Supplemental Information for

Single-molecule and in silico dissection of the interaction between Polycomb repressive complex 2 and chromatin

Rachel Leicher[a,b], Eva J. Ge[c], Xingcheng Lin[d], Matthew J. Reynolds[e], Wenjun Xie[d], Thomas Walz[f], Bin Zhang[d,*], Tom W. Muir[c], and Shixin Liu[a,*]

[a]Laboratory of Nanoscale Biophysics and Biochemistry, The Rockefeller University, New York, NY 10065; [b]Tri-Institutional PhD Program in Chemical Biology, New York, NY 10065; [c]Department of Chemistry, Princeton University, Princeton, NJ 08544; [d]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Laboratory of Structural Biophysics and Mechanobiology, The Rockefeller University, New York, NY 10065; and [f]Laboratory of Molecular Electron Microscopy, The Rockefeller University, New York, NY 10065

*Correspondence to: binz@mit.edu (B.Z.); shixinliu@rockefeller.edu (S.L.).

Table of Contents

www.pnas.org/cgi/doi/10.1073/pnas.2003395117

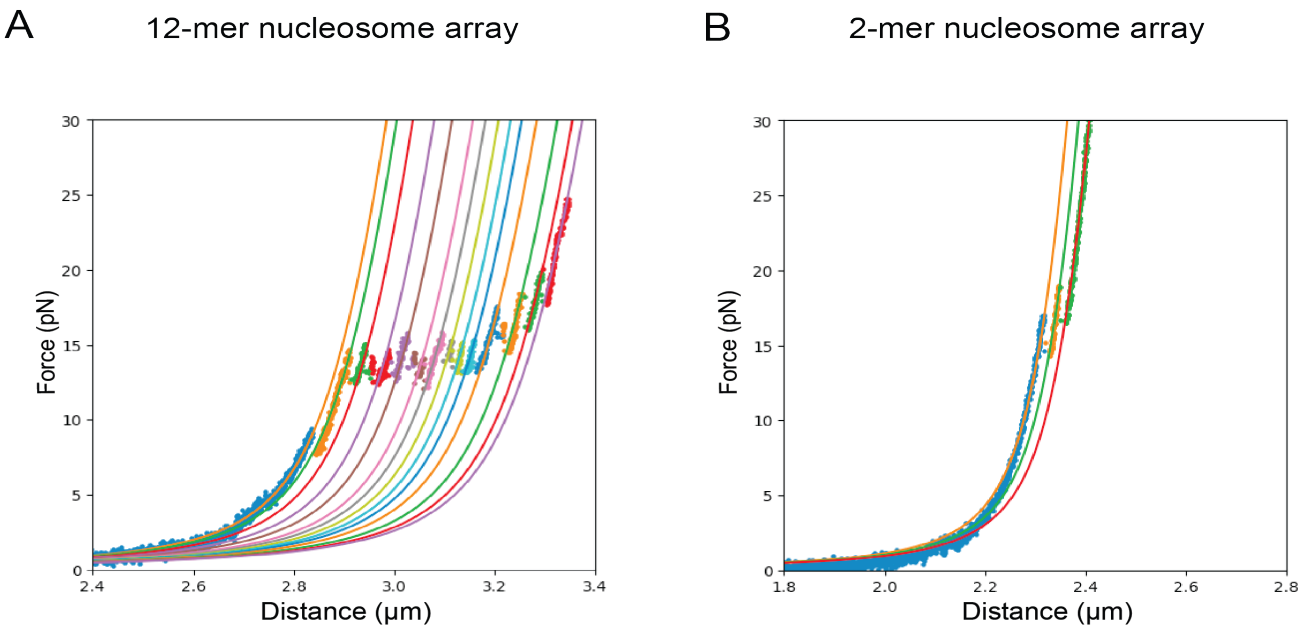

Fig. S1. Segmentation of force-extension curves. (*A*) A representative force-extension curve and its fitted segments for a 12-mer nucleosome array. See *Materials and Methods* for details of the segmentation analysis. (*B*) A representative force-extension curve and its fitted segments for a dinucleosome substrate.

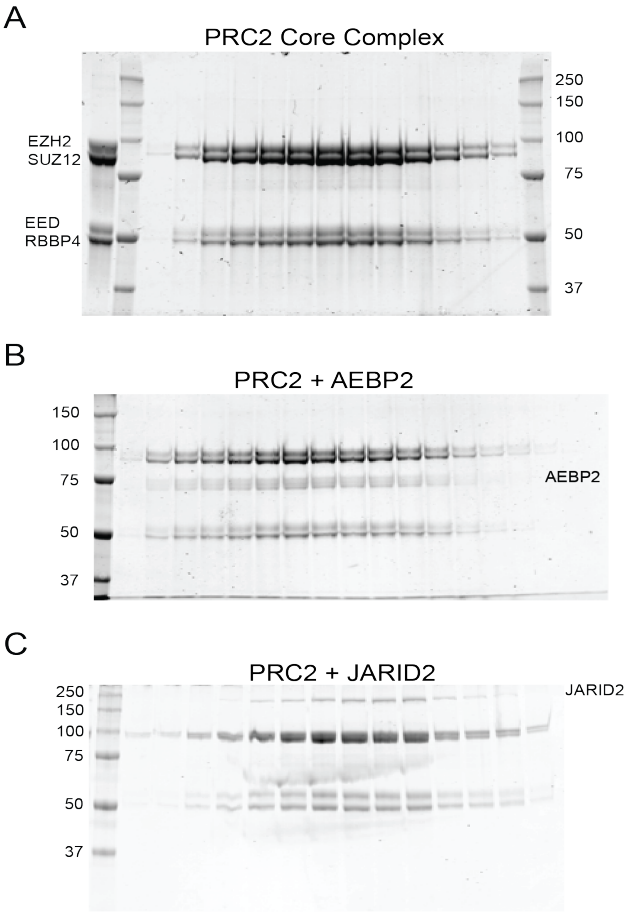

Fig. S2. SDS-PAGE analysis of PRC2 complexes. SDS-PAGE gels showing the fractions of PRC2 core (*A*), PRC2-AEBP2 (*B*), and PRC2-JARID2 (*C*) complexes from size-exclusion chromatography.

