Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Understanding base and backbone contributions of phosphorothioate DNA for molecular recognitions with SBD proteins

Jiayi Li, Shenggan Luo, Xingyu Ouyang, Geng Wu, Zixin Deng, Xinyi He, Yi-Lei Zhao*

State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

*To whom correspondence should be addressed: Prof. Yi-Lei Zhao Tel/Fax: +86-21-34207190; Email: yileizhao@sjtu.edu.cn

Contents

Fig. S1: Chalcogen-binding features in P-S···N^{P79} contact S4
Fig. S2: Root mean square deviation (*rmsd*) of 500ns MD trajectories S5
Fig. S3: The *rmsd* of structural alterations for each residue S6
Fig. S4: Overlays of snapshots and co-crystal structures of **SBD***Spr*/PT-DNA S7
Fig. S5: Overlays of snapshots and co-crystal structures of **SBD***Sco*/PT-DNA S8
Fig. S6: The bifurcated hydrogen bonds in bases ***5'-I***···H102 of **SBD***Spr*/PT-DNA S9
Fig. S7: The bifurcated hydrogen bonds in bases ***5''-III***···S105 in **SBD***Spr*/PT-DNA..... S10
Fig. S8: Counter-ions were discharged from **SBD**/PT-DNA interface S11
Fig. S9: Structural analyses of **SBD***Sco*-H116Y binding with PT-DNA S12
Fig. S10: Structural analyses of **SBD***Spr*-Q32R binding with PT-DNA S13
Fig. S11: Structural analyses of **SBD***Spr*-H102R binding with PT-DNA S14
Fig. S12: Structural analyses of **SBD***Spr*-G103R binding with PT-DNA S15
Fig. S13: Structural analyses of **SBD***Spr*-S105R binding with PT-DNA..... S16
Fig. S14: Structural analyses of **SBD***Spr*-D104Y binding with PT-DNA S17
Fig. S15: Structural analyses of **SBD***Spr*-Y31H binding with PT-DNA S18
Fig. S16: Structural analyses of **SBD***Sco*-R117Q binding with PT-DNA S19
Fig. S17: Structural analyses of **SBD***Sco*-R190H binding with PT-DNA S20
Fig. S18: Structural analyses of **SBD***Sco*-R191G binding with PT-DNA S21
Fig. S19: Structural analyses of **SBD***Sco*-R191S binding with PT-DNA..... S22
Fig. S20: Structural analyses of **SBD***Sco*-Y164D binding with PT-DNA..... S23
Fig. S21: Structural analyses of **SBD***Sco*-A107N binding with PT-DNA..... S24
Fig. S22: Structural analyses of **SBD***Sco*-R188T binding with PT-DNA S25
Fig. S23: Stereochemistry of PT-modification at backbones ***5'-II***..... S26
Fig. S24: Stereochemistry of PT-modification at backbones ***5'-I*** S27
Fig. S25: Stereochemistry of PT-modification at backbones ***3'-I*** S28
Fig. S26: Stereochemistry of PT-modification at backbones ***3'-II***..... S29
Fig. S27: Structural analyses of E156 in wild type **SBD***Sco*-G$_{PS}$GCC..... S30
Fig. S28: Structural analyses of **SBD***Sco*-E156K binding with PT-DNA S31
Fig. S29: Structural analyses of **SBD***Sco*-E156R binding with PT-DNA S32
Fig. S30: Structural analyses of **SBD***Sco*-E156L binding with PT-DNA..... S33
Fig. S31: Structural analyses of **SBD***Sco*-E156D binding with PT-DNA S34
Fig. S32: Structural analyses of **SBD***Sco*-E156Q binding with PT-DNA S35
Table S1: Calculated dV/dλ in the six TI cycles.....S36-S41
Table S2: **SBD***Spr* interacts with lateral bases S42
Table S3: **SBD***Sco* interacts with lateral bases S43
Table S4: Energy contribution of deoxyribose in **SBD***Spr*/PT-DNA complex..... S44
Table S5: Energy contribution of deoxyribose in **SBD***Sco*/PT-DNA complex..... S45
Table S6: The energy distributions of **SBD***Spr*/5'-GGC**G$_{PS}$GCC**C-3' S46
Table S7: The energy distributions of **SBD***Spr*/5'-GAT**G$_{PS}$ATC**C-3' S47

Table S8: The energy distributions of **SBD***Spr*/5'-GGC**G$_{PS}$AA**C-3' S48
Table S9: The energy distributions of **SBD***Sco*/5'-CC**G$_{PS}$GCC**G-3'..... S49
Table S10: The energy distributions of **SBD***Sco*/5'-CC**G$_{PS}$ATC**G-3' S50
Table S11: The energy distributions of **SBD***Sco*/5'-CC**G$_{PS}$AAC**G-3' S51
Table S12: Analysis of consensus sequences in *E. coli* B7A S52
Table S13: The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GGC**G$_{PS}$GCC**C-3' system with the H++ web server S53
Table S14: The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GAT**G$_{PS}$ATC**C-3' system with the H++ web server S54
Table S15: The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GGC**G$_{PS}$AAC**C-3' system with the H++ web server S55
Supplementary methods S56-S58

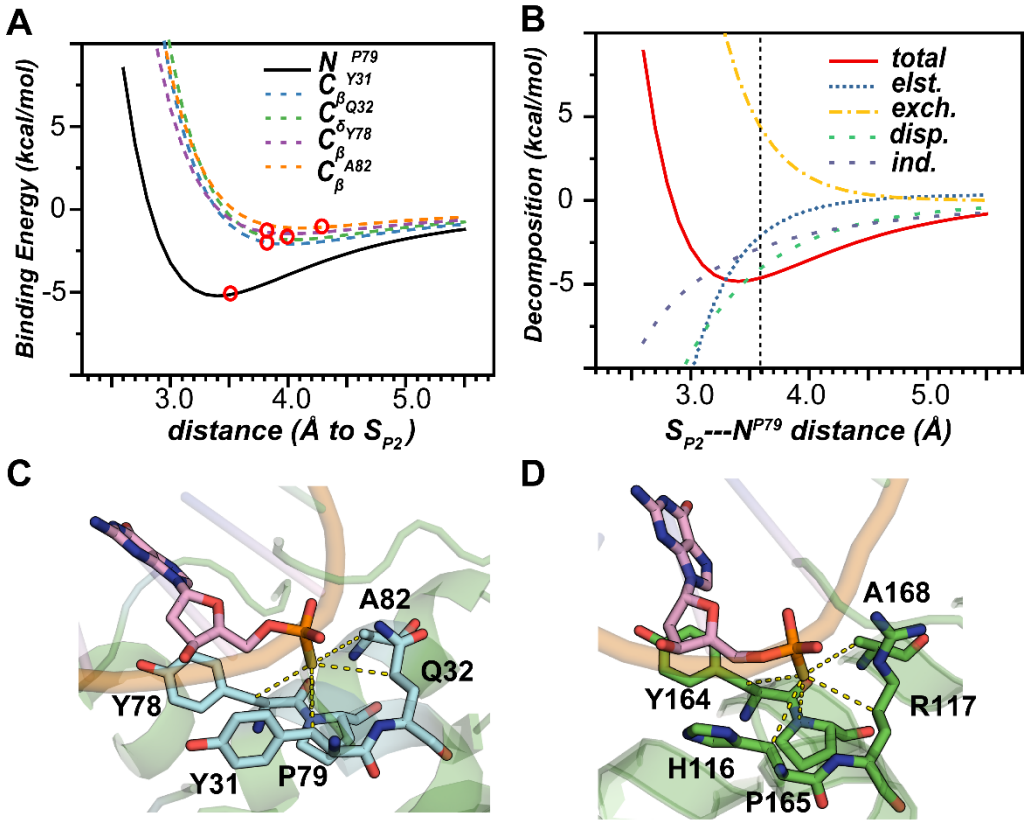

Fig. S1: Chalcogen-binding features in P-S···N^{P79} contact. (A) The binding energy of each contact for the PS moiety is scanned along the heavy atom distances from 2.5 to 5.5 Å. Their real distances in the co-crystal structure are marked with red circles. (B) The P-S···N^{P79} interaction is further analyzed by SAPT0/jun-cc-pVDZ calculations. Four types were used to categorize intermolecular interactions: electrostatics (*elst.*), exchange (*exch.*), dispersion (*disp.*), and induction (*ind.*). The exchange energy, which was repulsive, was not included in the percentage calculations for the attractive components. Binding pockets of the co-crystal structure of **SBD***Spr*-GPSGCC(C) and **SBD***Sco*-GPSGCC(D).

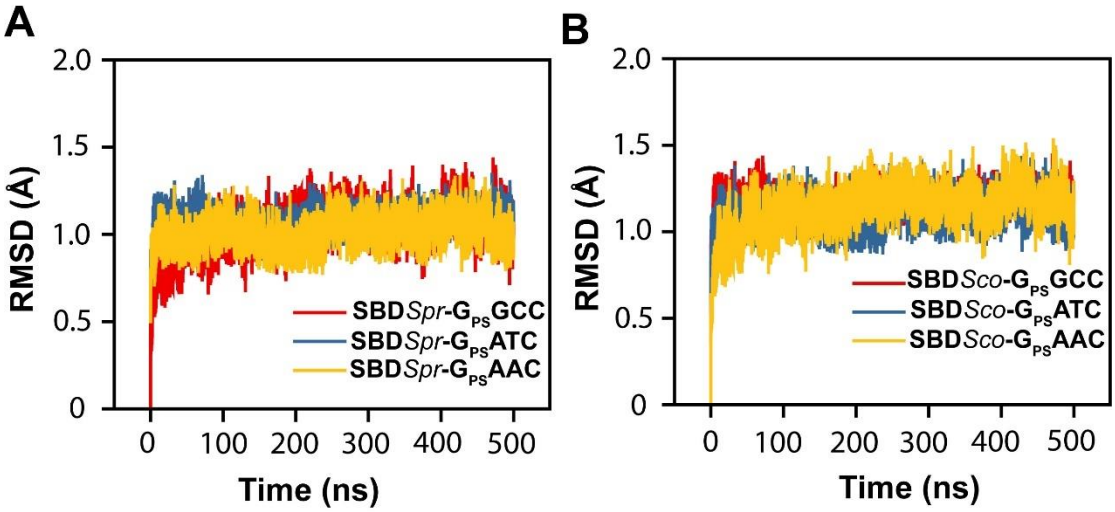

Fig. S2: Root mean square deviation (*rmsd*) of 500ns MD trajectories (A) **SBD***Spr* binding with 5'-GGC**G$_{PS}$GCC**C-3', 5'-GAT**G$_{PS}$ATC**C-3', and 5'-GGC**G$_{PS}$AAC**G-3'. (B) **SBD***Sco* binding with 5'-CC**G$_{PS}$GCC**G-3', 5'-CC**G$_{PS}$ATC**G-3', and 5'-CC**G$_{PS}$AAC**G-3'. The stability of the root-mean-square deviation (*rmsd*) values throughout the MD simulation suggests that the trajectories are suitable for further investigations.

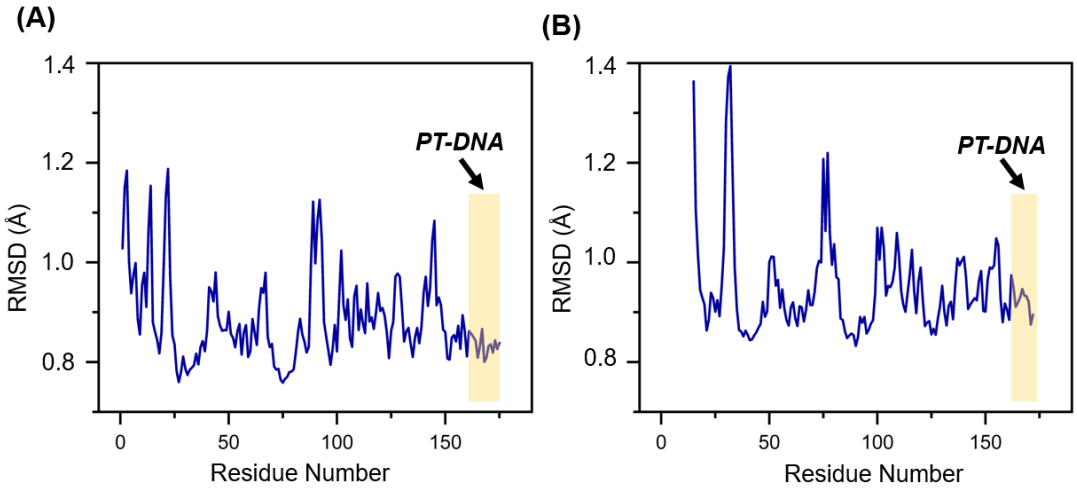

Fig.S3: The *rmsd* values in structural alterations for each residue.

