

## Molecular mechanism of methyl-dependent and spatial-specific DNA recognition of c-Jun homodimere

Li-Hua Bie · Jun-Wen Fei · Jun Gao\*

Received: date / Accepted: date

**Abstract** DNA methylation is important in regulation of gene expression and normal development because it alters the interplay between protein and DNA. Experiments have shown that a single 5-methylcytosine at different CpG sites (mCpG) might have different effects on specific recognition, but the atomistic origin and dynamic details are largely unclear. In this work, we investigated the mechanism of monomethylation at different CpG sites in the cognate motif and the cooperativity of full methylation. By constructing four models of c-Jun/Jun protein binding to the 5'-XGAGTCA-3' (X represents C or methylated C) motif, we characterized the dynamics of the contact interface using the all-atom molecular dynamics method. Free energy analysis of MM/GBSA suggests that regardless of whether the  $C^{12}pG^{13}$  site of the bottom strand is methylated, the effects from  $mC^{25}pG^{26}$  of the top strand are dominant and can moderately enhance the binding by  $\sim 31$  kcal/mol, whereas  $mC^{12}pG^{13}$  showed a relatively small contribution, in agreement with the experimental data. Remarkably, we found that this spatial-specific influence was induced by different regulatory rules. The influence of the  $mC^{25}pG^{26}$  site is mainly mediated by steric hindrance. The additional methyl group leads to the conformational changes in nearby residues and triggers an obvious structural bending in the protein, which results in the formation of a new T-ASN-C triad that enhances the specific recognition of TCA half-sites. The substitution of the methyl group at the  $C^{12}pG^{13}$  site of the bottom strand breaks the original H-bonds directly. Such changes in electrostatic interactions also lead to the remote allosteric effects of protein by multifaceted interactions but have negligible contributions to binding. Although these two influence modes are different, they can both fine-tune the local environment, which might produce remote allosteric effects through protein-protein interactions. Further analysis reveals that the discrepancies

---

Li-Hua Bie Jun-Wen Fei and Jun Gao  
Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, People's Republic of China

corresponding author: Jun Gao  
E-mail: gaojun@mail.hzau.edu.cn

---

in these two modes are primarily due to their location. Moreover, when both sites are methylated, the major determinant of binding specificity depends on the context and the location of the methylation site, which is the result of crosstalk and cooperativity.

**Keywords** Protein-DNA interaction · Molecular modelling · DNA methylation · Steric effect

## 1 Introduction

In mammalian DNA, methylation of the fifth position of cytosine (5-methylcytosine, 5mC) is a highly conserved epigenetic modification of DNA that frequently occurs at the CpG dinucleotides (mCpG) and can be read by a set of transcription factors (TFs) known as methyl CpG-binding proteins (MBPs) [44, 62]. Although it represents a small chemical change, this addition of a methyl group to cytosine can have a profound impact on regulation of gene expression and normal development by altering the interplay between protein and DNA [37, 14, 63, 39]. It has been found that in addition to the well-known methyl-CpG-binding domain (MBD) proteins, many TFs, including C2H2 zinc finger (ZF) proteins, basic-helix-loop-helix (bHLH), the basic leucine-zipper (bZIP) homodomain, and tumor suppressor protein (P53) can respond to different cytosine modification states [61, 35, 52].

In recent years, with the combination of ever growing computing power and highly optimized software packages, an array of molecular dynamics (MD) simulation applications have emerged and produced selected new insights on the effects of 5mC in many aspects [59, 30, 24, 58, 28, 48], especially 3D chromatin dynamics and phase separation [34, 33]. These studies bridge the gap between microscopic mechanisms and macroscopic regulation. However, because of limited temporal and spatial resolution and the sequence dependency of the structural diversity of DNA, too few high-resolution structures of corresponding complexes can be obtained. Systematic studies of 5mC using the all-atom molecular dynamic simulation method are scarce, and the exact mechanism by which an individual 5mC functions in different contexts remains largely a mystery.

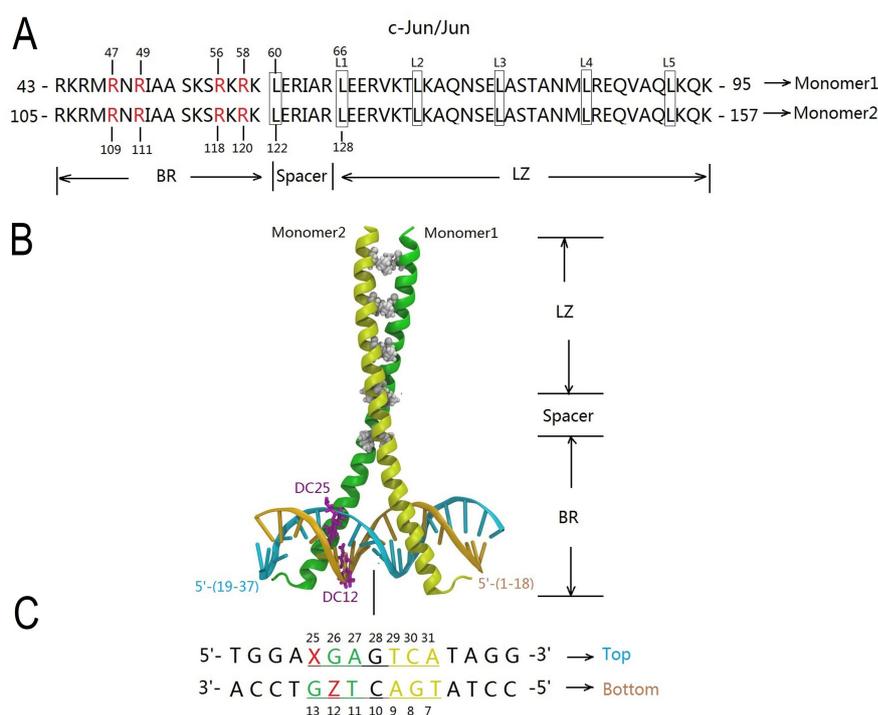
Evidence exists that each transcription factor might work independently, and thus the most important features of a cis-regulatory sequence should be the number of binding sites and how tightly it binds to those sites, which explains the hypomethylation or hypermethylation of DNA methylation observed in cancer [25]. For specificity and affinity of binding, regulation grammar has received extensive attention. It was reported that both the MBD and ZnF binding domain proteins recognized specific bases through classical hydrogen bonds and specific interactions unique to each protein [41, 19, 46, 3, 36]. In particular, it is considered that the bZIP protein AP-1/Zta, uses van der Waals contacts with 5mC to recognize the cognate motif [62, 20]. Certain other studies further proposed that although these proteins have structurally distinct DNA-binding domains, they share a common methyl-Arg-Gua triad mode to interact with 5mCpG and TpG [45, 22, 61, 35, 52]. In particular, the mammalian SRA domain recognizes methylated DNA by flipping 5mC from a hemi-methylated CpG out of the double helix and into a hydrophobic binding cage [48, 41, 19, 46]. Surprisingly, certain experiments have shown that selected mCpGs at different positions of the target motif have different effects on the binding of protein, such as Kaiso, and the major determinants of its remarkable specificity for methylated DNA are the interactions with the particular first single site  $mC^8pG$  in the  $5'C^8GC^{10}G$  motif [45].

Because the methyl group is positioned in the major groove, DNA methylation does not interfere with Watson-Crick base pairing [38], but it can alter the local DNA flexibility or conformation [30, 50, 47, 4], change the stability of nucleosomes and thus affect the local chromatin structure and access of the TFs to genomic DNA.

Computational studies have presented various results, e.g., cytosine methylation either stiffens DNA [51, 43, 10, 40, 57] or has no effect[24] on the structure of DNA, depending on the context. It has even been proposed that flexibility increases upon methylation [59, 58]. Other studies were conducted on the preferences of TFs for methylated over unmethylated CpGs from the perspective of free energy, and the results demonstrate that the preferences are highly variable, ranging from 2-fold to nearly 2 orders of magnitude[35, 45]. In addition, many studies have been conducted on the effect of 5mC on DNA bending, the energy and direction of binding, and the key interaction between proteins and DNA [56, 53].

In previous work, we explored how a single 5-methylation of cytosine regulates C/EBP  $\beta$ -DNA interactions and found that even a single 5mC can regulate protein-DNA recognition via the steric-hindrance effect of the methyl group and change the local atomic environment via the 5mC-Arg-G triad [2]. However, according to Hong's experiment [50], when the homodimer of the c-Jun protein (which is another component of the bZIP family) binds to DNA, the variation of the positions of 5mC in the target motif has a significantly effect on binding affinity. This observation leads us to propose the following questions: What is the reason for the spatial-specific phenomenon of monomethylation found in the experiment? Does the methylation at different sites use the same mechanism as in our previous findings? Can this mechanism be generalized across the bZIP TFs?

To answer these questions, we screened against the Protein Data Bank(PDB) and finally determined the complex of c-Jun homodimer binding to DNA as our target model system to conduct a further investigation. The c-Jun oncoprotein is one member of the AP-1 (activator protein 1) family, which is also an eukaryotic bZIP transcription factor, the simplicity and tractability of which make it an ideal system for molecular design and quantitative analysis of the minimal protein determinants for binding specificity and affinity[26, 32, 31]. It is well known that the c-Jun protein binds DNA in a sequence-specific manner that couples extracellular information in forms of growth factors, cytokines, hormones and stress[56, 1, 55]. and it can unleash its transcriptional activity due to its ability to recognize the pseudopalindromic TGACTCA via homodimerization. The consensus sequences TGACTCA and TGACGTCA are respectively referred to as the TPA (12-O-tetradecanoylphorbol-13-acetate) response element (TRE) and the cAMP response element (CRE) [60, 18, 49]. Experiments have revealed that the first mCpG step on the top strand of the TRE motif (i.e., MGACTCA) can enhance the binding affinity of the c-Jun protein, whereas the corresponding mCpG step on the bottom strand has a negligible effect [61, 1, 20]. Therefore, we extended our study to four models (see next section), involving different methylation sites and used the all-atom MD method to clarify the detailed dynamics.