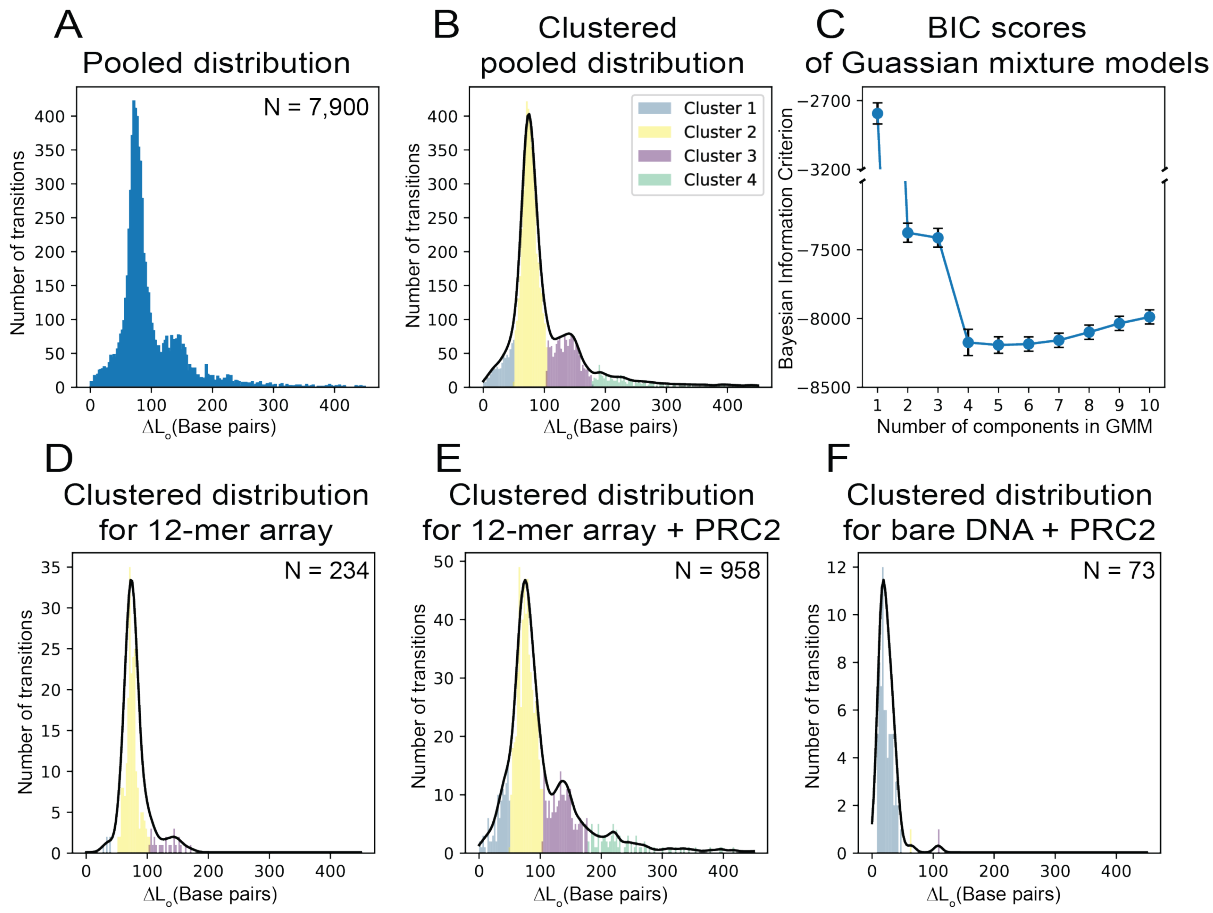

Fig. S3. Clustering of transitions identified from force-extension curves. (*A*) Histogram of contour length changes (ΔL_0) of transitions derived from the force-extension curves pooled from all conditions. (*B*) The same histogram as in (*A*) is grouped into four clusters based on a four-component Gaussian mixture model. The cluster boundaries are determined using a bootstrapping procedure. In black is the estimated kernel density of the underlying distribution. See *Materials and Methods* for details of cluster assignment, robustness evaluation, and kernel density estimation. (*C*) Bayesian information criterion (BIC) scores of *M*-component Gaussian mixture models (*M* ranges from 1 to 10) to fit the pooled experimental data of ΔL_0. A four-component model was selected for all cluster assignments in this study because the BIC score ceased to improve when more components were added to the model. (*D-F*) Clustered histograms of ΔL_0 distribution for the 12-mer array data (*D*), 12-mer + PRC2 data (*E*), and bare DNA + PRC2 data (*F*). Kernel density estimates are shown as black contours. *N* denotes the total number of transitions within each histogram.

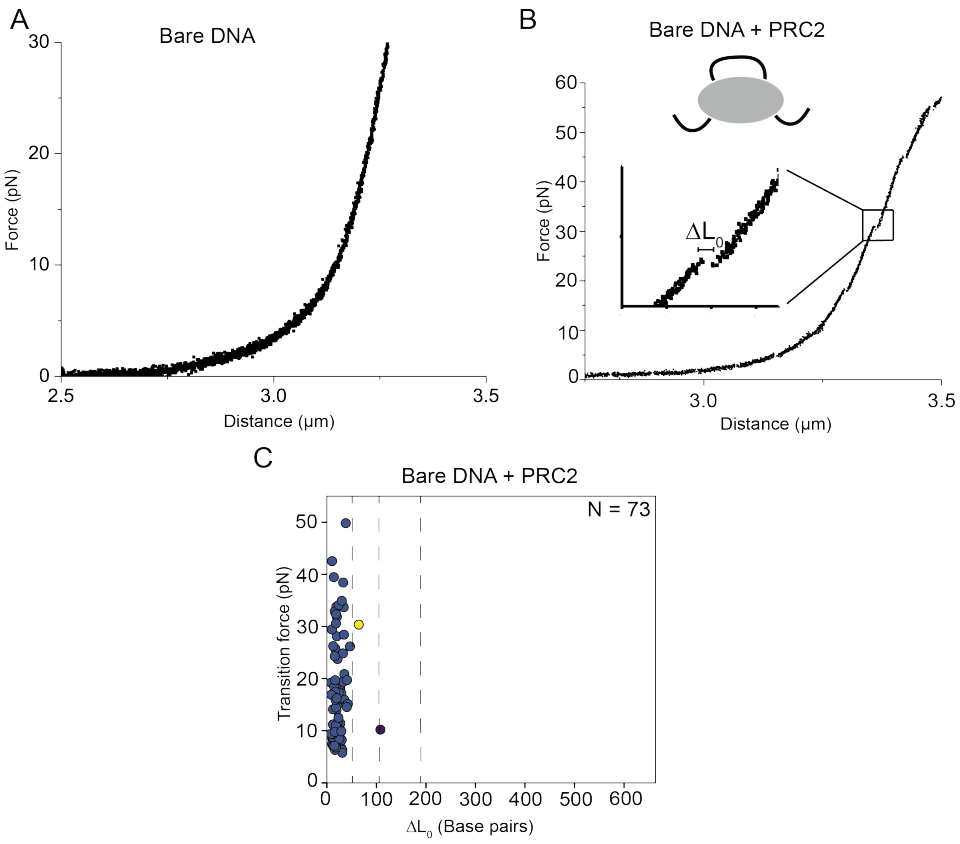

Fig. S4. Interaction of PRC2 with bare DNA. (*A*) A representative force-extension curve for a piece of 10-kbp-long bare DNA tethered between two optically trapped beads. (*B*) A representative force-extension curve for the same type of DNA bound with PRC2 core complexes. (*Inset*) Zoomed-in view of one example transition signifying PRC2 disengagement. (*C*) Cluster analysis of all transitions observed in the force-extension curves of PRC2-bound bare DNA. The average size of transitions for bare DNA + PRC2 is 25 ± 2 bp, somewhat lower than that of C1 transitions for 12-mer + PRC2 (36 ± 2 bp). This difference is likely due to the crosstalk between C1 and C2 clusters for the 12-mer + PRC2 condition (see Fig. S3*E*).

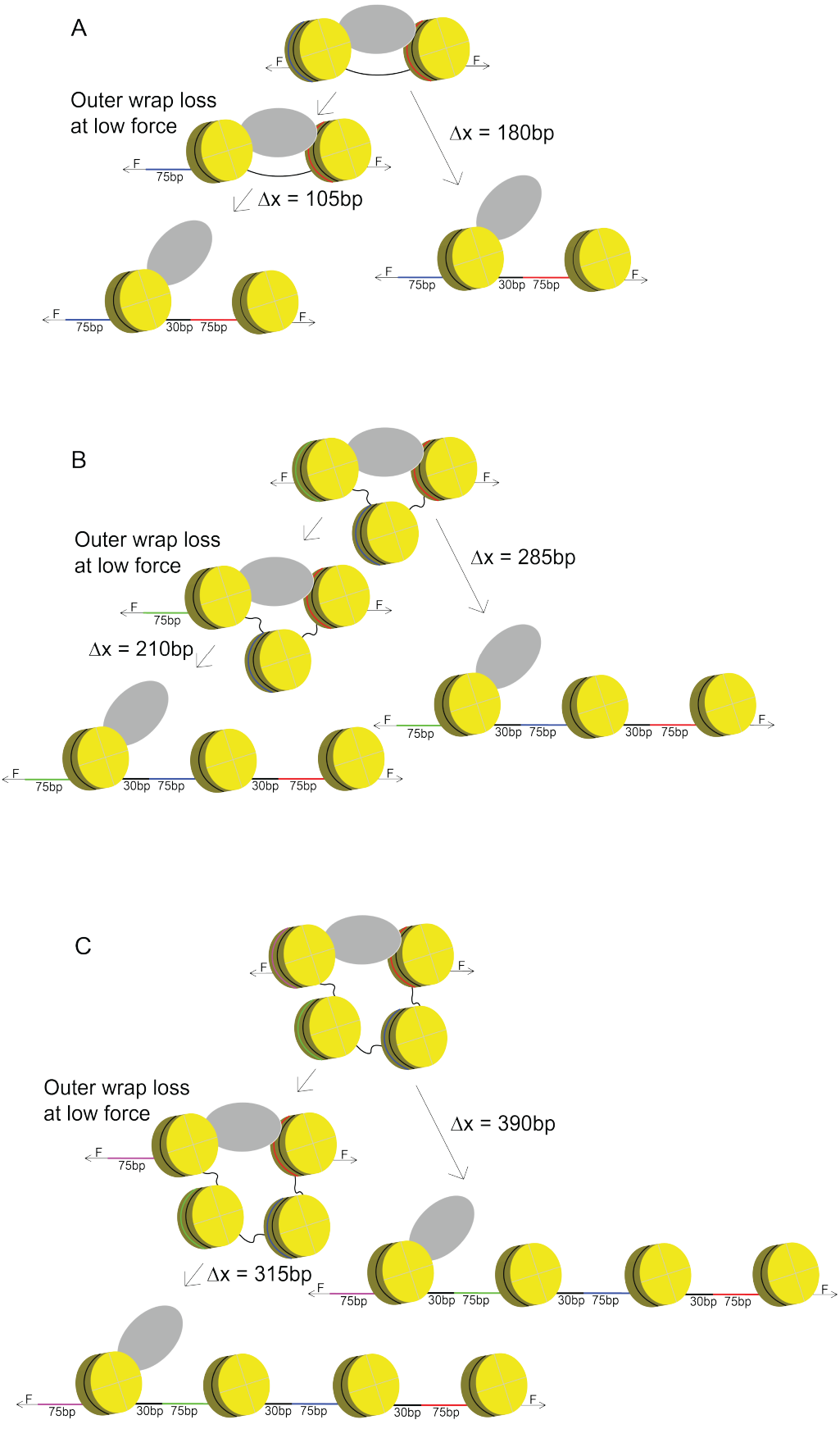
A
Outer wrap loss
at low force
Δx = 180bp
Δx = 105bp
F
75bp
30bp
B
Outer wrap loss
at low force
Δx = 285bp
Δx = 210bp
C
Outer wrap loss
at low force
Δx = 390bp
Δx = 315bp

