Supporting Information

Tables:

Table S1. Secondary structures content (in %) of the protein region in our enhanced sampling simulations. The results are reported for all five systems simulated here.

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
Helix	55.0	53.0	51.0	54.0	51.0
Coil	32.0	27.0	29.0	29.0	28.0
Bend	8.0	10.0	7.0	10.0	8.0
Turn	5.0	10.0	12.0	6.0	12.0
Beta	0.0	0.0	1.0	1.0	1.0

Table S2. Backbone Root-mean-square deviation (RMSD, in Å) of the protein and the **PtDNA** moieties in the five systems simulated here, relative to the X-ray structure.

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
Protein	1.8± 0.4	1.6± 0.3	1.5± 0.2	1.4± 0.2	1.6± 0.2
PtDNA	4.5± 0.7	3.3± 0.7	3.7± 0.7	6.1± 0.3	5.4± 0.9

Table S3. Root-mean-square fluctuations (RMSF, in Å) per residue of the protein terminals (residues 0 to 7 and 79 to 93), box-region (residues 8 to 78), and **PtDNA** in the five systems simulated here.

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
Terminals	0.8	0.5	0.5	0.3	0.4
Box-region	0.2	0.2	0.2	0.1	0.1
PtDNA	0.3	0.2	0.2	0.1	0.2

Table S4. Contributions (%) of **PtDNA**-protein van der Waals (VWC), stacking (STC), hydrogen bonding (HBC), and salt bridge contacts (SBC).

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
VWC	41.6	36.3	46.6	54.8	41.5
STC	37.9	49.8	36.9	27.6	41.7
HBC	19.4	12.7	15.5	16.8	15.2
SBC	1.1	1.2	1.0	0.7	1.6

Table S5. Average number and standard deviation (Std) of **PtDNA**-protein water-mediated hydrogen bonds for the five complexes simulated here.

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
Average	5	6	7	6	6
Std.	2	3	3	2	3

Table S6. Number of direct **PtDNA**-protein contacts in the synchronic ensemble, in the largest subgroup, and in all other subgroups summed together for the five systems simulated here. The percentages relative to the total contributions are reported in parentheses.

	PtDNA•noPTM	**PtDNA•A**	**PtDNA•B**	**PtDNA•C**	**PtDNA•D**
Entire synchronic ensemble	112	92	95	100	96
Largest Subgroup	74 (66%)	79(85%)	81(86%)	91(91%)	89(93%)
Other subgroups	38(34%)	13(15%)	14(14%)	9(9%)	7(7%)

Table S7. Selected structural models of **PtDNA•noPTM**. All residues' backbone dihedral angles are located in the favorable regions of the Ramachandran plot.(1) The residues feature also excellent DOPE score.(2) In addition, their RMSD (in Å) relative to the X-ray structure(3) is the smallest or within 0.01 Å from the smallest. They are sorted by their DOPE scores. Model #1 has been selected in this work.

DOPE	Model #	RMSD (backbone atoms of the box-region)
-6169.96	1	0.13
-6028.16	2	0.14
-5955.67	3	0.14
-5947.13	4	0.14
-5903.37	5	0.14

Table S8. Details of the MD simulations for the systems investigated here.

	# atoms	**# Na^+ ions**	**# Cl^- ions**	**# water molecules**	**Box size (nm^3)**
PtDNA•noPTM	81,920	93	77	26,392	10.2×9.7×8.4
PtDNA•A	82,205	99	77	26,477	10.2×9.7×8.4
PtDNA•B	83,785	103	78	26,998	10.3×9.7×8.4
PtDNA•C	81,994	103	77	26,405	10.3×9.7×8.4
PtDNA•D	92,103	100	77	26,443	10.2×9.7×8.4
noPTM	81,197	77	89	26,492	10.2×9.7×8.4
A	81,769	78	84	26,676	10.2×9.7×8.4
B	82,597	78	81	26,949	10.3×9.7×8.4
C	81,200	77	79	26,488	10.3×9.7×8.4
D	80,397	77	82	26,220	10.2×9.7×8.4

Supporting Figures:

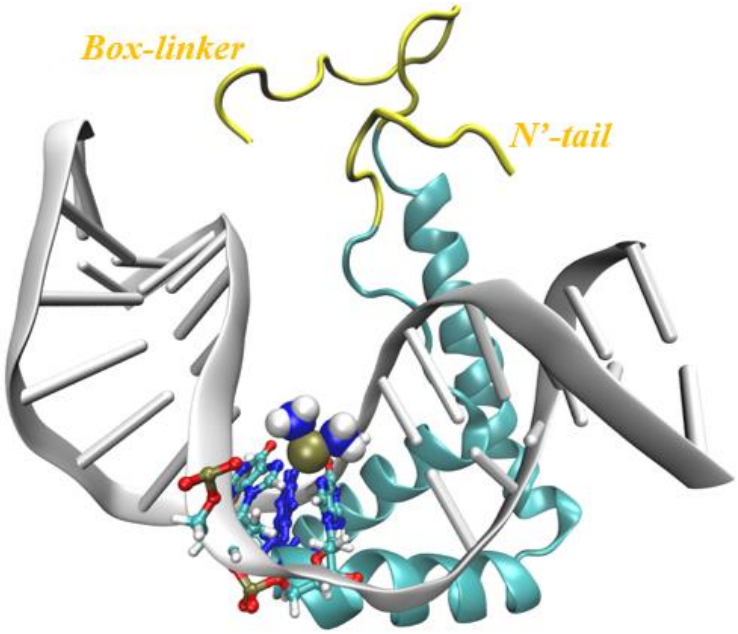


Figure S1. Initial conformation of MD simulations of **PtDNA•noPTM**. The box-region of the protein is colored in cyan; the two termini, N'-tail and box-linker, are colored in yellow; the hydrophobic anchor, F37, is colored in blue. The cross-linked dG8* and dG9* are in Licorice representation with P atoms in gold, N atoms in blue, O atoms in red, C atoms in Cyan, and H atoms in white. The cisplatin residue is in CPK representation with Pt in gold and N and H atoms in blue and white, respectively.