**Fig. 1** Structure of the c-Jun/Jun-DNA complex. A) Protein sequences of c-Jun/Jun homodimer. The bZIP domain is subdivided into an N-terminal basic region (BR), C-terminal leucine zipper (LZ) and a spacer region (also referred to as the hinge region). The signature leucines are boxed, and the conserved basic residues within the BR subdomains are highlighted in red. B) Crystal structure of complex (PDB ID: 5t01). The repeated leucines are depicted in gray, and the two methylated sites are shown in magenta. C) Nucleotide sequence of dsDNA oligo with the core motif underlined. The variations are labeled as X and Z, indicating C or 5mC.

## 2 Materials and Methods

### 2.1 Setup of the Simulation

The initial structure with a resolution of 1.89 Å was obtained from the Protein Data Bank (PDB ID: 5t01). The structure is a c-Jun homodimer in complex with hemimethylated DNA, as shown in Figure 1A and B. For convenience of description, the strand numbered from 1 to 18 was referred to as the bottom strand, and the strand numbered from 19 to 37 was referred to as the top strand. The positions 25 and 12 were the two methylated sites, which were manually mutated to 5mC by substituting the hydrogen atom of the 5-carbon site of cytosine with a methyl group (see Figure 1C).

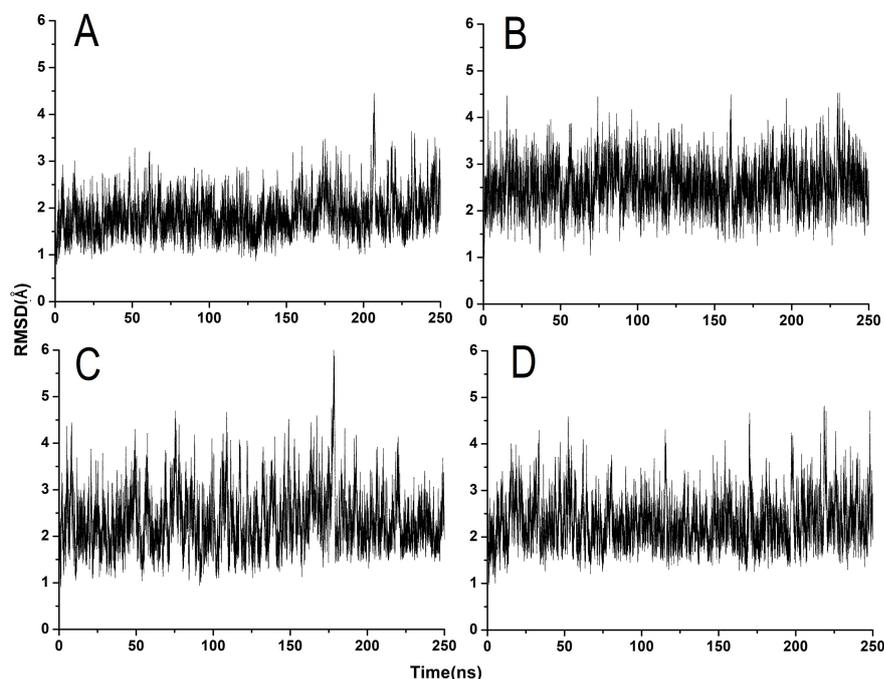
According to the methylation status of the two specific sites (labeled in Figure 1C as X or Z) in the cognate DNA motif binding by c-Jun/Jun protein, four models (namely, C/C, M/C, M/M and C/M) were constructed for comparative study using the

MD simulation method. The initial conformation of the protein-DNA complex of each model was solvated in a cubic box, ensuring a solvent shell of at least 10 Å around the solute. Sodium and chloride ions were added to neutralize the systems and reached a concentration of 150mM. Ultimately, the four solvated models contained 113657 (C/C), 113699 (M/C), 113702 (M/M) and 113699 (C/M) atoms, respectively.

## 2.2 Molecular Dynamic Simulation

All molecular dynamics simulations were performed using the AMBER 16 suite of programs with the ff14SB force field and TIP3P water model[23]. Force field parameters for 5-methylcytosine were obtained from the work of Lankas et al. [29] Electrostatic interactions were evaluated using a cutoff of 10.0 Å, and the particle mesh Ewald (PME) method was used in the long-range electrostatic interaction. Periodic boundary conditions were applied, and SHAKE constraints were applied to all covalent bonds involving hydrogens[9, 11, 54]. The integration time step was set to 2 fs in all MD simulations.

At first, each system was initially subjected to energy minimization for 1000 steps (500 steps of steepest descent followed by a further 500 steps of conjugate gradient) with harmonic restraints on the backbones. Second, the restraints were released, and the systems were further energy-minimized for 2500 steps (1000 steps of steepest descent followed by 1500 steps of conjugate gradient). Thereafter, the systems were heated to 300K during 500 ps with restraints on the backbones using Langevin dynamics, and the pressure was adjusted to 1 atm by the Berendsen weak-coupling algorithm [12]. Each system was subsequently equilibrated for 10 ns in the NVT ensemble without restraints [17, 5, 7]. To avoid a simulation time that was too short to sample the relevant protein changes, and to improve the reliability of molecular dynamics simulation [27], three 250 ns trajectories were replicated under the NPT ensemble as production runs [57]. Thus, the total combined simulation time reached 3 microseconds. For each model, the last three 100 ns parallel trajectories were collected for statistical averaging and further analysis. The backbone atoms C, C $\alpha$ , N and P of complexes were chosen to calculate the root mean square deviation(RMSD).The hydrogen bonds and distance were calculated with the Cpptraj command in AMBER Tools, and the MM/GBSA method was adopted to calculate the binding free energy, [21, 16, 13] and the entropic components (translational, rotational and vibrational) were estimated by Quasi-harmonic entropy approximation implemented in the ptraj module of Amber tools citeRNB4. The residue-based free energy decomposition method was applied to further compute the interactions of protein and DNA in the interface. The typical snapshots were extracted for more intuitive interpretation of our data . which can be easily determined by calculating the RMSDs of a certain duration trajectory, among which, the closest to the average structure was taken as the representative snapshot. VMD software was used to visualize the structures and trajectories [30, 7].



**Fig. 2** Time evolution of the RMSD values investigated from the 250ns trajectories of selected replicated runs. A, B C and D correspond to the C/C, M/C, M/M and M/C models, respectively. The average RMSD and corresponding error bar of each trajectory were provided in supplementary Table S2- S3 and the supplementary Figure S4 provided the RMSD plots of DNA using all the heavy atoms.

### 3 Results and discussion

#### 3.1 Equilibrium of models and binding free energy

The root mean square deviations (RMSD) of the backbone atoms of the twelve 250 ns trajectories were calculated to verify the equilibrium. As shown in Figure 2, the RMSDs of the four models exhibited similar fluctuation characteristics, which was also consistent with our previous finding on the C/EBP  $\beta$ -DNA complex [2]. Specifically, the overall structure appeared to have reached equilibrium, except for selected glitches and periodic transitions. Detailed analysis indicated that these fluctuations derived from the swing of the leucine zipper as well as the fluctuations of selected end residues (Figure S1). From Figure S1AB, we find that the subtle variations of each component accumulated into the hops near 200 ns, whereas no obvious conformation change was found by examining the snapshots of the system. Furthermore, when the RMSDs of the end residues in the disordered region of the protein were removed, the curves became smooth (shown in Figure S2) with average values of RMSD reduced from  $2.05 \pm 0.48$  Å and  $2.05 \pm 0.42$  Å to  $1.81 \pm 0.45$  Å and  $1.85 \pm 0.4$

**Table 1** Binding free energy of four systems under a temperature of 298K. The values of  $K_D$  come from the Hong et al. experiment [20] and the corresponding  $\Delta G^*$  are calculated accordingly. More details on per-residue decomposition and entropy contribution are provided in supplementary Table S4-S7. The units are kcal/mol.

System	$\Delta H$	$T\Delta S$	$\Delta G$ (MM/GBSA)	$K_D$ [20] ( $\mu M$ )	$\Delta G^*$
C/C	-167.24	-141.78	-25.45	$0.47 \pm 0.08$	-8.62
C/M	-168.19	-142.46	-25.73	$0.40 \pm 0.06$	-8.72
M/M	-173.78	-141.79	-31.99	$0.11 \pm 0.02$	-9.48
M/C	-175.96	-144.75	-31.21	$0.10 \pm 0.01$	-9.54

**Table 2** Energy decomposition of key residues in the interface for the M/C and C/M models. All data comes from the average of three repeats. The units are kcal/mol.

Residue	van der Waals		Electrostatic		Polar Solvation		Non-Polar Solv.		TOTAL	
	M/C	C/M	M/C	C/M	M/C	C/M	M/C	C/M	M/C	C/M
ASN48	-3.14	-2.53	-3.30	-1.17	5.00	3.25	-0.62	-0.42	-2.06	-0.86
DC12	-1.45	-2.83	-253.32	-272.02	252.83	273.69	-0.45	-0.73	-2.39	-1.89
DG13	-0.86	-0.70	-227.60	-240.82	226.93	241.34	-0.17	-0.20	-1.70	-0.39
SER55	-1.62	-0.95	-7.94	-6.43	7.56	6.90	-0.27	-0.19	-2.27	-0.66
DC25	-4.02	-1.62	-268.09	-258.29	271.09	260.29	-0.79	-0.47	-1.80	-0.09

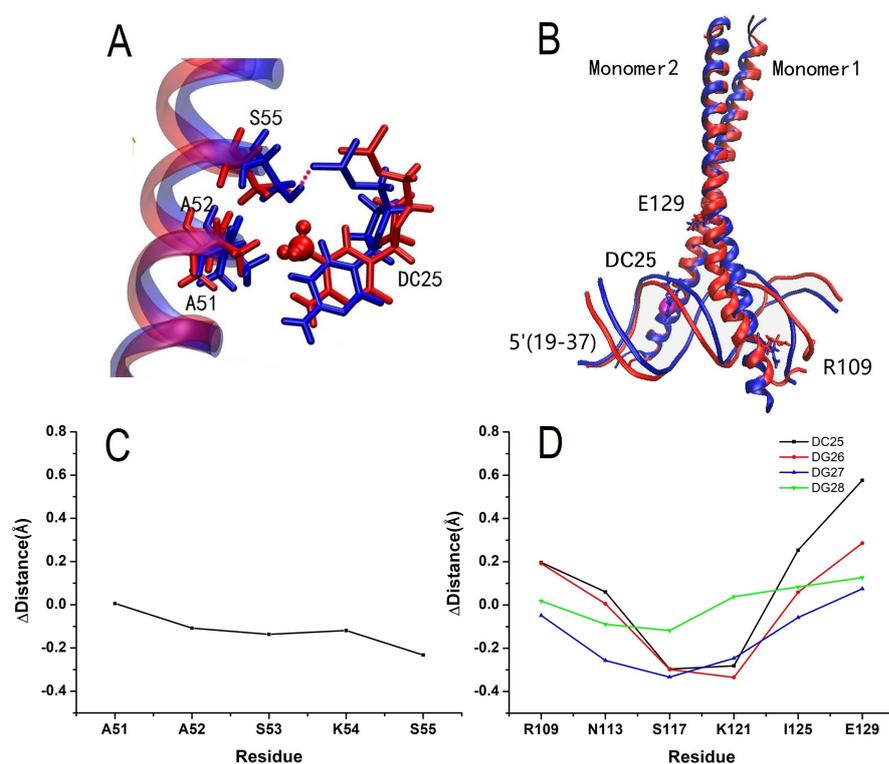
Å, respectively (see Table S2), which indicated that the phenomenon originated from the bias of the end effect. However, the free swing of the flexible protein chain at the zipper region resulted in the periodic wobble of 20 ns shown in Figure S1.C and D due to the lack of considering the cooperation and interactions of the c-Jun homodimer with other proteins. These results revealed the general conformational pattern of bZIP protein-DNA complex. Subsequently, in this work, we focused primarily on the contact interface between the core motif of DNA and the BR region of the protein.