Fig. S5. Schematic illustration of the amount of DNA released after disruption of PRC2-mediated bridging of nucleosome pairs. (*A*) Scenarios for force-induced disengagement of PRC2 initially engaged with two adjacent nucleosomes ($Nuc_{1\text{-}2}$ mode). In the left pathway, the outer wrap of the first nucleosome (blue segment) is not sequestered by PRC2, thus undone at low forces. Upon PRC2 unbinding that occurs at a higher force, the outer wrap of the second nucleosome (red segment, ~75 bp) and the 30-bp linker DNA—with a total length of ~105 bp—are released. The inner wraps of the nucleosomes have similar stabilities to PRC2 engagement. Therefore, they are expected to unravel later as independent transitions. Alternatively, as depicted in the right pathway, PRC2 sequesters both outer wraps. In this scenario, PRC2 disengagement would release two outer wraps (~150 bp) plus one linker DNA, totaling ~180 bp. If PRC2 sequesters part of the first outer wrap, a number between 105 and 180 bp is expected for the amount of DNA released. (*B*) Scenarios for force-induced disengagement of PRC2 bound to two nucleosomes that are separated by one spacer nucleosome ($Nuc_{1\text{-}3}$ mode). A total between 210 and 285 bp of DNA is expected to be released upon PRC2 disengagement. (*C*) Scenarios for force-induced disengagement of PRC2 bridging a pair of nucleosomes that are separated by two spacer nucleosomes ($Nuc_{1\text{-}4}$ mode). A total between 315 and 390 bp of DNA is expected to be released upon PRC2 disengagement.

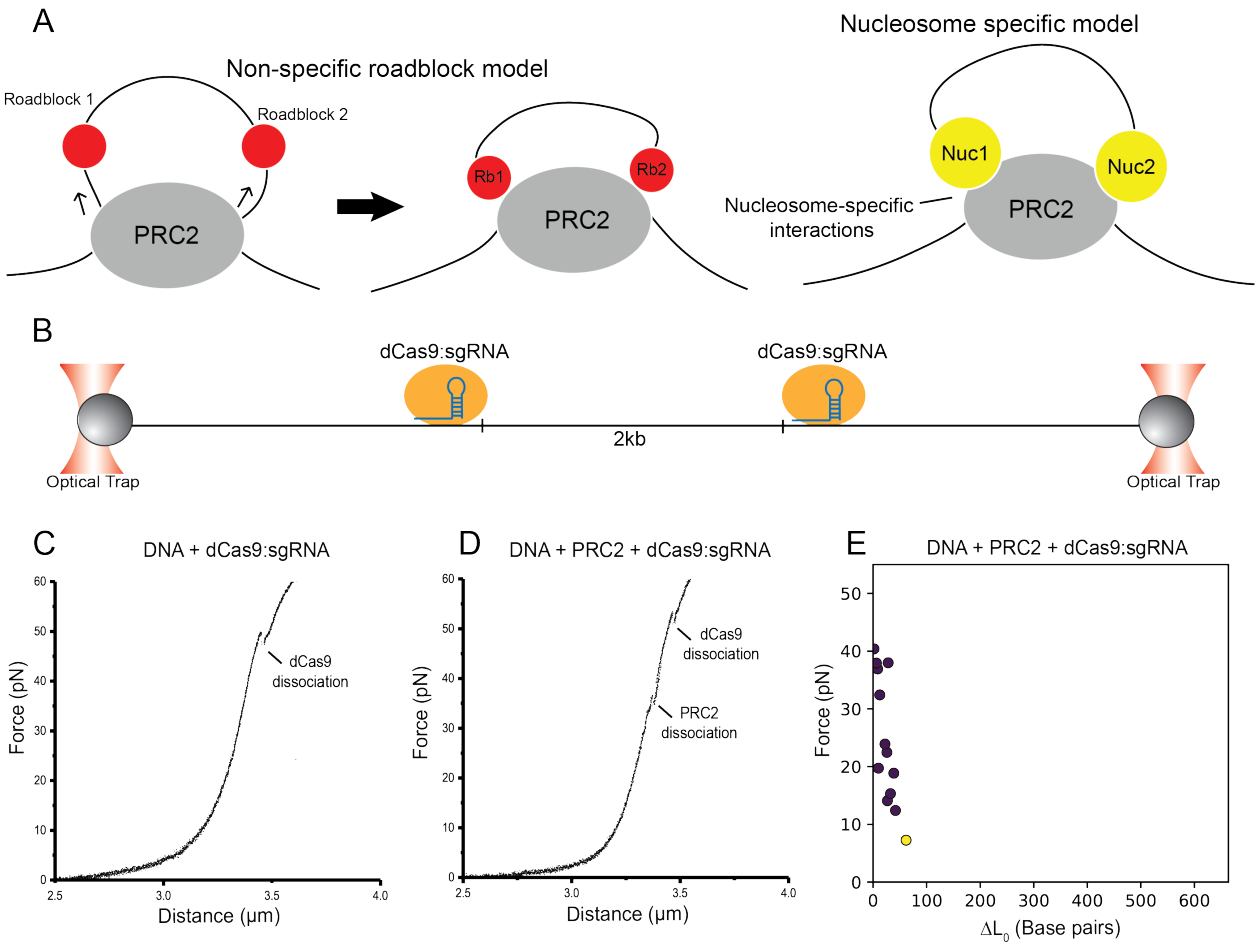

Fig. S6. Differentiating between alternative models for PRC2-mediated chromatin looping. (*A*) (*Left*) Schematic of a non-specific roadblock model in which PRC2 engages with two distal segments of DNA and slides on them—either via an inherent one-dimensional sliding activity of PRC2 or through force-induced threading—until running into a pair of steric blocks. (*Right*) Schematic of a nucleosome-specific model in which PRC2 makes specific contacts with a pair of nucleosomes. Both models could explain the observed changes in the end-to-end distance of DNA when PRC2 is mechanically dissociated by pulling the DNA. (*B*) Experimental design of two roadblocks engineered into a piece of tethered DNA based on dCas9:sgRNA targeting to specific DNA sites. The two dCas9 complexes are separated by ~2,000 bp. The same DNA was used to make 12-mer nucleosome arrays. (*C*) A representative force-extension curve for dCas9:sgRNA loaded DNA. Force-induced dCas9 dissociation can be observed at ~50 pN, indicating stable dCas9 binding. (*D*) A representative force-extension curve for dCas9:sgRNA loaded DNA incubated with PRC2. PRC2 disengagement events occurred at lower forces compared to dCas9 dissociation events. Importantly, no transition with a size close to 2,000 bp was observed, thus disfavoring the non-specific roadblock model. (*E*) Cluster analysis of PRC2-mediated transitions observed in the force-extension curves of PRC2-bound bare DNA in the presence of dCas9:sgRNA.