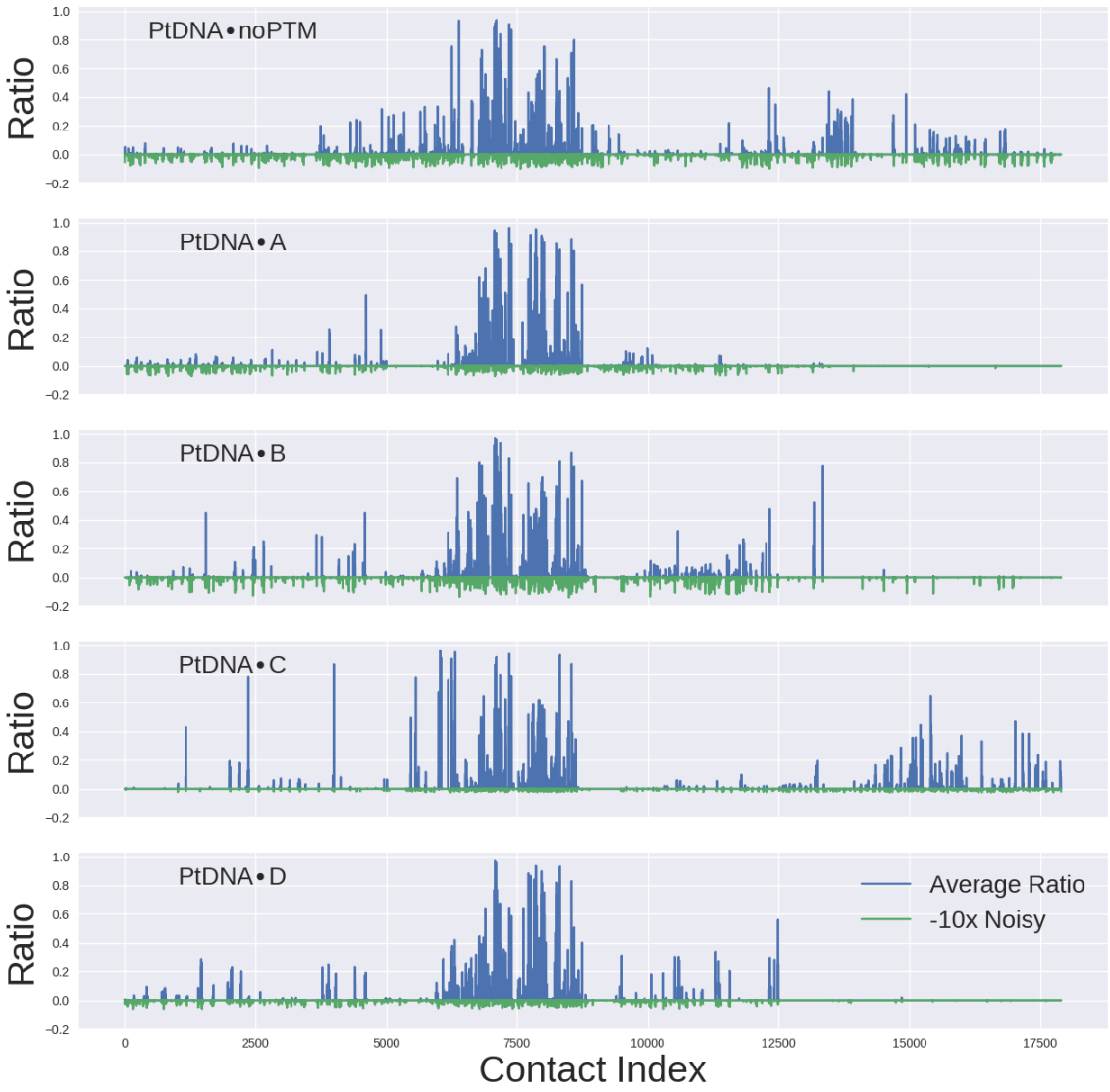


Figure S2. The average ratio and -10×noisy contacts are plotted as a function of contact index for the five systems investigated here. The averaged ratio is the occupancy of a **PtDNA**/protein contact in the whole conformational ensemble of a given system. Noisy contacts are those for which the average ratios are smaller than the interval between the upper and lower boundaries of 95% confidence interval in bootstrap sampling. Noisy contacts were discarded in the synchrony analysis hereafter. The contact index identifies different **PtDNA**/protein contacts (see Method section in main text) aligned across the five complexes. Therefore, the same contact index identifies the same atomistic contact between **PtDNA** and protein across different systems.

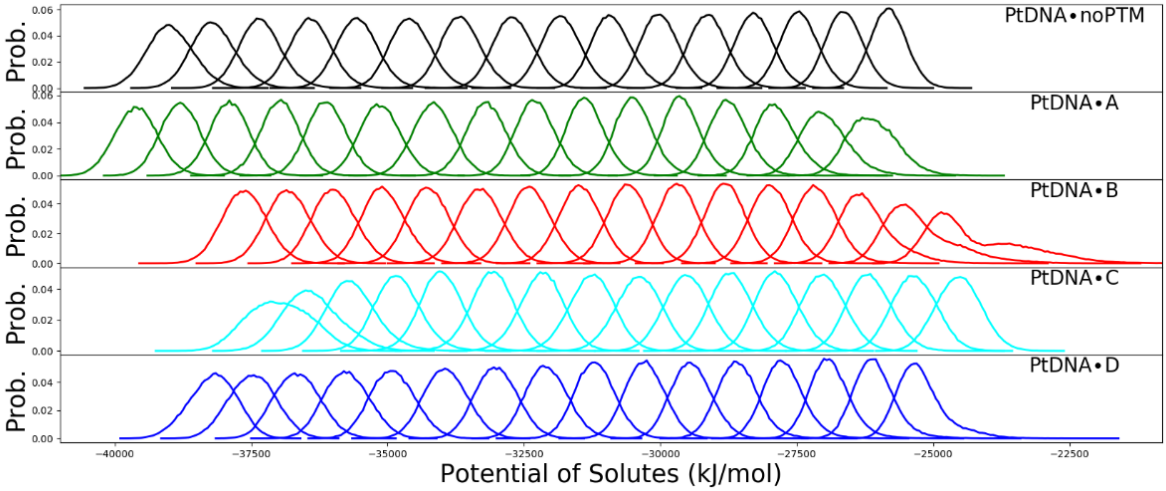


Figure S3. Overlapping of the potential of solutes of adjacent replicas along the chain of all replicas for complexes **PtDNA•noPTM**, **PtDNA•A**, **PtDNA•B**, **PtDNA•C**, and **PtDNA•D**. Curves referring to each complex are colored in black, green, red, cyan, and blue, respectively.

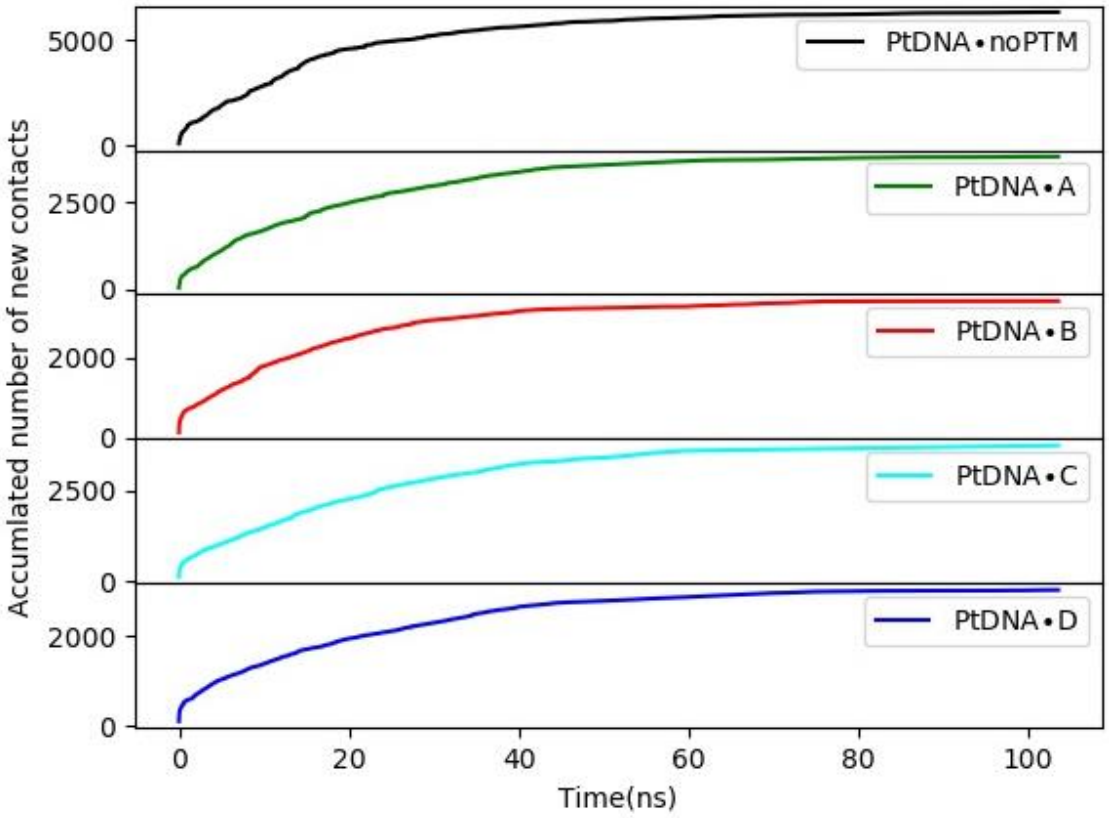


Figure S4. Convergence analysis of the enhanced sampling simulations by counting the number of new contacts between the protein and **PtDNA** along the simulation time. Here, a contact is considered new when it was not present in the previous steps. Color coding as in Fig. S3.