To gain insight into the impact of monomethylation at different locations on the binding of protein, every 10000 snapshots from each of the last three parallel 100 ns trajectories were extracted and calculated using the MM/GBSA method. The binding free energy of the four models were obtained by statistical averaging. Table 1 shows that the binding affinities of the M/M and M/C models were similar, with values near -31 kcal/mol, whereas the C/C and C/M models had an approximate binding free energy of -25 kcal/mol. The former indicates that regardless of whether the  $C^{12}pG^{13}$  site of the lower chain is methylated, methylation of  $C^{25}pG^{26}$  of the upper chain moderately enhances binding. The latter suggests that methylation at  $C^{12}pG^{13}$  site has nearly no effect on binding affinity. According to Hong's experiment[20], when compared with the M/C and M/M models, the C/M and C/C models showed comparable but 4- to 5- fold higher dissociation constants, which meant a  $\sim 3$ - fold loss of affinity. However, because the  $K_D$  values were only  $0.4 \pm 0.06 \mu M$  and  $0.1 \pm 0.01 \mu M$ , respectively, the binding free energies calculated based on these values were notably small, and the difference was within 1 kcal/mol. In this work, the difference quantified by MM/GBSA was 6 kcal/mol, which was within the reasonable error range. Our results confirm the different contributions of monomethylation at different sites to the binding affinity, and are in agreement with the experimental data.

Analysis of Table 1 demonstrates that the free energy contributions of the two mCpG sites have no additivity because free energy calculation is complicated by the competition of enthalpy and entropy, which usually move in opposite directions and compensate for changes to maintain a fairly stable overall free energy of binding. The effects of methylation on protein-DNA interactions cannot be understood merely in terms of their binding affinities[55]. Therefore, a more detailed investigation was performed by decomposing the apparent binding affinities of the four models into their corresponding enthalpy ( $\Delta H$ ) and entropy ( $T\Delta S$ ) components. The data in Table 1 indicate that the interactions are under strong enthalpic control and are accompanied by entropic penalty and by Quasi-harmonic entropy approximation, it suggests that the entropy contributions are mainly derived from vibration and show small difference across the four models ( supplementary Table S6 ). The favorable enthalpy change is most likely due to the formation of hydrogen bonds, hydrophobic contacts and electrostatic interactions between the BR region of the c-Jun/Jun homodimer and its target DNA duplex. This observation is consistent with our previous research[2]. Moreover, we noticed some different variations of the key residues in the interface:Asn48,SER55,DC12,DG13, and DC25 by the analysis of per-residue contributions (Table 2, supplementary Table S4 -S6).Obviously, compared with the C/M system, these residues contribute more to the binding in the M/C system. Wherein, the contribution difference of DC25, ASN48, SER55 and DG13 even exceeds 1.2 kcal/mol. In addition, it is worth noting that the contribution of DC25 increased from -0.09 to -1.8 kcal/mol after methylation, while in the C/M system, the methylation of DC12 reduced the contribution by 0.5kcal/mol, which implies that the mono-methylation at different sites seems to have opposite effects. These results are consistent with the hydrogen bond analysis which we will discuss in the subsequent sections.To further determine the influence mechanism of different mCpG sites, the production trajectories and the typical snapshots of each model were analyzed. Interestingly, the data show that the two mCpG sites adopt different interaction modes to exert their effects. The detailed conformational changes are discussed in the following sections.

### 3.2 Steric effect of methyl group drives conformational change in M/C model

The structures of the C/C and M/C models were compared to investigate the impact of mC<sup>25</sup>G on the local atomic environment. As presented in Figure 3A, the additional methyl group led to conformational changes of the surrounding residues. The side chain of serine S55 was orientated and obviously deflected away from DC25, thus resulting in an elongated distance between the hydroxyl of S55 and the backbone of DC25, which also indicated weakened interactions between DC25 and S55. Interestingly, the distances of backbone between the P atom of DNA and the C $\alpha$  atom of the protein decreased slightly within 0.25 Å (Figure 3C). This observation suggests that as the backbone of protein approaches, the side chains in the local region adopt significant adjustments. At the same time, it is surprising that there appears to be a bending in the BR region of monomer 2(Figure 3B). To verify this hypothesis, we calculated the distance variations in this region displayed in Figure 3D.As expected,



**Fig. 3** Local conformational changes and remote bending in protein by  $mC^{25}pG$ . A) Superposition of C/C(blue) and M/C(red) showing local adjustment of the protein side chain. The distance between the hydroxyl and P atom of DC25 is indicated by a red dashed line, and the methyl group is depicted by a red ball-stick model. B) Bending in the BR region of monomer2. C) Distance variations related to the local adjustment shown in A). The data come from the differences in distance between the P atom of DC25 and the C $\alpha$  atom of adjacent residues A51 to S55 compared with the C/C system. D) Distance variations related to the bending region of Monomer 2 shown in B). All data stems from the last three parallel 100 ns trajectories of system C/C and M/C.

it can be observed that the curves increased in both ends but decreased in the middle, suggesting that the two ends of the BR region withdrew while the middle approached the DNA backbone. Obviously, this distribution displayed a consistent bending trend and the variations ranged from  $-0.4 \text{ \AA}$  to  $0.6 \text{ \AA}$ , which confirmed the conjecture on bending.

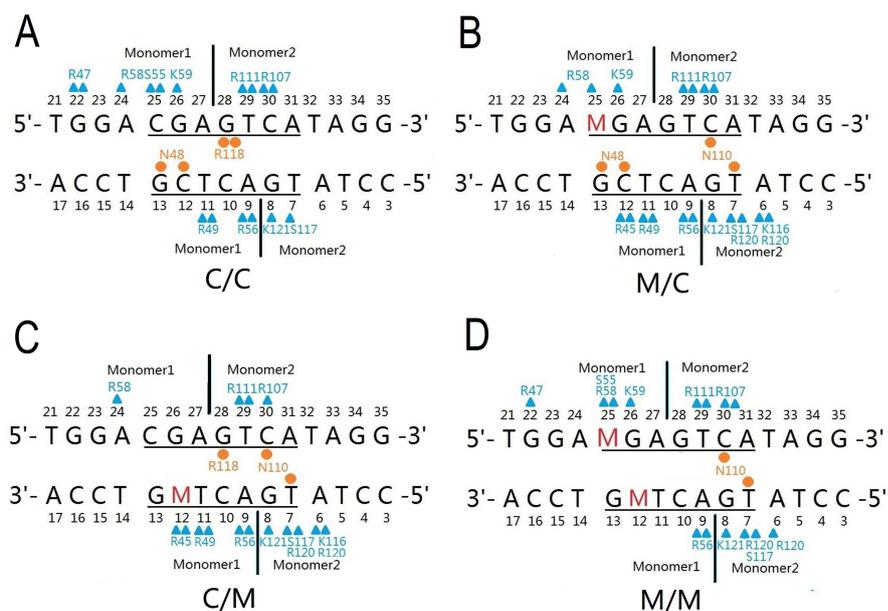
Furthermore, we determined the degree of bifurcation in the spacer region, which is generally considered to affect DNA binding ability due to its crucial role in bridging the two monomers and interacting with DNA. We calculated the average distance between two DNA binding segments, namely, the fifteen amino acids(residues 45-59/107-121) passing through the major groove of the cognate DNA in each basic region [6]. The results showed that the distance decreased by  $0.1 \text{ \AA}$  upon  $mC^{25}pG^{26}$  on the top strand (Table S1). Although subtle, this decrease can contribute significantly

to binding specificity, and in this case, it resulted in large distal changes, especially the structural bend of the protein. This result also evidenced the proximity of the BR region and the enhanced interactions between protein and DNA.

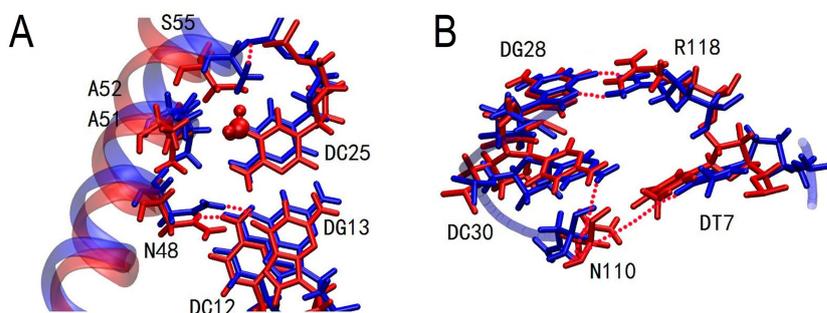
The question arises as to what drove these small local changes and thus result in the long-range allosteric effect in monomer 2? To explore this internal drive force, the hydrogen bonds (H-bonds) in the protein-DNA interface were calculated and are shown in Figure 4. For each system, 1000 snapshots were extracted from the last three replicated 100 ns trajectories to statistically obtain the changes of the H-bonds. The geometric criteria were applied with an angle cutoff of  $\geq 135^\circ$  and a distance cutoff of  $\leq 3.0 \text{ \AA}$ . For clarity, only the H-bonds were present greater than 40% were considered. It is generally accepted that specific recognition is primarily supplied by the interactions between bases and basic amino acids, whereas hydrogen bonds between the protein and DNA backbone atoms stabilize the protein-DNA interface, with the major contribution from the basic long chain residues including arginine and lysine. By comparing the H-bonds in Figure 4A and Figure 4B, we find that  $mC^{25}pG^{26}$  triggered a series of adjustments. First, the backbone H-bond between the methylation DC25 and serine 55 disappeared, which implied that the single methylation at  $C^{25}pG^{26}$  impaired the backbone interaction between DNA and protein. Moreover, it is significant that the specific recognition of H-bonds between R118 and DG28 disappeared but formed a new DC30-N110-DT7 triad mode in the M/C system. In contrast, the bidentate H-bond between asparagine(N48) and cytosine(C12) and guanine(G13) remained stable, indicating that the specific recognition of the TCG half-site remained uninfluenced upon  $mC^{25}pG^{26}$ .