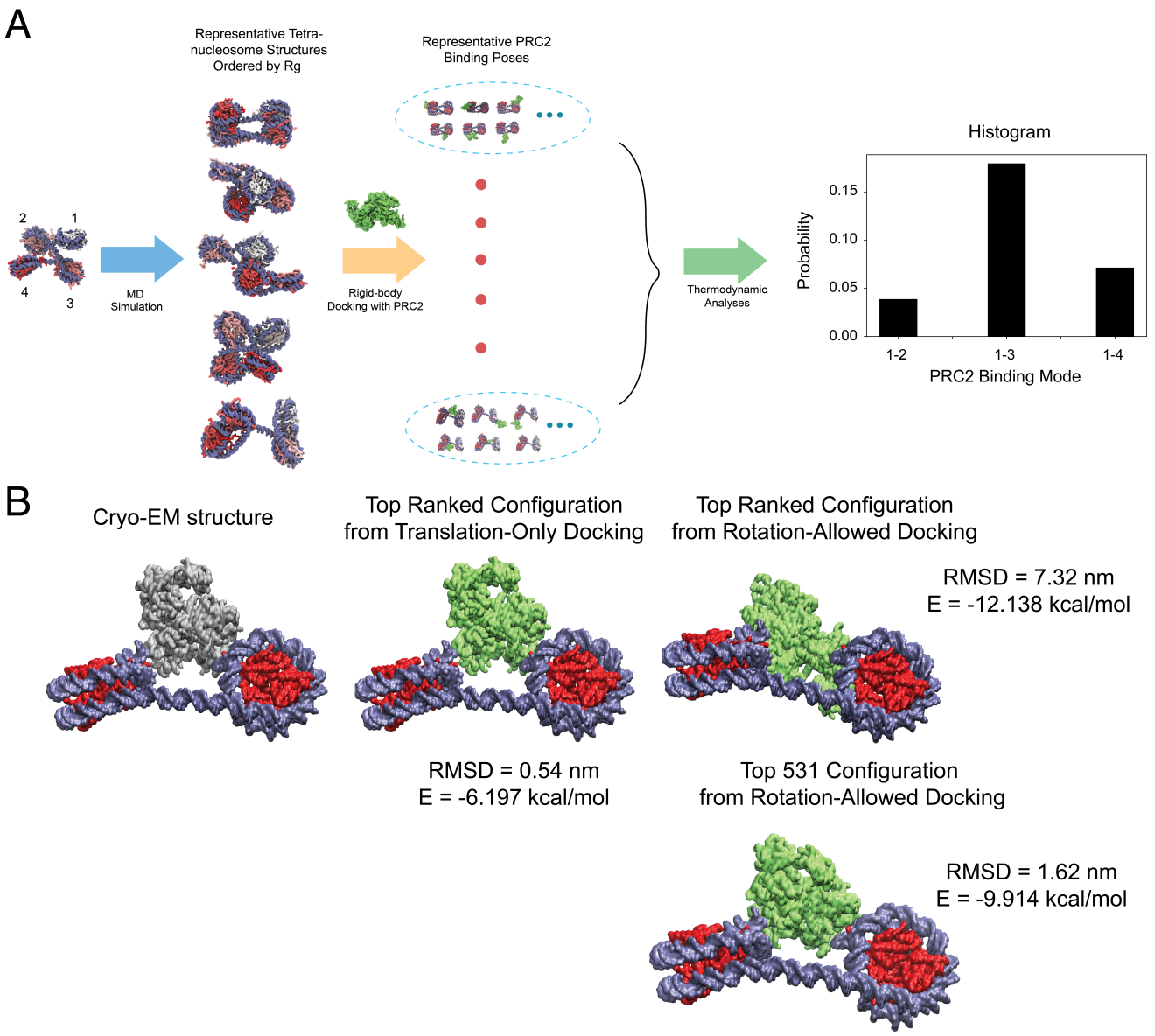

Fig. S7. Coarse-grained modeling of PRC2–chromatin interactions. (*A*) Flowchart of the algorithm used to quantify the various binding modes in which PRC2 engages with a tetranucleosome. See *Supplemental Methods* for details of the modeling. (*B*) Comparison between experimental and simulated PRC2 binding poses. (*Left*) Cryo-EM structure of the PRC2-dinucleosome complex determined by Ref. (1). (*Middle*) Lowest energy configuration predicted from rigid docking using a coarse-grained force field. See *Supplemental Methods* for details of the force field. The PRC2 structure used for docking includes SUZ12, EZH2, EED, and RBBP4 residues resolved in the cryo-EM structure. Only translational motion was allowed in these docking simulations and PRC2 was fixed in the same orientation as that found in the cryo-EM structure. The Cα RMSD between docked and cryo-EM structures is 0.54 nm. (*Right*) The lowest energy (*Top*) and a low RMSD (*Bottom*) configurations from docking simulations that allow both translational and rotational motions. In the lowest energy

configuration, PRC2 is again juxtaposed between the two nucleosomes as found in cryo-EM, but with a slightly tilted orientation. The difference in orientation between simulated and experimental configurations could be due to the omission of disordered regions in PRC2 and histones during docking. The accuracy of the force field could also potentially impact the PRC2 orientation. We found that the top 531st ranked configuration, though with an energy difference of less than 3 kcal/mol from the lowest energy configuration, is in much better agreement with the cryo-EM structure with a Cα RMSD of 1.62 nm. These results with the dinucleosome, therefore, support the use of the coarse-grained model for studying PRC2–chromatin binding modes.

