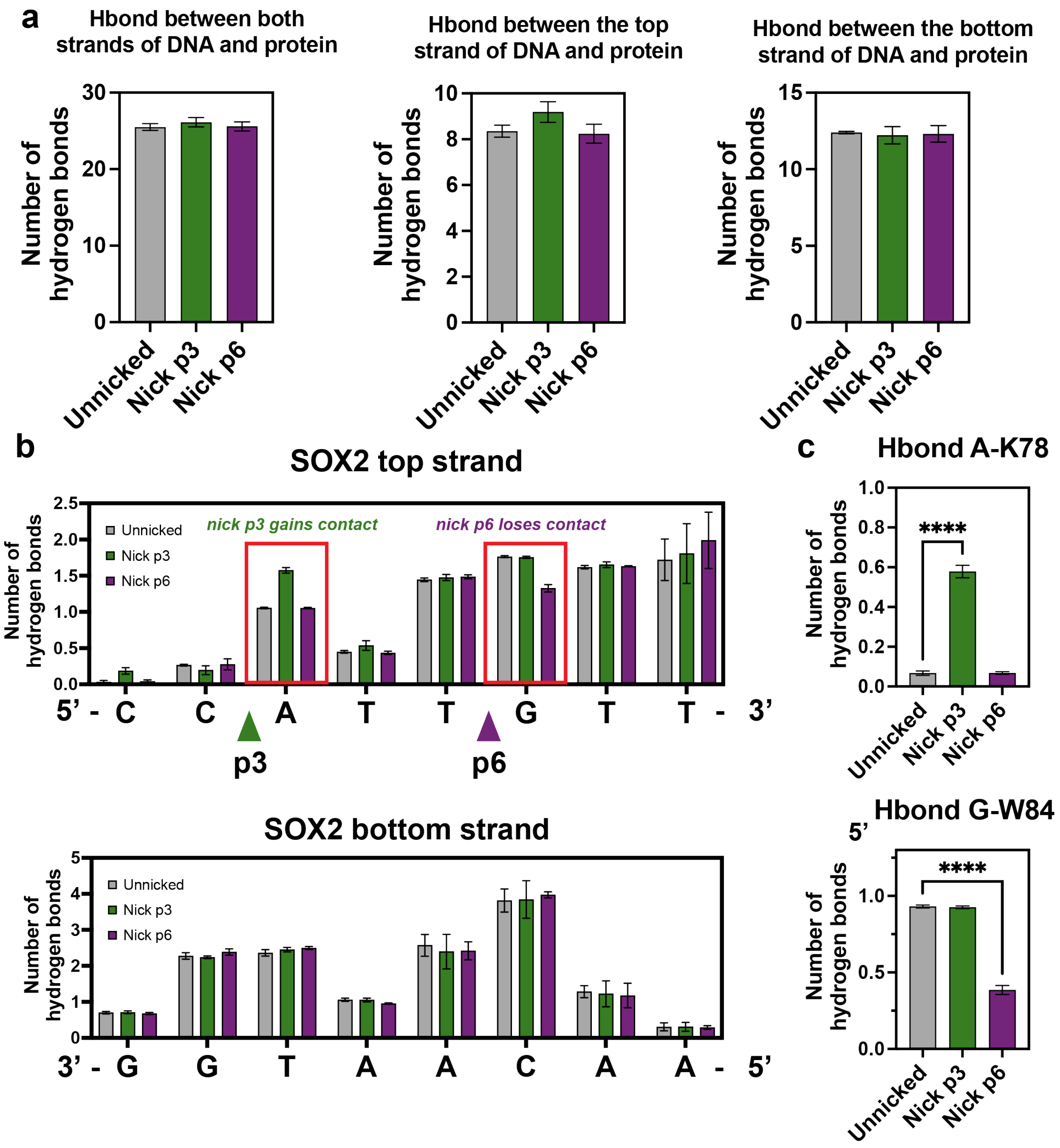
Extended Data Fig. 11 High-resolution crystal structure analysis of TBP in complex with nicked DNA



Extended Data Fig. 11 High-resolution crystal structure analysis of TBP in complex with nicked DNA

a. The binding level of TBP to the AG variant becomes almost identical to that of the TG variant when position 8 at the bottom strand is nicked with the retention of 5' phosphate. Two high-resolution crystal structures of nicked complexes corresponding to the AG variant and the TG variant were solved (Blue: AG variant complex; gold: TG variant complex). Structural overlay reveals a larger DNA bending angle in the TG complex compared to the same region in the AG complex. **b.** Structural overlay and zoom in of the TG nicked complex (with the removal of 5' phosphate at the break site) and TG intact complex (gray: intact TG complex; purple: nicked TG complex with no 5' phosphate). When nicked, the DNA geometry deviates significantly from that of intact DNA, introducing a larger bending angle and more deformation. The structural zoom in shows dramatic local shifts in the base pair geometry flanking the nick site, suggesting that the nick disrupts the local DNA architecture. **c.** Structural analysis of the four solved crystal structures shows that the TG nicked complex with no 5' phosphate has the largest deformation at the nick site. Representative base pair parameters (stretch and opening), base step parameters (tilt and twist) and helical parameters (helical rise and tip) were chosen to illustrate the parameter difference between nicked and intact complexes.

