Supporting Information:

Cooperative DNA Looping by PRC2 Complexes

Xingcheng Lin,[†] Rachel Leicher,[‡,§] Shixin Liu,[‡] and Bin Zhang[*,¶]

†*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*

‡*Laboratory of Nanoscale Biophysics and Biochemistry, The Rockefeller University, New York, NY 10065*

¶*Departments of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*

§*Tri-Institutional PhD Program in Chemical Biology, New York, NY 10065*

E-mail: binz@mit.edu

Adjusting the statistical potential for protein-DNA interactions

The statistical potential introduced by Skolnick and coworkers[S1] could improve the specificity of coarse-grained protein-DNA force fields. However, care must be taken when adopting it for molecular simulations to address the two following issues.

The uncertainty issue: The statistical potential was obtained from contact probabilities between pairs of amino acids and nucleotides estimated using protein-DNA complex structures in the PDB database. The corresponding absolute values of the pair-wise contact energy between amino acids are ill-defined. For example, arbitrarily shifting these values by a constant does not alter the contact probabilities. This uncertainty can be understood from a two-state analogy: the relative population of the two states only depends on their relative free energy difference, but not on the absolute value of the free energy for each state.

There is no easy way to address the uncertainty issue and we essentially chose the value that best fits the experimental binding free energy shown in Figure 2 of the main text. As shown in Figure S2, the simulated binding free energy using a statistical potential shifted with 0.1 kcal/mol agrees poorly with experimental values. The slope for the linear fit between simulated and experimental values is 0.78. The agreement between simulation and experiment improves if a shift of 0.05 kcal/mol was used to correct the statistical potential, and the slope increases to 1.01.

The double-counting issue: Since PDB structures are stabilized by all interactions that are present in the system, the statistical potential derived from them naturally accounts for contributions from electrostatic interactions. However, unlike hydrogen bonding and van der Waals interactions, electrostatic interactions between charged amino acids and phosphate atoms of the DNA are long ranged and sensitive to the screening effect of ions. They are better served with the Debye Hückel potential. Simply adding the Debye-Hückel potential on top of the statistical potential will inevitably introduce double counting of electrostatic contributions. To resolve this issue, we modified the statistical potential between charged pairs. Specifically, we decreased the attractive interactions between DNA phosphates and

positively charged residues by 0.05 kcal/mol, while increasing the corresponding values with negatively charged residues by 0.05 kcal/mol. The value 0.05 kcal/mol is the average strength of electrostatic interactions in the original statistical potential. The final values of the statistical potential are provided in Table S1.

We note that while quantitative values of the binding free energy depend on the parameters, the presented mechanisms on PRC2-DNA binding are robust. For example, even in simulations carried out without the statistical potential, the same protein-DNA interface as that shown in Figure 3 of the main text was observed (Figure S15A). The cooperative role of two PRC2 complexes in DNA looping was preserved as well (Figure S15B). The apparent insensitivity of mechanistic insights with respect to model parameters can be understood from the dominance of electrostatic interactions for PRC2-DNA binding.

Simulation details for protein-DNA binding free energy calculations

We carried out a series of umbrella-sampling simulations[S2] to compute protein-DNA binding affinity. A spring constant of 0.02 kcal/mol/Å^2 was used to restraint the center-of-mass (COM) distance between protein and DNA molecules at various windows. The centers of the umbrella windows vary from 0.0 to 140.0 Å with a spacing of 10.0 Å. The simulations were initialized from the corresponding crystal structures and carried out for 22 million steps with a timestep of 2 fs. The first two million steps were discarded as equilibration. A total of six independent trajectories initialized with different random seeds were performed for each umbrella window. Simulation data were combined using WHAM[S3] to compute the final free energy profile. A spherical correction term, $2k_{\mathrm{B}}T\log R$ was added to the free energy to account for the Jacobian arising from the transformation from Cartesian coordinate into the spherical coordinate.[S4] k_{B} is the Boltzmann Constant, T is the temperature, and R is the COM distance between protein and DNA.

Simulation details for PRC2-DNA binding

We rigidified the ordered regions of PRC2 to carry out simulations with a larger time scale of 5 fs, compared to 2 fs without rigidification. These regions include residues from EZH2: 10-182, 211-216, 220-249, 257-345, 422-479, 514-740; EED: 77-441; RBBP4: 3-93, 105-412. For the subunit SUZ12, we included 79-149, 154-167, 182-209, 211-224, 228-253, 295-322, 351-363, 423-548, 561-685 for PRC2b and 79-145, 156-166, 182-208, 228-254, 295-321, 329-337, 351-360, 368-378, 424-496, 505-548, 561-685 for PRC2f. The difference in SUZ12 arises from the different conformations adopted by the C2 domain (residue ID: 156-360). The same residues in PRC2f were applied to each monomer of PRC2d. The structure of these regions was shown to be stable with minimal changes in all-atom simulations (Fig. S1), supporting their rigidification. The loop regions reconstructed from homology modeling were kept flexible. Similar simulation approaches have been adopted elsewhere when studying protein complexes with both ordered and disordered domains.[S5,S6]

When building structure-based contact potentials for protein complexes, we applied slightly different procedures for the three PRC2 conformations to account for the variation of the C2 domain. The C2 domain was treated together with the main part of the core complex as a single rigid body in PRC2b. For PRC2f, it was separately modeled as an additional rigid body, while the loop region (residue ID: 146-155, 361-367) that connects C2 to the central part remains flexible. For simulations with two PRC2fs, we added interactions between a C2 domain and residues of the RBBP4 subunit from both the same and from the other complex. These interactions were based on the contacts seen in PRC2d. They could, in principle, stabilize dimeric structures, though dimerization was not observed in our simulations, and the C2 domain primarily binds with the DNA. A treatment similar to PRC2f was used for each monomer in PRC2d. The C2 domains were treated as rigid bodies independent from the rest of the protein complex. Structure-based interactions derived from the initial configuration help keep the dimeric structure intact during the simulations.

We performed umbrella simulations between PRC2 and a 147 bps long 601-sequence

DNA to compute its DNA binding affinity. The biasing coordinate, center of mass distance between PRC2 and DNA, was restricted to centers located at 0.0 to 390.0 Å with a spacing of 10.0 Å with a spring constant of 0.015 kcal/mol/Å^2. The simulations were performed for 12 million steps, and 6 trajectories were simulated with different starting random seeds. WHAM[S3] was used to compute the final free energy profile with spherical correction.

In addition to the three conformations for PRC2(core) 4mer, we built a structural model for the PRC2(AEBP2) 5mer to estimate the contributions from AEBP2 to DNA binding. Specifically, we made use of the most recently reported cryo-EM PRC2-AEBP2 structure[S7] that provides the most complete information to date. We further predicted the structure for the N-terminal part of AEBP2, which is missing from the cryo-EM structure, using Raptor X.[S8] RaptorX is a widely used deep-learning based structure-prediction tool. From these two structures, together with the ones used for building PRC2b, we constructed a complete PRC2-AEBP2 structure using Modeller.[S9] We carried out umbrella simulations to determined the DNA binding free energy profile for PRC2-AEBP2 5mer. Center of mass distance between PRC2 and DNA, was restricted to centers located at 0.0 to 600.0 Å with a spacing of 30.0 Å and a spring constant of 0.002 kcal/mol/Å^2. The simulations were performed for 12 million steps, and 6 trajectories were simulated with different starting random seeds.