We turned to extraction of the typical snapshots and clarification of the changes in the local atomic environment. As displayed in Figure 5A, the side chain of S55 moved away from DC25 while the distance between the hydroxyl and P atom of DC25 increased from  $1.92 \text{ \AA}$  to approximately  $3 \text{ \AA}$ , which was in accordance with the previous analysis on the breaking of this backbone H-bond. However, the bidentate hydrogen bonds between N48 and DC12 and DG13 were not influenced, and the related distances marked by red dashed line remained at  $\sim 1.75 \text{ \AA}$  and  $\sim 2.0 \text{ \AA}$ , respectively. We found that these interactions can remain unchanged because of the presence of residues A51 and A52. Although these two highly conserved residues did not supply the specific H-bond interactions with bases, they can tolerate the insertion of a methyl group for structural compatibility with the neighborhood. Therefore, the backbone of monomer 1 remained essentially uninfluenced, but the situation was far different in monomer 2, as mentioned above. With the bending in the BR region, R118 deviated from the major groove and destroyed the original hydrogen bond (Figure 5B), and simultaneously, N110 inserted deeper and formed the new triad mode with DC30 and DT7. The residues nearby, including R116 and R120, formed new additional interactions with the negatively charged DNA backbone. Consequently, the specific recognition was further enhanced, and the variations of the hydrogen bonds and the allosteric effect can be well explained.

From the above observations, the steric hindrance of the methyl group at  $mC^{25}pG^{26}$  resulted in a series of changes and adjustments of the local interactions, and therefore, the specific recognition of TCA half-site was enhanced. Furthermore, considering the crucial location of DC25 in the interface, which is right at the top edge of the ma-

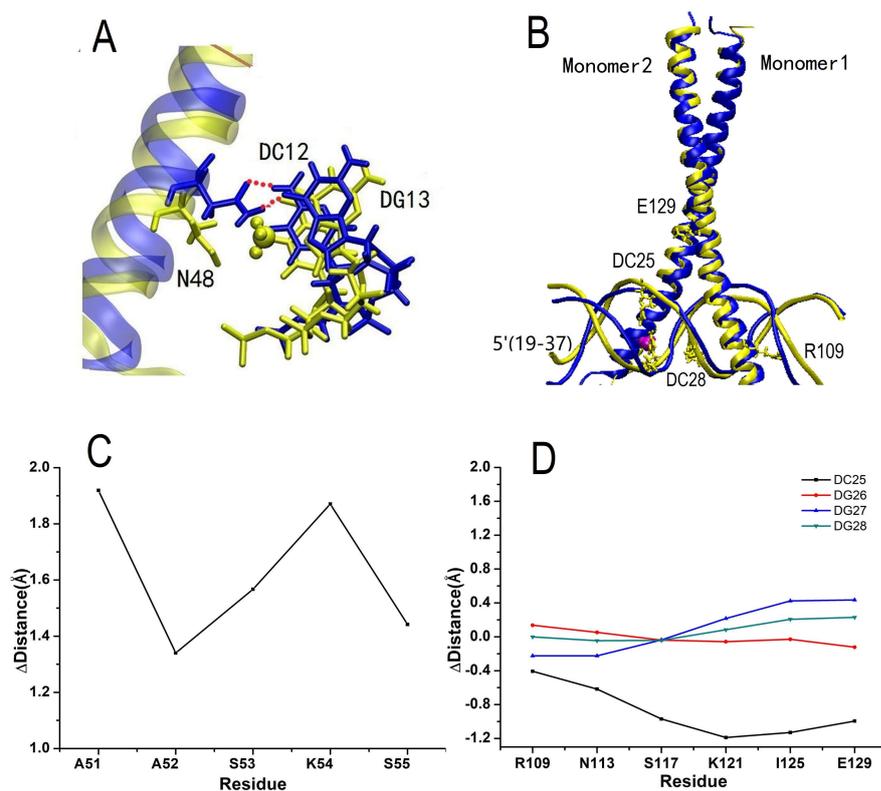


**Fig. 4** Comparison of H-bonds in the protein-DNA interface. A-D) H-bonds in the protein-DNA interface of the C/C, M/C, C/M and M/M systems, respectively. The blue triangles represent backbone H-bonds, and the orange solid circles represent the H-bonds with bases.



**Fig. 5** Superposition of the local atomic environment of C/C (blue) and M/C (red). A) H-bonds change near the  $mC^{25}pG$  site between DNA and monomer 1, with the methyl group depicted in the ball-stick model. B) H-bonds change between DNA and monomer 2. The dashed lines indicate the key H-bonds.

for groove close to the hinge region, it resulted in the long-range allosteric effect of protein through the multifaceted interactions, akin to the ripple effect.



**Fig. 6** Conformational changes in the local atomic environment and in the remote BR region of monomer 2 induced by  $mC^{12}pG$ . A) Superposition of C/C (blue) and C/M (yellow) showing local adjustment of the protein side chain. The H-bonds between N48 and bases DC12 and DG 13 in the C/C model are indicated by red dashed lines, and the methyl group is depicted in the yellow ball-stick model. B) Displacement in the BR region (marked by R109 to E129) of monomer 2, and the magenta sphere represents the methyl group of DC12. C) Distance variations related to the local adjustment shown in A). The figure shows the distance variations of the backbone atoms between DC25 and the surrounding residues N51-S55, taking C/C as reference. D) Distance variations between backbone atoms of DC25-DG28 and residues R105-E129, which correspond to the BR region of monomer 2 shown in B). All data stem from the last three parallel 100 ns trajectories of systems C/C and C/M.

### 3.3 Hydrogen bond interactions drive conformational change in C/M model

When methylation occurred at the  $C^{12}pG^{13}$  step on the bottom strand, a similar adjustment in the local environment and remote structural change of the protein can be found compared with the C/C system. However, surprisingly, the influence is mainly mediated not by the steric hindrance of the methyl group but by the change in the hydrogen bond interactions.

As shown in Figure 6A, residue N48 significantly adjusted its orientation and withdrew from DC12 and DG13. Combined with the previous data shown in Figure

4C, we can confirm that these bidentate hydrogen bonds were directly disrupted due to the substitution of the methyl group at DC12, and because there was no restraint of interactions between N48 and DC12-DG13, the backbone H-bond between DC25 and S55 disappeared accordingly. Additionally, it can be noted that the backbone of monomer 1 in this region displayed an overall displacement, which increased by 1.3 -1.9 Å upon methylation(Figure 6C). Following this displacement, the degree of bifurcation in the spacer region also increased. Even if the average distance of bifurcation increased by less than 0.1Å (Table S1), it led to a shift of the backbone in monomer 2 via protein-protein interactions and showed a subtle bending in the BR region (Figure 6D), implying a series of regulation of interactions with DNA.

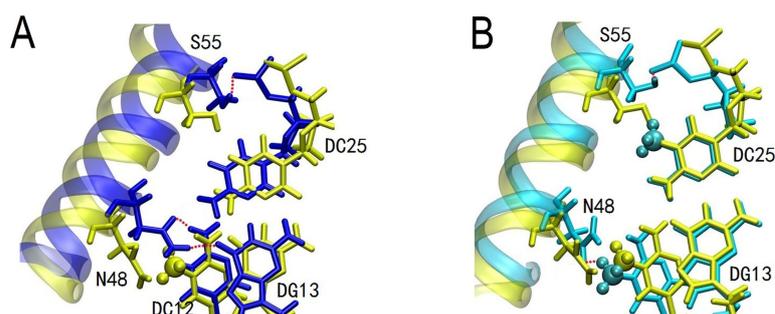
According to the H-bonds data depicted in Figure 4C, the interactions around the methylated DC12 are significantly reduced while the distal BR region in monomer 2 is enhanced, taking the C/C model as the reference. Specifically, the two original hydrogen bonds with N48 that supplied the specific recognition of the TCG half-site disappeared, together with the formation of a new DC30-N110-DT7 triad mode at the other TCA half-site. Among residues R120 and K116 and bases DT7 and DA6, several new backbone H-bonds emerged. In this case, although the  $mC^{12}pG^{13}$  altered the local electrostatic interactions, producing a series of subsequent adjustment, the weakening of local interactions was compensated to a certain extent by the enhancement of distal interactions. From the perspective of free energy analysis (see Table 1), this regulation of interactions by  $mC^{12}pG^{13}$  was negligible, and did not contribute to specific recognition, and compared with the M/C model, it caused nearly a 3-fold loss of binding affinity.

As discussed above, our findings suggest that the major determinants of this remarkable specificity for methylated DNA are spatially specific, and although  $mC^{12}pG^{13}$  and  $mC^{25}pG^{26}$  are both monomethylations, their positions in the interface are different, and their influence mechanisms are also quite different.

#### 4 Cooperativity of steric effect and hydrogen bond interactions in the M/M model

Based on the above results, we dissected the regulation mechanism of the two monomethylation sites at different positions. The results suggest that the protein is more sensitive to  $mC^{25}pG^{26}$  due to its specific location close to the spacer region. The additional bulk methyl group can lead to adjustment of the protein side chain with constraint by the H-bond between N48 and DC12 at the bottom of the groove. However, when DC12 is monomethylated, this restraint is broken. What if DC25 and DC12 are both methylated?