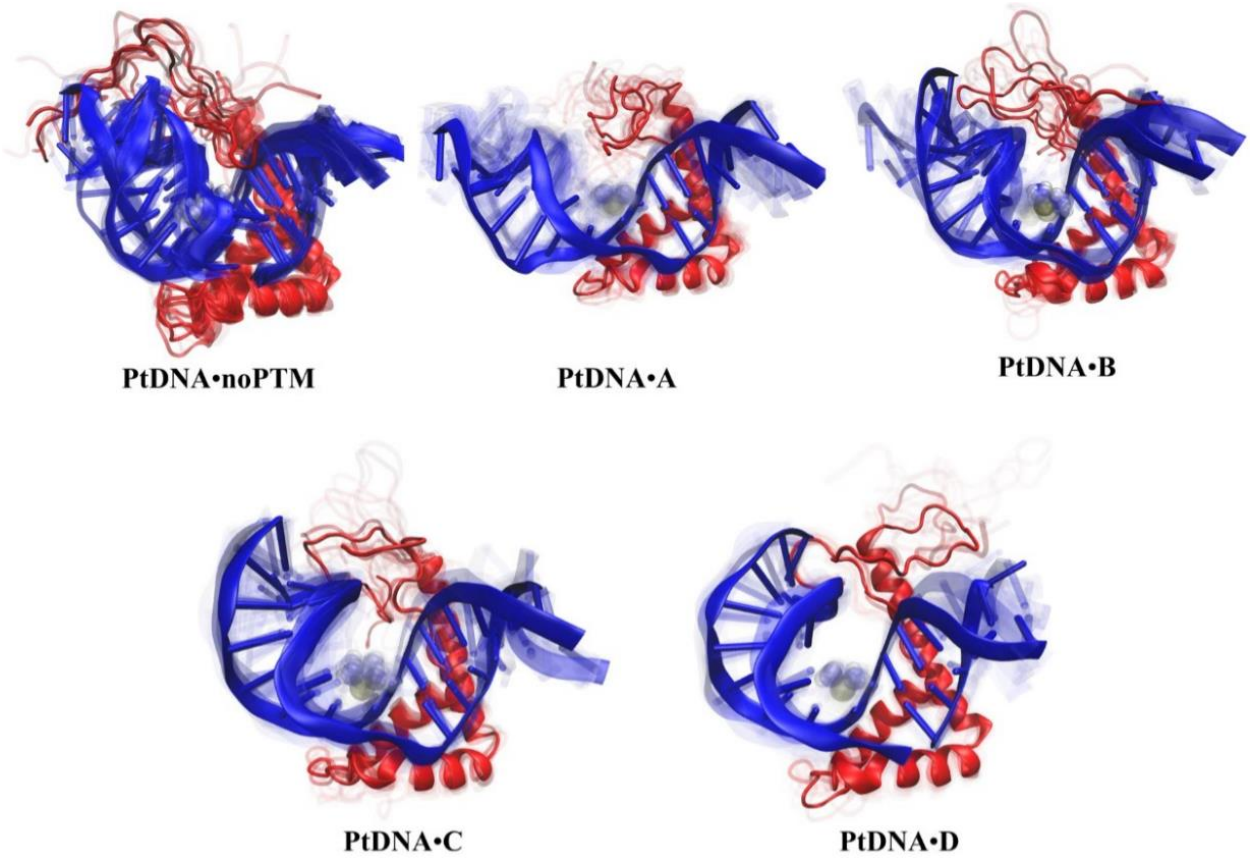


Figure S5. Structures representative of **PtDNA•noPTM** and **PtDNA•A-D** emerging from our enhanced samplings. The representative weight correlates with the opacity shown in the picture.

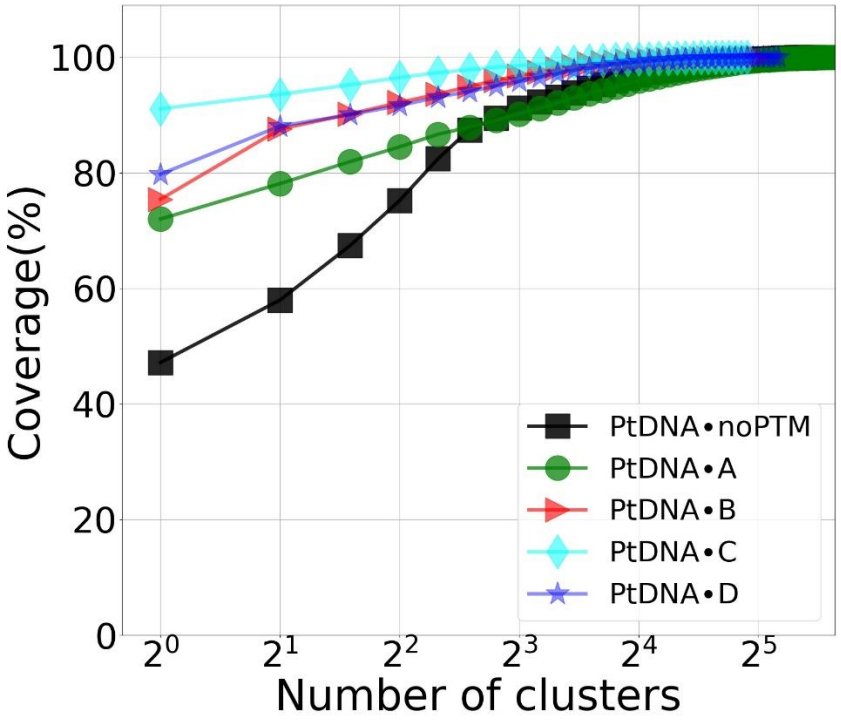


Figure S6. Evolution of the coverage of the number of conformations of the protein in the systems **PtDNA•noPTM** and **PtDNA•A-D** as a function of the number of clusters. Color coding as in Fig. S3.