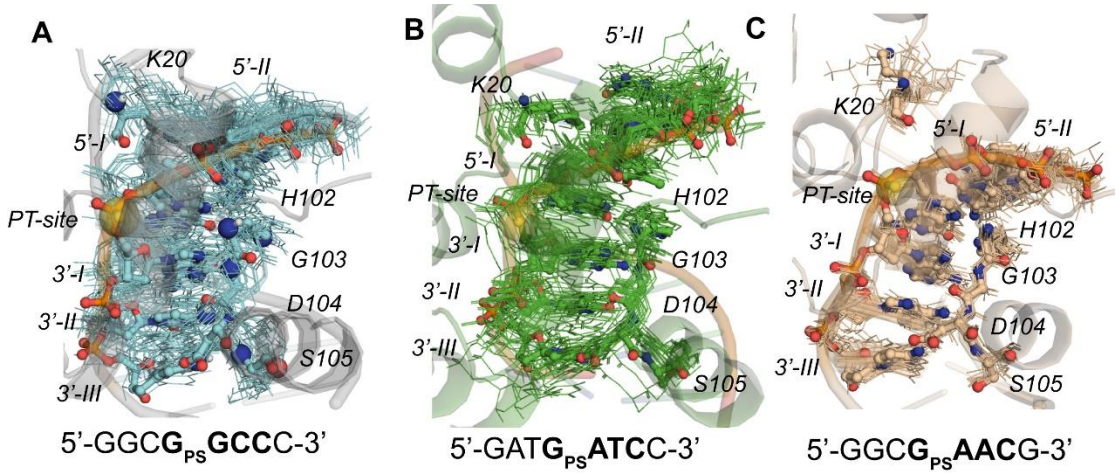

Fig. S4: Superposition of dynamic structures and co-crystal structures of **SBD***Spr* binding with (A) 5'-GGC**G$_{PS}$GCC**C-3', (B)5'-GAT**G$_{PS}$ATC**C-3', and (C) 5'-GGC**G$_{PS}$AAC**G-3'. The geometries of fifty snapshots are randomly extracted from the MD simulations. The co-crystal structures are depicted in stick representation, while the structures in the MD snapshot ensemble are displayed as lines.

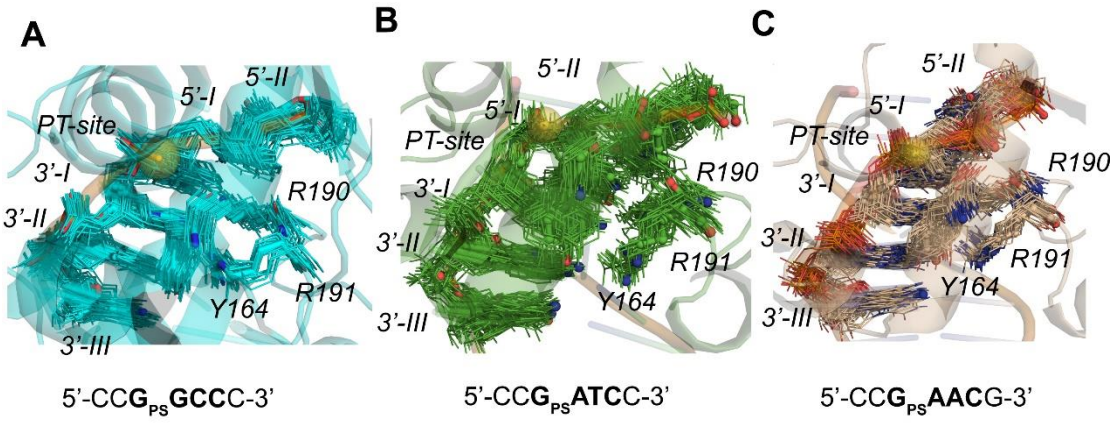

Fig. S5: Superposition of dynamic structures and co-crystal structures of **SBD***Sco* binding with (A) 5'-CC**G$_{PS}$GCC**G-3' (B) 5'-CC**G$_{PS}$ATC**G-3' and (C) 5'-CC**G$_{PS}$AAC**G-3'. The geometries of fifty snapshots are randomly extracted from of the MD simulation. The co-crystal structures are depicted in stick representation, while the structures in the MD snapshot ensemble are displayed as lines.

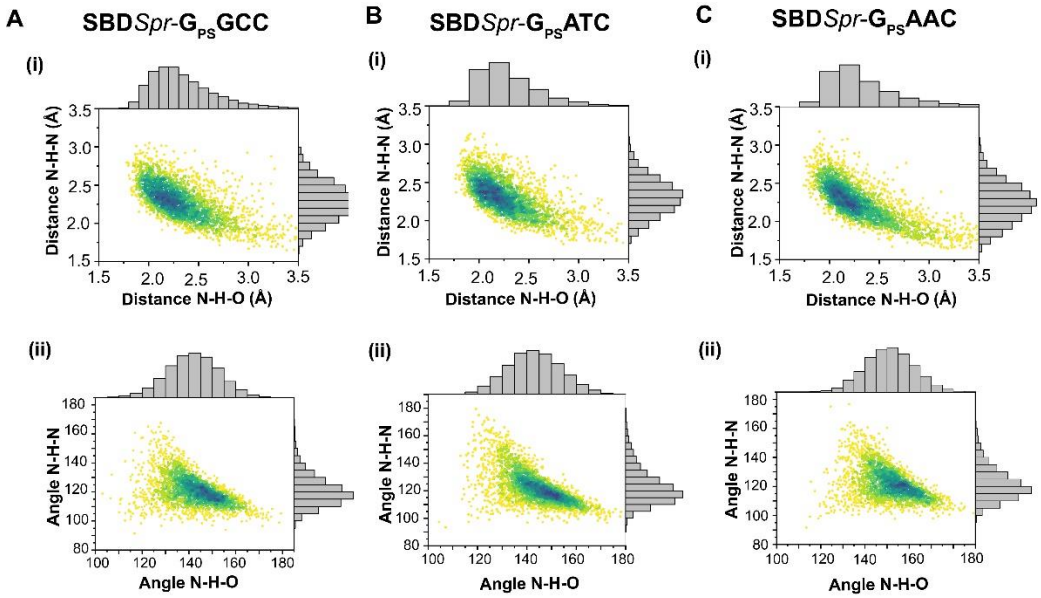

Fig. S6: The bifurcated hydrogen bonds formed between H102 imidazole rings (HB donors) and guanine O6, N7 atoms of bases ***5'-I*** **(**HB acceptors**)** in **SBD***Spr* bound with (A)5'-GGC**GPSGCC**C-3' (B) 5'-GAT**GPSATC**C-3' and (C) 5'-GGC**GPSAAC**G-3'. (i) distance and (ii) angle of N_{δ}-H···N7 and N_{δ}-H···O6 were sampled during the MD simulations. The population density was represented on the map with dark blue indicating the most populated areas and yellow for less populated areas.

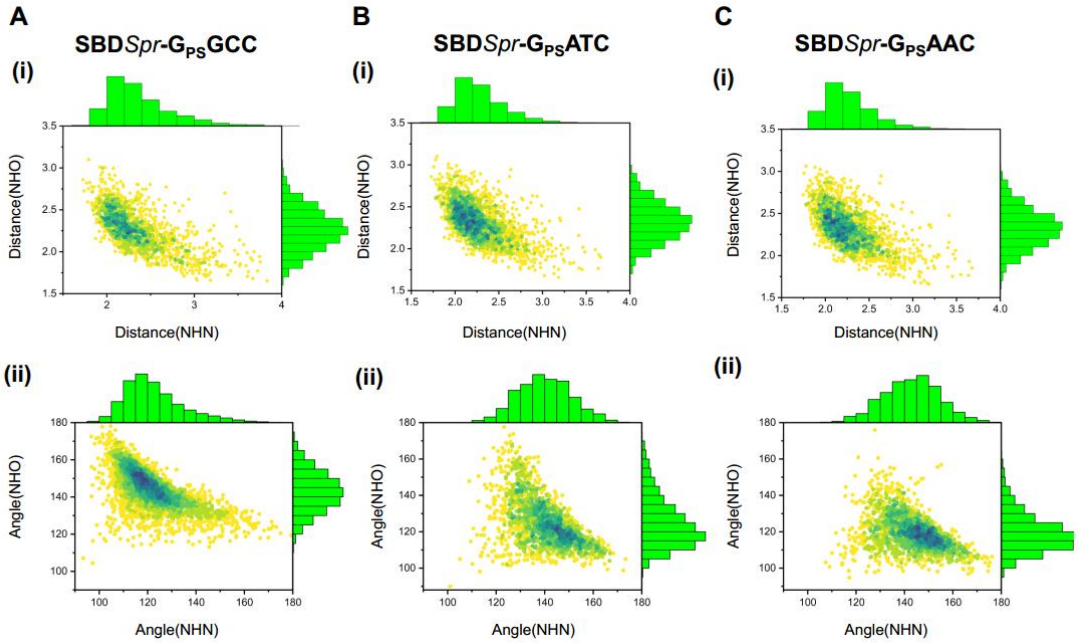

Fig. S7: The bifurcated hydrogen bonds between bases ***5''-III*** and S105 in **SBD***Spr* bound with (A)5'-GGC**G$_{PS}$GCC**C-3' (B) 5'-GAT**G$_{PS}$ATC**C-3' and (C) 5'-GGC**G$_{PS}$AAC**G-3'. The hydrogen bonds were formed between main chain's N atoms of S105 (HB donors) and guanine O6, N7 atoms of ***5''-III*** (HB accecptors). (i) distance and (ii) angle of N-H···N7 and N-H···O6 were sampled during the MD simulations. The population density was represented on the map with dark blue indicating the most populated areas and yellow for less populated areas.

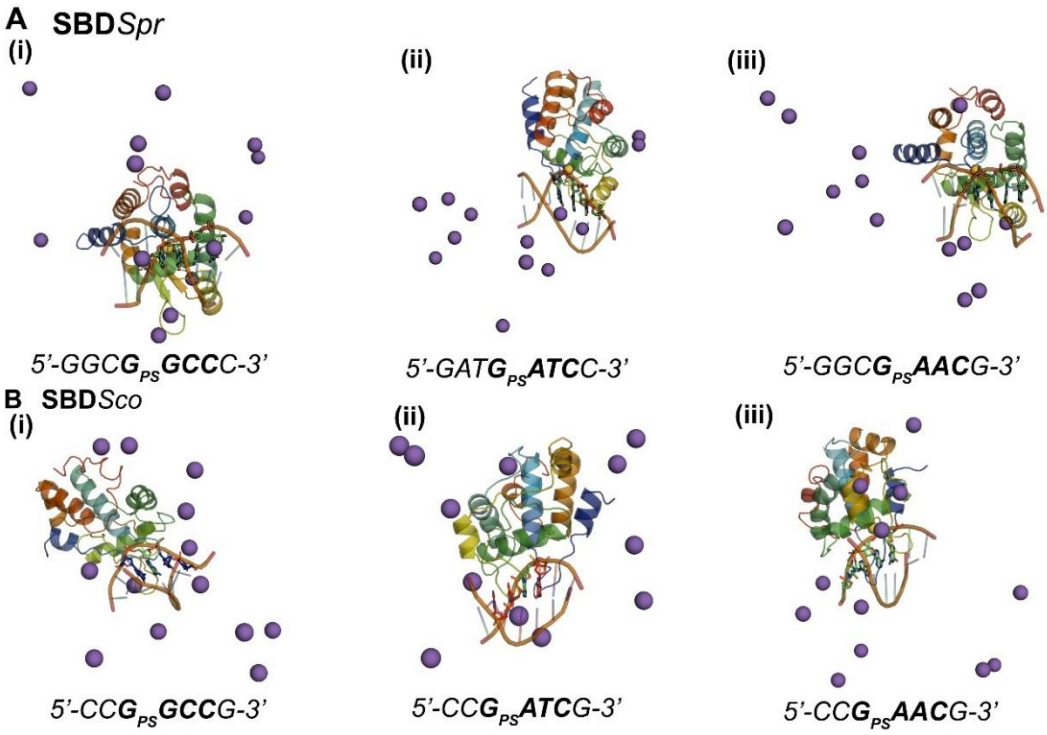

Fig. S8: Counter-ions were discharged from the **SBD**/PT-DNA interface, including the representative sturctures of (A) **SBD***Spr* binding (i) 5'-GGC**GPS**GCCC-3', (ii) 5'-GAT**GPS**ATCC-3', and (iii) 5'-GGC**GPS**AACG-3', and (B) **SBD***Sco* binding with (i) 5'-CC**GPS**GCCG-3', (ii)5'-CC**GPS**ATCG-3', and (iii)5'-CC**GPS**AACG-3'.