To investigate the cooperativity of full methylation, i.e., DC25 and DC12 are both methylated, we extracted the typical snapshots and constructed the superposition shown in Figure 7. The corresponding distances shown in Figure 8 were calculated using the same method as in the previous analysis. Figure 7A shows the situation in which DC12 is monomethylated. The H-bond interactions between ASN48 and bases DC12 and DG13 were disrupted, thus impairing the specific recognition of the TCG half-site. The H-bond between S55 and DC25 disappeared, but when the symmetrical

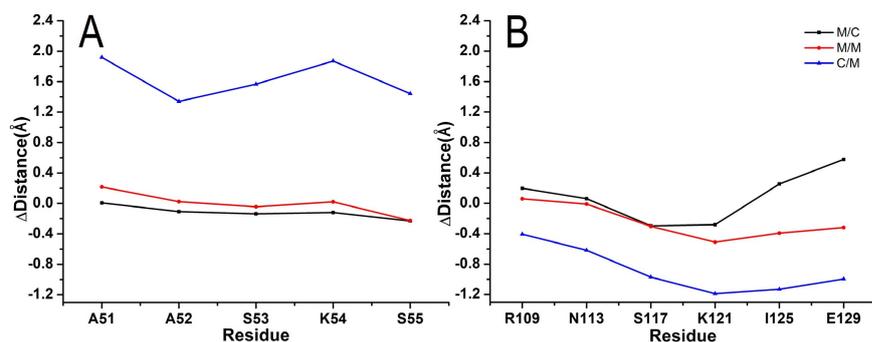


**Fig. 7** Superposition of local atomic environment. A) Alignment of C/C (blue) and C/M (yellow) . B) Alignment of C/M (yellow) and M/M (cyan). The additional methyl groups are displayed by the ball-stick model.

site (DC25) on the top strand was simultaneously methylated (Figure 7B), it reformed again and stabilized the backbone. We attribute this result to the steric hindrance of the bulk methyl group at DC25, which blocks the shift of S55. We can also conclude that the binding in the M/M model is dominant and is stronger than in the C/M model.

Taking the C/C model as the reference, we consider how the cooperativity can influence the variations of the backbone distance between the BR region of the protein and DNA. As depicted in Figure 8A, which is related to monomer1, it can be observed that the fluctuation trends of the M/M and M/C systems were similar, whereas that of C/M was relatively large (Figure S3). Specifically, when monomethylation occurred at DC12, the fluctuation of the backbone can reach  $\sim 2 \text{ \AA}$ , and when DC25 is monomethylated, its change is negligible. When both  $C^{12}pG^{13}$  and  $C^{25}pG^{26}$  were both methylated, we find that the large fluctuation by  $mC^{12}pG^{13}$  was partially offset. If we look at Figure 8B, which displays the displacement of the backbone in the distal BR region of monomer 2, a similar phenomenon can be found. These analyses demonstrate the cooperation of the two interaction modes and indicate that the methylation at  $mC^{25}pG^{26}$  on the top strand is dominant, which can compensate for the unfavorable effect by  $mC^{12}pG^{13}$  and account for the exceptional binding specificity.

To summarize, although the two methylation sites adopt different influence mechanisms to play the roles, their contributions are neither independent nor linearly additive. The two sites undergo crosstalk and cooperation, but which plays the main role depends on its location and the context. In this paper, the effect of  $mC^{25}pG^{26}$  is dominant and can promote specific binding. Because  $mC^{25}$  is closer to the spacer region, it more easily affects the protein-DNA interactions and can even drive long-range allosteric effects. This effect is achieved mainly through steric hindrance of the methyl group. In contrast, although  $mC^{12}pG^{13}$  breaks the H-bond directly, its effects are mostly limited to the local region, and the protein is less sensitive to its regulation because it locates at the bottom of the groove.



**Fig. 8** Comparison of distance changes of the backbone between protein and DNA in the contact interface. A) Distance variations between the P atom of DC25 and the C $\alpha$  atom of adjacent residues A51 to S55 in monomer1. B) Distance variations between the P atom of DC25 and the C $\alpha$  atom of residues R105 to E129 in monomer 2. All of these values were calculated by taking the C/C system as the reference.

#### 4.1 Conclusion

In this work, we explored the molecular mechanism of CpG methylation at different sites in the cognate DNA motif. The result suggests two different influence modes, namely, steric hindrance effect and electrostatic interactions. The effects of methylation are spatially specific. Although methylation does not affect the overall DNA structure, it can fine-tune the local environment, which might trigger remote allosteric effects through protein-protein interactions, depending on the sequence context and the location of the methylation site. When multiple sites are simultaneously methylated, the effect is nonlinear and cooperative. In this case, the dominant influences stem from  $mC^{25}pG^{26}$ , which is consistent with the experimental data. For specific recognition, we did not find the common methyl-Arg-Gua triad mode but the T-ASN-C triad mode, which supplied specific recognition of the TCA half-site, and the recognition of the other TCG half-site was supplied by the bidentate hydrogen bond between N48 and DC12 and G13. From this point, the methylation at this site obviously disrupts the specific binding. Finally, because our object system is composed of two monomers and two DNA strands, protein-DNA interactions and protein-protein interactions occur as well, which adds an additional layer of complexity in determining the TF-DNA interaction dynamics. For example, as we know, the choice of force fields for complex biomolecular systems, especially for protein-DNA system, is not a trivial matter. Notably, due to the good performance of "combined" ff14SB in our previous research, and the length of DNA strand in this work is limited, we did not introduce the updated force field OL15 or bsc1 for DNA [15, 8, 42]. The question whether or to what extent this choice of force field may lead to certain bias for different systems are beyond the scope of this article, and we will explore them in detail in subsequent work.

## Founding

This work was supported by the National Key R & D Program of China [Grant No. 2017YFB0203405], National Natural Science Foundation of China [No. 21873034], Fundamental Research Funds for the Central Universities [Project 2662018JC027 and 2662020XXPY02].

## Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

## Availability of data and material

N/A

## Code availability

N/A

## Authors' contribution

Conceptualization, J.G.; methodology, L.-H. B.; writing original draft L.-H.B.; Writing review and editing, L.-H.B. J.-W. F and J.G.; project administration, J. G.; funding acquisition, J.G. L.-H. B.

## References

1. Bejjani F, Evanno E, Zibara K, Piechaczyk M, Jariel-Encontre I (2019) The ap-1 transcriptional complex: Local switch or remote command? *Biochim Biophys Acta Rev Cancer* 1872(1):11–23, DOI 10.1016/j.bbcan.2019.04.003, URL <https://www.ncbi.nlm.nih.gov/pubmed/31034924>
2. Bie LH, Du L, Yuan Q, Gao J (2018) How a single 5-methylation of cytosine regulates the recognition of c/ebp  $\beta$  transcription factor: a molecular dynamic simulation study. *J Mol Model* 24(7):159, DOI 10.1007/s00894-018-3678-8, URL <https://www.ncbi.nlm.nih.gov/pubmed/29892907>
3. Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, Wright PE (2012) Molecular basis for recognition of methylated and specific dna sequences by the zinc finger protein Kaiso. *Proc Natl Acad Sci U S A* 109(38):15229–34, DOI 10.1073/pnas.1213726109, URL <https://www.ncbi.nlm.nih.gov/pubmed/22949637>

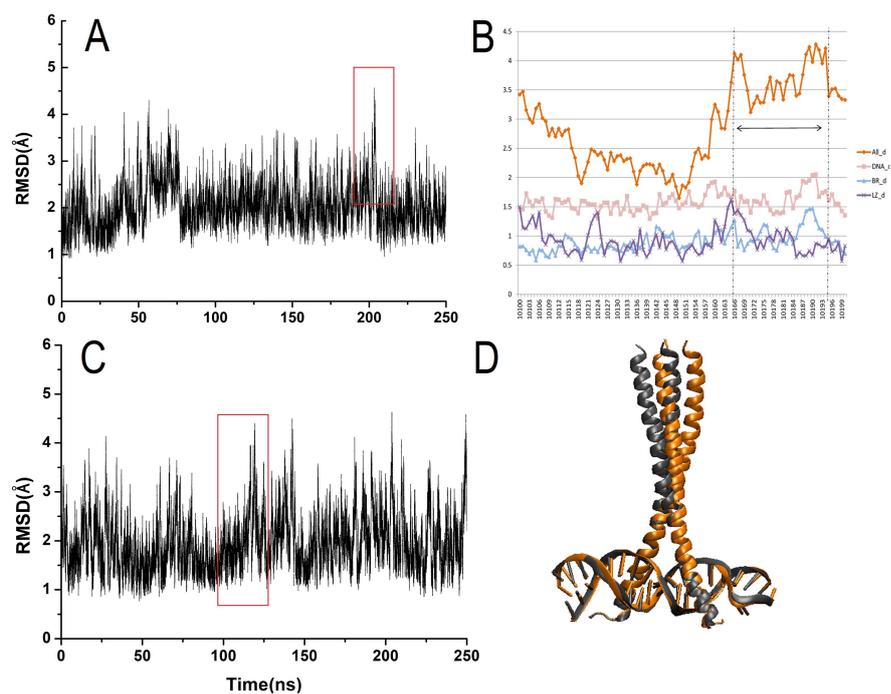
4. Carvalho ATP, Gouveia L, Kanna CR, Warmlander SKTS, Platts JA, Kamerlin SCL (2014) Understanding the structural and dynamic consequences of dna epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation. *Epigenetics* 9(12):1604–1612, DOI 10.4161/15592294.2014.988043, URL [;Go to ISI://WOS:000349364000006](#)
5. Chen L, Zheng QC, Zhang HX (2015) Insights into the effects of mutations on cren7-dna binding using molecular dynamics simulations and free energy calculations. *Physical Chemistry Chemical Physics* 17(8):5704–5711, DOI 10.1039/c4cp05413j, URL [;Go to ISI://WOS:000349697200023](#)
6. Choi YH, Yang CH, Kim HW, Jung S (1999) Leucine zipper as a fine tuner for the dna binding; revisited with molecular dynamics simulation of the fos-jun bzip complex. *Bull Korean Chem Soc* 20(11):1319
7. Cooper VR, Thonhauser T, Puzder A, Schroder E, Lundqvist BI, Langreth DC (2008) Stacking interactions and the twist of dna. *J Am Chem Soc* 130(4):1304–8, DOI 10.1021/ja0761941, URL <https://www.ncbi.nlm.nih.gov/pubmed/18163624>
8. Dans PD, Ivani I, Hospital A, Portella G, González C, Orozco M (2017) How accurate are accurate force-fields for B-DNA? *Nucleic Acids Research* 45(7):4217–4230, DOI 10.1093/nar/gkw1355, URL <https://doi.org/10.1093/nar/gkw1355>, <https://academic.oup.com/nar/article-pdf/45/7/4217/16965773/gkw1355.pdf>
9. Darden T, York D, Pedersen L (1993) Particle mesh ewald - an n.log(n) method for ewald sums in large systems. *Journal of Chemical Physics* 98(12):10089–10092, DOI 10.1063/1.464397, URL [;Go to ISI://WOS:A1993LG10100091](#)
10. Derreumaux S, Chaoui M, Tevanian G, Femandjian S (2001) Impact of cpg methylation on structure, dynamics and solvation of camp dna responsive element. *Nucleic Acids Res* 29(11):2314–26, DOI 10.1093/nar/29.11.2314, URL <https://www.ncbi.nlm.nih.gov/pubmed/11376150>
11. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. *Journal of Chemical Physics* 103(19):8577–8593, DOI 10.1063/1.470117, URL [;Go to ISI://WOS:A1995TE36400026](#)
12. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant-pressure molecular-dynamics simulation - the langevin piston method. *Journal of Chemical Physics* 103(11):4613–4621, DOI 10.1063/1.470648, URL [;Go to ISI://WOS:A1995RU11000023](#)
13. Fogolari F, Brigo A, Molinari H (2003) Protocol for mm/pbsa molecular dynamics simulations of proteins. *Biophys J* 85(1):159–66, DOI 10.1016/S0006-3495(03)74462-2, URL <https://www.ncbi.nlm.nih.gov/pubmed/12829472>
14. Fournier A, Sasai N, Nakao M, Defossez PA (2012) The role of methyl-binding proteins in chromatin organization and epigenome maintenance. *Brief Funct Genomics* 11(3):251–64, DOI 10.1093/bfpg/elr040, URL <https://www.ncbi.nlm.nih.gov/pubmed/22184333>
15. Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, Jurečka P, Cheatham TE (2016) Assessing the current state of amber force field modifications for dna. *Journal of Chemical Theory and Computation* 12(8):4114–4127,