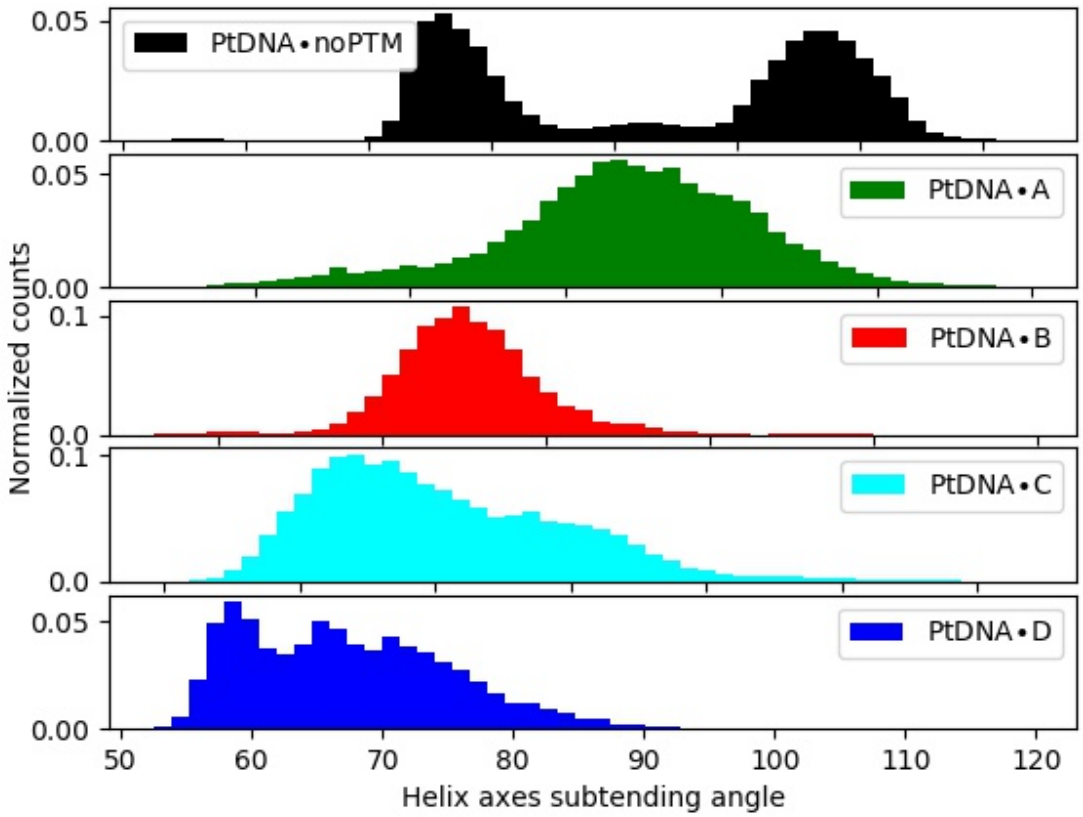
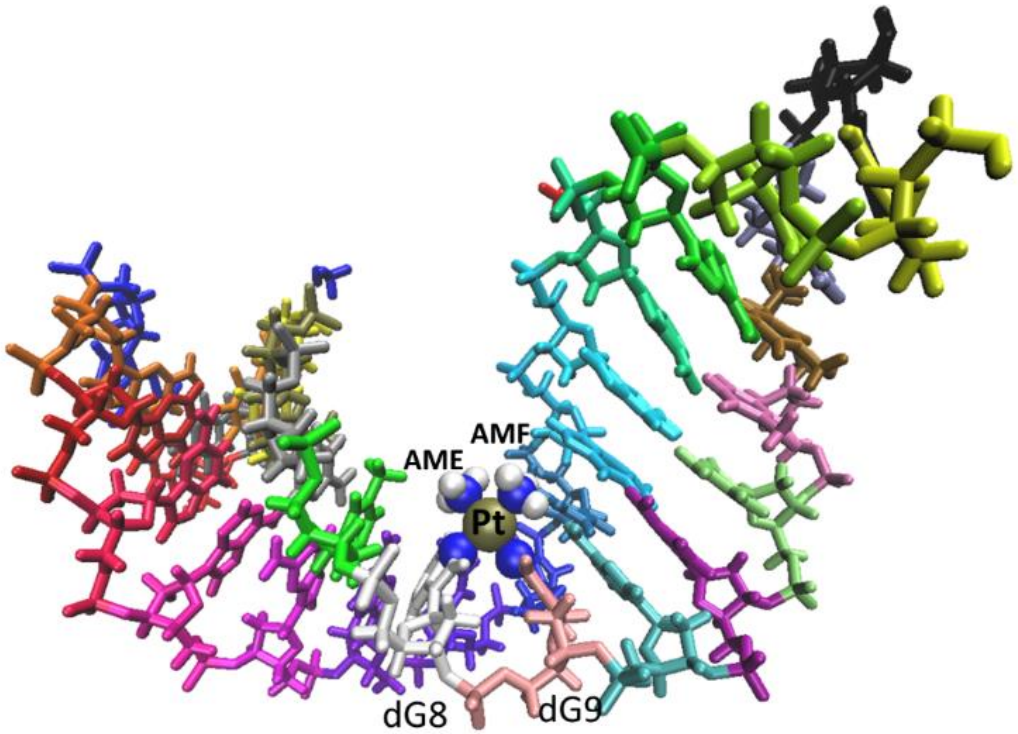


Figure S7. Histograms of helix axes subtending angles(4) of **PtDNA** in the five complexes investigated here. Color coding as in Fig. S3.



	1CKT	PtDNA•noPTM	PtDNA•A	PtDNA•B	PtDNA•C	PtDNA•D
N7'-Pt-N7''	90	88±3	89±3	88±3	88±3	88±3
NE-Pt-NF	89	90±3	89±3	89±3	89±3	89±3
N7'-Pt-NE	89	89±3	88±3	89±4	88±3	89±3
N7''-Pt-NF	92	93±3	93±3	93±3	92±3	92±3

Figure S8. 3D structure of the PtDNA (colored by residue number) and bond angles around the Pt atom in the X-ray structure(3) and in the five systems simulated here. The cisplatin residue is in CPK representation with Pt in gold and N and H atoms in blue and white, respectively. N7' and N7" are the nitrogen atoms N7 in dG8 and dG9, respectively. NE and NF are the nitrogen atoms in AME and AMF (see Figure above), respectively.