Simulation details for DNA bending free energy profiles

We carried out two-dimensional umbrella simulations to evaluate the impact of PRC2 binding on DNA bending. The first biasing coordinate is the end-to-end distance defined as the center of mass distance between the two 5-bp-DNA-segment at the two DNA ends. The contact number between DNA and PRC2 was used as the second umbrella collective variable. It was calculated as the number of Cα atoms within 8 Å of DNA beads. We only included every other ten beads of PRC2 and DNA in determining the contacts for computational efficiency. Additional biasing on the contact number helps the system to explore different

PRC2 binding poses by promoting transitions between dissociation and attachment.

We studied the DNA end-to-end distance from 100 to 900 Å with a spacing of 100 Å and the contact number from 2.0 to 8.0 with a spacing of 2.0. An independent umbrella simulation of 6 million steps was performed at every combination of the two variables using spring constants of 0.0002 kcal/mol/Å^2 and 0.3 kcal/mol. Replica exchange was applied in these simulations between temperatures from 300K to 410K with a spacing of 10K to further enhance conformational sampling. WHAM[S3] was used to process the biased simulation data from all temperatures and compute the final free energy profiles.

Similar protocols were used for simulations with two copies of PRC2. The DNA end-to-end distance umbrella center was capped at 800 Å due to the increased stability of more compact DNA configurations. The contact number is now defined as the contact between the DNA and both copies of the protein. The simulations lasted for 10.5 million steps for PRC2b and PRC2f, and 5.5 million steps for PRC2d.

Analysis of Simulated Structural Ensemble

PRC2-DNA structure clustering

The single-linkage algorithm was used to cluster structures obtained from umbrella simulations biased at 200 Å (for the two PRC2f case) and 300 Å (for the two PRC2b case). The clusters were generated by progressively adding new structures into the cluster if its distance to any existing element of the cluster is smaller than a cutoff.[S10] The distance metric was defined using the root-mean-square-deviation (RMSD) measured based on all protein and DNA beads. A cutoff value of 5.0 nm was used for clustering.

Definition of loop length

We introduced a new collective variable, loop length, to quantify PRC2-mediated DNA loop formation. For each PRC2 complex, we first identify its binding sites as two non-overlapping 10-bp-long DNA segments that contact with PRC2. Two molecules are in contact if any C_α

atom of PRC2 is within 8 Å of the DNA sugar beads. If multiple binding sites were found, the two that are farthest apart were chosen. We then determined the contour length of the DNA segments enclosed by each PRC2 as the sequence distance (in the unit of base pairs) between its two binding sites. The final loop length is defined as the larger contour length of the two.

Simulation details for in silico pulling

For direct comparison with results from single-molecule force spectroscopy experiments, we performed pulling simulations starting from a hairpin structure of a two PRC2f-DNA system. This structure was predicted to be the global minima on the two-dimensional free energy shown in Fig. 4B of the main text. A force defined as

$$F = -k(R_0 + vt - R_t) \tag{1}$$

was applied to the two DNA ends. R_0 and R_t are the DNA end-to-end distance at time 0 and t. We used a speed $v = 5.0\times10^{-12}$ mm/step and spring constant $k = 0.01$ kcal/mol/nm. This implementation is equivalent as pulling each end at a speed of $v = 2.5 \times 10^{-12}$ mm/step. Using the mapping introduced by Takada and coworkers[S11] that each coarse-grained step corresponds to 1ps, we have the pulling speed as $v = 2.5$ mm/s. A total of 12 trajectories were performed with different random seeds for generating velocities at 300K. The simulations lasted for 20 million steps. As shown in Fig. S14A,C, we observed rupture events that lead to the release of approximately 35 bp DNA. This amount of DNA released is comparable to the average experimental value and much smaller than the size of the most stable loop shown in Fig. 5B of the main text. The reduction in loop size is a result of PRC2f sliding along the DNA (Fig. S14B). In particular, as the DNA expands, the second PRC2f moves closer to the first protein to maintain multiple contacts with the DNA. The rupture occurs when the two proteins clash into each other, resulting in a PRC2f losing one of its binding sites. The

loss of PRC2-DNA contacts releases the loop and gives rise to the jump in DNA extension. In some cases, rather than crashing into the first PRC2, the second PRC2f rotates around the DNA while sliding across the first PRC2. The system thus relaxes without experiencing a rupture event.

We note that our pulling speed is still too fast compared to experimental values. It is plausible that the quantitative results from our simulations are sensitive to the pulling speed. While testing simulations at an even slower speed is impractical, we introduced a third PRC2 to the initial loop structure and carried out additional pulling simulations. Notably, the presence of three copies of PRC2 increases both the frequency of the rupture events and the magnitude of the rupture force (Fig. S14D). More PRC2 complexes could accumulate in the experimental setting due to their sliding along the DNA. The more evident rupture events ensure the robustness of simulation results that are less likely to be an artifact of pulling speed.

The initial structure for pulling with three PRC2fs was obtained from a REMD simulation.[S12] This simulations was initialized by placing the third PRC2 next to the hairpin structure, and a total of 12 temperatures from 300K to 410K with a spacing of 10K were used. Upon equilibration, the third PRC2 will bind to the DNA, and the bound configuration was used to perform 12 independent pulling simulations. These simulations were again initialized with different random seeds for generating velocities at 300K and carried out with the same pulling speed as that in the two PRC2f-DNA case. The simulations lasted for a minimum of 15 million steps.

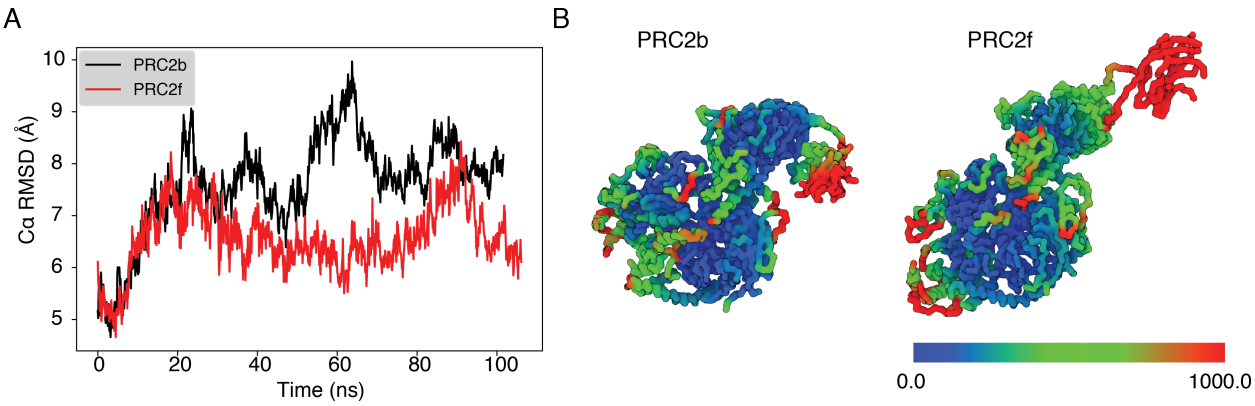


Figure S1: All-atom explicit-solvent simulations support the stability of PRC2 structures built from homology modeling. (A) Root mean squared deviations (RMSD) of PRC2b and PRC2f Cα atoms relative to the initial structure as a function of simulation time. Disordered regions and the flexible C2 domain were not included for calculating the RMSD. (B) B-factor values for each residue of PRC2b and PRC2f calculated based on the Cα atoms using the 100 ns all-atom simulations shown in part A. They were computed from the corresponding root mean squared fluctuation (RMSF) values using the expression $B_{\mathrm{i}} = \frac{8\pi^2 \mathrm{RMSF_i}^2}{3}$.[S13] Disordered loop regions are not shown for visual clarity.