- DOI 10.1021/acs.jctc.6b00186
16. Genheden S, Ryde U (2015) The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10(5):449–61, DOI 10.1517/17460441.2015.1032936, URL <https://www.ncbi.nlm.nih.gov/pubmed/25835573>
  17. Gu C, Zhang J, Yang YI, Chen X, Ge H, Sun Y, Su X, Yang L, Xie S, Gao YQ (2015) Dna structural correlation in short and long ranges. *J Phys Chem B* 119(44):13980–90, DOI 10.1021/acs.jpcc.5b06217, URL <https://www.ncbi.nlm.nih.gov/pubmed/26439165>
  18. Gustems M, Woellmer A, Rothbauer U, Eck SH, Wieland T, Lutter D, Hammerschmidt W (2014) c-jun/c-fos heterodimers regulate cellular genes via a newly identified class of methylated dna sequence motifs. *Nucleic Acids Res* 42(5):3059–72, DOI 10.1093/nar/gkt1323, URL <https://www.ncbi.nlm.nih.gov/pubmed/24371273>
  19. Ho KL, Mcnae LW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) Mecp2 binding to dna depends upon hydration at methyl-cpg. *Molecular Cell* 29(4):525–531, DOI 10.1016/j.molcel.2007.12.028, URL ;Go to ISI://WOS:000253693600015
  20. Hong S, Wang D, Horton JR, Zhang X, Speck SH, Blumenthal RM, Cheng X (2017) Methyl-dependent and spatial-specific dna recognition by the orthologous transcription factors human ap-1 and epstein-barr virus zta. *Nucleic Acids Res* 45(5):2503–2515, DOI 10.1093/nar/gkx057, URL <https://www.ncbi.nlm.nih.gov/pubmed/28158710>
  21. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the mm/pbsa and mm/gbsa methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 51(1):69–82, DOI 10.1021/ci100275a, URL <https://www.ncbi.nlm.nih.gov/pubmed/21117705>
  22. Hudson NO, Buck-Koehntop BA (2018) Zinc finger readers of methylated dna. *Molecules* 23(10), DOI 10.3390/molecules23102555, URL <https://www.ncbi.nlm.nih.gov/pubmed/30301273>
  23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79(2):926–935, DOI Doi 10.1063/1.445869, URL ;Go to ISI://WOS:A1983QZ31500046
  24. Karolak A, van der Vaart A (2014) Enhanced sampling simulations of dna step parameters. *J Comput Chem* 35(32):2297–304, DOI 10.1002/jcc.23751, URL <https://www.ncbi.nlm.nih.gov/pubmed/25303338>
  25. King DM, Hong CKY, Shepherdson JL, Granas DM, Maricque BB, Cohen BA (2020) Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. *eLife* 9:e41279, DOI 10.7554/eLife.41279, URL <https://doi.org/10.7554/eLife.41279>
  26. Kise KJ, Shin JA (2001) The contribution of the methyl groups on thymine bases to binding specificity and affinity by alanine-rich mutants of the bzip motif. *Bioorganic & Medicinal Chemistry* 9(9):2485–2491, DOI Doi 10.1016/S0968-0896(01)00216-4, URL ;Go to ISI://WOS:000170819000032

27. Knapp B, Ospina L, Deane CM (2018) Avoiding false positive conclusions in molecular simulation: The importance of replicas. *J Chem Theory Comput* 14(12):6127–6138, DOI 10.1021/acs.jctc.8b00391, URL <https://www.ncbi.nlm.nih.gov/pubmed/30354113>
28. Lankas F, Sponer J, Hobza P, Langowski J (2000) Sequence-dependent elastic properties of dna. *J Mol Biol* 299(3):695–709, DOI 10.1006/jmbi.2000.3781, URL <https://www.ncbi.nlm.nih.gov/pubmed/10835278>
29. Lankas F, Cheatham r Thomas E, Spacková N, Hobza P, Langowski J, Sponer J (2002) Critical effect of the n2 amino group on structure, dynamics, and elasticity of dna polypurine tracts. *Biophysical Journal* 82(5):2592–2609
30. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham r T, Dixit S, Jayaram B, Lankas F, Laughton C, Maddocks JH, Michon A, Osman R, Orozco M, Perez A, Singh T, Spackova N, Sponer J (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in b-dna. *Nucleic Acids Res* 38(1):299–313, DOI 10.1093/nar/gkp834, URL <https://www.ncbi.nlm.nih.gov/pubmed/19850719>
31. Li G, Quan Y, Wang XC, Liu R, Bie LH, Gao J, Zhang HY (2019) Trinucleotide base pair stacking free energy for understanding tf-dna recognition and the functions of snps. *Frontiers in Chemistry* 6:666, DOI 10.3389/fchem.2018.00666
32. Li Q, Xiong L, Gao J, Zhang HY (2019) Molecular mechanisms of the protein-protein interaction-regulated binding specificity of basic-region leucine zipper transcription factors. *J Mol Model* 25(8):246, DOI 10.1007/s00894-019-4138-9, URL <https://www.ncbi.nlm.nih.gov/pubmed/31342181>
33. Liu S, Zhang L, Quan H, Tian H, Meng L, Yang L, Feng H, Gao YQ (2018) From 1d sequence to 3d chromatin dynamics and cellular functions: a phase separation perspective. *Nucleic acids research* 46(18):9367–9383, DOI 10.1093/nar/gky633
34. Liu S, Quan H, Tian H, Zhou R, Yang L, Gao YQ (2020) 1d sequence based 3d chromatin phase separation: Forces, processes, and functions. *Acta Physico-Chimica Sinica* 36(1):1907010–0, DOI 10.3866/pku.Whxb201907010, URL <http://www.whxb.pku.edu.cn/EN/10.3866/PKU.WHXB201907010>
35. Liu Y, Zhang X, Blumenthal RM, Cheng X (2013) A common mode of recognition for methylated cpg. *Trends Biochem Sci* 38(4):177–83, DOI 10.1016/j.tibs.2012.12.005, URL <https://www.ncbi.nlm.nih.gov/pubmed/23352388>
36. Liu YW, Toh H, Sasaki H, Zhang X, Cheng XD (2012) An atomic model of zfp57 recognition of cpg methylation within a specific dna sequence. *Genes & Development* 26(21):2374–2379, DOI 10.1101/gad.202200.112, URL ;Go to ISI;://WOS:000310583900002
37. Luscombe NM, Austin SE, Berman HM, Thornton JM (2000) An overview of the structures of protein-dna complexes. *Genome Biol* 1(1):REVIEWS001, DOI 10.1186/gb-2000-1-1-reviews001, URL <https://www.ncbi.nlm.nih.gov/pubmed/11104519>
38. Ma N, van der Vaart A (2016) Anisotropy of b-dna groove bending. *J Am Chem Soc* 138(31):9951–8, DOI 10.1021/jacs.6b05136, URL <https://www.ncbi.nlm.nih.gov/pubmed/27435602>