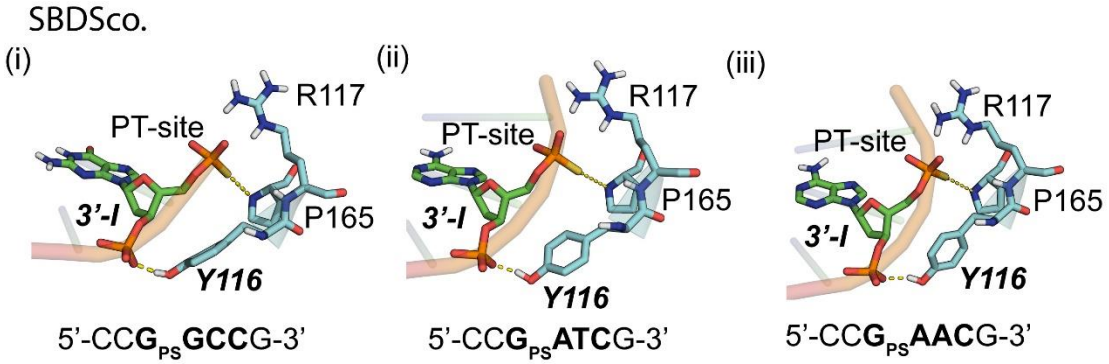

Fig. S9: Conformational dynamics analysis for **SBD***Sco* mutant H116Y binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3'*, (ii) *5'*-CC**G$_{PS}$ATC***G-3'*, and (iii) *5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

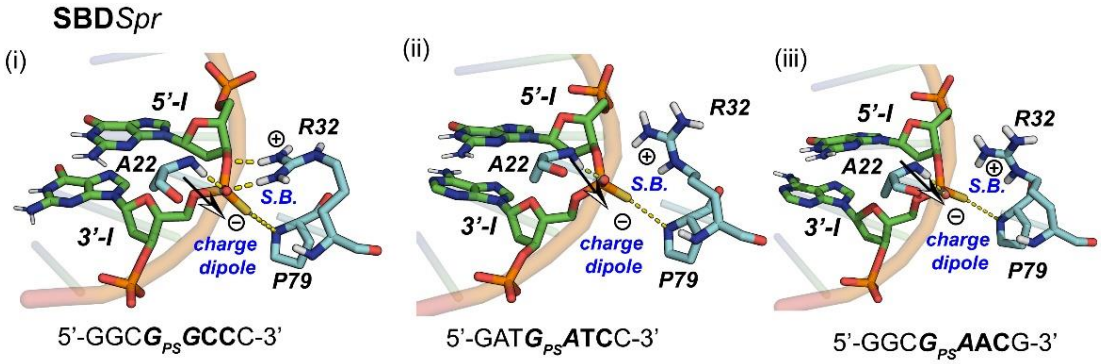

Fig.S10: Conformational dynamics analysis for **SBD***Spr* mutant Q32R binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5'*-GGC**G$_{PS}$GCC**C-*3'*, (ii) *5'*-GAT**G$_{PS}$ATC**C-*3'*, and (iii) *5'*-GGC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

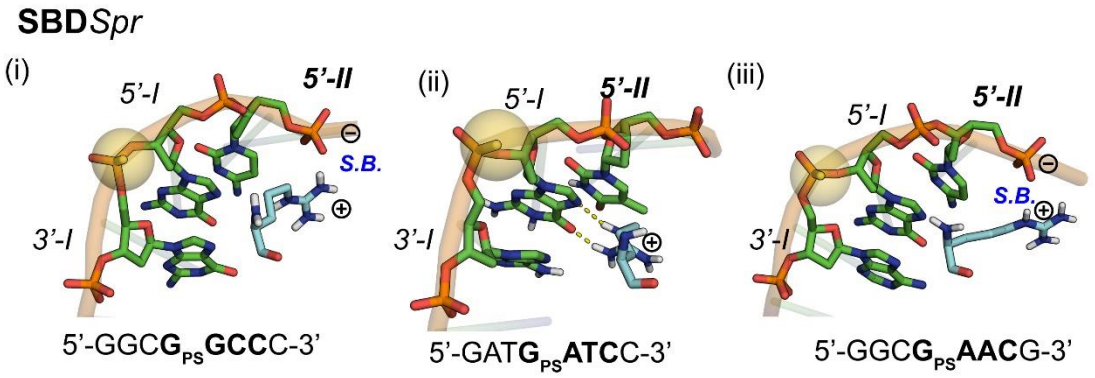

Fig. S11: Conformational dynamics analysis for **SBD***Spr* mutant H102R binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5'*-GGC**G$_{PS}$GCC**C-*3'*, (ii) *5'*-GAT**G$_{PS}$ATC**C-*3'*, and (iii) *5'*-GGC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

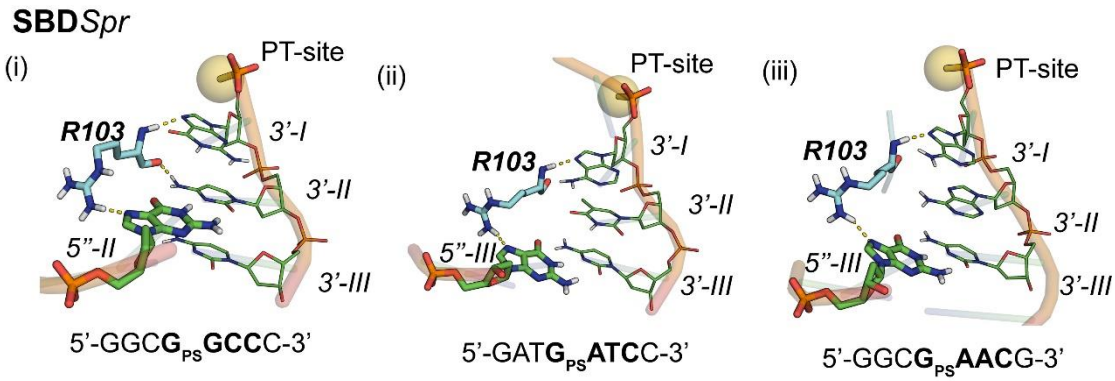

Fig. S12: Conformational dynamics analysis for **SBD***Spr* mutant G103R binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5'*-GGC**G$_{PS}$GCC**C-*3'*, (ii) *5'*-GAT**G$_{PS}$ATC**C-*3'*, and (iii) *5'*-GGC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

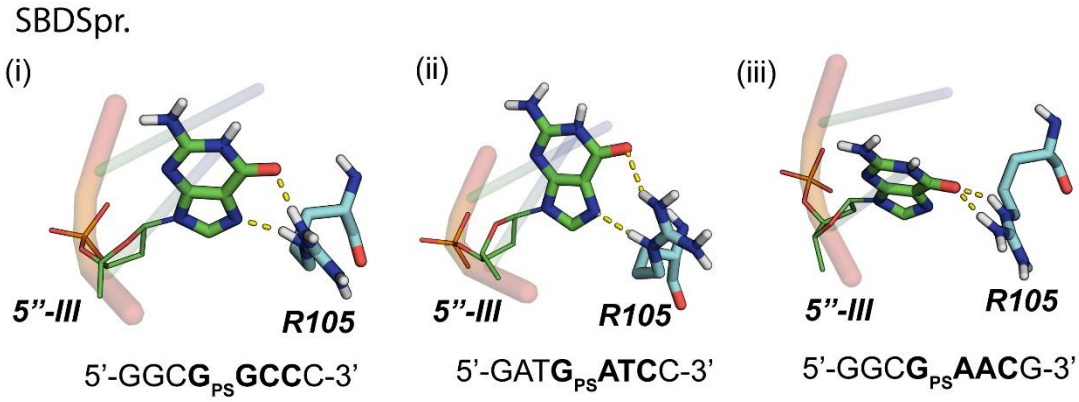

Fig. S13: Conformational dynamics analysis for **SBD***Spr* mutant S105R binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5'*-GGC**G$_{PS}$GCC**C-*3'*, (ii) *5'*-GAT**G$_{PS}$ATC**C-*3'*, and (iii) *5'*-GGC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

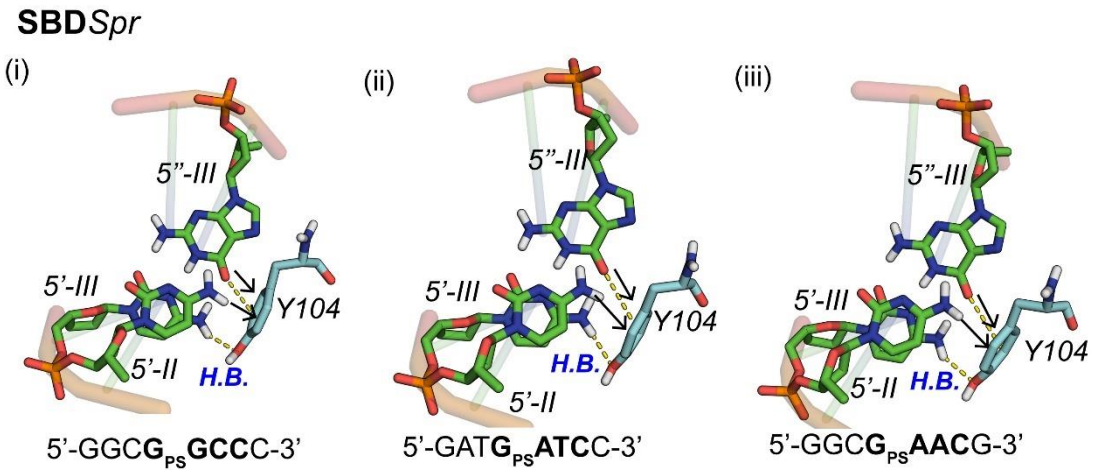

Fig. S14: Conformational dynamics analysis for **SBD***Spr* mutant D104Y binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5’*-GGC**GPSGCC**C-*3’*, (ii) *5’*-GAT**GPSATC**C-*3’*, and (iii) *5’*-GGC**GPSAAC**G-*3’* based on cluster analysis of the MD trajectories. The black arrows represent the T-shaped π-π interactions formed between the base groups and the phenyl rings of Y104.

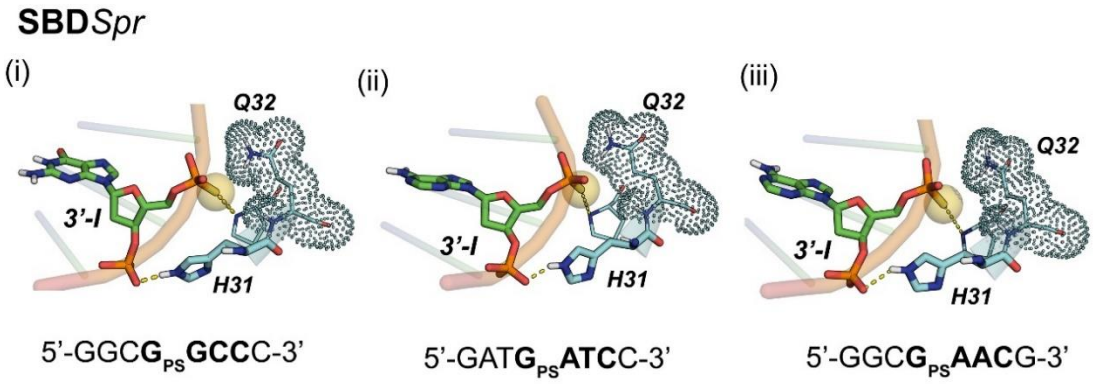

Fig. S15: Conformational dynamics analysis for **SBD***Spr* mutant Y31H binding with PT-DNA. The representative structures of **SBD***Spr* binding with (i) *5'*-GGC**G$_{PS}$GCC**C-*3'*, (ii) *5'*-GAT**G$_{PS}$ATC**C-*3'*, and (iii) *5'*-GGC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

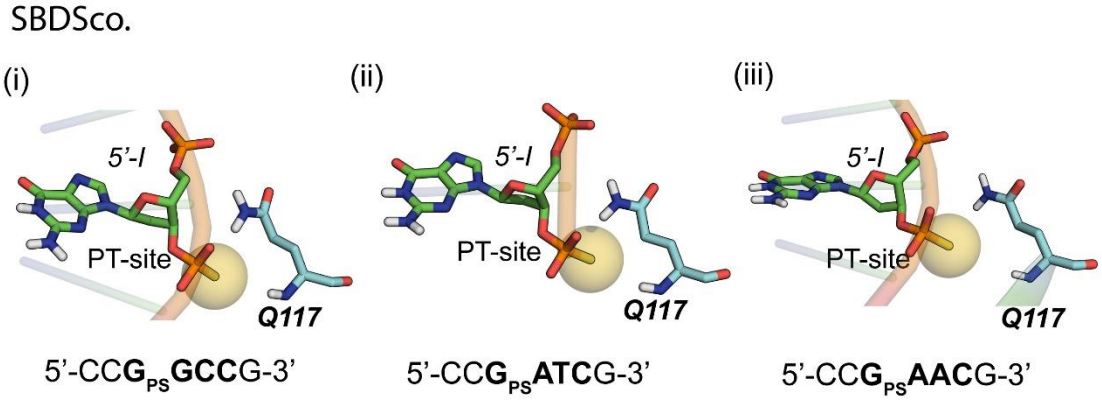

Fig. S16: Conformational dynamics analysis for **SBD***Sco* mutant R117Q binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) 5'-CC**G$_{PS}$GCC**G-3', (ii)5'-CC**G$_{PS}$ATC**G-3', and (iii)5'-CC**G$_{PS}$AAC**G-3' based on cluster analysis of the MD trajectories.