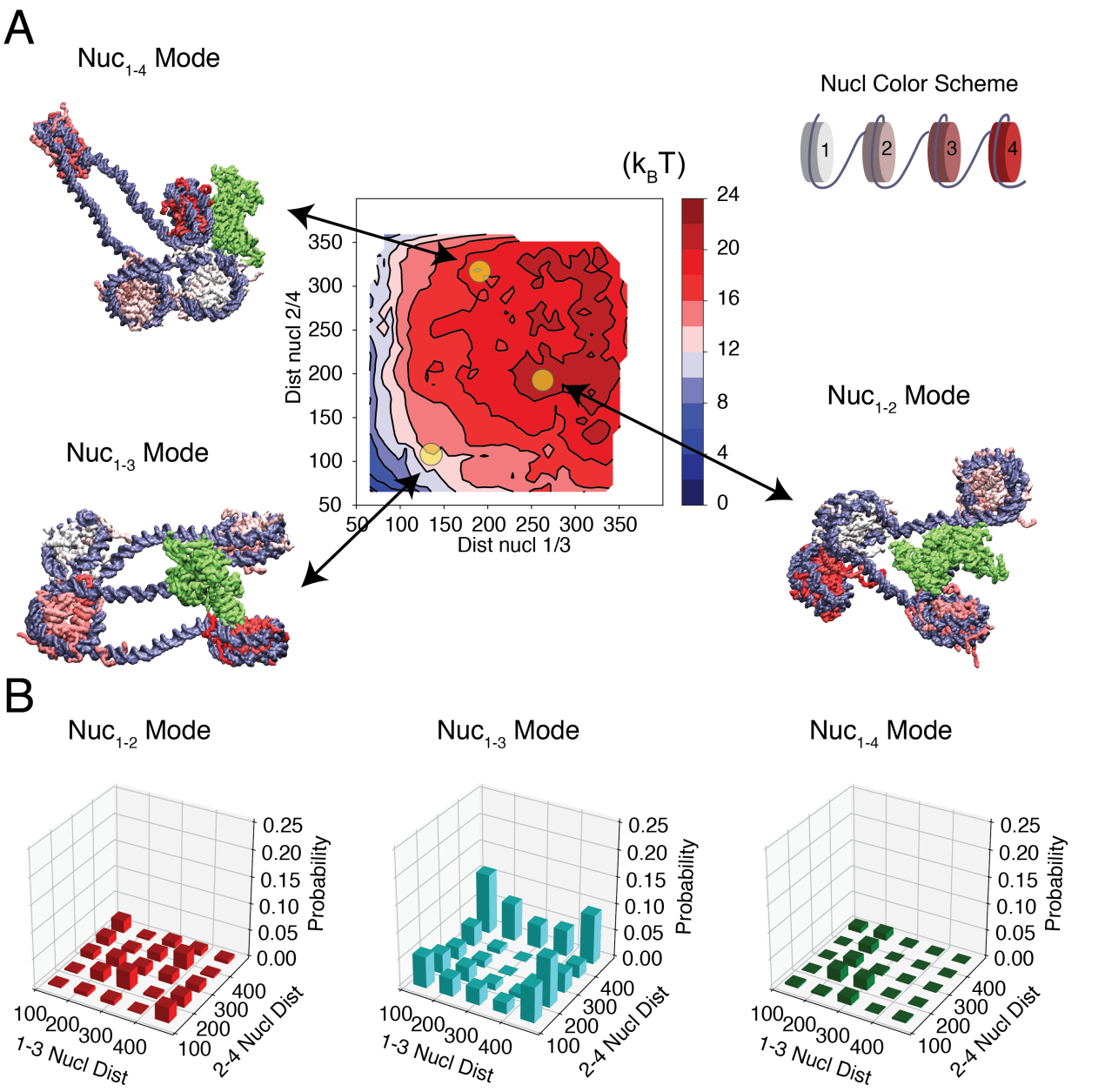

Fig. S8. In silico analysis of the interaction between PRC2 and tetranucleosomes with 50-bp linker. (*A*) Thermodynamic stability (free energy) for the tetranucleosome with a 50-bp-long linker DNA as a function of the spatial distances between the 1-3 and 2-4 nucleosome pairs. The impact of PRC2 was not included when computing the free energies. Example chromatin configurations are shown on the side, with DNA in purple and histone colors varying from white to red as the nucleosome index increases. PRC2 is colored in green and shown in its lowest energy pose in the three examples. (*B*) Fraction of PRC2 engaging in different nucleosome-bridging modes (Nuc_{1-2}, Nuc_{1-3}, and Nuc_{1-4}) at given tetranucleosome configurations grouped by its inter-nucleosome distances.

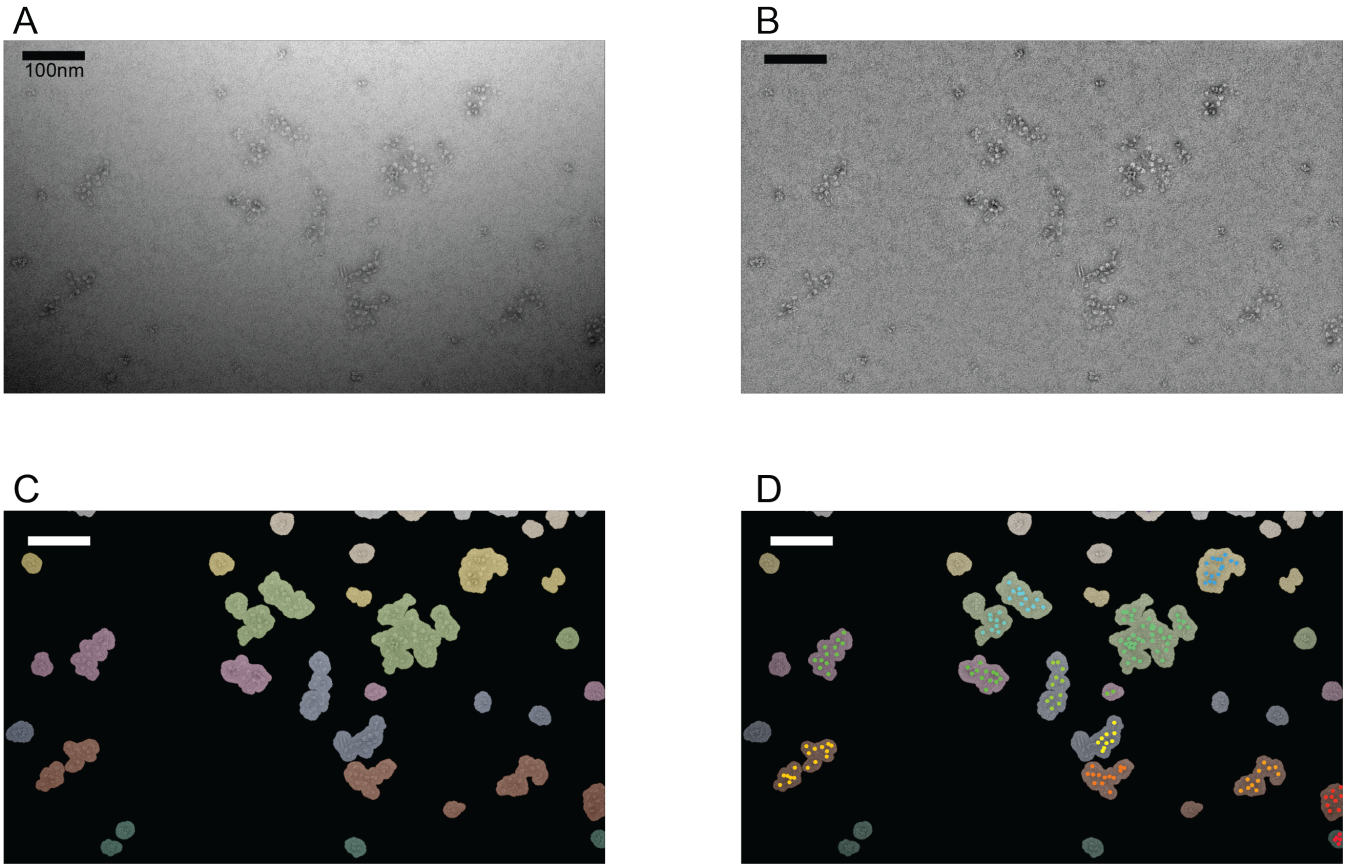

Fig. S9. Workflow for analyzing the electron micrographs of nucleosome arrays. (*A*) A representative negative-stain EM micrograph of 12-mer nucleosome arrays mixed with PRC2 at a 1:1 ratio (150 nM each). (*B*) The same micrograph after illumination correction to achieve a uniform background. (*C*) Binary masks of individualized nucleosome arrays. (*D*) Manually picked nucleosomes (colored dots) inside their corresponding masks. See *Materials and Methods* for details.

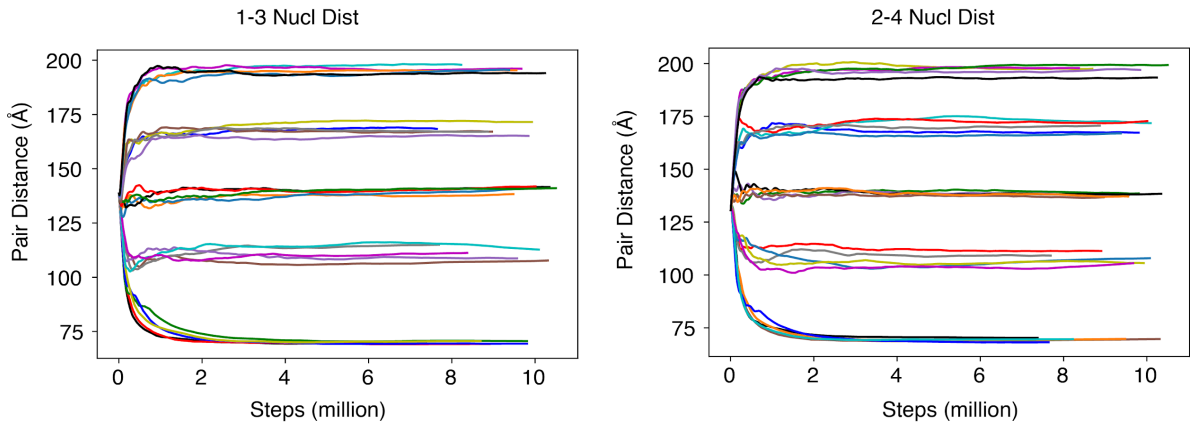

Fig. S10. Equilibration of tetranucleosome configurations in computer simulations. Cumulative averages of the distance between the 1-3 (*Left*) and 2-4 (*Right*) nucleosome pairs as a function of time steps for the various simulations performed to compute the tetranucleosome free energy profile.