39. Machado ACD, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, Rohs R (2015) Evolving insights on how cytosine methylation affects protein-DNA binding. *Briefings in Functional Genomics* 33(sup1):12–12
40. Nathan D, Crothers DM (2002) Bending and flexibility of methylated and unmethylated *ecori* DNA. *J Mol Biol* 316(1):7–17, DOI 10.1006/jmbi.2001.5247, URL <https://www.ncbi.nlm.nih.gov/pubmed/11829499>
41. Neel SJ, Webb HD, Ginder GD, Williams David C (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Research* 39(15):6741–52
42. Nerenberg PS, Head-Gordon T (2018) New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology* 49:129 – 138
43. Ngo TT, Yoo J, Dai Q, Zhang Q, He C, Aksimentiev A, Ha T (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat Commun* 7:10813, DOI 10.1038/ncomms10813, URL <https://www.ncbi.nlm.nih.gov/pubmed/26905257>
44. Nikolova EN, Stanfield RL, Dyson HJ, Wright PE (2018) CH<sub>2</sub> hydrogen bonds mediate highly specific recognition of methylated CpG sites by the zinc finger protein Kaiso. *Biochemistry* 57(14):2109–2120, DOI 10.1021/acs.biochem.8b00065, URL <https://www.ncbi.nlm.nih.gov/pubmed/29546986>
45. Nikolova EN, Stanfield RL, Dyson HJ, Wright PE (2018) CH<sub>2</sub> hydrogen bonds mediate highly specific recognition of methylated CpG sites by the zinc finger protein Kaiso. *Biochemistry* 57(14):2109–2020, DOI 10.1021/acs.biochem.8b00065
46. Ohki I, Shimotake N, Fujita N, Jee J, Ikegami T, Nakao M, Shirakawa M (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* 105(4):487–97, DOI 10.1016/S0092-8674(01)00324-5, URL <https://www.ncbi.nlm.nih.gov/pubmed/11371345>
47. Orozco M, Noy A, Perez A (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr Opin Struct Biol* 18(2):185–93, DOI 10.1016/j.sbi.2008.01.005, URL <https://www.ncbi.nlm.nih.gov/pubmed/18304803>
48. Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheatham r T, Dans PD, Jayaram B, Lankas F, Laughton C, Mitchell J, Osman R, Orozco M, Perez A, Petkeviciute D, Spackova N, Sponer J, Zakrzewska K, Lavery R (2014) muabc: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res* 42(19):12272–83, DOI 10.1093/nar/gku855, URL <https://www.ncbi.nlm.nih.gov/pubmed/25260586>
49. Paulson EJ, Speck SH (1999) Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J Virol* 73(12):9959–68, URL <https://www.ncbi.nlm.nih.gov/pubmed/10559309>
50. Perez A, Lankas F, Luque FJ, Orozco M (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res* 36(7):2379–94, DOI 10.1093/nar/gkn082, URL <https://www.ncbi.nlm.nih.gov/pubmed/18299282>
51. Perez A, Castellazzi CL, Battistini F, Collinet K, Flores O, Deniz O, Ruiz ML, Torrents D, Eritja R, Soler-Lopez M, Orozco M (2012) Impact of methylation on the physical properties of DNA.

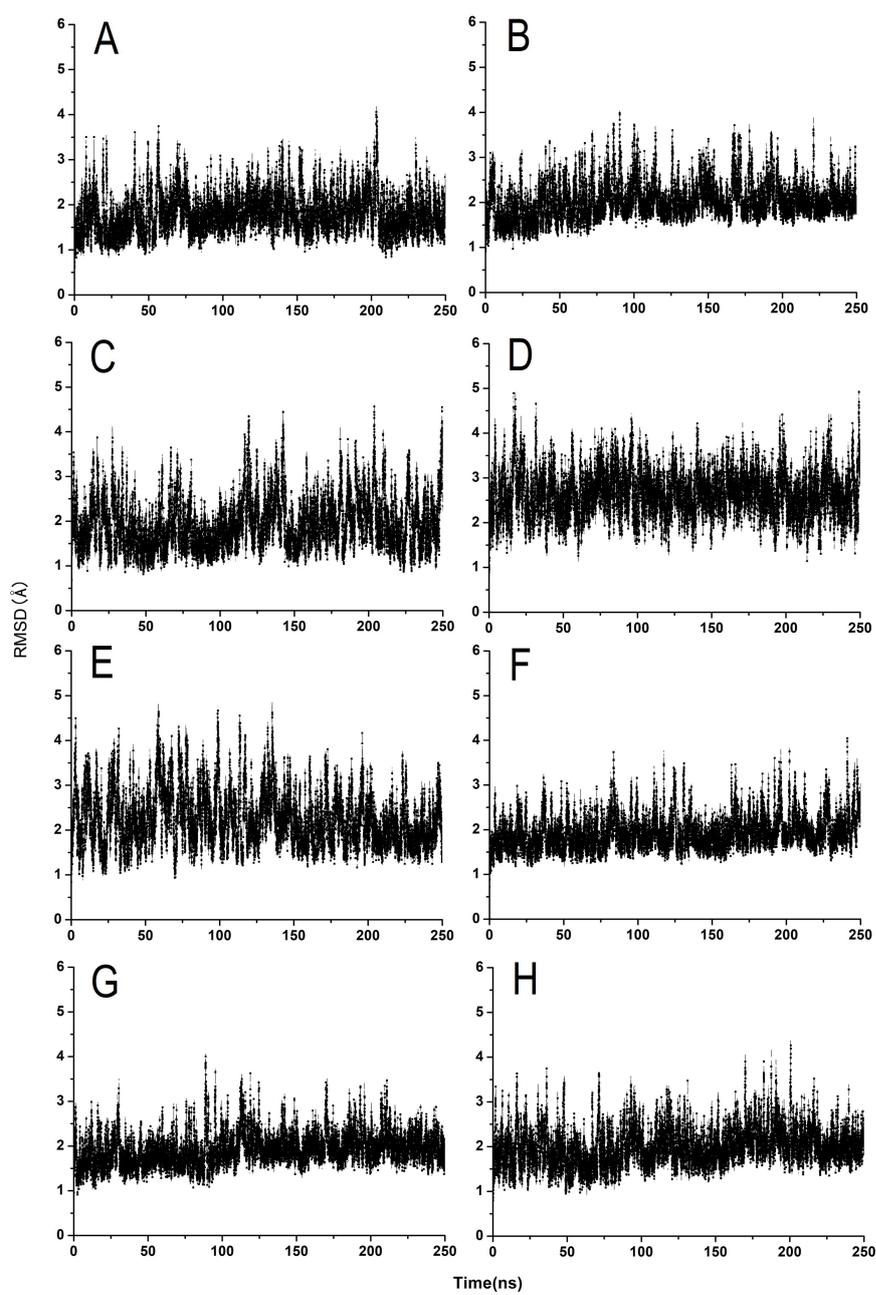
- Biophys J 102(9):2140–8, DOI 10.1016/j.bpj.2012.03.056, URL <https://www.ncbi.nlm.nih.gov/pubmed/22824278>
52. Ren R, Horton JR, Zhang X, Blumenthal RM, Cheng X (2018) Detecting and interpreting dna methylation marks. *Curr Opin Struct Biol* 53:88–99, DOI 10.1016/j.sbi.2018.06.004, URL <https://www.ncbi.nlm.nih.gov/pubmed/30031306>
  53. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B (2009) The role of dna shape in protein-dna recognition. *Nature* 461(7268):1248–53, DOI 10.1038/nature08473, URL <https://www.ncbi.nlm.nih.gov/pubmed/19865164>
  54. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23(3):327–341
  55. Seldeen KL, McDonald CB, Deegan BJ, Farooq A (2008) Evidence that the bzip domains of the jun transcription factor bind to dna as monomers prior to folding and homodimerization. *Arch Biochem Biophys* 480(2):75–84, DOI 10.1016/j.abb.2008.10.010, URL <https://www.ncbi.nlm.nih.gov/pubmed/18940179>
  56. Seldeen KL, McDonald CB, Deegan BJ, Farooq A (2009) Single nucleotide variants of the tgactca motif modulate energetics and orientation of binding of the jun-fos heterodimeric transcription factor. *Biochemistry* 48(9):1975–83, DOI 10.1021/bi802068s, URL <https://www.ncbi.nlm.nih.gov/pubmed/19215067>
  57. Temiz NA, Donohue DE, Bacolla A, Luke BT, Collins JR (2012) The role of methylation in the intrinsic dynamics of b- and z-dna. *PLoS One* 7(4):e35558, DOI 10.1371/journal.pone.0035558, URL <https://www.ncbi.nlm.nih.gov/pubmed/22530050>
  58. Teng X, Hwang W (2016) Elastic energy partitioning in dna deformation and binding to proteins. *ACS Nano* 10(1):170–80, DOI 10.1021/acsnano.5b06863, URL <https://www.ncbi.nlm.nih.gov/pubmed/26638896>
  59. Teng X, Hwang W (2018) Effect of methylation on local mechanics and hydration structure of dna. *Biophys J* 114(8):1791–1803, DOI 10.1016/j.bpj.2018.03.022, URL <https://www.ncbi.nlm.nih.gov/pubmed/29694859>
  60. Tulchinsky EM, Georgiev GP, Lukanidin EM (1996) Novel ap-1 binding site created by dna-methylation. *Oncogene* 12(8):1737–45, URL <https://www.ncbi.nlm.nih.gov/pubmed/8622894>
  61. Yang J, Horton JR, Wang D, Ren R, Li J, Sun D, Huang Y, Zhang X, Blumenthal RM, Cheng X (2019) Structural basis for effects of cpa modifications on c/ebp $\beta$  binding of dna. *Nucleic Acids Res* 47(4):1774–1785, DOI 10.1093/nar/gky1264, URL <https://www.ncbi.nlm.nih.gov/pubmed/30566668>
  62. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C, Taipale J (2017) Impact of cytosine methylation on dna binding specificities of human transcription factors. *Science* 356(6337), DOI 10.1126/science.aaj2239, URL <https://www.ncbi.nlm.nih.gov/pubmed/28473536>



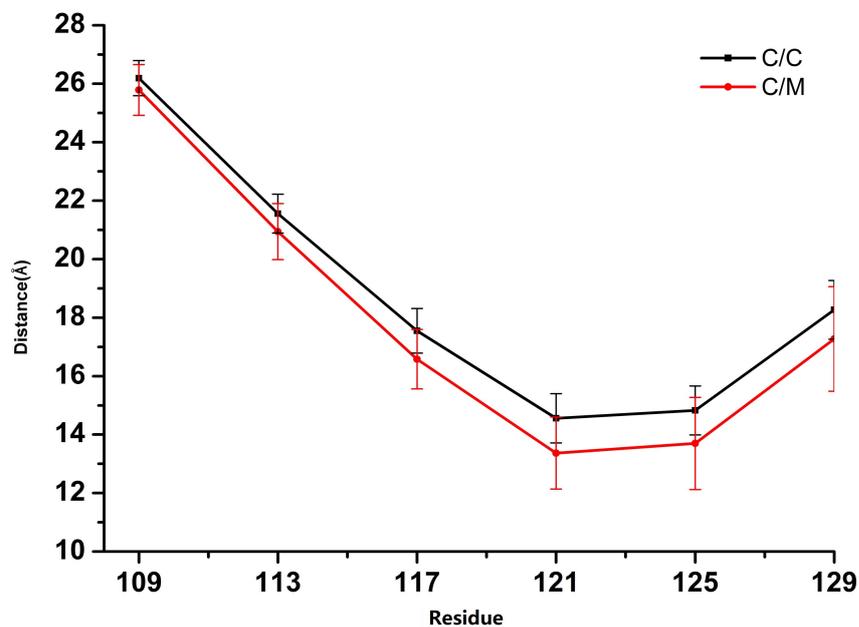
**Fig. S1** General pattern of the RMSD change of the bZIP protein-DNA complex. A) One curve of the RMSDs of the three 250ns trajectories of the C/C model. The jump at 200 ns in the rectangular area is decomposed as described in the right panel. B) One of the RMSDs of one of the three 250ns trajectories of the M/C model. The periodic wobbles of 20 ns originate from the swing of the LZ region described in the right panel.