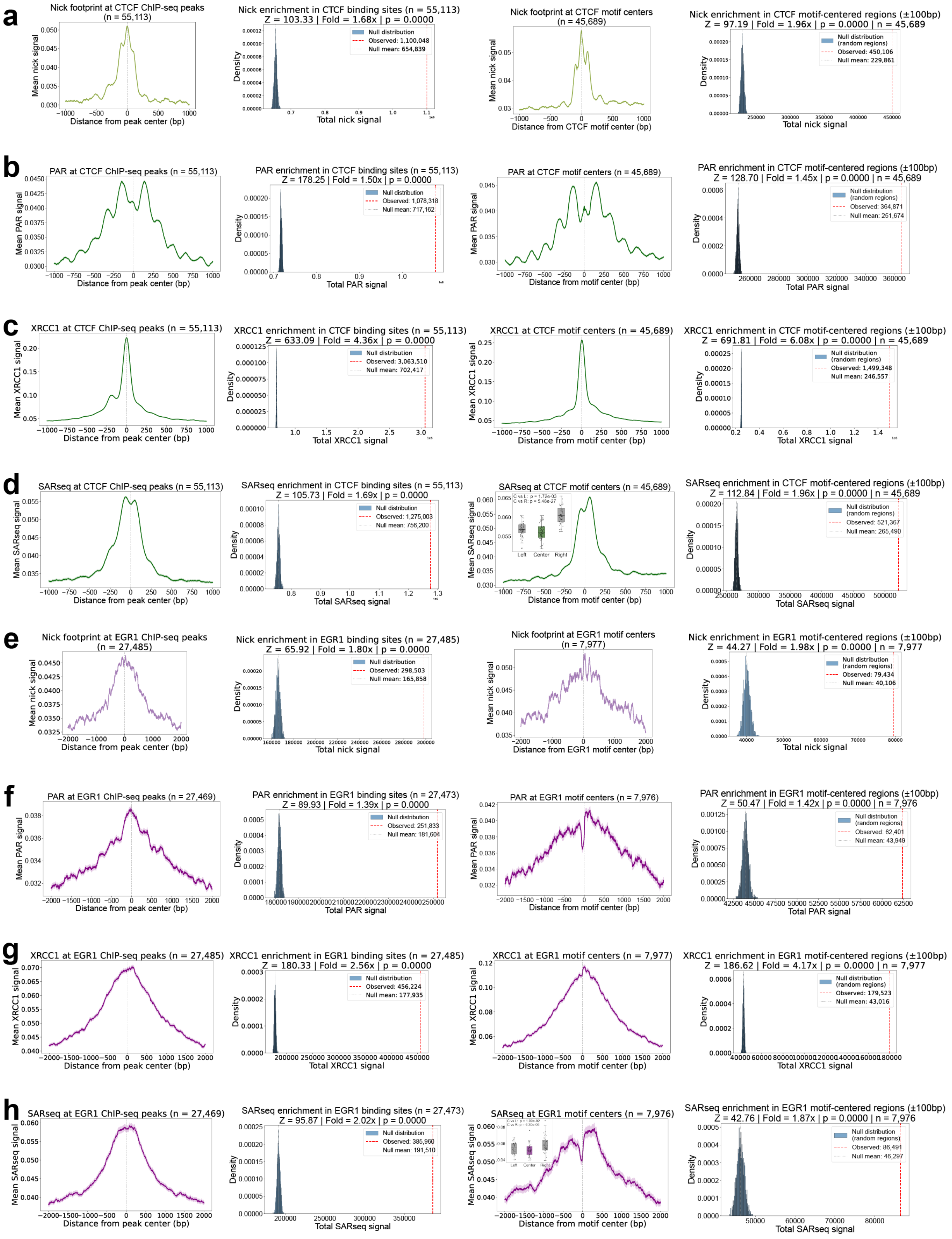
Extended Data Fig. 12 All atom Molecular Dynamic (MD) simulations for SOX2



