Nucleic Acids Research, 2026, **54**, gkag952
https://doi.org/10.1093/nar/gkag952
Gene regulation, Chromatin and Epigenetics

OXFORD

Exploring the bistable equilibrium of methylated CpG DNA recognition by the MBD2 protein

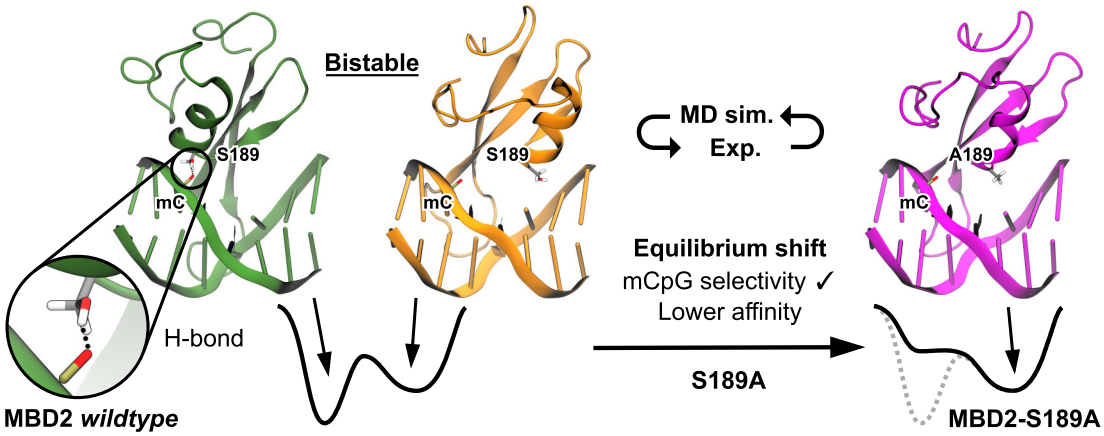
Senta Volkenandt[1], Julia Belyaeva[2], Torry Li[3], Matthias Elgeti[2], Ralf Metzler[4], Petra Imhof[1], David C. Williams, Jr.[3], Mahdi Bagherpoor Helabad[2,5,6,*]

[1]Computer Chemistry Center, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany
[2]Institute for Drug Discovery and Institute for Medical Physics and Biophysics, Leipzig University Medical School, 04103 Leipzig, Germany
[3]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, United States
[4]Institute of Physics and Astronomy, University of Potsdam, Potsdam 14476, Germany
[5]Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany
[6]Institute of Chemistry, Martin Luther-University Halle-Wittenberg, Weinbergweg 22, Halle (Saale) 06120, Germany
*To whom correspondence should be addressed. Email:mahdi.b-helabad@chemie.uni-halle.de

Abstract

Methyl-CpG binding domain 2 (MBD2) is a critical epigenetic regulator that selectively binds methylated CpG dinucleotides, key marks controlling gene regulation and chromatin organization. Understanding the interactions and conformational dynamics underlying this high selectivity is essential to elucidate MBD2's regulatory role. Here, using extensive classical MD simulations, we explored the formation of the MBD2–mCpG recognition complex. By initially positioning MBD2 one base pair downstream of its target, we observed its transition to the target site within microseconds. Notably, upon binding, MBD2 adopts two distinct stable conformations: a primary state closely resembling the X-ray crystal structure, and a secondary state of reduced affinity that nevertheless retains comparable selectivity for mCpG. Our results establish S189 as a key macro-switch; loss of its interaction with the methylcytosine backbone shifts the equilibrium toward the secondary state. This is corroborated by MD simulations of the S189A mutant, which preferentially adopts the secondary state-like conformation. Complementary NMR experiments confirm that S189A mutation does not alter mCpG selectivity, while fluorescence polarization measurements reveal a reduced binding affinity, consistent with our MD simulations results. Together, these findings indicate that MBD2 binding to methylated CpG involves a bistable equilibrium, providing new insights into how high affinity and adaptability are balanced in epigenetic recognition. In a broader context, our findings suggest that such alternative bound-state equilibria may represent an inherent feature of specific protein–DNA complexes.

Graphical abstract



Introduction

In vertebrates, CpG dinucleotides—cytosines preceding guanines—are almost exclusively the sites of cytosine methylation, commonly represented as mCpG. In humans, approximately 75% of CpG sites are methylated [1, 2]. In cells, 5-methylcytosine (5mC) represents a primary epigenetic modification, where a methyl group is added to the 5-carbon position of the cytosine base [3–5]. This modification is involved in processes such as aging [6], embryonic development and differentiation [7, 8], nuclear reprogramming [9], and tissue-specific gene expression [10]. The symmetric nature of methylated cytosines in mCpG provides a distinct