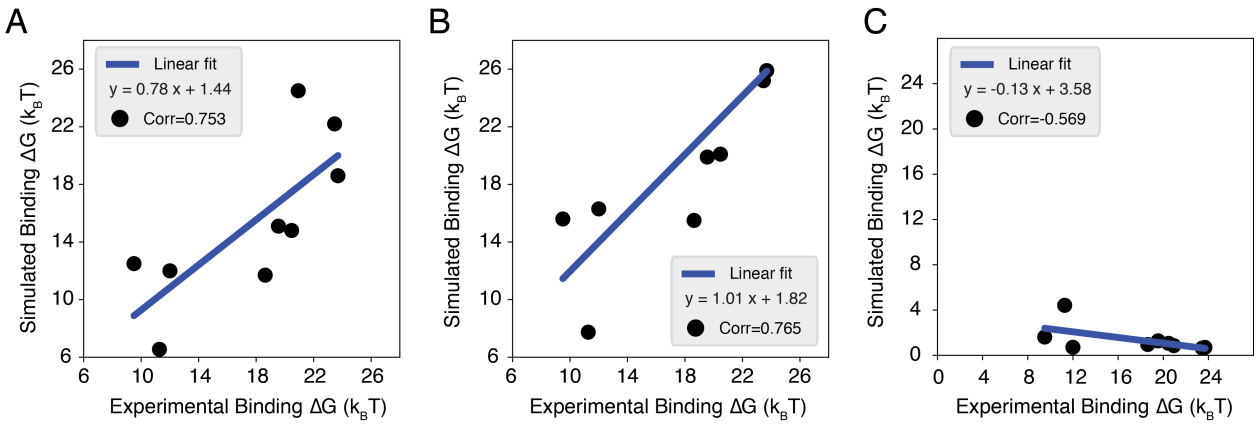


Figure S2: Binding free energy for various protein-DNA complexes computed using different modifications of the coarse-grained force field. See text *Section: Adjusting the statistical potential for protein-DNA interactions* for further discussion. (A, B) Results obtained from simulations that combine the statistical potential with the Debye-Hückel potential to model protein DNA interactions. The statistical potential was shifted by 0.1 (A) or 0.05 (B) kcal/mol with no corrections on the double counting issue. (C) Results obtained from simulations that solely employed the statistical potential, i.e., without the Debye-Hückel potential, for modeling protein DNA interactions.

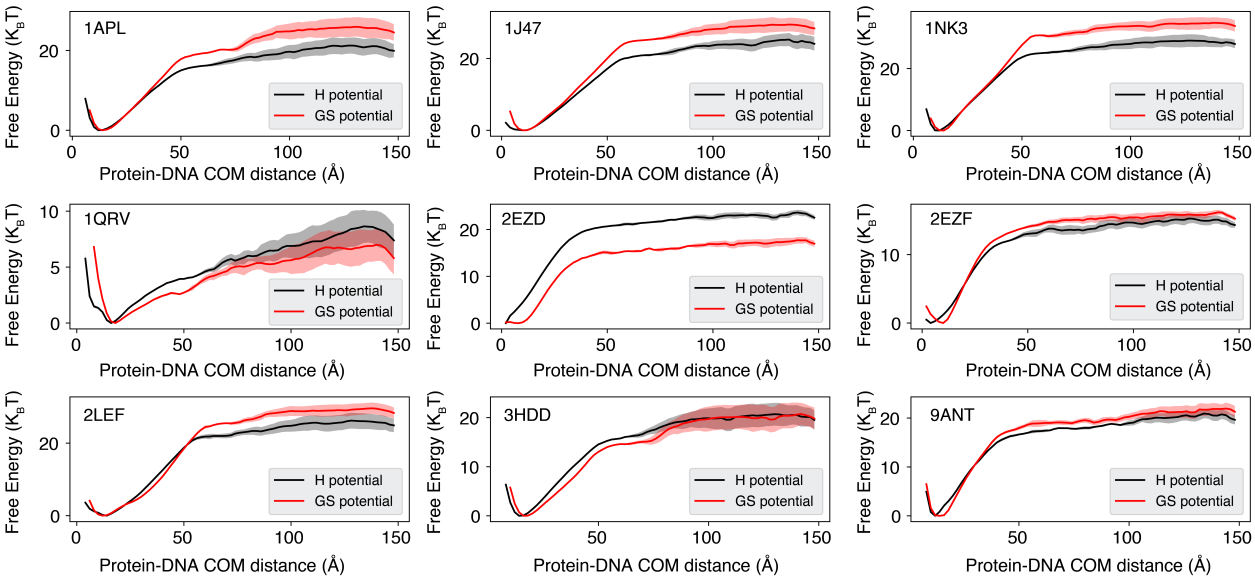


Figure S3: Binding free energy curves for the nine protein-DNA complexes used for benchmarking the force field. The PDB ID for each complex is shown in the top left corner. Both binding curves produced with a homogeneous potential (H potential, black) and with the newly calibrated force field (GS potential, red,[S1]) are shown. For the H potential, the interaction defined in Eq. 1 of the main text was replaced with the LJ potential $U(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}]$ with $\epsilon = 0.03$ kcal/mol and $\sigma = 5.7$Å. The lines correspond to mean values from three independent simulations and shaded regions represent standard deviations. See text *Simulation details for protein-DNA binding free energy* for simulation details.

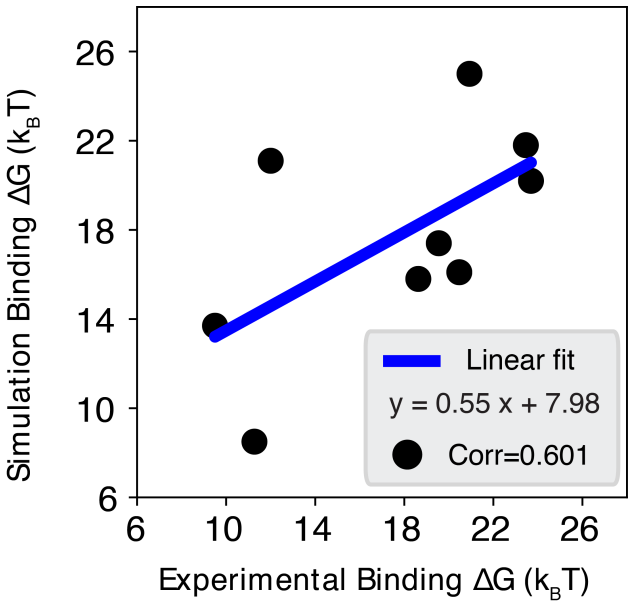


Figure S4: Comparison between simulated and experimental binding free energy for nine protein-DNA complexes. The simulated values were obtained using a homogeneous potential (H potential). See Figure S3 for details of these simulations.