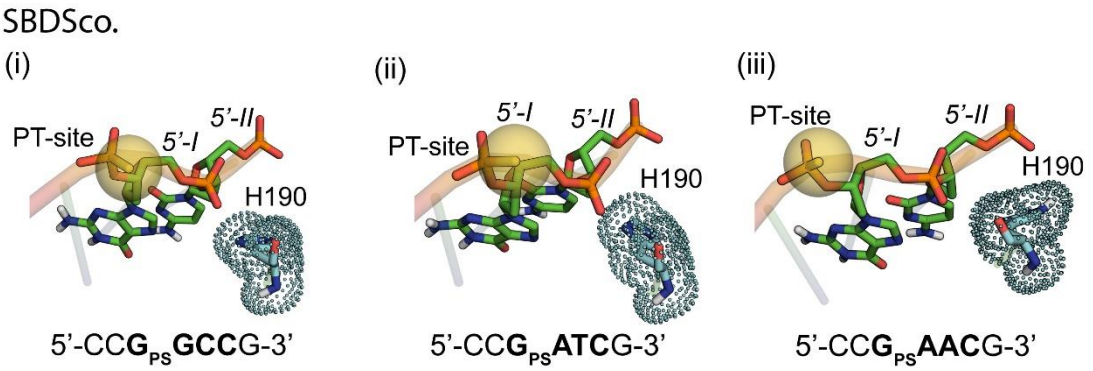

Fig. S17: Conformational dynamics analysis for **SBD***Sco* mutant R190H binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) 5'-CC**G$_{PS}$GCC**G-3', (ii)5'-CC**G$_{PS}$ATC**G-3', and (iii)5'-CC**G$_{PS}$AAC**G-3' based on cluster analysis of the MD trajectories.

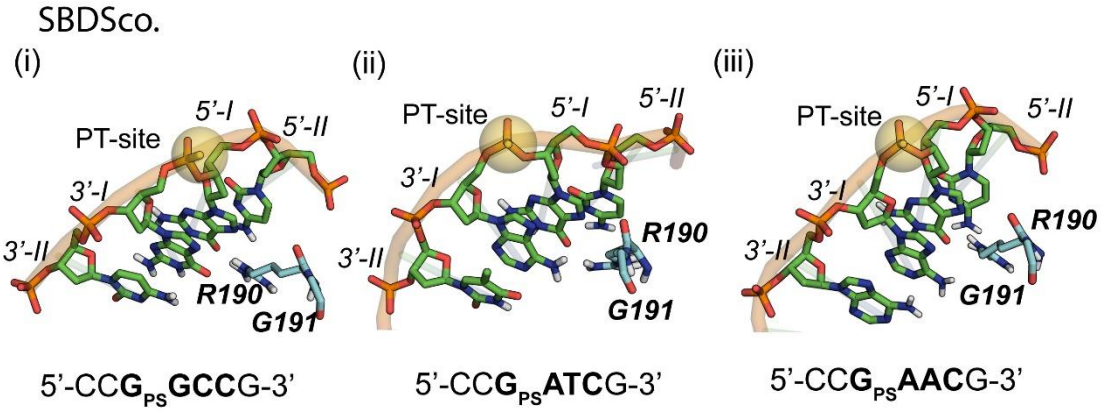

Fig.S18: Conformational dynamics analysis for **SBD***Sco* mutant R191G binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3'*, (ii)*5'*-CC**G$_{PS}$ATC***G-3'*, and (iii)*5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

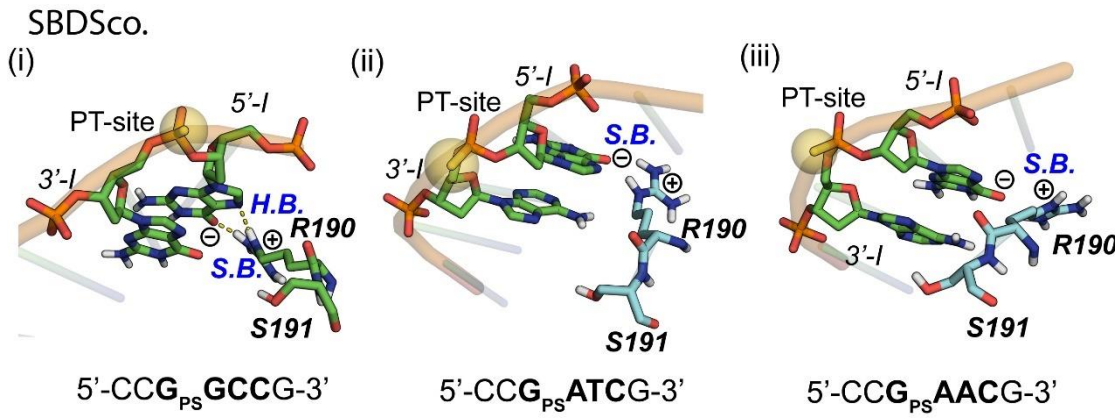

Fig.S19: Conformational dynamics analysis for **SBD***Sco* mutant R191S binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3'*, (ii)*5'*-CC**G$_{PS}$ATC***G-3'*, and (iii)*5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

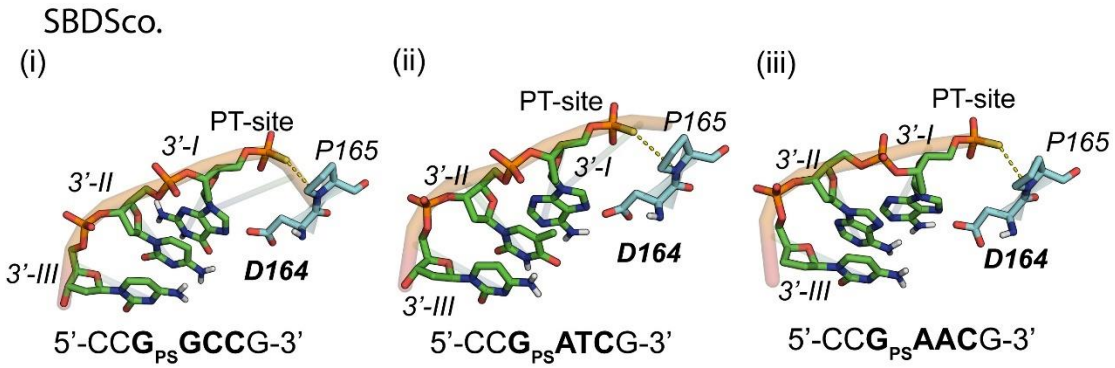

Fig. S20: Conformational dynamics analysis for **SBD**Sco mutant Y164D binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3'*, (ii)*5'*-CC**G$_{PS}$ATC***G-3'*, and (iii)*5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

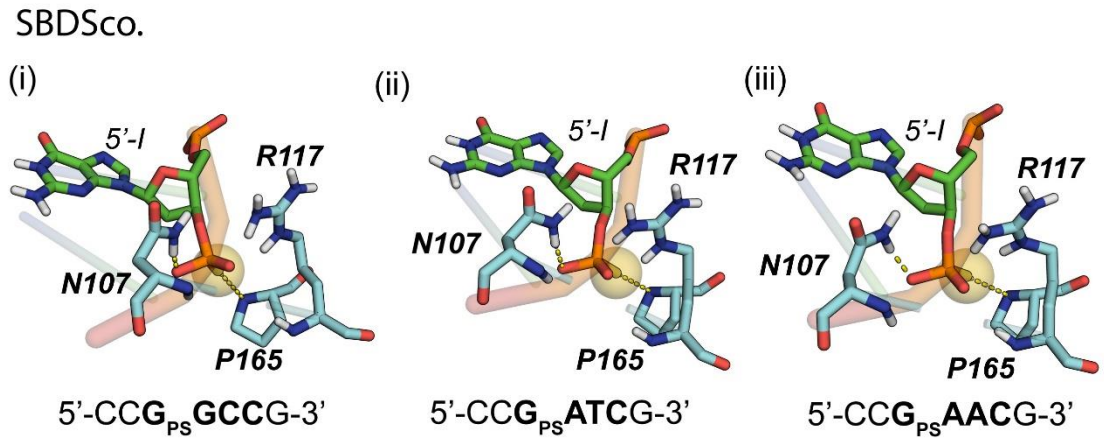

Fig.S21: Conformational dynamics analysis for **SBD***Sco* mutant A107N binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3'*,* (ii)*5'*-CC**G$_{PS}$ATC***G-3'*, and (iii)*5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

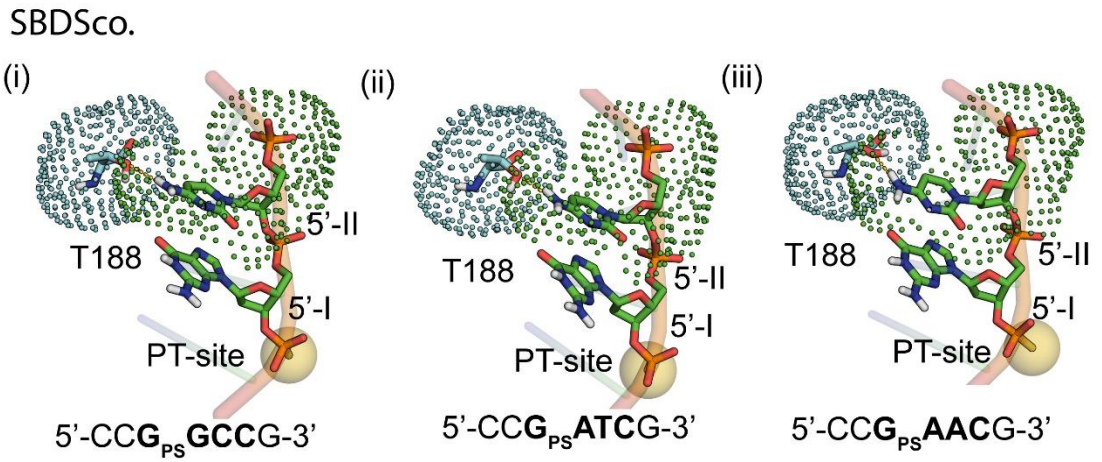

Fig. S22: Conformational dynamics analysis for **SBD***Sco* mutant R188T binding with PT-DNA. The representative structures of **SBD***Sco* binding with (i) *5'*-CC**G$_{PS}$GCC***G-3',* (ii)*5'*-CC**G$_{PS}$ATC***G-3'*, and (iii)*5'*-CC**G$_{PS}$AAC**G-*3'* based on cluster analysis of the MD trajectories.

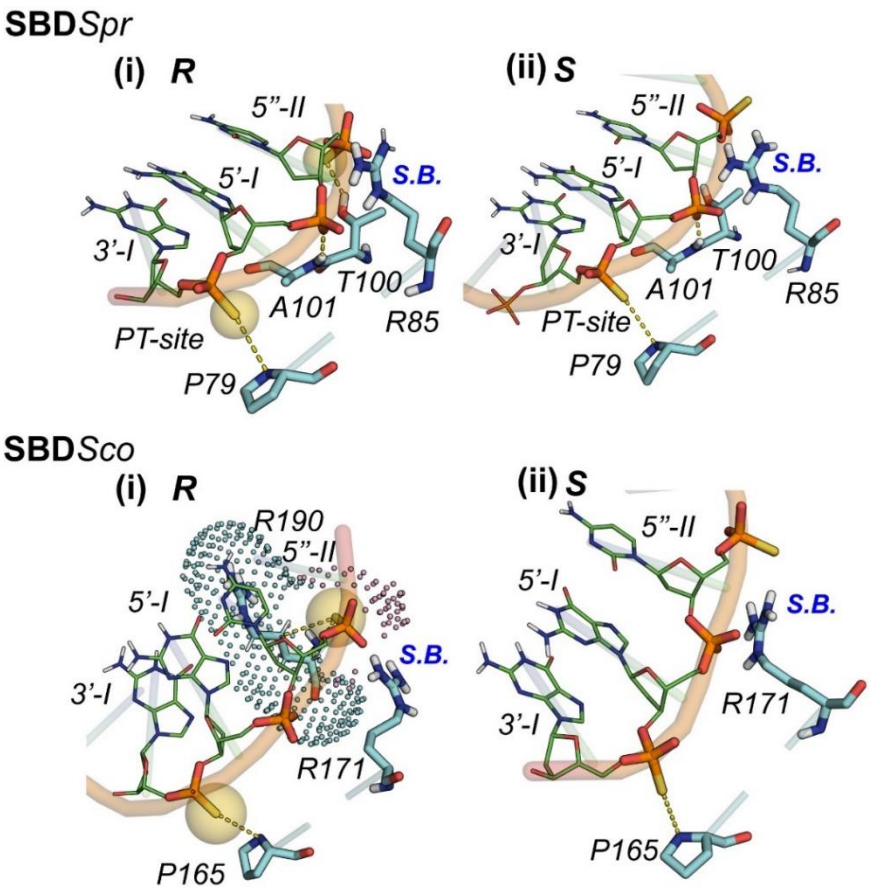

Fig. S23: Stereochemistry of PT-modification at backbones ***5’-II*** in **SBD**/PT-DNA. The representative structures of (A) **SBD***Spr* and (B) **SBD***Sco* binding with PT in *R*p (i) and *S*p (ii) configurations based on cluster analysis of the MD trajectories.