Received: July 2, 2025. **Revised:** August 5, 2026. **Accepted:** August 22, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

conformational signal that allows specific domains of some proteins to selectively recognize it [11]. The main class of proteins that selectively recognize mCpG dinucleotides is the methyl-CpG binding domain (MBD) family. The MBD family, which includes MeCP2, MBD1, MBD2, MBD3, and MBD4, is primarily known for its role in epigenetic regulation and its involvement in signaling events through various protein complexes, with a key function in driving gene repression [12, 13]. The DNA-binding domain, consisting of about 70 residues, is a conserved domain among MBD proteins and exhibits notable structural similarity throughout the family [14, 15]. However, mCpG selectivity depends on the sequence context of individual MBDs. For instance, MBD3, which is highly homologous to MBD2 among MBD proteins, contains a phenylalanine (F34 in MBD3) instead of the highly conserved tyrosine (Y178 in MBD2) found in other MBD family members, resulting in significantly reduced selectivity and binding affinity for mCpG [16, 17]. A pair of conserved arginine residues (R166 and R188 in MBD2), known as arginine fingers, is essential in MBD family members for forming base-specific hydrogen bonds with the symmetric guanine residues within the mCpG dinucleotide and for engaging in cation–π interactions with the corresponding methylcytosines [15, 18].

Among MBD proteins, MBD2 exhibits the highest affinity and selectivity for mCpG sites over non-methylated CpG sites (up to 100-fold) [19–22]. The function of MBD2 has been linked to various conditions, including cancer, immune system dysfunction, and neurodevelopmental disorders [2, 23–25]. In cells, MBD2 is an essential component of the nucleosome remodeling and deacetylase (NuRD) complex, which is responsible for chromatin modification and gene silencing by directing the NuRD complex to mCpG islands [26, 27]. For instance, the role of the MBD2–NuRD complex and DNA methylation in silencing the fetal γ globin gene in human erythroid cell models is well established [28, 29]. Moreover, MBD2, through its function in chromatin modification, strongly mediates the transition of chromatin from an open to a compacted state [27]. Studies have shown that MBD2 can engage mCpG sites across a range of DNA conformations—from linear DNA to bent or nucleosome-bound configurations [18, 22, 30]. Therefore, an important question is how MBD2 adjusts its high-affinity binding to specific DNA target sites in response to variations in DNA geometry.

MBD2's function is closely tied to its structural properties and precise mCpG recognition. In the MBD2–mCpG specific complex, in addition to base-specific interactions involving R166, R188, and Y178, there are other residues in MBD2 that establish hydrogen bonds with the DNA backbone [15, 31]. For instance, residue S189, located in the loop connecting β3 and α1 (Fig. 1A), forms a direct hydrogen bond with the backbone of methylcytosine. Mutational studies on this residue in other MBD proteins, such as MBD1 (S45A) and MeCP2 (S134A), show that this mutation reduces their affinity for mCpG by approximately half [32, 33] and likely leads to altered protein conformation relative to DNA, due to the loss of interactions between the protein loop region and the DNA backbone. Importantly, the S134C mutation in MeCP2 has been shown to be pathogenic, causing Rett syndrome; similarly to S134A, S134C reduces the protein's binding affinity for mCpG—consistent with a loss of function [33–35].

Studies of MBD2 dynamics on DNA have shown that methylation significantly restricts MBD2 diffusion, whereas on non-methylated DNA, it diffuses more rapidly [22]. It has been shown that MBD2 remains bound to mCpG islands and cannot escape for timescales ranging from seconds to minutes [36]. Nevertheless, MBD2 rapidly exchanges between closely spaced mCpG sites on a sub-millisecond timescale, therefore showing flexibility of MBD2 bound to mCpG [22, 37]. For instance, decreasing the mCpG separation from twelve to eight base pairs accelerates MBD2 exchange between the two mCpG sites, shifting it from an intermediate to a fast NMR timescale [36]. In such a short distance between mCpG sites, sliding is the most likely motion enabling such efficient exchange between the two mCpG sites. In the target search process of DNA-binding proteins, three key driving factors are (i) the protein's exploration of nonspecific sites, primarily via sliding and hopping [38], (ii) the transition from nonspecific search to specific recognition, and (iii) the formation of the specific bound state.

Consistent with experimental studies on the kinetics of the target search by DNA-binding proteins [39–44], computational studies have characterized the sliding process on different DNA-binding proteins at atomic and coarse-grained levels [38, 45–56]. However, at the atomic level, the search-to-recognition transition and the formation of the specific complex remain poorly understood. Understanding the molecular-level dynamics of such transitions, by using molecular dynamics (MD) simulations (now feasible at microsecond time scales thanks to powerful computational tools), can help predict specific protein–DNA complex structures that are difficult to resolve experimentally.

Therefore, elucidating the underlying dynamics and binding events governing MBD2's recognition of mCpG sites is crucial for understanding its role in epigenetic regulation.

Here, using extensive MD simulations (in total, 277 μs, Supplementary Table S1), we investigated the recognition of the mCpG dinucleotide by MBD2. Our results reveal that MBD2 adopts two stable binding states with mCpG DNA, consisting of a primary state equivalent to the X-ray-resolved structure and an alternative stable secondary state. The MD results, validated by NMR experiments, reveal that the key residue distinguishing these states is S189, whose interaction with or detachment from the methylated cytosine's backbone directs the MBD2–mCpG complex toward the primary or secondary state, respectively. We suggest that the bistable binding behavior at the mCpG site may serve as a mechanism by which MBD2 optimizes the balance between binding specificity and dynamic adaptability.