Extended Data Fig. 12 All atom Molecular Dynamic (MD) simulations for SOX2

(a) Total number of protein–DNA hydrogen bonds across 3 independent MD replicates for unnicked DNA (grey), DNA nicked at position p3 (green), and DNA nicked at position p9 (magenta), counted across both strands (left), the top strand only (middle), and the bottom strand only (right). **(b)** Per-nucleotide hydrogen-bond occupancy along the SOX2 binding site, top strand (upper) and bottom strand (lower). Green and magenta triangles mark the p3 and p6 nick sites, respectively. Red boxes highlight the two positions where contact patterns diverge: SOX2 gains hydrogen bond interactions significantly at the A adjacent to the p3 nick when p3 is nicked, and loses hydrogen bond interactions significantly at the G adjacent to the p6 nick when p6 is nicked. **(c)** A closer investigation at the two positions highlighted in (c), confirming a significant gain of interaction with K78 at A in the p3-nicked state and a significant loss of interaction with W84 at G in the p6-nicked state.

Extended Data Fig. 13 SSBs accumulate at TF binding sites and engage the DNA repair machinery *in vivo*



Extended Data Fig. 13 SSBs accumulate at TF binding sites and engage the DNA repair machinery *in vivo*.

Genomic analyses integrating four complementary datasets in induced neurons (iNeurons): S1-END-seq (genome-wide SSB detection), PAR ChIP-seq (PARP1 activity), XRCC1 ChIP-seq (SSB repair scaffold recruitment), and SAR-seq (active repair synthesis). Data are shown for CTCF (green, panels a–d) and EGR1 (purple, panels e–h). **(a, e)** Aggregate S1-END-seq signal (mean ± 95% CI) centered on ChIP-seq peak summits, with permutation test comparing observed nick signal at TF peaks versus 1,000 random genomic shuffles of equivalent size. **(b, f)** Aggregate PAR ChIP-seq signal (mean ± 95% CI) at ChIP-seq peak summits, with permutation test. **(c, g)** Aggregate XRCC1 ChIP-seq signal (mean ± 95% CI) at ChIP-seq peak summits, with permutation test. **(d, h)** Aggregate SAR-seq signal (mean ± 95% CI) at ChIP-seq peak summits, with permutation test. For each panel, aggregate profiles are shown at both ChIP-seq peak level (left) and motif-centered regions ±100 bp (right), with corresponding permutation tests. Fold enrichment, Z-scores, and p-values are indicated in each permutation panel. Detailed methods and data sources are provided in Supplementary Methods. *Insets*: Motif sites for CTCF and EGR1 were randomly partitioned into 50 disjoint subgroups, and the mean signal at the dip center and the two flanking borders was computed per subgroup. Boxplots show these per-subgroup means; individual subgroups are shown as black dots. P-values are from one-sided paired t-tests (center < border). Position labels are in bp relative to the motif center. The center bin is the average signal across ±10 bp around the motif center; the left and right border bins are single positions at ±100 bp for EGR1 and ±50 bp for CTCF.

Extended Data Table. 1 Comparison of PIC-NIC with prior experimental approaches for probing DNA backbone contributions to protein–DNA recognition

Approach	Main perturbation	Base identity	Phos identity	Backbone continuity disrupted?	Separates phos chemistry from backbone mechanics?	Throughput / scope	Readout
a. *Classical chemical / footprinting approaches*							
Missing contact (PMID: 2958845)	Removes base via depurination/depyrimidation	Removed at probed positions	Retained	Mostly no	No — loss of base H-bonds and local geometry change are coupled	Low; selected positions; one protein at a time	Indirect (gel shift + sequencing)
Missing nucleoside (PMID: 2611245)	Hydroxyl radical removes base + sugar	Removed at probed positions	Often indirectly altered	Locally destabilized	No — base chemistry, stacking, and local geometry all perturbed simultaneously	Low; one protein/site at a time; not position-specific	Indirect (gel shift + sequencing)
Phosphate ethylation interference (PMID: 4033762)	Bulky ethyl group added to non-bridging phosphate oxygen	Retained	Chemically modified; charge and H-bonding	Mostly no	No — electrostatic, steric, and geometric perturbations are coupled	Low; selected proteins/sites	Limited / indirect
Methylphosphonate substitution (PMID: 11929991)	Neutralization or alteration of phosphate charge; minimal steric change	Retained	Altered (charge neutralized)	Mostly no	No — charge and geometry coupled	Low; selected systems	Quantitative in selected cases
Prior systematic nicking (PMID: 10373594; 11689571; 2174540)	Site-specific nick introducing free 3′-OH and 5′-OH termini	Varies by approach	Removed	Yes	No — backbone discontinuity and phosphate loss inseparable	Low; one or few positions per study; single protein	Limited / indirect (gel shift or competition assay)
Crystal structures of TF–nicked DNA complexes (e.g. PMID: 1653449)	Nick at defined position; structural snapshot	Retained	Depends on construct	Yes	Partial — depends on the construct	Very low; single TF; single position	No binding quantification
b. *PIC-NIC (our method)*							
PIC-NIC — phosphate-removed nick	Site-specific nick lacking 5′ phosphate	Retained	Removed	Yes	Partial — ambiguity remains, but directly comparable to phosphate-preserving condition	High; all motif positions; 15 TFs spanning 8 structural families	Quantitative binding; selected kinetics and crystal structures
PIC-NIC — phosphate-preserving nick	Site-specific nick retaining 5′ phosphate	Retained	Retained	Yes	**All chemical moieties intact, cleaner separations than existing methods**		