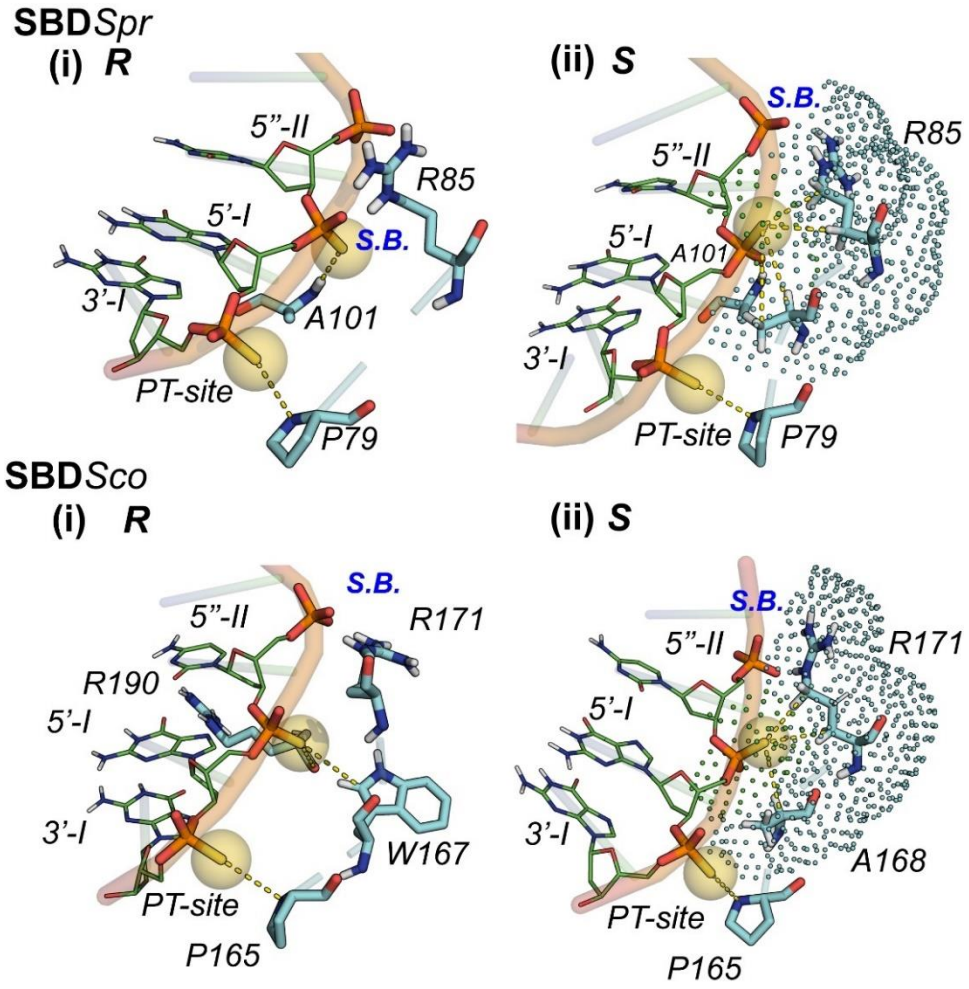

Fig. S24: Stereochemistry of PT-modification at backbones ***5’-I*** in **SBD**/PT-DNA. The representative structures of (A) **SBD***Spr* and (B) **SBD***Sco* binding with PT in *R*p (i) and *S*p (ii) configurations based on cluster analysis of the MD trajectories.

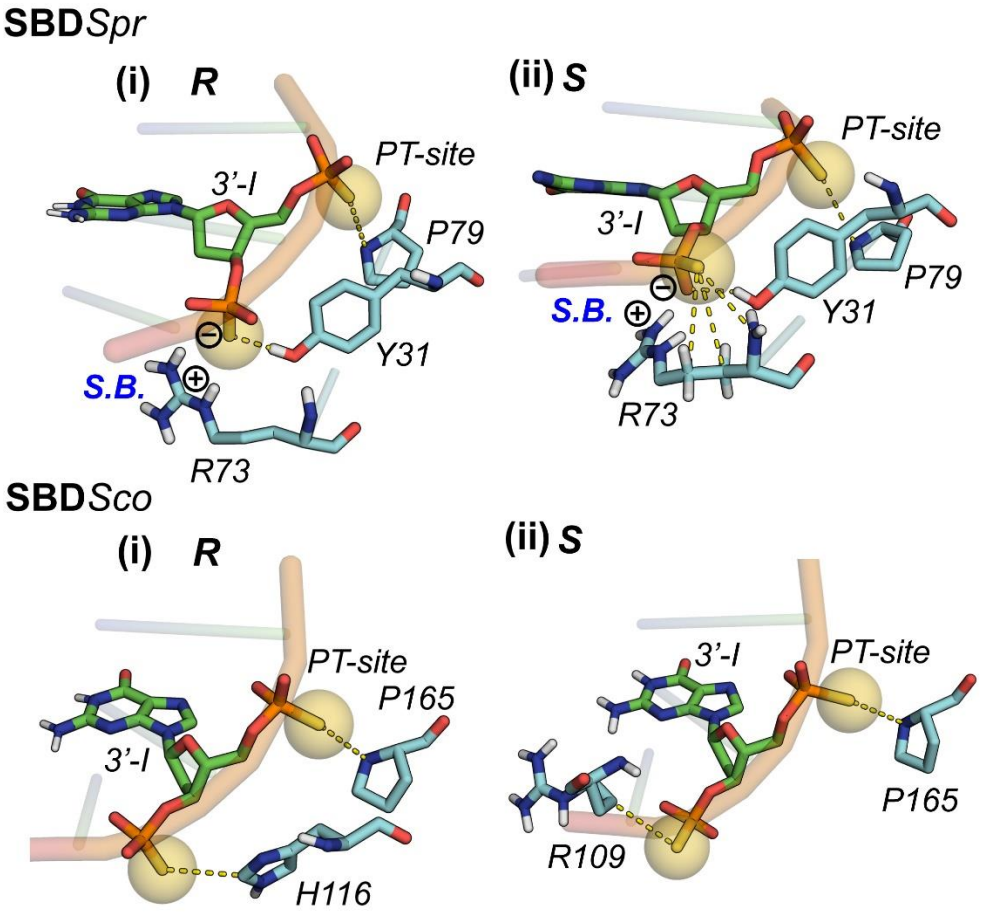

Fig. S25: Stereochemistry of PT-modification at backbone ***3'-I*** in **SBD**/PT-DNA. The representative structures of (A) **SBD***Spr* and (B) **SBD***Sco* binding with PT in *R*p (i) and *S*p (ii) configurations based on cluster analysis of the MD trajectories.

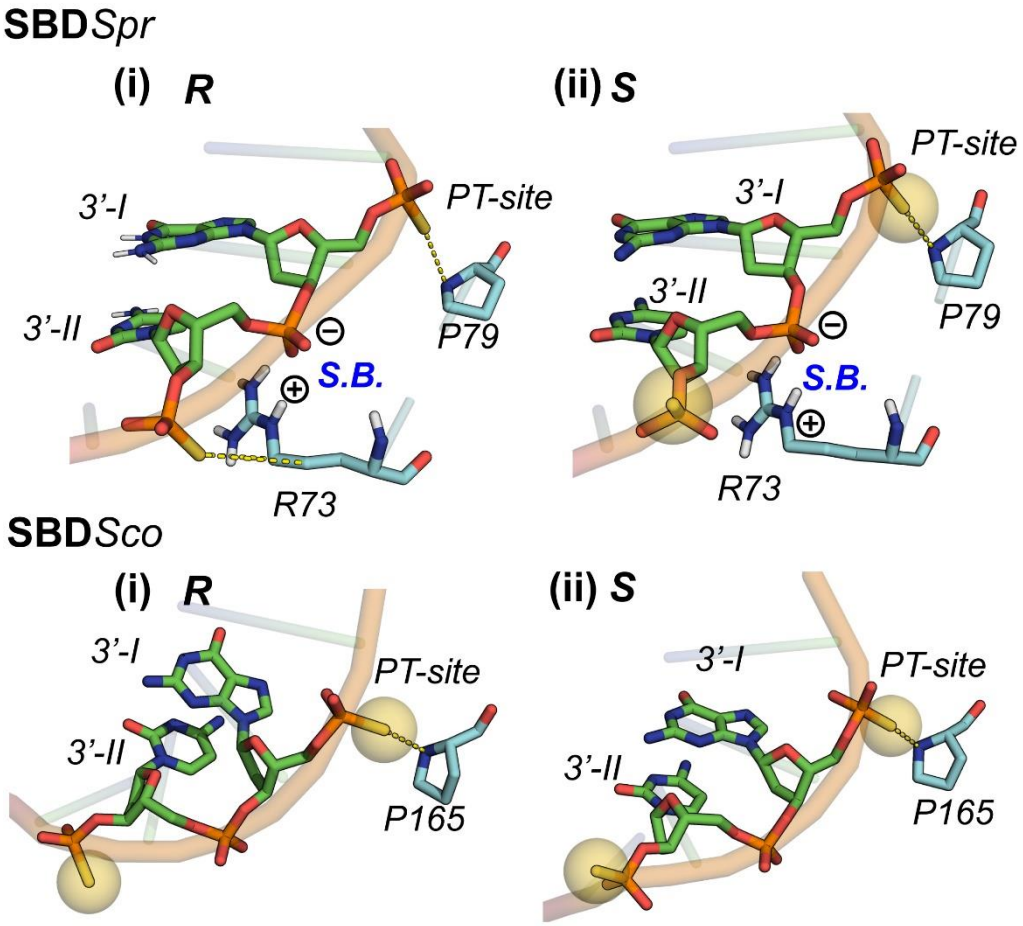

Fig. S26: Stereochemistry of PT-modification at backbone ***3'-II*** in **SBD**/PT-DNA. The representative structures of (A) **SBD***Spr* and (B) **SBD***Sco* binding with PT in *R*p (i) and *S*p (ii) configurations based on cluster analysis of the MD trajectories.

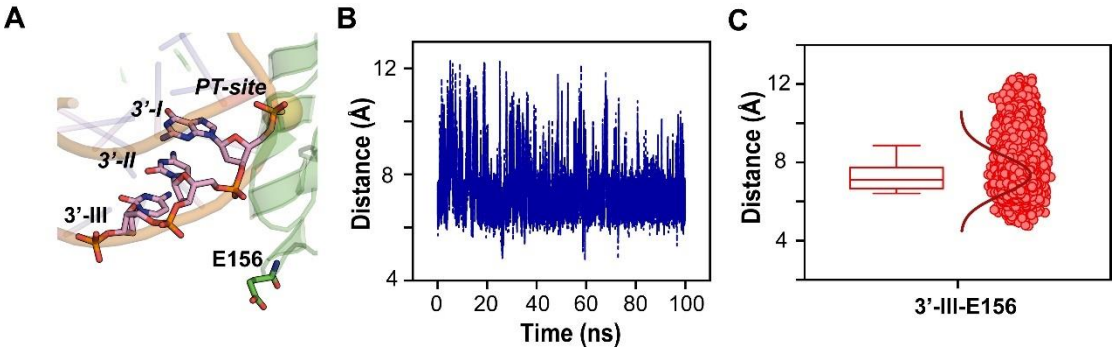

Fig. S27: Conformation dynamic analysis of E156 in **SBD***Sco*-**GPSGCC** during MD simulations. The representative MD frames of **SBD***Sco*-**GPSGCC** complex after cluster analysis of MD trajectories (A). The distances between the E156 and backbone 3'-II were sampled during the MD simulations (B) & (C).

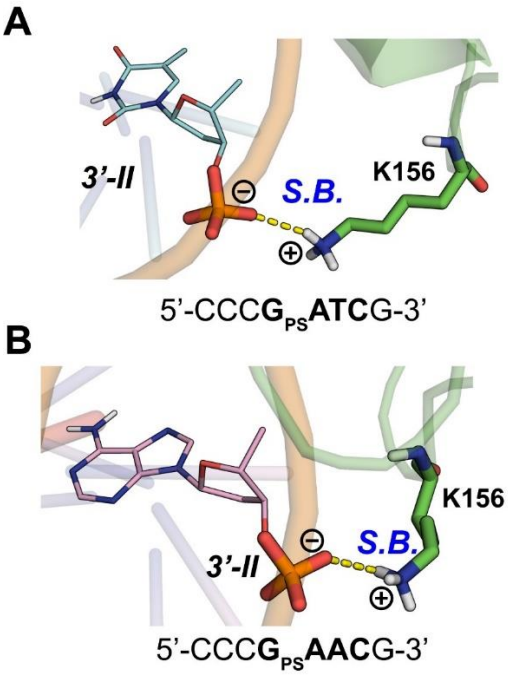

Fig. S28: The representative structures of E156K mutant of **SBD***Sco* binding with PT-DNA (A) *5'*-CC**G$_{PS}$ATC**G-*3'* and (B) *5'*-CC**G$_{PS}$AAC**G-*3'* during MD simulations. The backbone phosphate segments ***3'-II*** form a salt bridge with K156.