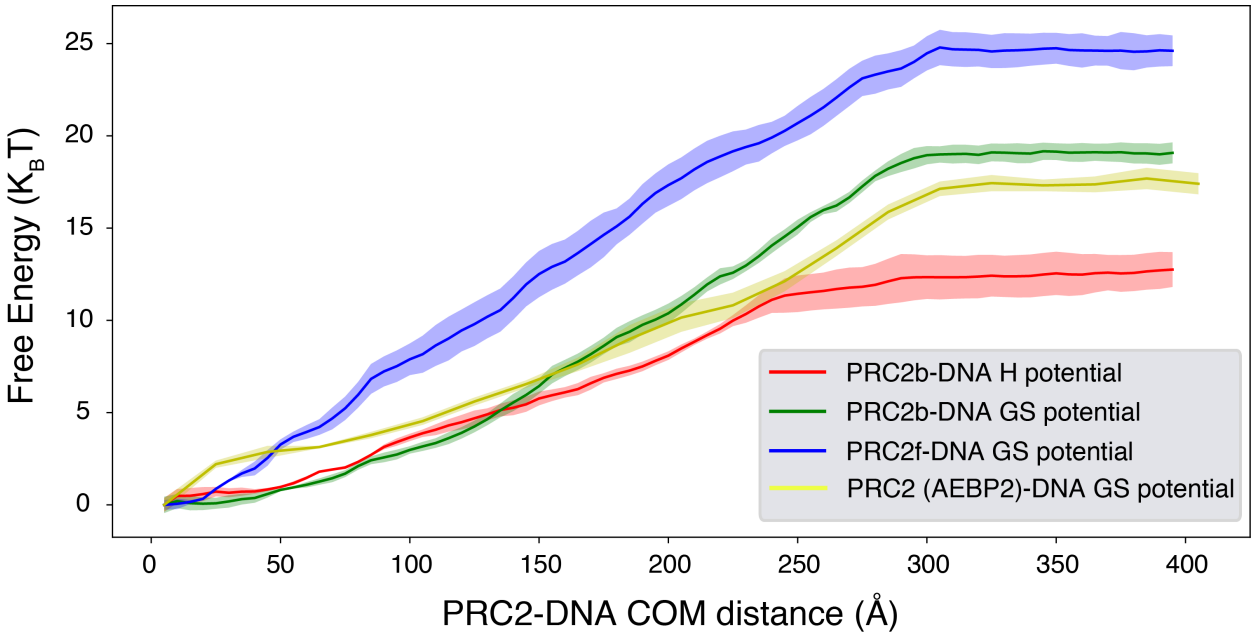


Figure S5: Binding free energy curves for various PRC2-DNA complexes. The results for the PRC2b-DNA complex calculated using the homogeneous potential (H potential) and the newly calibrated force field (GS potential) are shown in red and green. The curves for the PRC2f-DNA and PRC2(AEBP2)-DNA complex calculated with the newly calibrated force field are shown in blue and yellow. The lines correspond to mean values from three independent simulations and shaded regions represent standard deviations. See text *Section: Simulation details for PRC2-DNA binding* for additional discussion.

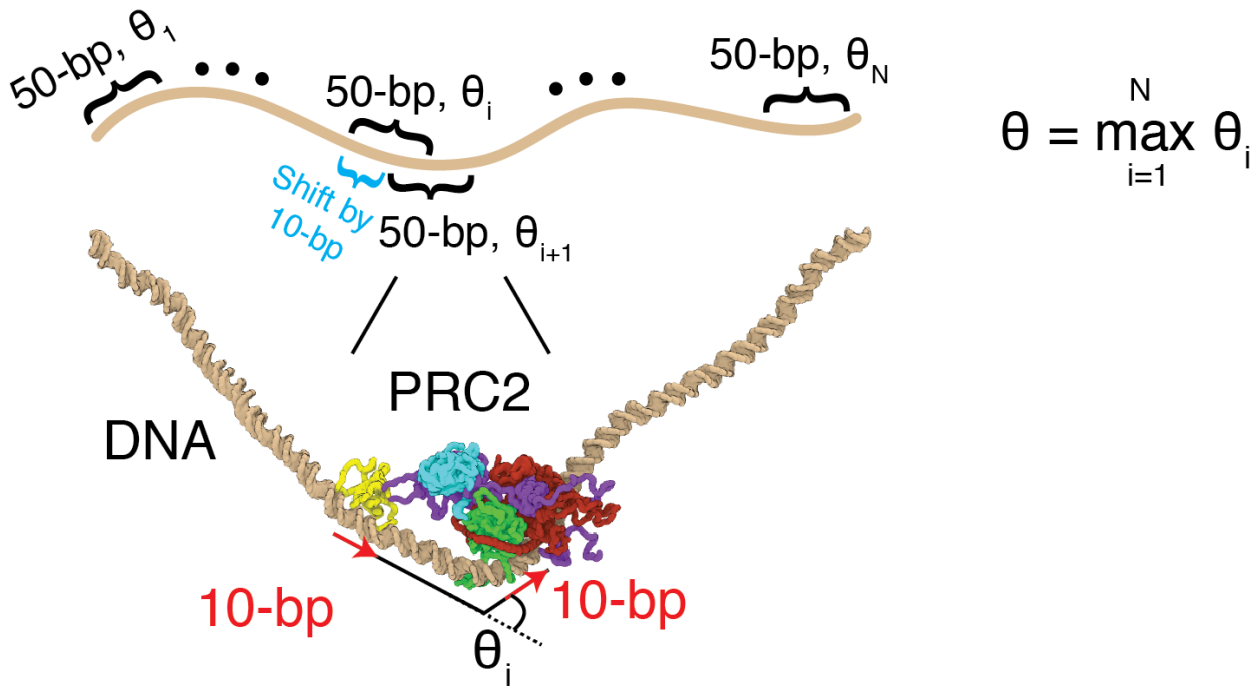


Figure S6: Illustration of the bending angle definition. A 50-bp long sliding window (black) was used to scan through the entire DNA and locate the most curved region with the largest bending angle. An increment of 10 bps was used when sliding along the DNA. As shown in the bottom panel, at the i-th sliding window and for each strand of DNA, we identified two vectors (red) using the two 10-bp segments at the ends of the 50-bp DNA segment. The vector in each 10-bp segment was defined using the position of the sugar beads of the first and last nucleotide. The angle between the two vectors was determined, and the average from the two strands was used to measure the bending angle θ_{i} of the sliding window. The final angle θ for the entire DNA was calculated as the maximum value of all θ_{i}.

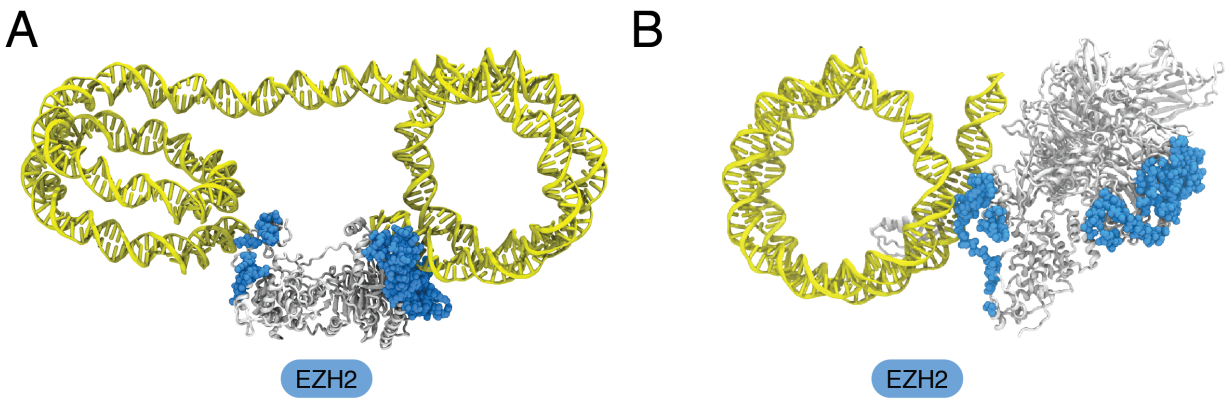


Figure S7: Cryo-EM structures of PRC2 bound to nucleosomes[S7,S14] support EZH2 as the primary interaction domain with DNA. The catalytic domains of PRC2 that include EZH2, EED and C-terminal SUZ12 are shown in white and nucleosomal DNA is drawn in yellow. Histone proteins are not shown for visual clarity. The residues predicted by our coarse-grained simulations to contact DNA (with contact probability shown in Fig. 3C of the main text larger than 0.1) are highlighted in blue spheres. They are seen to engage in DNA binding in the cryo-EM structures.