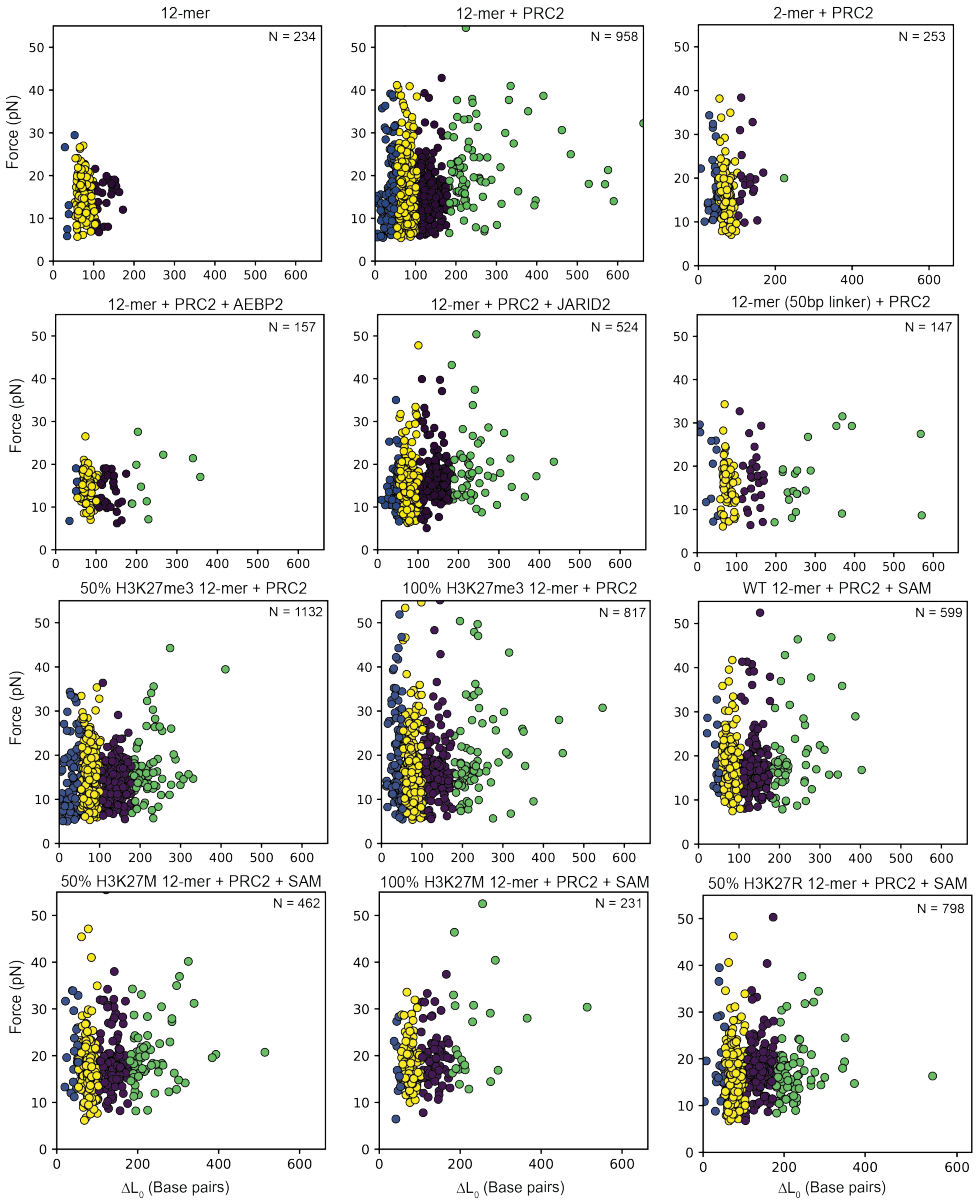

Fig. S11. Cluster analysis of force-induced transitions under different experimental conditions. X-axis and y-axis represent the contour length change and the rupture force of each transition, respectively. Blue, yellow, purple, and green symbols represent Cluster 1, Cluster 2, Cluster 3, and Cluster 4 transitions, respectively. *N* denotes the number of transitions collected under each condition.

Table S1. Statistics of force-induced transitions collected with different PRC2–chromatin assemblies. The nucleosome arrays all have 30-bp-long linker DNA unless noted otherwise. Force values are presented as mean ± SEM. Individual data points are plotted in Fig. S11.

Assembly	# of transitions	Average transition force (pN)	**Cluster 1** transitions		**Cluster 2** transitions		**Cluster 3** transitions		**Cluster 4** transitions	
			#	Transition force (pN)	#	Transition force (pN)	#	Transition force (pN)	#	Transition force (pN)
12-mer	234	16.0 ± 0.3	6	15.6 ± 4.1	203	16.2 ± 0.3	25	14.7 ± 0.9	0	-
12-mer + PRC2	958	17.0 ± 0.3	119	16.8 ± 0.8	541	16.6 ± 0.3	212	16.3 ± 0.4	86	21.4 ± 1.0
2-mer + PRC2	253	17.5 ± 0.3	36	17.4 ± 1.0	198	17.3 ± 0.3	18	20.1 ± 1.8	1	-
12-mer + PRC2 + AEBP2	157	14.5 ± 0.3	4	13.8 ± 2.7	112	14.7 ± 0.3	31	13.3 ± 0.7	10	16.3 ± 2.1
12-mer + PRC2 + JARID2	524	16.7 ± 0.3	43	15.3 ± 0.9	277	16.1 ± 0.3	149	17.2 ± 0.5	55	19.0 ± 1.1
12-mer (50-bp linker) + PRC2	147	17.1 ± 0.6	10	19.4 ± 2.7	88	16.1 ± 0.5	27	17.3 ± 1.3	22	19.6 ± 2.9
50% H3K27me3 12-mer + PRC2	1132	15.0 ± 0.1	112	13.7 ± 0.7	801	15.1 ± 0.1	316	14.6 ± 0.2	83	16.9 ± 0.8
100% H3K27me3 12-mer + PRC2	817	17.3 ± 0.4	103	20.3 ± 1.3	475	16.2 ± 0.4	163	17.2 ± 0.7	76	21.0 ± 1.3
WT 12-mer + PRC2 + SAM	599	17.8 ± 0.3	17	18.5 ± 1.6	350	17.4 ± 0.3	175	17.8 ± 0.5	57	19.8 ± 1.2
50% H3K27M 12-mer + PRC2 + SAM	462	18.3 ± 0.3	22	21.5 ± 1.6	252	17.2 ± 0.4	126	19.1 ± 0.6	62	20.3 ± 0.9
100% H3K27M 12-mer + PRC2 + SAM	231	19.6 ± 0.5	10	18.1 ± 2.2	126	18.5 ± 0.4	71	19.6 ± 0.7	24	26.1 ± 2.6
100% H3K27R 12-mer + PRC2 + SAM	798	17.2 ± 0.3	23	19.3 ± 1.7	495	16.8 ± 0.2	198	18.0 ± 0.4	82	17.7 ± 0.7

SUPPLEMENTAL METHODS

dCas9 roadblock experiments

dCas9 was purified as described previously (2). Two sgRNA sequences (IDT) were designed to target two sites flanking the 12 repeats of 601 nucleosome positioning sequence within the DNA template that was also used for generating the 12-mer nucleosome arrays. dCas9 and sgRNA were incubated for 10 min at room temperature at a concentration of 1 μM (3). The dCas9:sgRNA complexes were diluted to a final concentration of 25 nM and flown into Channel 4. DNA tethers were incubated with dCas9:sgRNA for 3 min before pulling. For experiments with PRC2, tethers were incubated with dCas9:sgRNA in Channel 4 and subsequently moved to Channel 5 that contained 500 nM PRC2 in the imaging buffer.

In silico modeling and simulations

Structural modeling for PRC2 and tetranucleosomes. We built a structural model for the full-length human PRC2 protein using homology modeling (4). Only the four core subunits, SUZ12, EZH2, EED, and RBBP4 were included in the structure. Three partially solved EM/X-ray structures of PRC2: PDBID 6C23 (5), 5WAI (6) and 5HYN (7) were used as templates for structural modeling. Missing residues that cannot be found in any of the PDB structures were built as random loops.