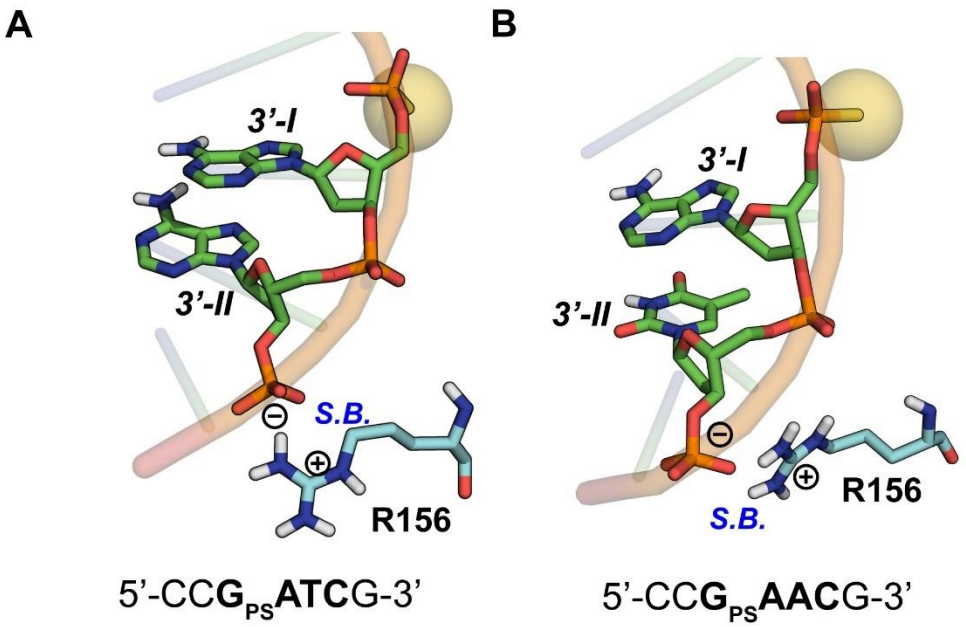

Fig. S28: The representative structures of E156R mutant of **SBD***Sco* binding with PT-DNA (A) *5'*-CC**G$_{PS}$ATC**G-*3'* and (B) *5'*-CC**G$_{PS}$AAC**G-*3'* during MD simulations. The backbone phosphate segments ***3'-II*** form a salt bridge with R156.

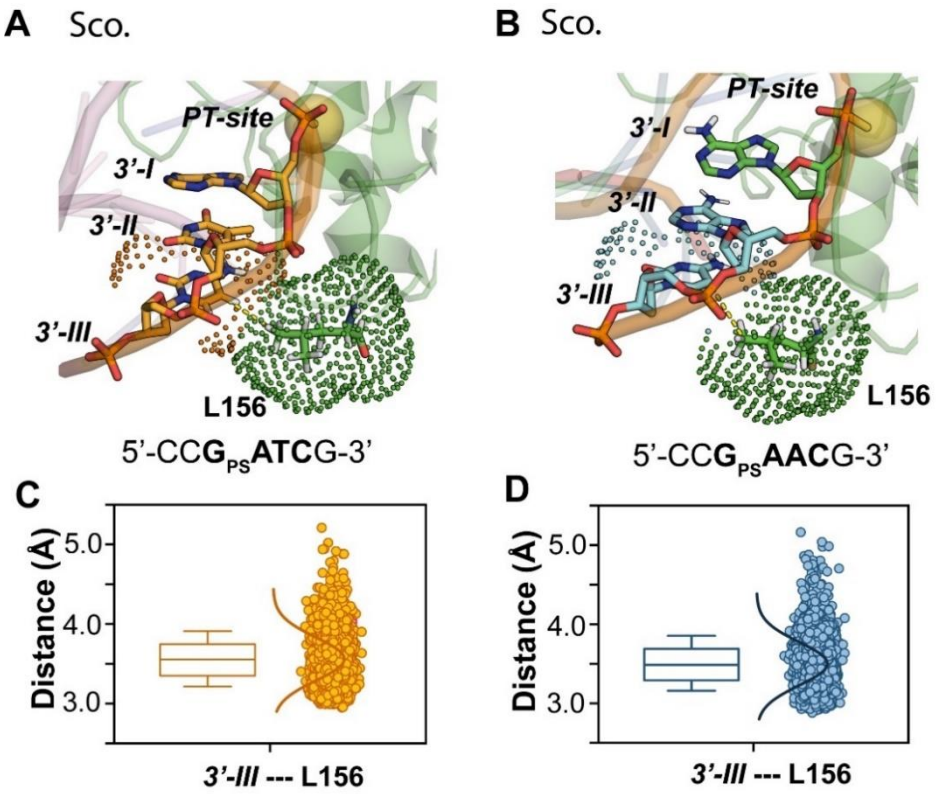

Fig. S30: Conformational dynamics **SBD***Sco* mutant E156L binding with PT-DNA. (A) and B) exhibit the representative structures for variant E156L binding with *5'*-CC**G$_{PS}$ATC**G-*3'* and 5'-CC**G$_{PS}$AAC**G-*3'*, respectively. Distributions of distances of $C_{\delta 1}{}^{L156}$ and bases ***3'-III*** in (C) E156L-G$_{PS}$ATC and (D) E156L-G$_{PS}$AAC, respectively.

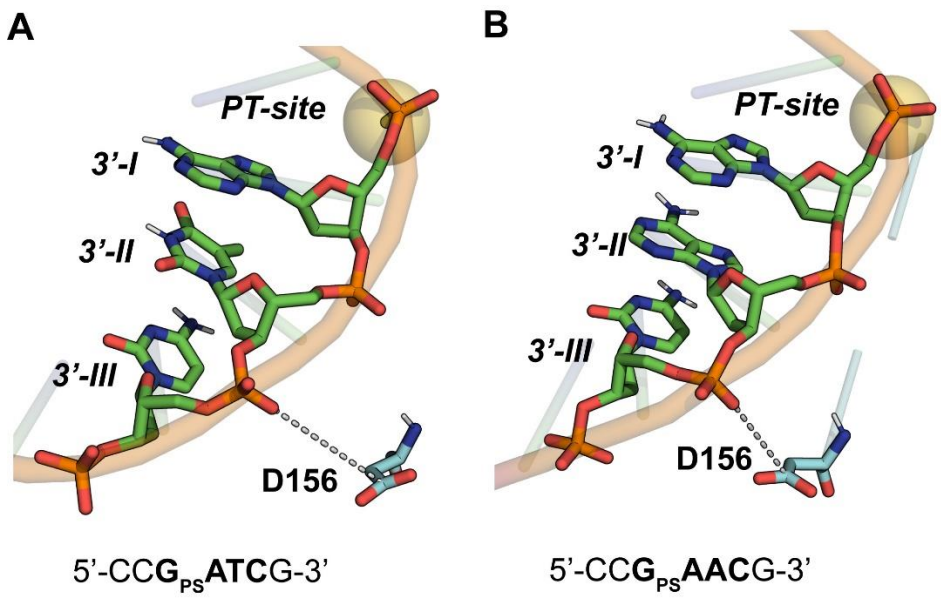

Fig. S31: Conformational dynamics **SBD***Sco* mutant E156D binding with PT-DNA. (A) and (B) exhibit the representative structures for variant E156D binding with *5'*-CC**G$_{PS}$ATC**G-*3'* and 5'-CC**G$_{PS}$AAC**G-*3'*, respectively. The backbone ***3'-II*** and D156 are too distant from each other to form any interaction.

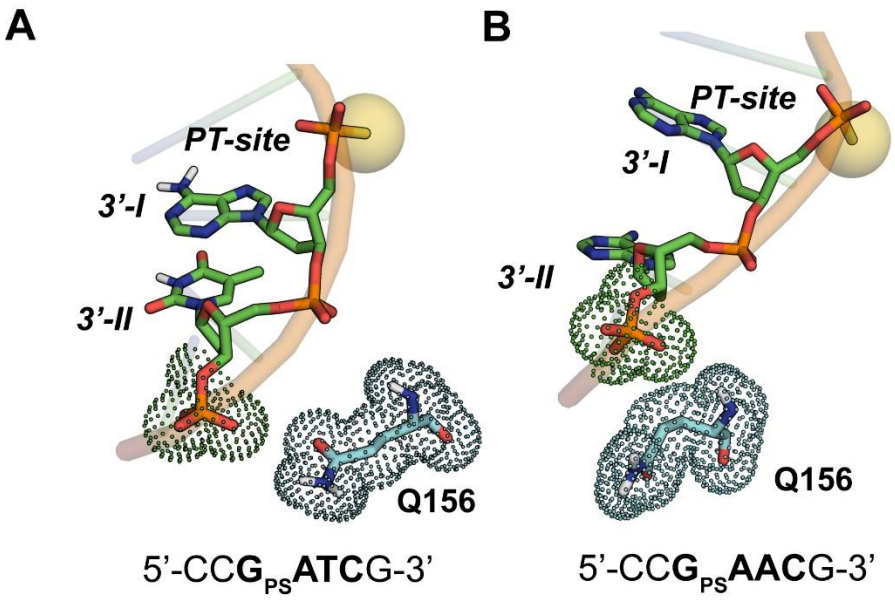

Fig. S32: Conformational dynamics **SBD***Sco* mutant E156Q binding with PT-DNA. (A) and (B) exhibit the representative structures for variant E156Q binding with *5'*-CC**G$_{PS}$ATC**G-*3'* and 5*'*-CC**G$_{PS}$AAC**G-*3'*, respectively. The backbone ***3'-II*** does not form a specific interaction with Q156.

Table S1 Calculated dV/dλ in the four TI cycles. Calculated dV/dλ values and errors (deviation) of the 21 curves in the six different thermodynamic cycle in thermodynamic integration (TI) calculation.

(1) Thermodynamic cycle of **SBD***Spr* binding with *5'*-GGC**G$_{PS}$GCC**C-3'.

Lambda	hemi-modified-G$_{PS}$GCC	Standard deviation	hemi-modified **SBD***Spr*-**G$_{PS}$GCC**	Standard deviation
0	9.9±0.2	4.0	24.7±0.2	6.3
0.05	5.6±0.2	4.9	17.8±0.1	5.5
0.10	5.8±0.2	3.6	13.1±0.1	5.4
0.15	4.5±0.2	4.0	8.1±0.1	5.3
0.20	2.2±0.2	3.9	5.9±0.1	4.8
0.25	2.1±0.1	3.5	3.8±0.1	4.7
0.30	0.5±0.1	3.7	-0.6±0.1	5.3
0.35	-0.5±0.1	3.2	-2.1±0.1	5.1
0.40	-0.7±0.1	2.9	-4.9±0.1	5.2
0.45	-0.8±0.1	2.9	-9.9±0.2	6.6
0.50	-1.9±0.1	2.7	-8.5±0.1	5.3
0.55	-2.0±0.1	2.5	-10.1±0.2	4.4
0.60	-3.1±0.1	2.5	-11.8±0.2	4.0
0.65	-4.9±0.1	3.1	-14.6±0.2	4.2
0.70	-6.6±0.1	2.3	-16. 3±0.1	4.3
0.75	-7.4±0.1	2.3	-16.9±0.1	4.7
0.80	-8.7±0.1	2.0	-22.7±0.2	6.0
0.85	-12.9±0.1	3.1	-33.0±0.2	7.9
0.90	-21.5±0.1	3.8	-46.6±0.2	9.1
0.95	-29.7±0.3	3.9	-64.3±0.2	12.8
1.00	-41.4±0.9	6.1	-115.1±0.4	18.2

(2) Thermodynamic cycle of **SBD***Spr* binding with *5'*-GAT**G_{PS}ATC**C-3'.

Lambda	hemi-modified G_{PS}ATC	Standard deviation	hemi-modified - **SBD***Spr*/ **G_{PS}ATC**	Standard deviation
0	8.8±0.2	4.9	25.7±0.2	7.1
0.05	6.8±0.2	4.1	18.3±0.1	5.7
0.10	5.9±0.2	4.1	13.6±0.1	5.4
0.15	3.5±0.2	3.5	8.1±0.1	5.3
0.20	3.0±0.2	3.9	5.9±0.1	4.8
0.25	2.9±0.1	3.5	3.3±0.1	5.0
0.30	0.9±0.1	3.7	-0.3±0.1	5.3
0.35	0.2±0.1	3.2	-4.2±0.1	5.7
0.40	-0.7±0.1	2.9	-5.3±0.1	5.2
0.45	-0.8±0.1	2.9	-7.4±0.2	5.6
0.50	-2.6±0.1	2.7	-10.1±0.1	6.1
0.55	-3.3±0.1	2.5	-11.2±0.2	4.6
0.60	-3.5±0.1	2.5	-12.3±0.2	4.4
0.65	-3.9±0.1	3.1	-13.2±0.2	3.9
0.70	-3.9±0.1	2.3	-15.7±0.1	4.5
0.75	-6.1±0.1	2.3	-18.0±0.1	4.9
0.80	-8.8±0.1	2.1	-22.5±0.2	6.5
0.85	-10.9±0.1	2.4	-30.0±0.2	8.0
0.90	-15.0±0.1	3.2	-45.3±0.2	8.5
0.95	-22.8±0.3	4.3	-66.9±0.2	11.7
1.00	-33.9±0.7	7.5	-113.0±0.4	17.6

(3) Thermodynamic cycle of **SBD***Spr* binding with *5'*-GGC**G_{PS}AAC**G-3'.