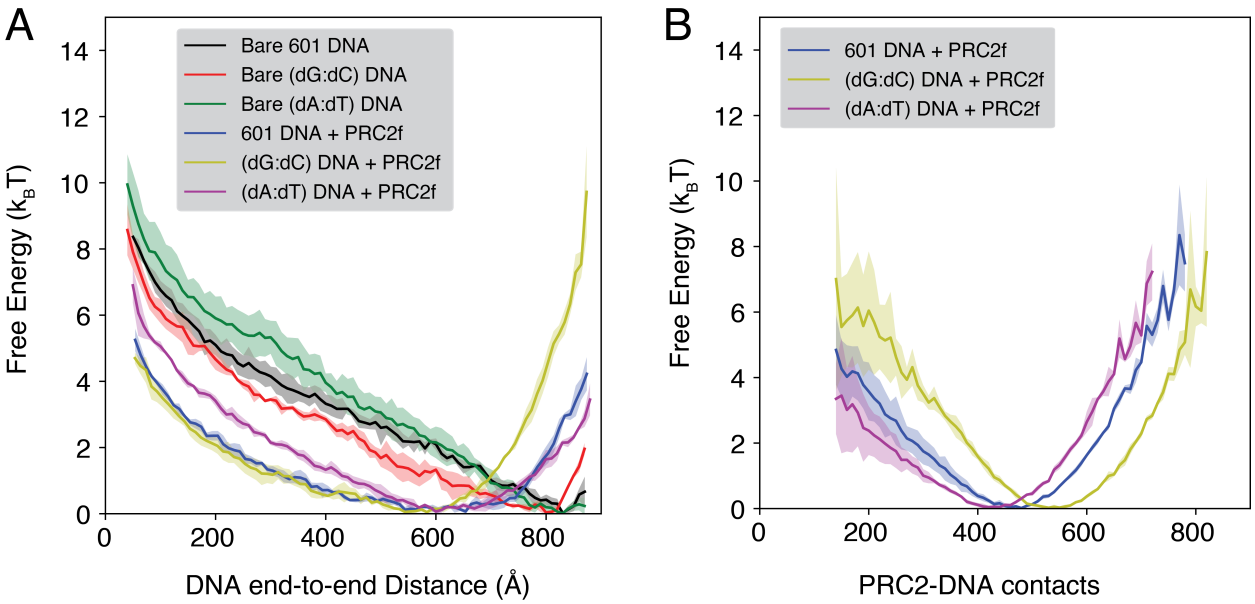


Figure S8: PRC2f preferentially binds and bends DNA with GC-rich sequence. (A) Free energy profiles as a function of the DNA end-to-end distance for the bare DNA with 601 (black), poly-dG:dC (red), poly-dA:dT (green) sequence. The corresponding results for DNA bound with PRC2f are shown in blue, yellow, and magenta, respectively. Results for the 601 sequence are identical to those shown in Fig. 3 of the main text. The shaded regions represent standard deviation of the mean. poly-dG:dC and poly-dA:dT DNA are of 147 bp long, and consists of $(G)_n$ or $(A)_n$ in one stand, and $(C)_n$ or $(T)_n$ in the complementary strand. (B) Free energy profiles as a function of the number of contacts between PRC2 and DNA. Coloring scheme is the same as in part A. PRC2f shows most binding with the poly-dG:dC sequence, indicating stronger binding. The lines correspond to mean values from three independent simulations and shaded regions represent standard deviations.