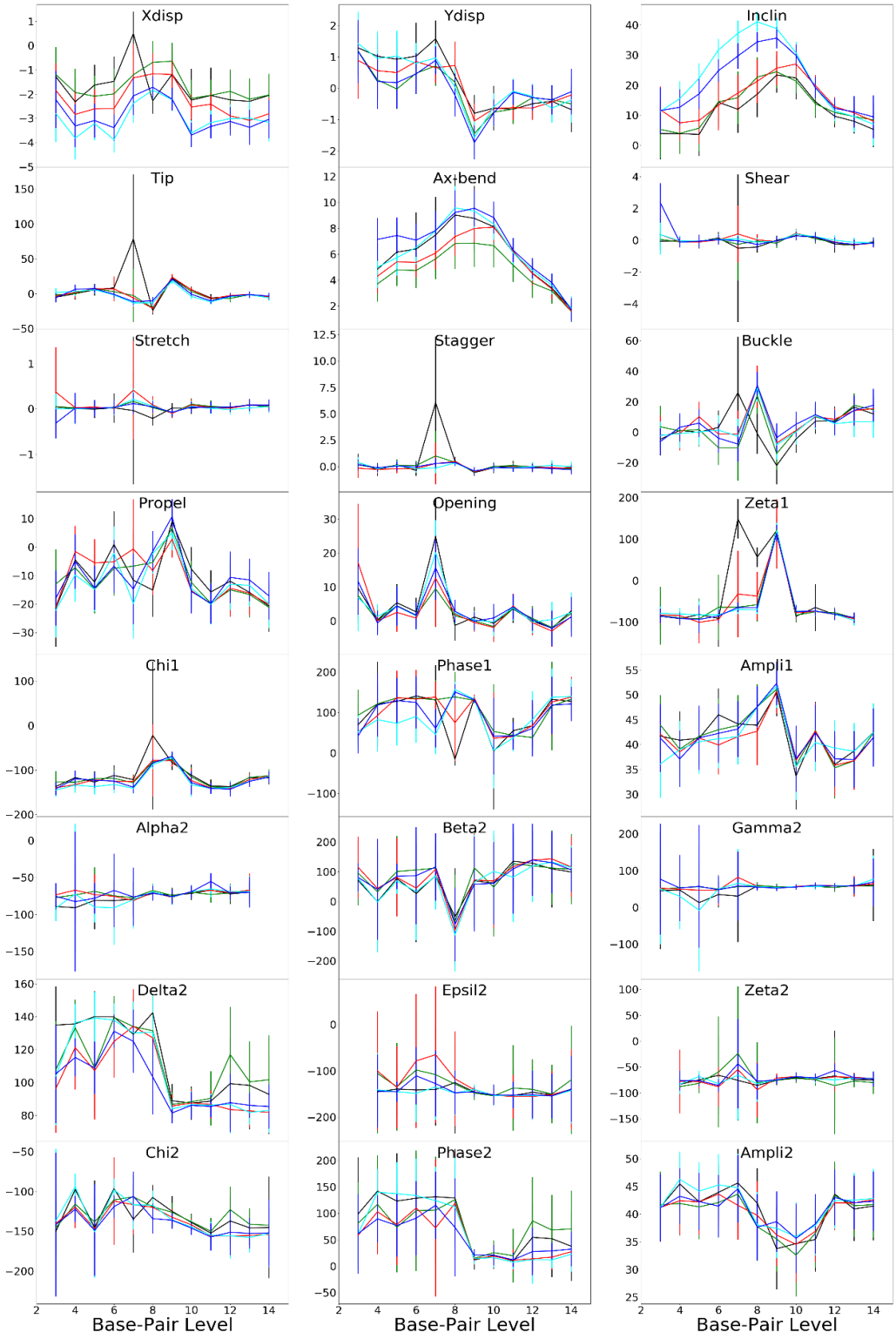


Figure S9. Selected structural parameters of **PtDNA**. These include the shear (in Å), stretch (in Å), stagger (in Å), translations (Xdisp and Ydisp, in Å) and rotations (Tip, in °), as well as other parameters including Inclin (in °), Ax-bend (in °), Buckle (in °), Propel (in °), Opening (in °),

Zeta1 (in °), Chi1 (in °), Phase1 (in °), Ampli1 (in °), Alpha2 (in °), Beta2 (in °), Gamma2 (in °), Delta2 (in °), Epsil2 (in °), Zeta2 (in °), Chi2 (in °), Phase2 (in °), and Ampli2 (in °). These properties have been defined in and calculated by Curves+,(5) for the five complexes simulated here. Color code: **PtDNA•noPTM** (black), **PtDNA•A** (green), **PtDNA•B** (red), **PtDNA•C** (cyan), and **PtDNA•D** (blue). The terminal capping base-pair levels 1-2 and 15-16 are not included in the calculations. Color coding as in Fig. S3.

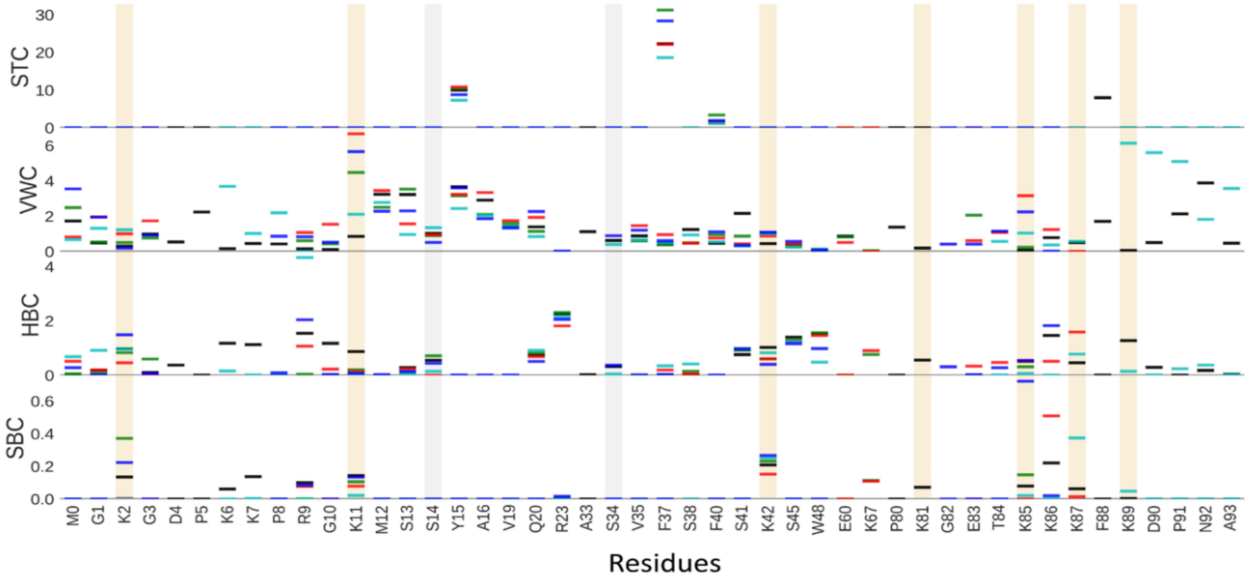


Figure S10. Distribution of stacking contacts (STC), Van der Waals contacts (VWC), hydrogen bond contacts (HBC), and salt bridge contacts (SBC) of each residue in the five complexes simulated here. The values are averaged during the simulation. Color coding as in Fig. S3. Acetylated and phosphorylated residues are identified by yellow and grey color bands, respectively.

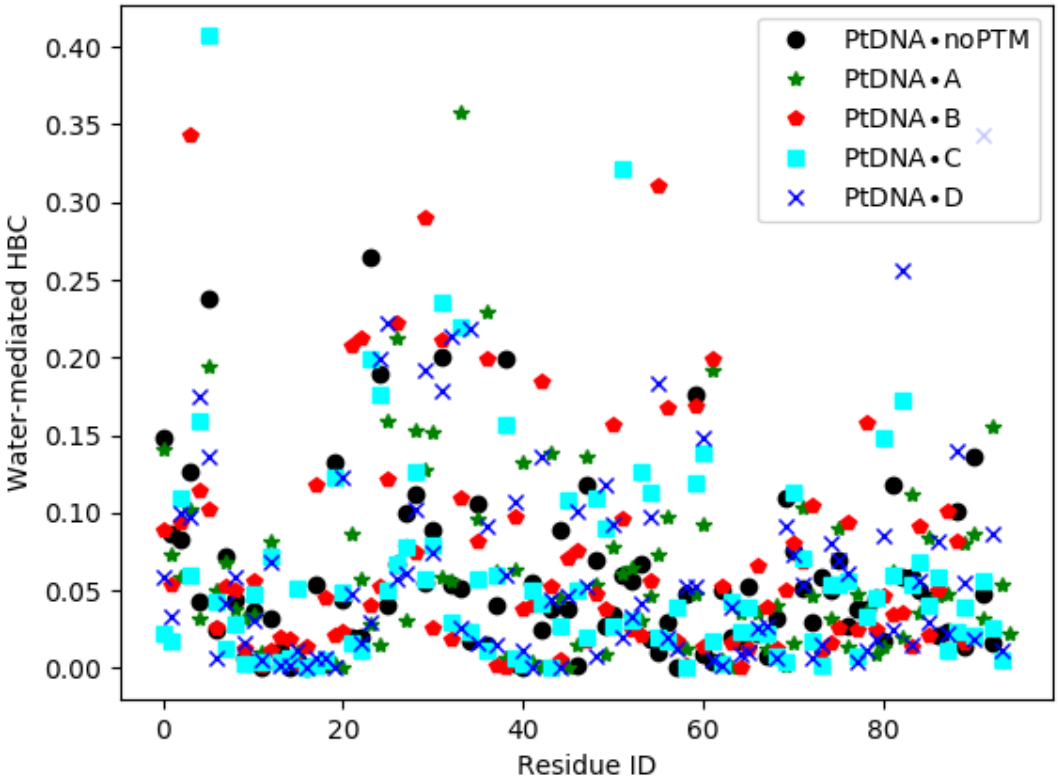


Figure S11. Distribution of water-mediated HBC of each residue across the five systems simulated here. Color coding as in Fig. S3.