Lambda	hemi-modified G_{PS}AAC-	Standard deviation	hemi-modified - **SBD***Spr*/ **G_{PS}AAC**	Standard deviation
0	8.5±0.2	4.5	24.8±0.2	6.1
0.05	6.4±0.2	3.9	17.9±0.1	5.5
0.10	5.7±0.2	4.1	12.1±0.1	5.4
0.15	4.5±0.2	3.5	8.4±0.1	5.5
0.20	2.5±0.2	3.7	4.5±0.1	4.8
0.25	2.1±0.1	3.0	1.4±0.1	5.0
0.30	1.4±0.1	3.0	-0.3±0.1	5.1
0.35	0.0±0.1	3.5	-3.9±0.1	6.6
0.40	-0.4±0.1	2.7	-4.4±0.1	5.3
0.45	-0.9±0.1	2.9	-9.3±0.2	6.9
0.50	-1.2±0.1	2.7	-8.7±0.1	5.0
0.55	-2.9±0.1	2.5	-11.7±0.2	5.0
0.60	-2.6±0.1	2.2	-11.8±0.2	4.2
0.65	-3.2±0.1	1.8	-15.0±0.2	4.5
0.70	-5.6±0.1	2.6	-14.8±0.1	4.0
0.75	-5.7±0.1	2.9	-17.0±0.1	5.0
0.80	-6.4±0.1	2.3	-23.1±0.2	6.0
0.85	-9.5±0.1	2.6	-32.7±0.2	7.7
0.90	-10.6±0.2	2.1	-42.3±0.2	9.3
0.95	-13.1±0.3	3.5	-67.1±0.2	13.2
1.00	-25.1±0.5	7.0	-116.2±0.4	17.2

(4) Thermodynamic cycle of **SBD***Sco* binding with *5'*-CC**G$_{PS}$GCC**G-3'.

Lambda	hemi-modified **G$_{PS}$GCC**	Standard deviation	hemi-modified - **SBD***Sco* **G$_{PS}$GCC**	Standard deviation
0	11.5±0.2	4.8	25.2±0.2	7.2
0.05	9.4±0.2	3.9	17.9±0.1	5.5
0.10	8.6±0.2	4.1	13.1±0.1	5.4
0.15	7.5±0.2	3.9	8.1±0.1	5.5
0.20	6.8±0.2	3.2	4.5±0.1	4.8
0.25	3.7±0.1	3.3	1.4±0.1	5.0
0.30	2.4±0.1	3.4	-0.3±0.1	5.1
0.35	1.2±0.1	3.1	-3.9±0.1	6.6
0.40	-0.1±0.1	2.3	-4.4±0.1	5.3
0.45	-0.9±0.1	2.4	-9.3±0.2	6.9
0.50	-1.2±0.1	2.8	-8.7±0.1	4.8
0.55	-2.5±0.1	2.2	-11.7±0.2	5.0
0.60	-3.0±0.1	2.3	-11.8±0.2	4.7
0.65	-3.2±0.1	2.1	-15.0±0.2	4.1
0.70	-5.6±0.1	2.0	-14.8±0.1	3.5
0.75	-6.7±0.1	2.9	-17.0±0.1	4.7
0.80	-7.4±0.1	2.3	-23.1±0.2	5.7
0.85	-8.5±0.1	2.6	-32.7±0.2	7.1
0.90	-9.6±0.1	2.3	-42.3±0.2	8.8
0.95	-10.1±0.3	3.8	-65.1±0.2	12.8
1.00	-17.4±0.7	6.9	-112.4±0.3	15.4

(5) Thermodynamic cycle of **SBD***Sco* binding with *5'*-CC**G$_{PS}$ATC**G-3'.

Lambda	hemi-modified **G$_{PS}$ATC**	Standard deviation	hemi-modified **SBD***Sco*/**G$_{PS}$ATC**	Standard deviation
0	9.9±0.2	4.5	26.2±0.2	6.1
0.05	7.5±0.2	4.4	17.2±0.1	5.5
0.10	6.1±0.2	4.1	12.5±0.1	5.4
0.15	5.1±0.2	3.6	9.2±0.1	5.5
0.20	4.0 ±0.2	3.4	5.2±0.1	4.8
0.25	3.1±0.1	2.9	3.4±0.1	5.0
0.30	2.4±0.1	3.6	-1.3±0.1	5.1
0.35	0.9±0.1	3.2	-2.7±0.1	6.6
0.40	-0.2±0.1	2.6	-4.1±0.1	5.3
0.45	-1.1±0.1	3.1	-8.2±0.2	6.9
0.50	-1.6±0.1	2.2	-10.7±0.1	5.0
0.55	-2.2±0.1	2.8	-11.2±0.2	5.0
0.60	-2.5±0.1	2.5	-12.8±0.2	4.2
0.65	-3.7±0.1	2.3	-14.0±0.2	4.1
0.70	-3.8±0.1	3.0	-15.9±0.1	4.7
0.75	-4.7±0.1	3.2	-17.0±0.1	5.2
0.80	-6.5±0.1	3.0	-24.1±0.2	6.0
0.85	-8.0±0.1	2.1	-34.7±0.2	6.7
0.90	-8.8±0.1	2.7	-48.3±0.2	7.3
0.95	-11.5±0.3	3.8	-68.3±0.3	12.6
1.00	-16.1±0.7	6.5	-115.8±0.4	14.7

(6) Thermodynamic cycle of **SBD***Sco* binding with *5'*-CC**G$_{PS}$AAC**G-3'.

Lambda	hemi-modified **G$_{PS}$AAC**	Standard deviation	hemi-modified **SBD**Sco/ **G$_{PS}$AAC**	Standard deviation
0	9.5±0.3	7.9	28.5±0.2	10.1
0.05	6.4±0.3	5.4	16.8±0.1	8.5
0.10	5.7±0.2	4.6	12.6±0.1	7.4
0.15	4.5±0.2	3.8	8.8±0.1	6.2
0.20	2.5±0.2	3.2	4.4±0.1	4.8
0.25	2.1±0.1	3.4	1.5±0.1	5.0
0.30	1.7±0.1	3.6	-0.9±0.1	5.1
0.35	0.2±0.1	3.8	-4.0±0.1	4.6
0.40	-0.3±0.1	2.9	-4.7±0.1	5.5
0.45	-0.7±0.1	2.7	-8.1±0.2	4.9
0.50	-1.2±0.1	2.8	-9.9±0.1	5.0
0.55	-2.7±0.1	2.9	-12.4±0.2	5.1
0.60	-3.6±0.1	2.4	-13.8±0.2	4.2
0.65	-4.5±0.1	2.1	-16.0±0.2	4.7
0.70	-6.2±0.1	2.2	-17.8±0.1	4.3
0.75	-7.7±0.1	2.9	-21.2±0.1	5.8
0.80	-9.4±0.1	3.3	-26.1±0.2	6.4
0.85	-12.5±0.1	4.2	-30.4±0.2	7.3
0.90	-16.9±0.1	5.4	-51.1±0.2	8.3
0.95	-18.7±0.2	5.2	-73.2±0.3	13.2
1.00	-23.6±0.4	6.5	-112.5±0.4	18.5

Table S2: **SBD***Spr* interacts with lateral bases

Site	Base type	Residues in SBD	Distance(Å)	Type of interaction
5'-III	(1) Nε atom of guanine in *5'*-G*G*C**G$_{PS}$GCC**C-*3'* and *5'*-G*G*C**G$_{PS}$AAC**G-*3'* (2) C2 atom of adenine in in *5'*-G*A*T**G$_{PS}$ATC**C-*3'*	H102's Cε atom	4.1-4.5	hydrophobic
5'-II	(1) C2 atom of cytosine *5'*-GG*C***G$_{PS}$GCC**C-*3'* and *5'*-GG*C***G$_{PS}$AAC**G-*3'*; (2) C5 atom of thymine in *5'*-GA*T***G$_{PS}$ATC**C-*3'*	imidazole side chain of H102	3.3-3.7	π-π stacking
5'-I	guanine	H102 aromatic nitrogen atom	1.9 - 2.3	hydrogen bond
3'-I	N7 atom of guanine and adenine	nitrogen atom of G103	1.9 - 2.2	hydrogen bond
5''-I	cytosine N4 of 5'-I thymine O4 of 5''-I	imidazole ring of H102	4.0-4.4	hydrophobic
3'-II	(1) N4 atom of cytosine in *5'*-GGC**G$_{PS}$G*C***C-*3'* (2) C7 atom of thymine in *5'*-GAT**G$_{PS}$A*T***C-*3'* (3) N6 atom of adenine in *5'*-GGC**G$_{PS}$A*A***CG-*3'*	oxygen atom of G103	(1) 2.0 (2) 2.6 (3) 2.5	(1) hydrogen bond (2)&(3) hydrophobic
3'-III	N4 atom of cytosine	$O_δ$ atom of D104	2.0	hydrogen bond
5'-III	(1) complementary strand of 5'-GGC**G$_{PS}$GC*C***C-3' (guanine O6 and N7) (2) complementary strand of 5'-GAT**G$_{PS}$AT*C***G-3' (guanine O6) (3) complementary strand of 5'-GGC**G$_{PS}$AA*C***C-3'	main chain N-H group of S105	(1) 1.9-2.1 (2) 2.0-2.2 (3) 3.7-4.2	(1) hydrogen bond (2) hydrogen bond (3) hydrophobic

Table S3: **SBD***Sco* interacts with lateral bases

Site	Base type	Residues in SBD	Distance(Å)	Type of interaction
5'-II	cytosine	guanidyl $N_{\eta 2}$ atom of R180	3.8-4.1	hydrophobic
5'-I	O6 atom of guanine	guanidinium group of R190	1.9 - 2.1	hydrogen bond
3'-I	guanidyl N6 atom of *5'*-*CCG***GPS*G*CC***G-3'* adenine in *5'*-CC**GPS*A*AC**G-*3'* or *5'*-CC**GPS*A*TC**G-*3'*	R191	1.9 - 2.2	hydrogen bond hydrophobic
3'-II	(1) N4 atom of cytosine in *5'*-GGC**GPSG*C*C**C-*3'* (2) thymine in *5'*-GAT**GPSA*T*C**C-*3'* (3) N6 atom of adenine in *5'*-GGC**GPSA*A*C**G-*3'*	(1) Y164 (2) Y164 and R191 (3) Y164	(1) 1.9 (2) 2.6 & 2.0 (3) 1.9	hydrogen bond hydrophobic & hydrogen bond hydrogen bond
5''-II	(1) the complementary strand of *5'*-CC**GPSG*C*C**G-*3'* (guanine O6) and *5'*-CC**GPSA*A*C**G-*3'* (thymine O6) (2) adenine of ***5''-II*** in *5'*-CC**GPSA*T*CC**G-*3'*	R191	2.0 3.3	hydrogen bond hydrophobic
3'-III	N4 atom of cytosine	Y164	3.4-3.8	hydrophobic
5''-III	guanine N7 of ***5''-III*** in 5'-GGC**GPSGC*C***C-3' and 5'-GAT**GPSAA*C***G-3' guanine of ***5''-III*** in 5'-GGC**GPSAT*C***C-3'	(1) $N_{\eta 2}$ atom of R191 (2) R191	(1) 1.9-2.1 (2) 3.7-4.2	(1) hydrogen bond (2) hydrophobic

Table S4**:** Binding energy contribution of deoxyribose in **SBD***Spr*/PT-DNA complex

Deoxyribose	**Residue**	**Contribution**
5'-III		negligible
5'-II	T100	~4.8-5.3%
5'-I	A82	~2.3-3.1%
3'-I	Y31	~1.5-2.2%
3'-II	R29	~2.4-3.3 %
3'-III		negligible
Sum		~11.0-13.9%

Table S5**:** Binding energy contribution of deoxyribose in **SBD***Sco*/PT-DNA complex

Deoxyribose	Residue	Contribution
5'-III		negligible
5'-II	R171	~4.1-4.7%
5'-I	A168	~2.4-3.2%
3'-I	Y164, H116	~2.9-3.8%
3'-II	R109	~2.1-2.9 %
3'-III		negligible
Sum		~11.5-14.6%

Table S6: The energy distributions of **SBD***Spr*/5'-GGC**G$_{PS}$GCC**C-3' **(**Binding energy in calculation = -14.5 kcal/mol**)**

	PT-strain		**Complementary**	
base		*backbone*	*base*	
		~0.4%		
5'-III	~0.9%			
		~1.1%		
5'-II	~1.8%			
		~14.8%		
5'-I	~8.9%			
		~16.2% (PT)		
3'-I	~6.5%		**5"-I**	~4.5%
		~18.9%		
3'-II	~7.8%		**5"-II**	~0.3%
		~3.9%		
3'-III	~6.4%		**5"-III**	~10.6%
4 core bases	~29.7%	~39.0%		

Table S7: The binding energy distributions in the case of **SBD***Spr*/5'-GAT**G$_{PS}$ATC**C-3' **(**Binding energy in calculation = -13.3 kcal/mol**)**

PT-strain			**Complementary**	
base		*backbone*	*base*	
		~0.5%		
5'-III	~1.0%			
		~1.5%		
5'-II	~2.1%			
		~15.9%		
5'-I	~8.9%			
		~18.1% (PT)		
3'-I	~7.0%		**5"-I**	~3.6%
		~18.9%		
3'-II	~5.2%		**5"-II**	~0.5%
		~3.1%		
3'-III	~8.2%		**5"-III**	~8.0%
4 core bases	~29.3%	~40.1%		

Table S8: The binding energy distributions of **SBD***Spr*/5'-GGC**G$_{PS}$AAC**C-3' **(**Binding energy in calculation = -12.7 kcal/mol**)**

	PT-strain			**Complementary**	
base		*backbone*		*base*	
		~0.6%			
5'-III	~1.1%				
		~1.6%			
5'-II	~2.5%				
		~17.0%			
5'-I	**~10.7%**				
		~19.2%			
		(PT)			
3'-I	~7.4%			**5"-I**	3.0%
		~22.2%			
3'-II	~3.0%			**5"-II**	~0.5%
		~2.8%			
3'-III	~7.8%			**5"-III**	~3.3%
4 core bases	~28.8%	~43.2%			

Table S9: The binding energy distributions in the case of **SBD***Sco*/5'-CC**G$_{PS}$GC**CG-3' **(**Binding energy in calculation = -12.6 kcal/mol**)**

PT-strain			**Complementary**	
base		*backbone*	*base*	
		~0.6%		
5'-III	~0.3%			
		~22.2%		
5'-II	~3.3%			
		~3.7%		
5'-I	**~15.6%**			
		~24.2% (PT)		
3'-I	~0.6%		**5"-I**	negligible
		~4.6%		
3'-II	**~15.9%**		**5"-II**	~3.3%
		~4.4%		
3'-III	~1.0%		**5"-III**	~4.0%
4 core bases	~32.1%	~33.2%		

Table S10: The binding energy distributions in the case of **SBD***Sco*-5'-CC**G$_{PS}$ATC**G-3' (Binding energy in calculation = -11.4 kcal/mol).