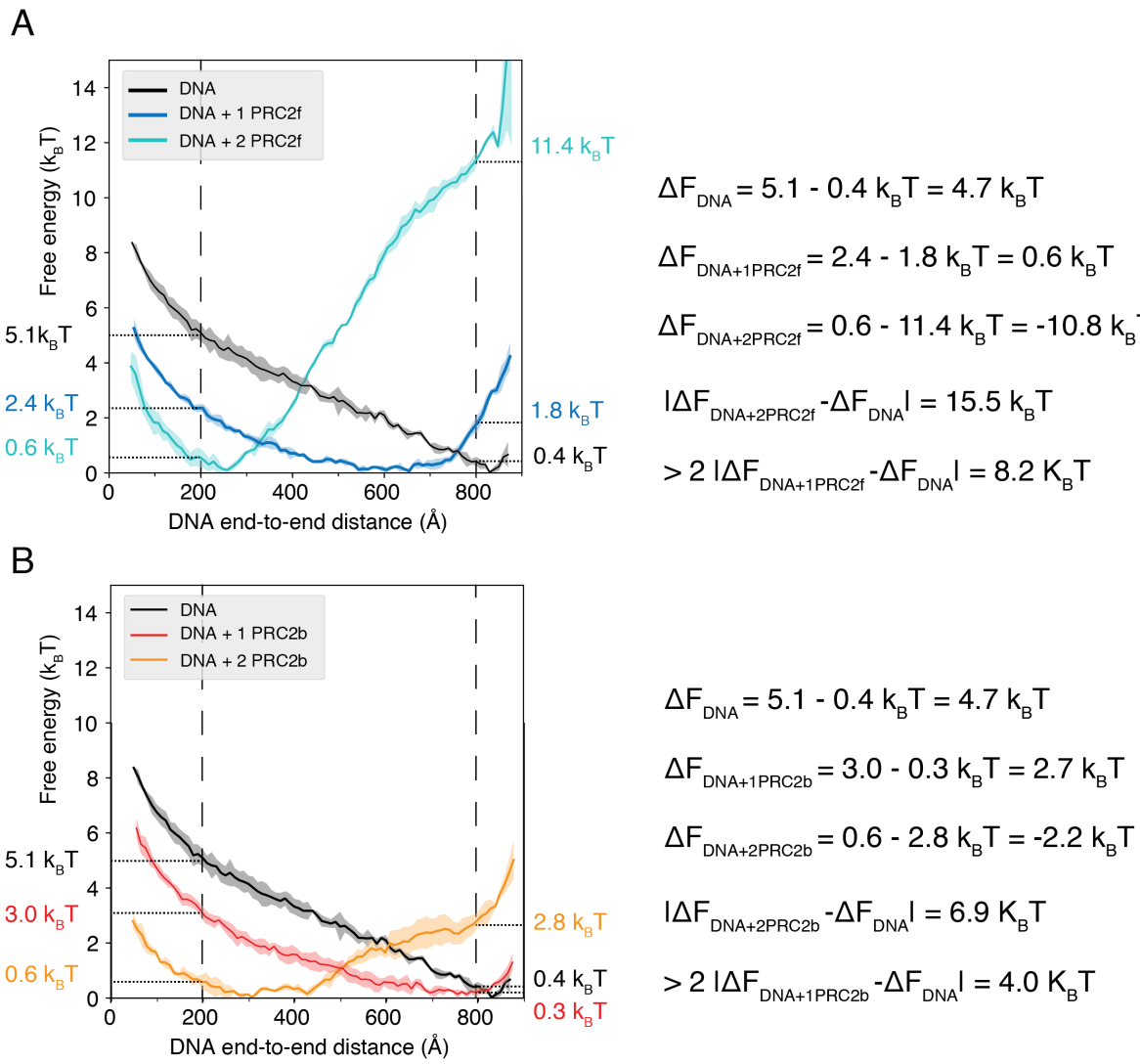


Figure S9: Illustration for the calculation of the free energy difference between open and closed DNA states for DNA bound with PRC2f (A) and with PRC2b (B). The open and closed states were defined at end-to-end distance of 800 and 200 Å as indicated by the dashed lines. The free energy values of the states can be read off from the intersections and were provided on the side with consistent colors. The right panel of each plot details the calculation of the free energy difference between open and closed states. Because the effect of two PRC2 in stabilizing the closed DNA state ($\Delta\Delta F$) is more than twice that of one PRC2, the two complexes are cooperative in DNA binding and bending.

Figure S10: Representative structures of the top 2 most populated clusters obtained from simulations performed with two PRC2b and a 294-bp-long DNA. The first cluster shows a state where two PRC2b independently bend a DNA, similar to the second cluster of simulations with two PRC2f (Fig. 4B of the main text). The second cluster shows a state where two PRC2b cooperatively loop a DNA, similar to the first cluster of the simulations with two PRC2f. The switched ordering in the relative population of the two clusters is consistent with the free energy profile of loop length shown in Fig. S11.

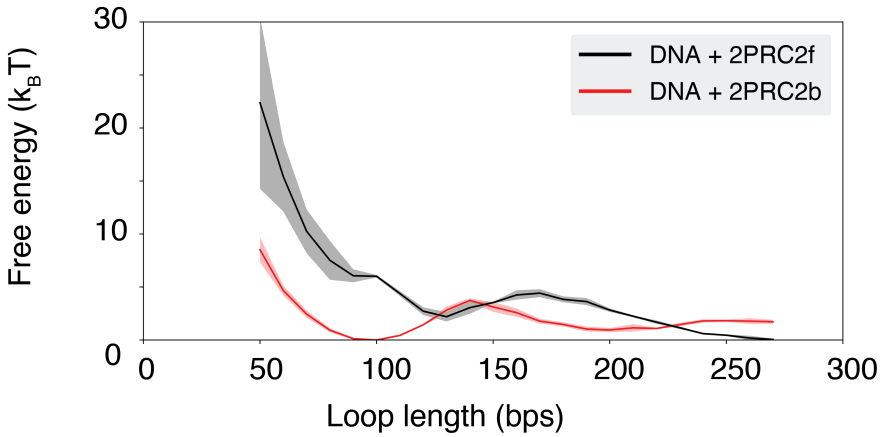


Figure S11: Free energy profile as a function of loop length, defined in Fig. 5A of the main text, for simulations with two PRC2f (black) and two PRC2b (red). The two PRC2f system exhibits more cooperativity to favor large loop formation, as is supported by the global free energy minimum at large loop length (∼250 bp)

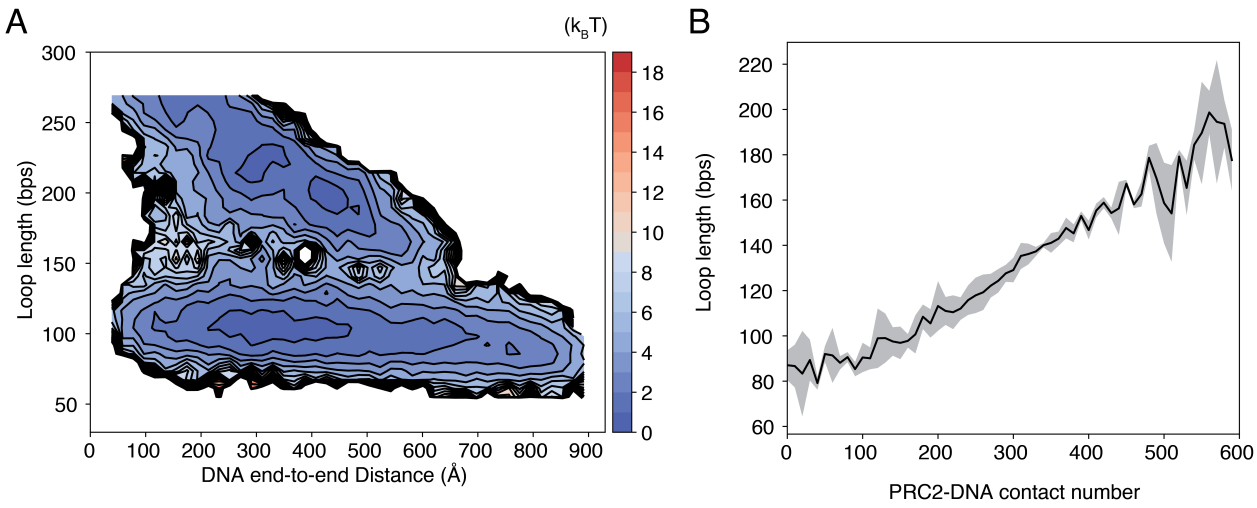


Figure S12: Cooperativity between the two PRC2b complexes is weaker compared to that between two PRC2f seen in Fig. 5 of the main text. (A) Two-dimensional free energy profile as a function of the DNA end-to-end distance and the loop length. (B) The average length of the dominant loop as a function of the number of contacts between the DNA and the PRC2f that forms the smaller loop. The shaded region represents standard deviation of the mean.