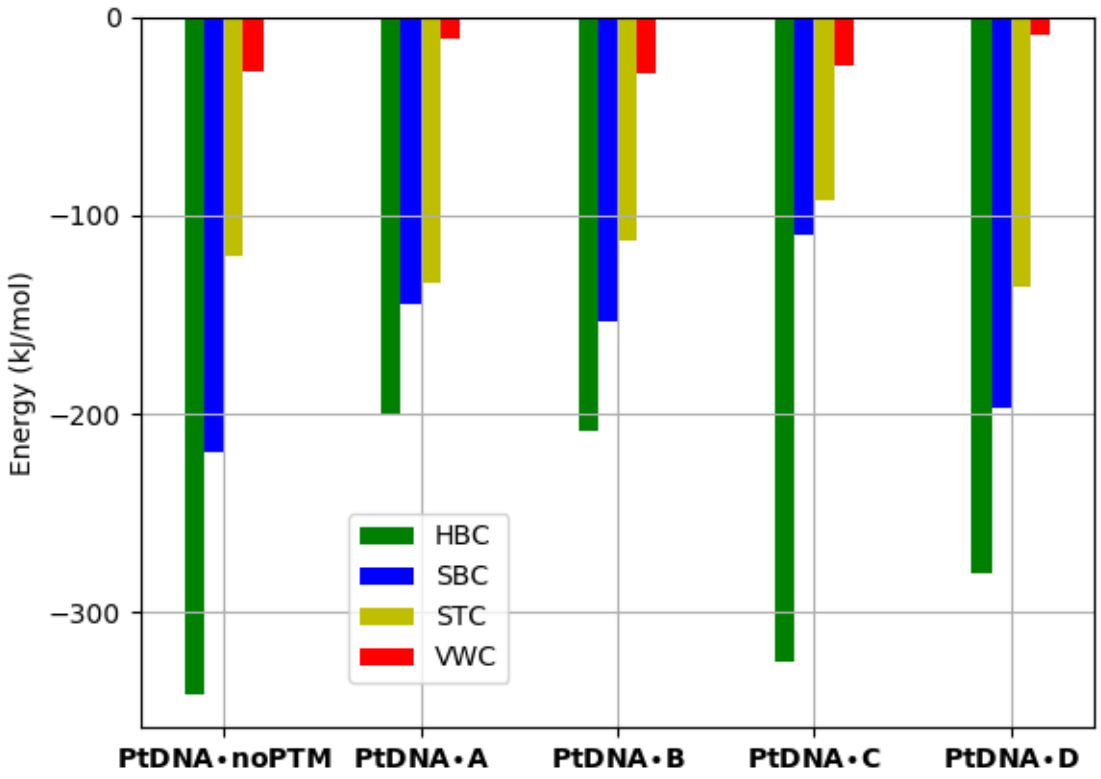


Figure S12. Energetics associated with STC, HBC, SBC, VWC in the five systems investigated here. Only direct contacts are considered. The values are reported here only for relative comparisons and they provide no information as regarding the free energy of platinated DNA/protein binding.

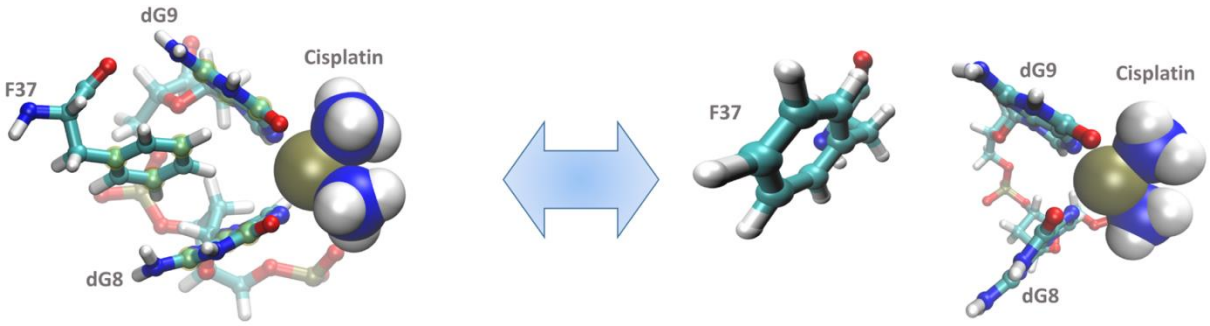


Figure S13. Rearrangement of the phenyl ring of F37 in the hydrophobic notch of PtDNA in **PtDNA•noPTM**. This is associated with a change of the dG8* -F37 distance (**R**) (see Figure 3 in the main text). The structures associated with **R** = 0.58 nm and **R** = 1.0 nm are shown on the left- and right-hand side, respectively. Color coding by atom type: N, blue; H, white; Pt, gold; C, cyan; O, red; P, gold.

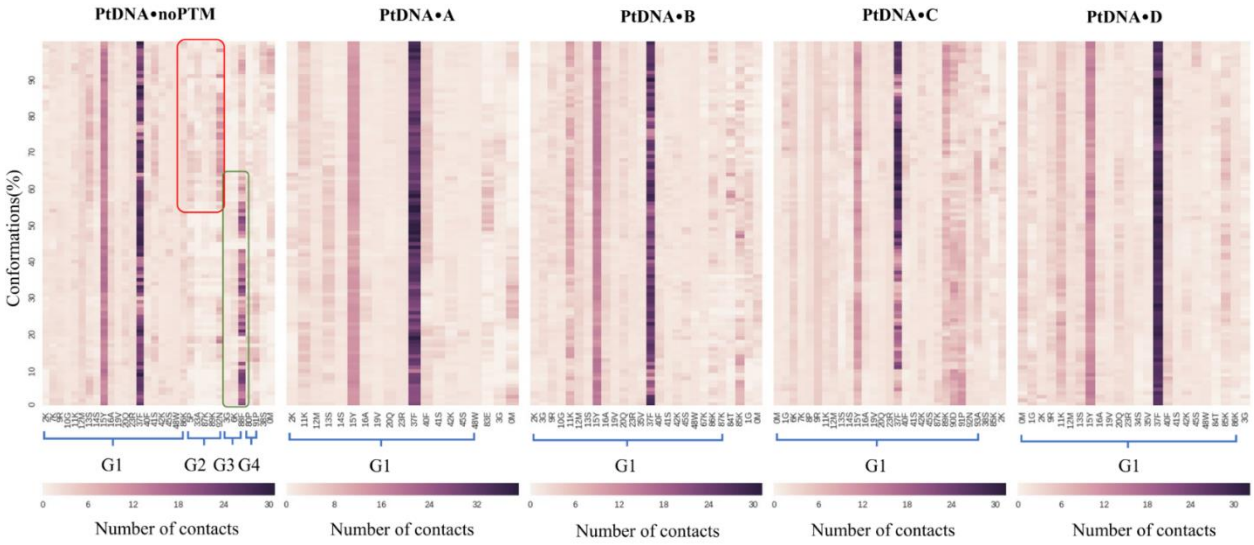


Figure S14. **PtDNA**-protein contact heat map across the five systems simulated here. The contacts associated with different synchronic subgroups are grouped together manually. Non-isolated subgroups (those having more than one member) are labeled by G*i* (*i* is the index of the subgroup). For **PtDNA•noPTM**, the asynchronic contacts of subgroups G2 and G3 are highlighted in red and green boxes, respectively.