63. Zhu H, Wang GH, Qian J (2016) Transcription factors as readers and effectors of dna methylation. *Nature Reviews Genetics* 17(9):551–565, DOI 10.1038/nrg.2016.83, URL ;Go to ISI;://WOS:000381510700012

## 5 Supplementary Material



**Fig. S2** Time evolution of the RMSD values investigated from the 250ns trajectories of the other two replicated runs of each model, among which AB, CD, EF AND GH correspond to the C/C, M/C, M/M and M/C models, respectively.



**Fig. S3** Distance values and standard deviations (vertical bars) between the backbone P atom of DC25 and the C $\alpha$  atom of residues 109 to 129 in comparing the C/C system with the C/M system.

**Table S1** Distance variations of bifurcation in the spacer region compared with the C/C model.

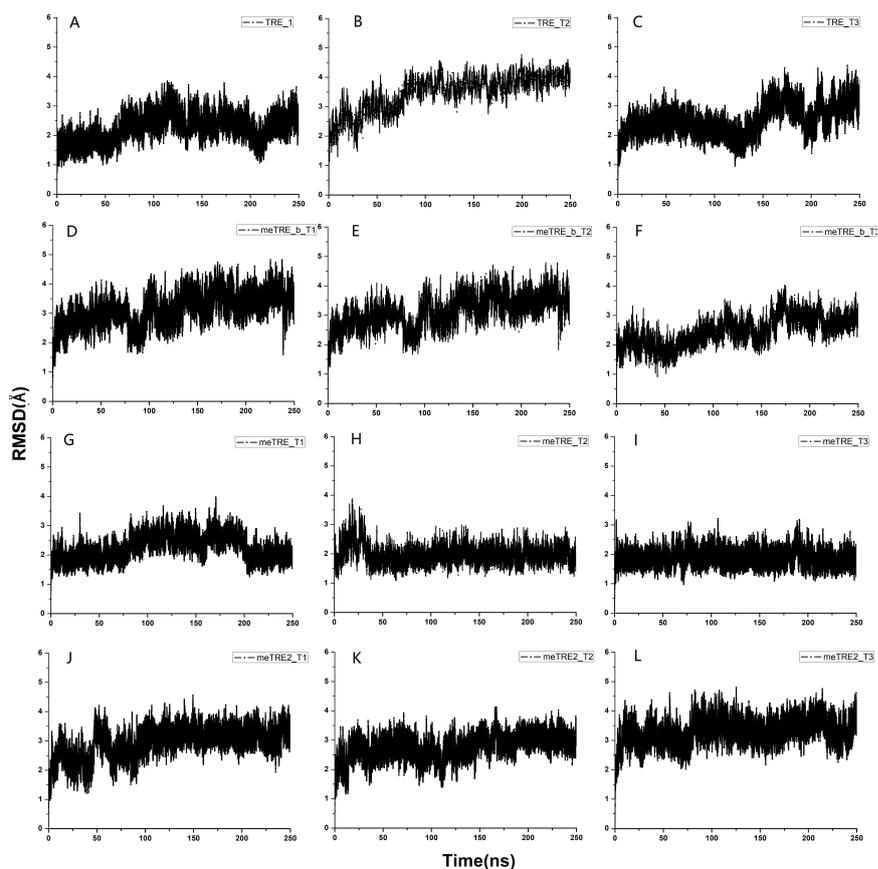
Dsiance	Sapcer region	BR region
$\Delta M/C$	-0.08	-0.09
$\Delta M/M$	0.00	0.28
$\Delta C/M$	0.07	-0.01

**Table S2** Comparison of average RMSD values of four models.

System	P1	Std.	P2	Std.	P3	Std.
C/C	2.054	0.479	2.027	0.397	2.057	0.417
C/M	2.522	0.651	2.741	0.495	2.987	0.562
M/M	2.562	0.604	2.654	0.644	2.663	0.505
M/C	2.513	0.579	2.590	0.544	2.542	0.635

**Table S3** Comparison of average RMSD values of four models with two terminal residues removed. The three trajectories of each system are represented by p1, p2, and p3, respectively.

System	P1	Std.	P2	Std.	P3	Std.
C/C	1.808	0.440	2.027	0.397	1.852	0.405
C/M	1.941	0.566	2.538	0.457	2.643	0.502
M/M	2.244	0.609	2.339	0.600	1.930	0.374
M/C	1.925	0.356	2.313	0.502	1.992	0.415



**Fig. S4** Time evolution of the RMSD values for DNA backbone in the four models investigated from the 250ns trajectories.

**Table S4** Per-residue decomposition of protein in the interface for the four models. All data comes from the average of three repeats. The units are kcal/mol.

Monomer1	C/C	M/C	M/M	C/M	Monomer2	C/C	M/C	M/M	C/M
ARG45	-7.28	-9.18	-6.15	-9.70	ARG107	-9.11	-9.04	-9.43	-8.23
MET46	-0.07	-0.08	-0.13	0.03	MET108	-0.03	0.06	0.03	0.04
ARG47	-10.25	-7.77	-10.54	-8.03	ARG109	-7.28	-4.86	-6.43	-6.82
ASN48	-2.37	-2.06	-0.75	-0.86	ASN110	-1.55	-2.54	-1.96	-2.47
ARG49	-11.58	-11.31	-9.89	-12.70	ARG111	-11.75	-11.97	-12.16	-11.67
ILE50	-0.47	-0.40	-0.49	-0.43	ILE112	-0.41	-0.24	-0.39	-0.29
ALA51	-1.29	-1.30	-1.67	-0.96	ALA113	-1.57	-1.03	-1.53	-1.24
ALA52	-1.79	-1.89	-1.87	-1.51	ALA114	-1.58	-1.83	-1.81	-1.78
SER53	-0.21	-0.23	-0.19	-0.18	SER115	-0.34	-0.23	-0.28	-0.26
LYS54	-2.84	-2.71	-2.55	-3.16	LYS116	-2.75	-3.52	-2.90	-3.29
SER55	-2.35	-2.27	-3.13	-0.66	SER117	-2.89	-4.03	-3.93	-4.04
ARG56	-7.82	-7.69	-7.78	-7.79	ARG118	-3.84	-6.13	-6.38	-4.99
LYS57	-0.74	-0.57	-0.63	-0.56	LYS119	-0.76	-0.60	-0.81	-0.93
ARG58	-9.22	-9.35	-8.16	-8.91	ARG120	-8.03	-10.27	-9.83	-9.67
LYS59	-2.42	-3.25	-3.46	-2.31	LYS121	-3.18	-3.19	-3.44	-3.31

**Table S5** Per-residue decomposition of DNA in the interface for the four models. All data comes from the average of three repeats. The units are kcal/mol.

Top	C/C	M/C	M/M	C/M	Bottom	C/C	M/C	M/M	C/M
DA20	-1.38	-0.46	-1.32	-0.43	DG13	-1.96	-1.70	-0.85	-0.39
DT21	-1.89	-1.05	-4.01	-2.22	DC12	-1.66	-2.39	-0.26	-1.89
DG22	-2.41	-1.80	-1.58	-1.54	DT11	-3.46	-4.49	-1.77	-3.67
DG23	-0.48	-0.78	-0.20	-1.57	DC10	0.57	0.26	-0.50	0.48
DA24	-2.81	-2.76	-1.41	-1.02	DA9	-1.46	-0.82	-1.62	-1.41
DC25	-1.26	-1.80	-2.26	-0.09	DG8	0.14	-2.05	-0.28	-0.34
DG26	0.43	0.24	0.18	0.87	DT7	-3.22	-5.52	-5.58	-5.37
DA27	0.60	-0.58	-0.57	-0.17	DA6	-0.55	-4.27	-2.65	-3.59
DG28	-4.40	-1.88	-2.20	-2.69	DT5	-1.71	-0.94	-1.31	-1.29
DT29	-3.67	-4.17	-4.31	-4.31	DC4	0.04	-0.74	-0.26	-1.18
DC30	-0.99	-1.83	-1.70	-1.65	DC3	-1.18	0.63	-2.01	-1.02
DA31	0.28	0.31	0.11	0.32	DT2	-0.96	-1.63	-1.61	-1.02

**Table S6** The components of entropy contributions. Entropy results are obtained by Quasi-harmonic approximation calculated with pthaj. All data comes from the average of three repeats. The units are kcal/mol and temperature is 298.15 K.

		Translational	Rotational	vibrational	Total
C/C	Complex	16.77	17.88	3488.1	3522.76
	Protein	16.26	17.21	2521.4	2554.88
	DNA	16.03	16.34	1077.29	1109.66
	$\Delta S$	-15.52	-15.67	-110.59	-141.78
M/C	Complex	16.77	17.89	3475.2	3509.87
	Protein	16.26	17.2	2512.3	2545.77
	DNA	16.03	16.34	1076.47	1108.85
	$\Delta S$	-15.52	-15.66	-113.57	-144.75
M/M	Complex	16.77	17.89	3481.22	3515.88
	Protein	16.26	17.21	2523.16	2556.64
	DNA	16.03	16.34	1068.66	1101.04
	$\Delta S$	-15.53	-15.66	-110.6	-141.79
C/M	Complex	16.77	17.88	3489.65	3524.3
	Protein	16.26	17.2	2527.6	2561.06
	DNA	16.03	16.34	1073.32	1105.7
	$\Delta S$	-15.52	-15.66	-111.28	-142.46