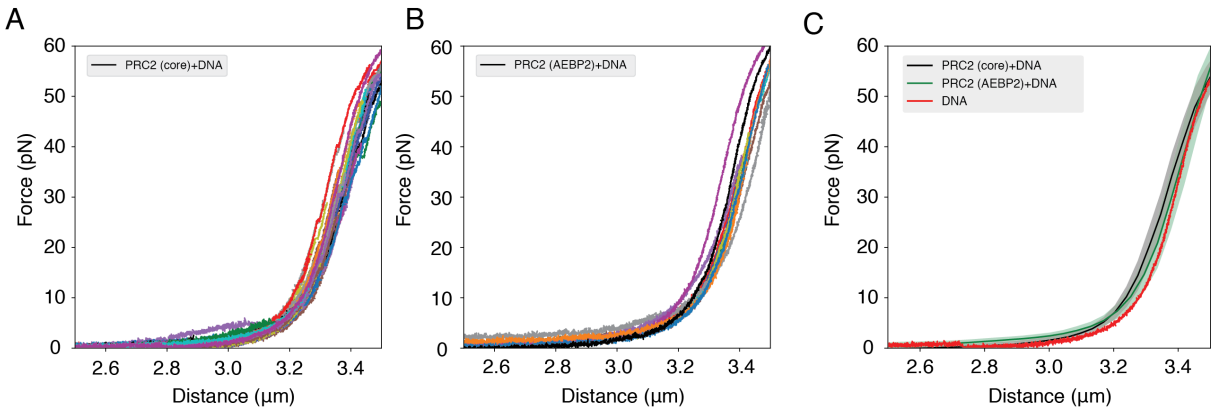
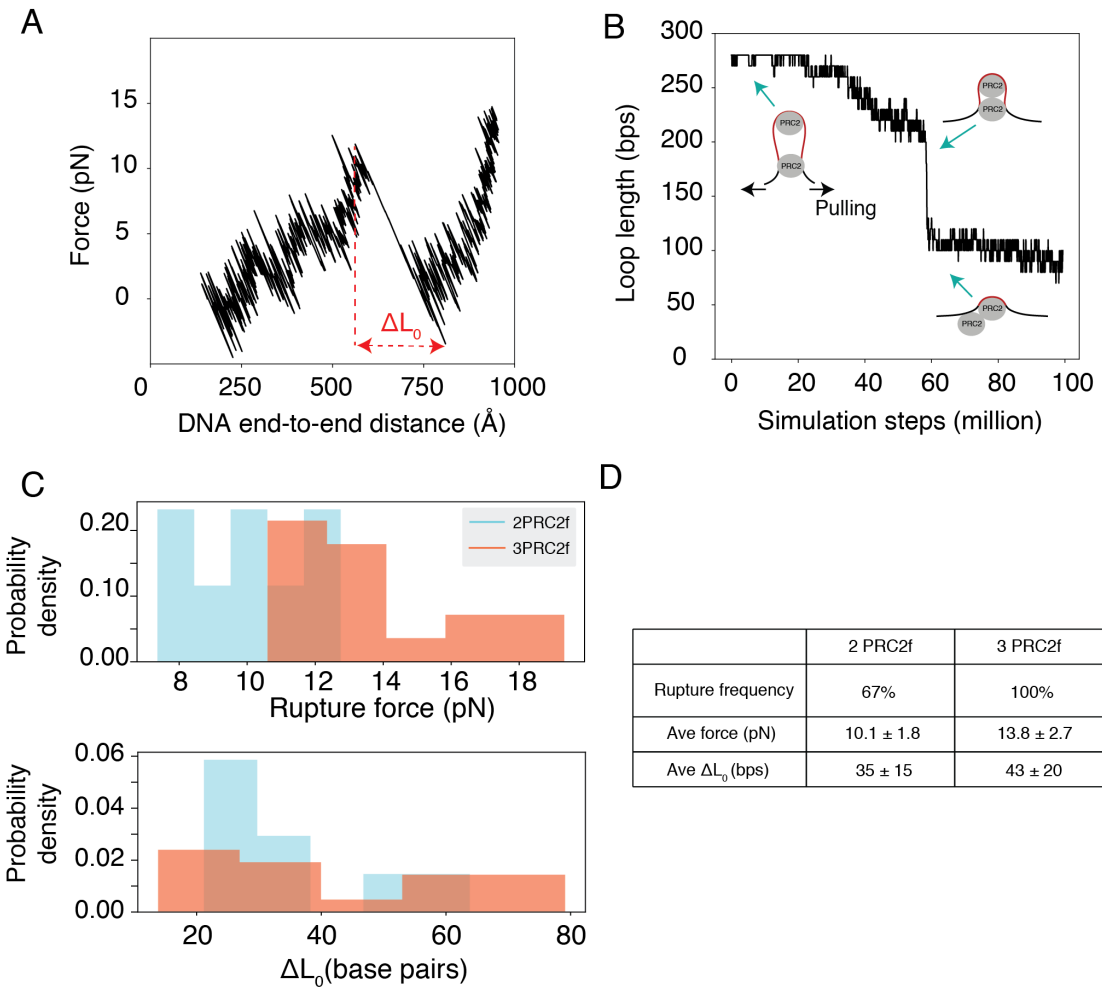


Figure S13: Force extension curves from single-molecule force spectroscopy experiments. (A,B) Individual traces for DNA incubated with PRC2 4mer (A) and PRC2(AEBP2) 5mer. (C) Average traces for PRC2 4mer (black) and PRC2(AEBP2) 5mer (green). The shaded regions represent standard deviations. The trace for bare DNA (red) is shown as red for reference.



	2 PRC2f	3 PRC2f	Experiment
Rupture frequency	67%	100%	n.a.
Ave force (pN)	10.1 ± 1.8	13.8 ± 2.7	18.6 ± 10.2
Ave ΔL_0 (bps)	35 ± 15	43 ± 20	24.3 ± 14.3

Figure S14: Quantitative comparison between in silico pulling simulations and the single-molecule force spectroscopy experiments. See text *Simulation details for in silico pulling* for more details. **(A)** Example simulated force-extension curve obtained from pulling the looped DNA configuration bound with two PRC2f complexes. ΔL_0 measures the change of the DNA contour length during a rupture event. **(B)** Time trace of corresponding loop length along the pulling simulation shown in part A. Example cartoon representations of the PRC2-DNA complex are shown on the side for illustration. **(C)** Probability distributions of the rupture force (top) and change in DNA contour length (ΔL_0, bottom) during rupture events for in silico simulations with two (cyan) and three (orange) PRC2fs. **(D)** Summary of the statistics of the rupture events observed from 12 independent simulations and from experiments. The rupture frequency was defined as the number of rupture events observed over all 12 simulation trajectories divided by 12. This number is not available (n.a.) for experimental measurements.

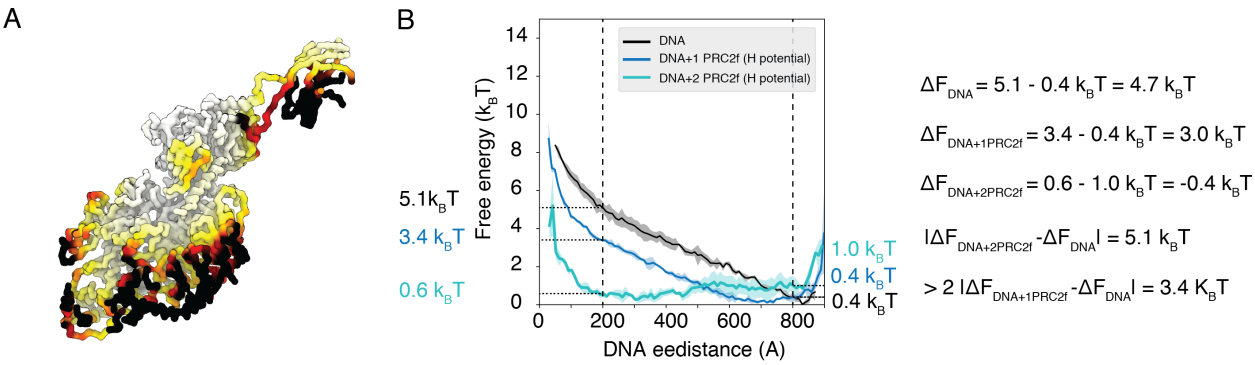


Figure S15: Main conclusions of the manuscript on DNA-binding interface and cooperative binding between multiple complexes are robust to the potential used for protein-DNA interactions. (A) Residue specific probability map for PRC2f contacting the DNA. This plot is a counterpart of the one provided in Figure 3C of the main text. (B) Illustration for the calculation of the free energy difference between open and closed DNA states. The open and closed states were defined at end-to-end distance of 800 and 200 Å as indicated by the dashed lines. The free energy values of the states can be read off from the intersections and were provided on the side with consistent colors. The right panel details the calculation of the free energy difference between open and closed states. Because the effect of two PRC2 in stabilizing the closed DNA state (F) is more than twice that of one PRC2, the two complexes are cooperative in DNA binding and bending. See Figure S9 for similar plots calculated using results presented in the main text.