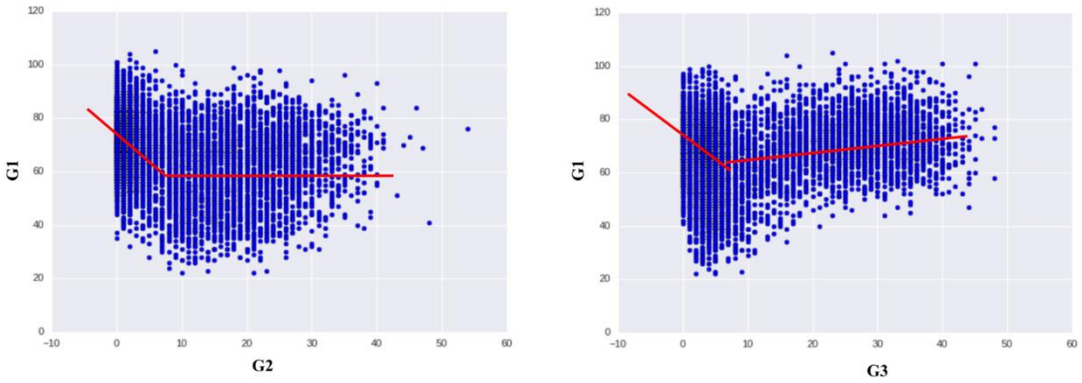


Figure S15. Left: Scattering plots of the number of PtDNA-protein contacts associated with the first (G1) and the second (G2) synchronic subgroups in **PtDNA•noPTM** (see Fig. S13). Right: same plot for the first (G1) and the third (G3) synchronic subgroups (see Fig. S13).

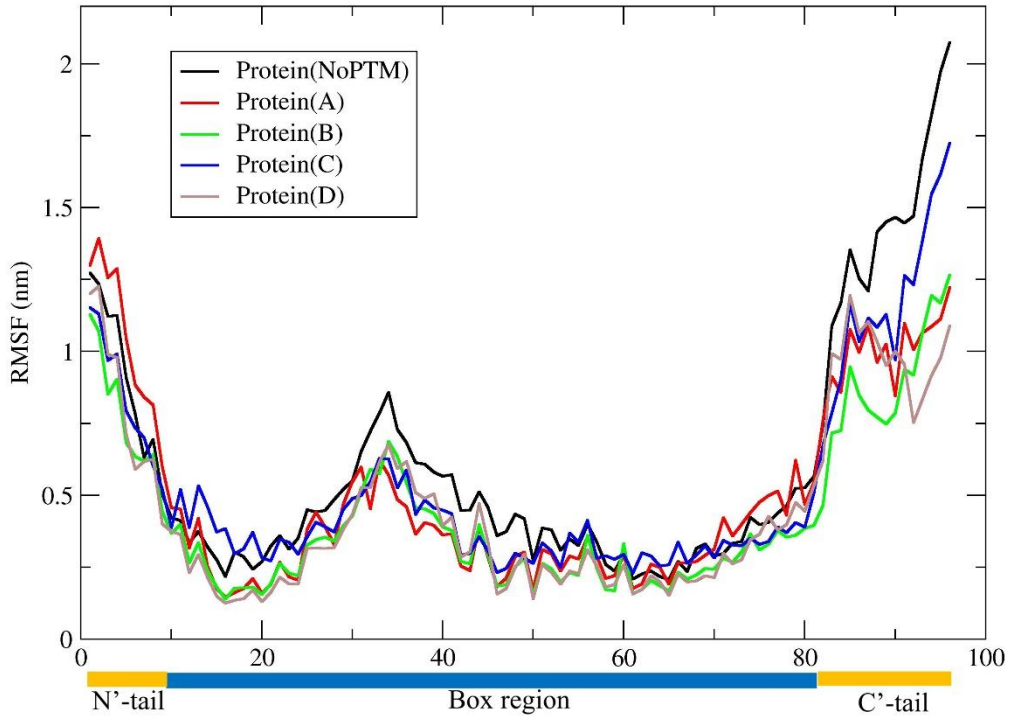


Figure S16. Root-mean-square fluctuations (RMSFs) of the residues in proteins bearing PTMs as in **A-D,** in the free state in water solution, along with the protein, also in the free state, without the PTMs. All these proteins were simulated under identical sampling conditions. In particular, we performed100 ns enhanced sampling simulations at 0.15 M NaCl.

Energetics associated with the protein/PtDNA contacts

The potential energy associated with direct (that is, no-water mediated) HBC, SBC, STC, and VWC interactions was calculated based on the AMBER ff99SB-ILDN(6-8) and Parmbsc1(9) force fields (Fig. S12). Specifically, the energy was calculated as a sum of the non-bonded terms between the moieties forming the H-bonds, salt bridges, pi-staking and van der Waals interactions, respectively.

The water-mediated HBC and SBC contacts were not included in the calculations, as the water (and possibly counterions) involved in such interactions may vary during the dynamics, affecting the number of overall interactions at each simulation snapshot. Even considering this fact, HBC and SBC contributions are already the largest (Fig. S12). As expected, HBC and SBC are larger in **PtDNA•noPTM** relative to **PtDNA•A-D**: indeed, in the latter, several lysines have been acetylated and some serines phosphorylated (see Tab. 1). Obviously, this calculation has no implication on the overall free energy of binding, as the latter depends on difference in enthalpy and entropy on passing from the unbound to the bound state.

Methods:

Homology modeling: The deposited X-ray structure of HMGB1 in complex with **PtDNA** ($[Pt(NH_3)_2]^{2+}$-d(CCUCTCTG*G*ACCTTCC)-d(GGAGAGACCTGGAAGG (PDBID 1CKT(3)) includes the box-region A from residues 8 to 78. To include the N'-tail (residues 0-7) and the box-linker (residues 79-93) of the protein, we performed a multi-templates and full-H modeling using MODELLER version 9.(10) Homology modeling of the missing part was based on the full-length NMR structure of HMGB1 (PDBID 2YRQ)). In total, 800 candidate structures were generated. The modeled structure used in the start of the MD simulation was selected based on its root-mean-square-deviation (RMSD) with respect to the X-ray structure, its Ramachandran Plot, and its built-in Discrete Optimized Protein Energy (DOPE) scores.(2) A structure with very low RMSD, excellent Ramachandran plot, and low DOPE score (See Tab. S7) was selected as the initial structure of our simulations (Fig. S1).

MD parameters: AMBER ff99SB-ILDN,(6-8) Forcefield PTM,(11) Parmbsc1,(9) TIP3P,(12) and Joung-Cheatham(13) forcefields were used for the protein's standard residues, the protein residues

undergoing PTM, DNA , water, and counterions, respectively. The force field for the platinated lesion, fully compatible with the AMBER force field, was taken from literature.(14) This force field has already been successfully applied to the structure of **PtDNA•noPTM** truncated at the terminal region, for which the X-ray structure has been solved.(14)

Long-range electrostatic interactions were evaluated using the particle mesh Ewald method.(15) A Fourier spacing of 0.12 nm combined with a fourth-order cubic spline interpolation was used. A 1.5 nm and a 1.3 nm cutoff were used for van der Waals interactions and the real-space part of the electrostatic interactions, respectively. All bond lengths were constrained with LINCS.(16) The time step was set to 2 fs.