Extended Data Table. 2 Crystallographic data collection and refinement statistics: EGR1-DNA complexes

	EGR1-n7	**EGR1-ren5-P**	**EGR1-ren7**
Data collection			
PDB code	9RIC	9RI6	9RJ6
Space group	$C222_1$	$C222_1$	$C222_1$
Wavelength (Å)	1.34	1.34	1.34
Resolution (Å)	17.23-1.9	17.43-2.0	18.38-1.95
	(2.09-1.90)	(2.07-2.00)	(2.15-1.95)
Cell dimensions			
a, *b*, *c* (Å)	44.56,56.05,130.21	44.77,55.51,129.88	44.45,56.16,129.79
α, β, γ (°)	90,90,90	90,90,90	90,90,90
R-merge	0.053 (0.207)	0.073 (0.237)	0.100 (0.458)
R-pim	0.021 (0.085)	0.028 (0.098)	0.037 (0.169)
Mean I/sigma (I)	23.3 (7.5)	19.11 (6.5)	14.33 (5.2)
Completeness (%)	99.82 (99.95)	99.88 (100.00)	99.85 (100.00)
Redundancy	7.4 (7.0)	7.2 (6.7)	8.4 (7.8)
$CC_{1/2}$	0.999 (0.977)	0.998 (0.953)	0.997 (0.945)
Refinement			
No. reflections	13,206 (3,254)	11,283 (2,773)	12,198 (2,997)
R_{work} / R_{free} (%)	19.2/24.1	18.6/21.7	19.9/21.2
R.m.s. deviations			
Bond lengths (Å)	0.007	0.008	0.007
Bond angles (°)	0.89	0.93	0.91
Ramachandran			
Favored (%)	100.00	98.8	100.00
Disallowed (%)	0.00	0.00	0.00
Clashscore	0.48	1.44	0.48

Extended Data Table. 3 Crystallographic data collection and refinement statistics: TBP-DNA complexes

	TBP-TG-P	**TBP-TG**	**TBP-AG-P**	**TBP-AG**
Data collection				
PDB code	9OW8	9OW7	9OWZ	9OWI
Space group	$P2_1$	$C222_1$	I2	$P2_1$
Wavelength (Å)	1.01	1.01	1.01	1.01
Resolution (Å)	48.6-2.60 (2.69-2.60)*	65.7-2.30 (2.45-2.30)	88.3-2.91 (3.10-2.91)	48.3-2.49 (2.52-2.48)
Cell dimensions				
a, *b*, *c* (Å)	69.1,117.5,102.1	77.6,123.4,66.4	122.7,106.2,161.9	68.6,117.3,102.1
α, β, γ (°)	90.0,106.9,90.0	90.0,90.0,90.0	90.0,101.1,90.0	90.0,107.0,90.0
R-merge	0.083 (0.508)	0.053 (0.644)	0.076 (0.595)	0.086 (0.550)
R-pim	0.055 (0.321)	0.027 (0.350)	0.049 (0.384)	0.055 (0.342)
Mean I/sigma (I)	9.6 (2.3)	13.6 (1.4)	8.3 (1.9)	9.0 (2.2)
Completeness (%)	97.7 (99.5)	98.6 (90.5)	97.5 (98.4)	99.1 (98.3)
Redundancy	3.3 (3.4)	5.8 (5.3)	3.3 (3.4)	3.5 (3.5)
$CC_{1/2}$	0.994 (0.794)	0.998 (0.851)	0.994 (0.679)	0.995 (0.759)
Refinement				
No. reflections	46965 (4418)	17896 (328)	43821 (1466)	54270 (4434)
R_{work} / R_{free} (%)	20.2/25.3	18.9/22.9	23.5/25.3	20.7/25.1
R.m.s. deviations				
Bond lengths (Å)	0.009	0.008	0.012	0.008
Bond angles (°)	1.05	1.00	1.28	0.965
Ramachandran				
Favored (%)	96.0	99.0	91.6	95.9
Disallowed (%)	0.00	0.00	0.00	0.00
Clashscore	18.2	9.69	21.93	14.0

*Values in parentheses are for the highest-resolution shell.