An initial configuration for the tetranucleosome was obtained by sequentially extending the dinucleosome cryo-EM structure (1). A 30-bp-long linker DNA (sequence: TATGACAGTGCATCACGGGGTGAGATCGCT) was used to connect the 2^{nd}/3^{rd} and the 3^{rd}/4^{th} nucleosomes. The linker DNA segments were constructed in the perfect helical form using 3DNA tool kit (8), and their orientations were dictated by the exiting nucleosomal DNA configuration to minimize bending and twisting. Configurations for the 3^{rd} and 4^{th} nucleosomes were taken from the PDB structure (ID: 3LZ1) and we replaced the protein coordinates with those from 1KX5 to model disordered histone tails. In the final construct, all nucleosomes share the same 601 sequence. We followed a similar procedure to model the tetranucleosome with 50-bp-long linker DNA (sequence: TATGACAGTGCATCACGGGGTGTGACAGTGCATCACGGGGTGAGATCGCT). We note that our results and conclusions are independent of this initial configuration because of the use of extensive simulations for equilibration.

Coarse-grained simulation of tetranucleosomes. We combined the 3SPN.2C DNA model (9) and the structure-based Cα model (10, 11) to create a coarse-grained force field for accurate and efficient modeling of protein–DNA interactions. We represent each DNA base with three beads and every α-carbon with one bead. The energy function of the system includes contributions from intra-DNA, intra-protein, inter-protein and protein–DNA interactions. Parameters from 3SPN.2C were directly applied to model intra-DNA interactions for the tetranucleosome sequence studied here. When simulating proteins with the structure-based model, we treated each histone octamer as a single unit. Intra-protein interactions therefore refer to all interactions within an octamer, while inter-protein interactions correspond to those between octamers. To ensure the stability of the histone octamer during simulation, we included a list of native contacts for intra-protein interactions. These contacts were generated from the PDB structure (ID: 1KX5) using the Shadow contact map (12). Two residues were considered in contact if their minimal atomic distance is 6 Å or less, regardless of whether they are from the same protein chain or not. We scaled the energy of the structure-based model by a factor

of 2.5 to 0.6 kcal/mol to keep the protein complex from unfolding at a temperature of 300 K. Detailed expressions of the energy function for the protein and DNA models can be found in Ref. (9, 13). Electrostatic interactions modeled at the Debye-Hückel level were included between charged beads, including DNA phosphates, Lys, Arg, Glu, and Asp residues. A salt concentration of 150 mM was used for the screening effect. In addition, a weak, nonspecific Lennard-Jones potential was applied between all protein–DNA beads. Detailed expressions for these potentials can be found in Ref. (14). Our model treats water molecules implicitly, and the solvation effect was accounted for when parameterizing the DNA model and protein–DNA interactions. Similar treatments are widely used in coarse-grained modeling of protein–DNA complexes and were shown to accurately model different properties of chromatin, including the energetic cost of nucleosomal DNA unwrapping (15), the twisting of DNA relative to core histones (16), and the interaction strength between a pair of nucleosomes (17). Since the stability of chromatin conformation and PRC2–chromatin interaction is mostly driven by electrostatic interactions (18), we anticipate the results shown here are robust with respect to our coarse-grained modeling approach. Further studies are needed to investigate DNA-sequence-specific PRC2 binding and PRC2–histone interactions.

To further improve the computational efficiency, we modeled the core region of each nucleosome as rigid bodies. This region includes the folded segments of the histone octamer and the central 107 bp of nucleosomal DNA. The rigid units consist of six degrees of freedom for translation and rotation and all atoms within the same unit move concurrently. Our setup maintains the flexibility of the linker DNA, part of the outer-wrap nucleosomal DNA (20 bp from the entry/exit site of a nucleosome), and disordered histone tails. Since the inner DNA wrap is known to bind tightly to the well-folded histone core in resting nucleosomes under no stress, we anticipate the rigid-body treatment to be a good approximation.

The most stable tetranucleosome configuration is expected to be collapsed under a salt concentration of 150 mM. To explore more expanded configurations in which the tetranucleosome might bind more favorably to PRC2, we carried out 25 independent simulations of the tetranucleosome in the presence of harmonic biases. These biases were implemented to restrain the distances between the non-adjacent nucleosomes (1-3 and 2-4 nucleosomes) at specified values and adopt the following expression:

$$V_{bias} = \frac{k}{2}[(d_{13} - d_{13}^0)^2 + (d_{24} - d_{24}^0)^2].$$

d_{13} and d_{24} stand for the distance between the 1-3 and 2-4 non-neighboring nucleosomes. d_{13}^0 and d_{24}^0 are the corresponding target values and range from 80 to 200 Å with an increment of 30 Å. The spring constant k = 0.01 kcal/mol/Å^2 was chosen to overcome the free energy barrier at large distances while ensuring sufficient overlap among umbrella windows.

Molecular dynamics simulations were carried out with a time step of 5 fs using the LAMMPS software package to explore tetranucleosome configurations. The Nosé-Hoover thermostat was applied to maintain the simulations at a temperature of 300 K. Periodic boundary condition was enforced with a cubic box of 2,000 Å × 2,000 Å × 2,000 Å in size. The box length is much larger than the size of the tetranucleosome (~300 Å) in our simulation. A total of at least 7.5 million steps were carried out for each simulation and we saved the configurations along the trajectory at every 5,000 steps. Tetranucleosome configurations were well equilibrated in these simulations as evidenced by the

convergence of the cumulative averages of the distance between 1-3 and 2-4 nucleosomes (Fig. S10). Only data after the first three million steps were used for constructing the free energy profile.

Rigid-body docking for PRC2–tetranucleosome binding. In principle, one can study PRC2 binding with the tetranucleosome via molecular dynamics simulations. However, the slow timescale associated with diffusion and the rugged energy landscape make an exhaustive exploration of different binding modes challenging. To more efficiently study the binding, we applied a rigid docking procedure as detailed below.

First, to account for the conformational flexibility of the tetranucleosome, we selected 1,000 structures from biased simulations introduced in the previous section. The structures were chosen based on a *K*-means clustering over all the simulated configurations to include both collapsed and extended configurations. Each configuration was represented with the six inter-nucleosome distances for clustering.

For each one of the 1,000 structures, we then determined the set of most stable PRC2 binding configurations by evaluating the energy of a large set of structures. Specifically, we selected 577 PRC2 orientations from a uniform sampling of the three Euler angles. For each orientation, we then searched for every possible position on a grid of ~800 Å × 800 Å × 800 Å in size with a spacing of 2 Å using discrete Fourier transform. Since the size of PRC2 is roughly 150 Å, the grid is large enough to enclose all possible PRC2-bound tetranucleosome configurations. Interaction energy between PRC2 and tetranucleosome was evaluated using the same force field introduced in the previous section. To avoid steric clashes, we did not use the disordered regions of PRC2 (SUZ12: residue# 1-78, 150-153, 168-181, 210, 224-227, 254-294, 323-350, 364-422, 549-560, 686-739; EZH2: residue# 1-9, 183-210, 217-219, 250-256, 346-421, 480-513, 741-746; EED: residue# 1-76; RBBP4: residue# 1-2, 94-104, 413-425) for energy evaluation.

Population estimation of PRC2 binding modes. From the docking simulations performed in the previous section, we can estimate the fraction of different PRC2 binding modes at a given tetranucleosome configuration. These data, however, cannot be combined straightforwardly to estimate an overall probability for various binding modes as the tetranucleosome configurations were collected from biased simulations. Proper thermodynamic reweighting must be carried out before averaging as detailed below (19).

The probability of PRC2 binding in between adjacent nucleosomes ($\text{Nuc}_{1\text{-}2}$ mode) at distances d_{13}^o, d_{24}^o for 1-3 and 2-4 nucleosomes can be defined as:

$$p_{12}(d_{13}^o, d_{24}^o) = \frac{1}{Z}\int e^{-\beta H}\delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$
$$= \frac{1}{Z}\int e^{-\beta(E_N+E_P+E_{NP})}\,\delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}.$$

where R_N and R_P correspond to the nucleosome and PRC2 degrees of freedom and $Z = \int e^{-\beta H}\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$ is the partition function. $H = E_\mathrm{N} + E_\mathrm{P} + E_\mathrm{NP}$ is the potential energy, with E_N, E_P, and E_NP being the internal energy of the tetranucleosome, the internal energy of PRC2, and the binding energy between PRC2 and nucleosomes, respectively. $\delta[C_{12}(R_P)]$ represents all the configurations in which PRC2 binds simultaneously with the first and second nucleosomes. $\delta[d_{13}(R_N) - d_{13}^o]$ and $\delta[d_{13}(R_N) - d_{24}^o]$ select

out tetranucleosome configurations with 1-3 nucleosome distance at d_{13}^o or 2-4 nucleosome distance at d_{24}^o, respectively.