	PT-strain		**Complementary**	
base		*backbone*	*base*	
		~0.6%		
5'-III	~0.3%			
		~20.0%		
5'-II	~3.0%			
		~3.5%		
5'-I	**~15.6%**			
		~21.8% (PT)		
3'-I	~8.1%		**5"-I**	~4.2%
		~4.2%		
3'-II	~4.5%		**5"-II**	~0.6%
		~5.4%		
3'-III	~8.4%		**5"-III**	~9.3%
4 core bases	~36.6%	~31.4%		

Table S11. The binding energy distributions in the case of **SBD***Sco*-5'-CC**G$_{PS}$AAC**G-3' (Binding energy in calculation = -11.8 kcal/mol).

	PT-strain			**Complementary**	
base		*backbone*		*base*	
		~0.6%			
5'-III	~0.5%				
		~19.4%			
5'-II	~2.4%				
		~3.8%			
5'-I	**~11.3%**				
		~22.0% (PT)			
3'-I	~8.0%			**5"-I**	~4.2%
		~4.8%			
3'-II	~4.5%			**5"-II**	~0.6%
		~5.2%			
3'-III	~8.3%			**5"-III**	~9.2%
4 core bases	~32.2%	~32.0%			

Table S12: Analysis of upstream bases of 360 PT-modifications 5′-G_{PS}AAC-3′/5′-G_{PS}TTC-3′ with a frequency higher than 30% in *E. coli* B7A.

		5'-III				
		A	**C**	**G**	**T**	**TOTAL**
5'-II	**A**	(41/360) 11.39%	(39/360) 10.83%	(45/360) 12.50%	(5/360) 1.39%	36.11%
	C	(19/360) 5.28%	(25/360) 6.94%	(32/360) 8.89%	(9/360) 2.50%	23.61%
	G	(48/360) 13.33%	(24/360) 6.67%	(29/360) 8.06%	(18/360) 5.00%	33.06%
	T	(6/480) 1.67%	(13/360) 3.61%	(5/360) 1.39%	(2/360) 0.56%	7.22%
	TOTAL	31.67%	28.06%	30.83%	9.44%	100.00%

Table S13. The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GGC**G**$_{\mathbf{PS}}$**GCC**C-3' system with the H++ web server

Residue	$pK_{1/2}$	Residue	$pK_{1/2}$	Residue	$pK_{1/2}$
Asp7	2.4	Asp55	3.3	Asp104	2.5
Arg8	>12.0	Asp57	4.6	Arg108	>12.0
Glu10	4.4	Glu58	10.4	Glu113	4.1
Asp11	>12.0	Lys65	12.0	Arg115	>12.0
Asp14	3.1	Arg66	7.0	Glu122	3.1
Asp15	5.2	His67	>12.0	His125	5.2
Arg17	>12.0	Arg70	2.9	Asp126	1.1
Lys20	11.7	Glu72	12.0	His129	7.1
Arg23	>12.0	Arg73	6.1	Arg130	>12.0
Arg29	>12.0	Arg75	>12.0	His133	6.6
Tyr31	>12.0	Asp77	1.3	Arg135	>12.0
Arg42	>12.0	Tyr78	>12.0	Glu140	5.1
Arg44	>12.0	His84	6.1	Tyr146	>12.0
Arg45	>12.0	Arg85	12.0	Asp151	3.3
Glu47	2.1	Glu92	3.8	Glu157	5.0
Arg49	>12.0	His94	5.9	Asp158	1.5
Arg55	3.3	Glu97	4.4	Tyr162	10.8
Asp49	>12.0	***His102***	***10.5***		

Table S14. The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GAT**G$_{PS}$ATC**C -3' system with the H++ web server

Residue	$pK_{1/2}$	Residue	$pK_{1/2}$	Residue	$pK_{1/2}$
Asp7	1.4	Asp55	3.5	Asp104	1.9
Arg8	>12.0	Asp57	2.6	Arg108	>12.0
Glu10	4.0	Glu58	4.2	Glu113	4.3
Asp11	2.6	Lys65	10.5	Arg115	>12.0
Asp14	>12.0	Arg66	>12.0	Glu122	3.5
Asp15	>12.0	His67	6.5	His125	4.7
Arg17	>12.0	Arg70	>12.0	Asp126	2.6
Lys20	>12.0	Glu72	3.1	His129	6.5
Arg23	>12.0	Arg73	>12.0	Arg130	>12.0
Arg29	>12.0	Arg75	>12.0	His133	6.2
Tyr31	>12.0	Asp77	1.9	Arg135	>12.0
Arg42	>12.0	Tyr78	>12.0	Glu140	4.4
Arg44	>12.0	His84	6.9	Tyr146	>12.0
Arg45	>12.0	Arg85	>12.0	Asp151	3.3
Glu47	2.2	Glu92	3.3	Glu157	5.0
Arg49	>12.0	His94	4.8	Asp158	2.5
Arg55	3.5	Glu97	4.1	Tyr162	10.8
Asp49	>12.0	***His102***	***8.3***		

Table S15. The $pK_{1/2}$ values predicted for the titratable residues in the **SBD***Spr*-5'-GGC**G**$_{\mathbf{PS}}$**AAC**G-3' system with the H++ web server

Number	pK(1/2)	Number	pK(1/2)	Number	pK(1/2)
Asp7	2.6	Asp55	2.9	Asp104	1.4
Arg8	>12.0	Asp57	2.6	Arg108	>12.0
Glu10	4.3	Glu58	4.6	Glu113	4.6
Asp11	1.6	Lys65	10.3	Arg115	11.7
Asp14	>12.0	Arg66	>12.0	Glu122	3.2
Asp15	>12.0	His67	6.7	His125	5.3
Arg17	>12.0	Arg70	>12.0	Asp126	1.3
Lys20	10.7	Glu72	3.3	His129	6.8
Arg23	>12.0	Arg73	>12.0	Arg130	>12.0
Arg29	>12.0	Arg75	>12.0	His133	6.6
Tyr31	>12.0	Asp77	1.9	Arg135	>12.0
Arg42	>12.0	Tyr78	>12.0	Glu140	4.3
Arg44	>12.0	His84	6.0	Tyr146	>12.0
Arg45	>12.0	Arg85	>12.0	Asp151	3.4
Glu47	2.2	Glu92	3.4	Glu157	5.1
Arg49	>12.0	His94	5.3	Asp158	1.9
Arg55	2.9	Glu97	4.4	Tyr162	10.7
Asp49	>12.0	***His102***	***10.1***		

SUPPLEMENTARY METHOD

Alchemical Binding Free Energy Calculation details

We explored the alchemical transformations of PT-DNA into normal DNA via sulfur-oxygen swaps. The relative binding free energies were calculated from equilibrium simulations using thermodynamic integration (TI) formulations with Bennett Acceptance Ratio(BAR) and its multistate generalization (MBAR)(1-3). The free energy differences were calculated by gradually perturbing from one to another in a series of discrete steps, represented by λ values. A series of artificial states, parametrized by λ, connecting the unmodified and modified forms, were created by interpolating the force field parameters. We used the notation that λ = 0 corresponds to unmodified states (normal DNA) and λ = 1 corresponds to modified states (PT-DNA). The basic free energy protocol will be a three-step protocol with a decharging step of A, transformation of (partially) discharged A to (partially) discharged B, and finally recharging of B. First, the atoms in the softcore region (those atoms that will transform into dummy atoms) are fully decharged. Next, these decharged atoms undergo a LJ transformation using softcore potentials, while at the same time the charges of the nonsoftcore atoms are also transformed. Last, the atoms in the softcore region are recharged to the final state. This protocol is generally quite robust, since the softcore LJ transformations occur after the partial charges of the softcore atoms have been eliminated. In total, 21 λ-windows for each DNA transformation were performed. Each minimization step consisted of 2000 cycles using the steepest descent method. Afterward, the system was heated from 0 to 300 K gradually over one ns with a coupling restraint of 5 kcal/mol·Å^2 on the solute, followed by equilibration at 300 K using the NPT ensemble for 100 ps with the same restraint. Then another 100 ps of NPT equilibration with a weaker restraint (2 kcal/mol·Å^2) was performed. Finally, the restraint was released and the system was equilibrated using NPT conditions for 200 ps. With these settings, the simulations successfully finished, and the structures appeared fine after visual inspection. Then a 2 ns NPT production run was performed. A time step of 1 fs is used together with the SHAKE. TI calculations in water were performed with explicit solvent (TIP3P, minimum 12 Å to the box side) and under periodic boundary conditions with PME. The conformation of the **SBD**/PT-DNA was taken from the most populated conformation sampled during standard MD simulation. To improve convergence, “soft-core” potentials were applied to the Lennard-Jones and the Coulombic potentials as implemented in AMBER 18. Both the charge and vdW interactions between the disappearing (or appearing) unique atoms with the surrounding atoms were described by softcore potentials. Softcore atoms are treated in a dual topology fashion, while the other atoms are considered in a single topology. Free energy derivatives ($\partial V/\partial\lambda$) were collected independently for each λ from the production run. In the TI method, the free energy difference is calculated from the integral of $\partial V(\lambda)/\partial\lambda$ from 0 to 1, where V is the potential energy.

MM/GBSA binding energy calculation details

The MM/GBSA calculations were performed under IGB=2 and ionic strength of 100 mM, as in the previous literature (4). The binding energies between DNA and protein were calculated with 1000 snapshots extracted from MD trajectories by molecular mechanics/generalized Born surface area (MM/GBSA) approach for binding energy-residue decomposition analysis.

using the following equations:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{DNA}} + G_{\text{Protein}}),$$

$$\Delta G_{\text{bind}} = \Delta H - T \cdot \Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S),$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{vdW}} + \Delta E_{\text{ele}},$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{GB}} + \Delta G_{\text{SA}},$$

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{DNA}} + G_{\text{Protein}}),$$

where ΔE_{int} can be completely canceled because the single trajectory strategy was used for the MM/GBSA calculations. In the current Amber codes, this is taken to be proportional to the total solvent accessible surface area (SA) of the molecule, with a proportionality constant derived from experimental solvation energies of small non-polar molecules, and uses a fast LCPO algorithm: $\Delta G_{SA} = \gamma \times SASA + \beta$, to compute an analytical approximation to the solvent accessible area of the molecule (5). where the surface tension constants γ and β were set to 0.0072 and 0, respectively. The polar part of the solvation energy (ΔGB) was estimated using the Generalized Born (GB) model proposed by Onufriev et al. (igb = 2) (6). The ΔE_{vdw}, ΔE_{ele}, ΔG_{GB}, and ΔG_{SA} terms were computed based on the 1000 snapshots extracted from the last 50 ns MD trajectories. Results from each trajectory were calculated individually and analyzed statistically. In the per-residue energy decomposition analysis, the contribution of each residue was estimated with $i_{decomp} = 2$, by which the 1-4 EEL interaction energies were added to the electrostatic potential term and the 1-4 VDW interaction energies to the van der Waals potential term.

Interaction calculation via SAPT methods

Symmetry-Adapted Perturbation Theory (SAPT) method enables direct computation of interaction energy between monomers. Additionally, SAPT calculations can provide an interaction energy decomposition into four different, physically meaningful terms: electrostatic, exchange, induction, and dispersion. We performed energy decomposition and evaluate the interaction energies applying the SAPT method at the SAPT0/jun-ccpVDZ theory using the PSI4 package (7,8).

Reference:

1. Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* 2008, 129, 124105.

2. Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys*. 1977, 23, 187-199.

3. Shirts, M. R.; Pande, V. S. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys*. 2005, 122, 144107.

4. Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Genet.* 2004, 55, 383-394.

5. Weiser, J.; Shenkin, P.S.; Still, W.C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.*, 1999, *20*, 217-230.

6. Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B,* 2000, 104, 3712-3720.

7. Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* 1994, 94, 1887-1930.

8. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: An open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci*. 2012, 2, 556.