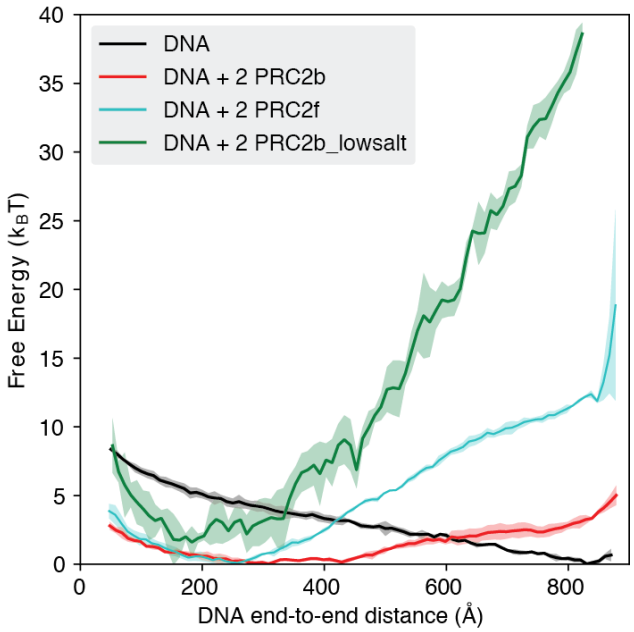


Figure S16: Dependence of PRC2-DNA interactions on salt concentration. The black, red, and blue curves are identical to those shown in Figure 4 of the main text. The green curve was obtained with two copies of PRC2b but at a salt concentration of 25 mM, which is the value used in the Perkins/Cech study.[S15] Comparing the green and the red curve, one can see that the stability of the looped configurations with end-to-end distance around 200 Å increases substantially upon lowering salt concentration. The increased stability could potentially facilitate loop detection in AFM imaging.

Table S1: Final values for the statistical potential in the unit of kcal/mol adjusted to address the uncertainty and double-counting issues. See text *Section: Adjusting the statistical potential for protein-DNA interactions* for further discussion.

DNA site	Residue	Value	DNA site	Residue	Value	DNA site	Residue	Value	DNA site	Residue	Value	DNA site	Residue	Value	DNA site	Residue	Value
P	ALA	0.09	S	ALA	0.14	A	ALA	0.33	T	ALA	0.11	C	ALA	0.24	G	ALA	0.32
P	ARG	-0.11	S	ARG	-0.15	A	ARG	-0.06	T	ARG	-0.17	C	ARG	-0.16	G	ARG	-0.18
P	ASN	-0.04	S	ASN	-0.02	A	ASN	-0.0005	T	ASN	-0.07	C	ASN	-0.04	G	ASN	0.03
P	ASP	0.07	S	ASP	0.18	A	ASP	0.37	T	ASP	0.22	C	ASP	0.05	G	ASP	0.24
P	CYS	0.11	S	CYS	0.24	A	CYS	0.29	T	CYS	0.10	C	CYS	0.22	G	CYS	0.40
P	GLU	0.16	S	GLU	0.28	A	GLU	0.43	T	GLU	0.24	C	GLU	0.10	G	GLU	0.43
P	GLN	0.002	S	GLN	0.02	A	GLN	0.09	T	GLN	0.04	C	GLN	0.03	G	GLN	0.04
P	GLY	0.006	S	GLY	0.04	A	GLY	0.28	T	GLY	0.10	C	GLY	0.13	G	GLY	0.18
P	HIS	-0.05	S	HIS	0.02	A	HIS	0.18	T	HIS	0.01	C	HIS	-0.02	G	HIS	-0.02
P	ILE	0.10	S	ILE	0.16	A	ILE	0.36	T	ILE	0.26	C	ILE	0.24	G	ILE	0.35
P	LEU	0.19	S	LEU	0.25	A	LEU	0.47	T	LEU	0.22	C	LEU	0.40	G	LEU	0.45
P	LYS	-0.07	S	LYS	-0.05	A	LYS	0.14	T	LYS	0.05	C	LYS	0.01	G	LYS	-0.02
P	MET	0.10	S	MET	0.10	A	MET	0.25	T	MET	0.14	C	MET	0.13	G	MET	0.26
P	PHE	0.07	S	PHE	0.12	A	PHE	0.32	T	PHE	0.16	C	PHE	0.25	G	PHE	0.22
P	PRO	0.08	S	PRO	0.12	A	PRO	0.26	T	PRO	0.14	C	PRO	0.35	G	PRO	0.32
P	SER	-0.06	S	SER	-0.01	A	SER	0.21	T	SER	-0.02	C	SER	0.13	G	SER	0.15
P	THR	-0.06	S	THR	-0.03	A	THR	0.22	T	THR	-0.02	C	THR	0.07	G	THR	0.19
P	TRP	-0.02	S	TRP	0.08	A	TRP	0.26	T	TRP	0.19	C	TRP	0.14	G	TRP	0.21
P	TYR	-0.05	S	TYR	-0.03	A	TYR	0.21	T	TYR	0.006	C	TYR	0.005	G	TYR	0.11
P	VAL	0.10	S	VAL	0.18	A	VAL	0.34	T	VAL	0.16	C	VAL	0.29	G	VAL	0.43

References

(S1) Gao, M.; Skolnick, J. DBD-Hunter: A Knowledge-Based Method for the Prediction of DNA–Protein Interactions. *Nucleic Acids Res.* **2008**, *36*, 3978–3992.

(S2) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(S3) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(S4) Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem* **2007**, *8*, 162–169.

(S5) Kim, Y. C.; Hummer, G. Coarse-Grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. *J Mol Biol* **2008**, *375*, 1416–1433.

(S6) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence Determinants of Protein Phase Behavior from a Coarse-Grained Model. *PLOS Comput. Biol.* **2018**, *14*, e1005941.

(S7) Kasinath, V.; Beck, C.; Sauer, P.; Poepsel, S.; Kosmatka, J.; Faini, M.; Toso, D.; Aebersold, R.; Nogales, E. JARID2 and AEBP2 Regulate PRC2 in the Presence of H2AK119ub1 and Other Histone Modifications. *Science* **2021**, *371*.

(S8) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nat Protoc* **2012**, *7*, 1511–1522.

(S9) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinforma.* **2006**, *Chapter 5*, Unit–5.6.

(S10) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(S11) Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* **2015**, *48*, 3026–3035.

(S12) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(S13) Kuzmanic, A.; Zagrovic, B. Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *Biophys. J.* **2010**, *98*, 861–871.

(S14) Poepsel, S.; Kasinath, V.; Nogales, E. Cryo-EM Structures of PRC2 Simultaneously Engaged with Two Functionally Distinct Nucleosomes. *Nat. Struct. Mol. Biol.* **2018**, *25*, 154–162.

(S15) Heenan, P. R.; Wang, X.; Gooding, A. R.; Cech, T. R.; Perkins, T. T. Bending and Looping of Long DNA by Polycomb Repressive Complex 2 Revealed by AFM Imaging in Liquid. *Nucleic Acids Res.* **2020**, *48*, 2969–2981.