The above definition, though formally exact, is difficult to calculate in practice, as it requires simulations in the presence of both PRC2 and tetranucleosomes. To make progress, we invoke the mean field approximation and replace E_{NP} with the average value at the given distances d_{13}^o, d_{24}^o for the $\text{Nuc}_{1\text{-}2}$ mode, $\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}}$. The probability can then be simplified as:

$$p_{12}(\, d_{13}^o, d_{24}^o) \approx \frac{1}{Z} e^{-\beta\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}}} \int\int e^{-\beta(E_N+E_P)}\, \delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$

$$= \frac{1}{Z} e^{-\beta[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}} + F(d_{13}^o, d_{24}^o) + E_P]} \int \delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$

$$= \frac{1}{Z} e^{-\beta[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}} + F(d_{13}^o, d_{24}^o) + E_P]} \frac{N_{12}(d_{13}^o,\ d_{24}^o)}{N},$$

where $N_{12}(d_{13}^o,\ d_{24}^o)$ is the number of docked PRC2 structures engaging in the $\text{Nuc}_{1\text{-}2}$ mode at distances d_{13}^o, d_{24}^o, while *N* is the total number of docked PRC2 configurations. $F(d_{13}^o,\ d_{24}^o)$ is the free energy of tetranucleosomes with different 1-3, 2-4 nucleosome distances and was calculated with the weighted histogram analysis method (WHAM) from the biased simulations. To obtain the final population estimation independent of inter-nucleosome distances, we integrate over d_{13}^o and d_{24}^o

$$p_{12} = \frac{1}{Z} \int e^{-\beta[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}} + F(d_{13}^o, d_{24}^o) + E_P]} \frac{N_{12}(d_{13}^o,\ d_{24}^o)}{N} \mathrm{d}d_{13}^o \mathrm{d}d_{24}^o$$

Population of $\text{Nuc}_{1\text{-}3}$ and $\text{Nuc}_{1\text{-}4}$ binding modes can be similarly defined as:

$$p_{13} = \frac{1}{Z} \int e^{-\beta[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{13}} + F(d_{13}^o, d_{24}^o) + E_P]} \frac{N_{13}(d_{13}^o,\ d_{24}^o)}{N} \mathrm{d}d_{13}^o \mathrm{d}d_{24}^o$$

$$p_{14} = \frac{1}{Z} \int e^{-\beta[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{14}} + F(d_{13}^o, d_{24}^o) + E_P]} \frac{N_{14}(d_{13}^o,\ d_{24}^o)}{N} \mathrm{d}d_{13}^o \mathrm{d}d_{24}^o$$

To determine the numerical values of the various populations, the integration was converted into a summation over the 1,000 tetranucleosome structures. For each structure, we used the top 1,000 lowest PRC2 binding configurations determined from docking to estimate $\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle$ and $N_{12}(d_{13}^o,\ d_{24}^o)$. When determining the binding modes, the minimum atomic distance between PRC2 and each of the four nucleosomes was calculated. A cutoff distance of 0.8 nm (approximately one Debye length) was used to determine whether PRC2 and the corresponding nucleosome are in contact.

Information theoretic modeling of ChIP-seq data

We obtained the genome-wide profile of H3K27me3 for IMR90 cells from the ROADMAP epigenomics project (20) and binarized the data at a resolution of 200 bp (21). From this data, the mean mark occupancy, $\frac{1}{N}\sum_{n=1}^{N_g} s_n$, and the correlation coefficient between two marks that are separated by l nucleosomes, $\frac{1}{N_g - l}\sum_{n=1}^{N_g - l} s_n s_{n+l}$, were calculated. In these notations, the binary variable s_n indicates whether the histone mark is present (= 1) at the n^{th} nucleosome or not (= 0), and N_g is the genome length. We note that the experimental data used here were obtained from a bulk average, but epigenetic marks are under constant remodeling in individual cells. The correlation coefficients

between histone marks determined here, therefore, only represent a mean-field approximation to the actual values. Single-cell epigenomics data are becoming available and could provide more accurate estimations for the correlation between histone marks that are present on the same chromosome at the same time.

As emphasized in the main text, the correlation between histone marks from different nucleosomes may arise from a transient effect. To differentiate this transient effect from direct contacts that could support PRC2-mediated inter-nucleosome interactions and chromatin looping, we parameterized an information theoretic model. The model's potential energy adopts the following form:

$$E(s, L) = \sum_{n} \left[h s_n + \sum_{l=1}^{L} K_l s_n s_{n+L} \right],$$

where n indexes over a total of *N* = 201 nucleosomes. The parameters h and K_l measure the overall propensity for the appearance of the histone mark and the coupling strength between two marks separated by l nucleosomes, respectively. Values for these parameters were determined to ensure that the information theoretic model reproduces the mean mark occupancy and correlation coefficients estimated from the ChIP-seq data. We included inter-nucleosome interactions for up to $L = 35$ nucleosomes. Values for *N* and *L* were chosen such that the model can account for the full range of correlation between histone marks but remains computationally efficient. We note that, as shown in Ref. (22), the above information theoretic model provides the simplest and most plausible mechanism for the observed correlation between histone marks. Similar approaches have been used to analyze protein sequences and filter out indirect correlations to identify amino acid pairs that are in direct 3D contact (23, 24).

SUPPLEMENTAL REFERENCES

1. S. Poepsel, V. Kasinath, E. Nogales, Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat Struct Mol Biol* **25**, 154-162 (2018).
2. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
3. M. D. Newton *et al.*, DNA stretching induces Cas9 off-target activity. *Nat Struct Mol Biol* **26**, 185-192 (2019).
4. N. Eswar *et al.*, Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* **Chapter 2**, Unit 2 9 (2007).
5. V. Kasinath *et al.*, Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science* **359**, 940-944 (2018).
6. S. Chen, L. Jiao, M. Shubbar, X. Yang, X. Liu, Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol Cell* **69**, 840-852 e845 (2018).
7. N. Justin *et al.*, Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun* **7**, 11316 (2016).
8. X. J. Lu, W. K. Olson, 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**, 5108-5121 (2003).
9. G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, J. J. de Pablo, Coarse-grained modeling of DNA curvature. *J Chem Phys* **141**, 165103 (2014).
10. C. Clementi, H. Nymeyer, J. N. Onuchic, Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* **298**, 937-953 (2000).
11. J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, J. N. Onuchic, SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* **38**, W657-661 (2010).
12. J. K. Noel, P. C. Whitford, J. N. Onuchic, The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J Phys Chem B* **116**, 8692-8702 (2012).
13. J. K. Noel *et al.*, SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput Biol* **12**, e1004794 (2016).
14. B. Zhang, W. Zheng, G. A. Papoian, P. G. Wolynes, Exploring the Free Energy Landscape of Nucleosomes. *J Am Chem Soc* **138**, 8126-8133 (2016).
15. J. Lequieu, A. Cordoba, D. C. Schwartz, J. J. de Pablo, Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent Sci* **2**, 660-666 (2016).
16. J. Lequieu, D. C. Schwartz, J. J. de Pablo, In silico evidence for sequence-dependent nucleosome sliding. *Proc Natl Acad Sci U S A* **114**, E9197-E9205 (2017).
17. J. Moller, J. Lequieu, J. J. de Pablo, The Free Energy Landscape of Internucleosome Interactions and Its Relation to Chromatin Fiber Structure. *ACS Cent Sci* **5**, 341-348 (2019).
18. X. Wang *et al.*, Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Struct Mol Biol* **24**, 1028-1038 (2017).
19. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011-1021 (1992).

20. C. Roadmap Epigenomics *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
21. P. F. Kuan *et al.*, A Statistical Framework for the Analysis of ChIP-Seq Data. *J Am Stat Assoc* **106**, 891-903 (2011).
22. W. J. Xie, B. Zhang, Learning the Formation Mechanism of Domain-Level Chromatin States with Epigenomics Data. *Biophys J* **116**, 2047-2056 (2019).
23. F. Morcos *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* **108**, E1293-1301 (2011).
24. R. R. Stein, D. S. Marks, C. Sander, Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. *PLoS Comput Biol* **11**, e1004182 (2015).