Constant temperature (298.5 K) and pressure (1 bar) were achieved by coupling with Nose-Hoover(17) thermostat and Andersen-Parrinello-Rahman(18) barostat. The MD simulations were performed with GROMACS version 5.1.(19)

Simulations. The five systems were solvated by water with ~0.154 M sodium chloride (NaCl) to mimic physiological saline condition (for details see Tab. S8). Zero-total charge is achieved by adding 16, 22, 25, 26, and 23 extra sodium ions to systems passing from **PtDNA•noPTM** to **PtDNA•A-D**, respectively. We used an octahedral box under three-dimensional periodic boundary conditions (3D-PBC), with a minimum distance of 2 nm between the solutes and the box surface to avoid the artificial effects between two PBC images.

The systems underwent energy minimizations by 5,000 steps steepest descent(20) optimization and 5,000 steps conjugate gradient(21) optimization, successively. The systems were then heated to room temperature (298.5 K) by 1-ns long simulated annealing from the temperature of 2 K. Then, they underwent 100-ns long plain MD simulations followed by replica exchange with solute scaling sampling simulations,(22) at room temperature (298.5 K) and atmospheric pressure (1 bar). Sixteen replicas were employed for each **PtDNA**-protein complex, with initial structures corresponding to the cluster representative of MD simulation. The 16 replicas have scaled-down Hamiltonian of solutes,(23) *i.e.* the **PtDNA**-Protein complexes here, from room temperature (T_0) of 298.5 K to an effective high temperature (T_h) of 433.18 K. The effective temperatures of replicas were first initialized by an online temperature predictor,(24) and then a precise optimization was performed “on-the-fly” in the replica exchange simulations to minimize the variance of exchange ratios between adjacent replicas. The temperature serials were fixed before data collection. These

sampling strategies guarantee a low memory effect from initial structure and an efficiency of conformation sampling along the chain of all replicas (Fig. S3). 100-ns long enhanced sampling simulations were carried out for each system, based on a convergence analysis of the number of accumulated new contacts between **PtDNA** and protein (Fig. S4). For the free state sampling of the proteins, we adopt the same sampling conditions and simulation lengths as employed in the simulations of **PtDNA**-Protein complexes.

Maximization of *S(L)*

A log-domain version of the max-product algorithm, so-called affinity propagation,(25) was used to search over configurations of the labels $L=(L_1,\ldots, L_R)$ that maximize S(L) of $[R\times R]$ pairwise synchrony (see eq. 2 in main text) matrix $\boldsymbol{P}$. R is the number of nodes in each network. It represents the number of amino acids interacting with **PtDNA** of a system. For the synchrony matrix $\boldsymbol{P}$ of each system, an array $\lambda=(\lambda_1, ..., \lambda_m)$ is created from the min($\boldsymbol{P}$) to the max($\boldsymbol{P}$), with interval of 0.01. Where, $m = \lceil 100(\max(\boldsymbol{P}) - \min(\boldsymbol{P})\rceil$. For each $\lambda_x \in \lambda$, we perform the affinity propagation by doing the following steps to record the configurational labels L of the $\max(S^{\lambda_x}{}_{\lambda_x\in\lambda}(L) = \sum_{i=1}^{R} P_{i,L_i})$:

(i) The elements of two matrices $\boldsymbol{R}[R\times R]$ and $\boldsymbol{A}[R\times R]$) are set to zero.

(ii) The elements are updated by two message-passing steps iteratively:

$$\boldsymbol{R}_{i,k} = \begin{cases} \lambda_x - \max\limits_{k':\,k'\neq k}(\boldsymbol{A}_{i,k'}+\boldsymbol{P}_{i,k'}), & \text{for } i=k \\ \boldsymbol{P}_{i,k} - \max\limits_{k'\,k'\neq k}(\boldsymbol{A}_{i,k'}+\boldsymbol{P}_{i,k'}), & \text{for } i\neq k \end{cases} \quad (1)$$

$$\boldsymbol{A}_{i,k} = \begin{cases} \sum_{i':i'\neq k}\max(0,\boldsymbol{R}_{i',k}), & \text{for } i=k \\ \min\{0,\, \boldsymbol{R}_{k,k}+\sum_{i':i'\in\{i,k\}}\max(0,\boldsymbol{R}_{i',k})\}, & \text{for } i\neq k \end{cases} \quad (2)$$

Here, $P_{i,k}$ are the elements of $\boldsymbol{P}$. $\boldsymbol{R}$ represents the log-probability matrix of each node to serve as the synchronic center of other nodes. $\boldsymbol{A}$ is the log-probability matrix of each node to select other nodes as its synchronic centers. For a detailed explanation of these formulas see ref.(25)

(iii) the value of a configuration label L_i reads:

$$L_i = \underset{j}{\operatorname{argmax}}(A_{i,j} + R_{i,j}) \quad (3)$$

(iv) Go to (ii) until L does not change anymore.

(v) Calculate $S^{\lambda_x}(L) = \sum_{i=1}^{R} P_{i,L_i}$.

Supporting references:

1. Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol*, **7**, 95-99.
2. Shen, M.Y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci*, **15**, 2507-2524.
3. Ohndorf, U.M., Rould, M.A., He, Q., Pabo, C.O. and Lippard, S.J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature*, **399**, 708-712.
4. Kozelka, J., Petsko, G.A., Quigley, G.J. and Lippard, S.J. (1986) High-Salt and Low-Salt Models for Kinked Adducts of Cis-Diamminedichloroplatinum(Ii) with Oligonucleotide Duplexes. *Inorg Chem*, **25**, 1075-1077.
5. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*, **37**, 5917-5929.
6. Wang, J., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, **21**, 1049-1074.
7. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, **78**, 1950-1958.
8. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics*, **65**, 712-725.
9. Ivani, I., Dans, P.D., Noy, A., Perez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goni, R., Balaceanu, A. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat Methods*, **13**, 55-58.
10. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **29**, 291-325.
11. Khoury, G.A., Thompson, J.P., Smadbeck, J., Kieslich, C.A. and Floudas, C.A. (2013) Forcefield_PTM: Ab Initio Charge and AMBER Forcefield Parameters for Frequently Occurring Post-Translational Modifications. *J Chem Theory Comput*, **9**, 5653-5674.
12. Mark, P. and Nilsson, L. (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A*, **105**, 9954-9960.
13. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B*, **112**, 9020-9041.
14. Nguyen, T.H., Rossetti, G., Arnesano, F., Ippoliti, E., Natile, G. and Carloni, P. (2014) Molecular Recognition of Platinated DNA from Chromosomal HMGB1. *J Chem Theory Comput*, **10**, 3578-3584.
15. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys*, **103**, 8577-8593.
16. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, **18**, 1463-1472.
17. Martyna, G.J., Klein, M.L. and Tuckerman, M. (1992) Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J Chem Phys*, **97**, 2635-2643.

18. Nose, S. and Klein, M.L. (1983) Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol Phys*, **50**, 1055-1076.
19. Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*, **4**, 435-447.
20. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845-854.
21. Zimmermann, K. (1991) Oral - All Purpose Molecular Mechanics Simulator and Energy Minimizer. *J Comput Chem*, **12**, 310-319.
22. Wang, L.L., Friesner, R.A. and Berne, B.J. (2011) Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J Phys Chem B*, **115**, 9431-9438.
23. Wang, L., Friesner, R.A. and Berne, B.J. (2011) Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2) (vol 115, pg 9431, 2011). *J Phys Chem B*, **115**, 11305-11305.
24. Patriksson, A. and van der Spoel, D. (2008) A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, **10**, 2073-2077.
25. Frey, B.J. and Dueck, D. (2007) Clustering by Passing Messages Between Data Points. *Science*, **315**, 972-976.