Materials and methods

Molecular dynamics simulation

Specific protein–DNA complexes

The initial specific protein–mCpG DNA complex was based on atomic resolution coordinates adapted from an NMR structure (PDB ID: 2KY8 [31]) and an X-ray crystal structure (PDB ID: 7MWK) of the MBD2 protein. In the NMR structure, two EDTA–thymidine residues that are used for paramagnetic resonance enhancement purposes are mutated to standard thymine residues. The flanking sequence to the mCpG in the X-ray structure was also mutated to that of the NMR structure sequence for consistency. Therefore the DNA sequence used in our simulations is 5′-CTGGGAT**mCpG**GCTCAGC-3′. Two termi-

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

nal base pairs are also added in the form of a B-DNA model to the DNA sequence to diminish the terminal effects on the central protein–DNA complex. Simulations of the S189A mutation were conducted starting from the X-ray structure. The central eight nucleotides are indexed as $G_{-4}A_{-3}T_{-2}mC_{-1}G_{+1}G_{+2}C_{+3}T_{+4}$ with the same index numbers in its complementary strand. Starting from the constructed MBD2–DNA specific complex, we performed eight independent 1-μs MD replicas for each system: the X-ray-derived complex (MD runs Ri, $i = 1, \ldots, 8$), the NMR-derived complex (MD runs Ii, $i = 1, \ldots, 8$) and the X-ray-derived S189A mutant complex (MD runs S189A$_i$, $i = 1, \ldots, 8$). Independent replicas were initiated from equilibrated configurations, with initial velocities randomly assigned from the Maxwell–Boltzmann distribution (see Supplementary Table S1).

Nonspecific protein–DNA complexes

We constructed two 18 bps B-DNA molecules with the sequences **DNA 1**: 5′-TCCTGGGAT **mCpG** GCTCAGCC-3′ and **DNA 2**: 5′-AGGGGCmCGG **mCpG** GCTGGCTA-3′ using Chimera [57]. After aligning the central six AT**mCpG**GC base pairs of this B-DNA model with the equivalent region in the NMR structure, we introduced mutations to mimic the effect of 1 bp shifting the DNA sequence downstream, thereby mimicking the condition where the MBD2 protein is located one base pair upstream from the mCpG site. With the MBD2–DNA nonspecific complex constructed, we conducted seven independent replicas with DNA 1 (named Si; $i = \{1, \ldots, 7\}$) and nine independent replicas with DNA 2 (named S$_N i$; $i = \{1, \ldots, 9\}$) of long unbiased MD simulations of ~ 224 μs in total (see Supplementary Table S1). Independent replicas were initiated from equilibrated configurations, with initial velocities randomly assigned from the Maxwell–Boltzmann distribution.

MD simulations protocol

All molecular dynamics simulations were performed with the GROMACS 2018.4 and GROMACS 2022.5 package [58], using AMBER14SB [59] and AMBER parmbsc1 [60] force fields for protein and DNA, respectively. For the methylated cytosine (5mC), we used the combined protocol suggested by Liebl *et al.* [61] in order to increase the compatibility of the whole DNA force field. For the partial charges of the mCpG base atoms, we used the parameters by Carvalho [62]. The systems were solvated with TIP3P water [63] and potassium (K^+) ions were placed randomly to neutralize the system. We also added a salt concentration of 150 mM KCl to mimic physiological conditions. All simulations were initially energy minimized for 5000 steps using the steepest descent algorithm, followed by 100 ps MD simulation in the NVT ensemble with positional restraints of 1 000 kJ mol^{-1} nm^{-2} on all heavy atoms of protein and DNA. Subsequently, positional restraints were gradually lowered to 100 kJ mol^{-1} nm^{-2} in three 100 ps steps, followed by MD production simulations without positional restraints in the NPT ensemble with a 2 fs time step and trajectory output each 20 ps. In order to preserve a base-paired geometry consistent with B-DNA and prevent melting and terminal fraying, we applied 1 000 kJ mol^{-1} nm^{-2} distance restraints between hydrogen bond partners and the N1/N9 atoms of the terminal bases on opposing strands. The v-rescale [64] and Parrinello-Rahman [65] algorithms were used to keep the temperature and pressure at 300 K and 1 bar, respectively. The van der Waals interactions were switched off between 1.1 and 1.2 nm and long-range electrostatics were treated with particle-mesh Ewald [66, 67] and periodic boundary conditions. The covalent bonds of hydrogen atoms were constrained during the simulations via the LINCS algorithm [68].

Biased MD simulations

Setup of the funnel potential

To prevent artificial sliding of the protein along the DNA during steered molecular dynamics (SMD) simulations, we applied a funnel potential using the FUNNEL_PS module in PLUMED, following the official tutorial guidelines [69, 70]. A schematic of the funnel potential configuration is shown in Supplementary Fig. S25A. Defining funnel potential in PLUMED we work with two components of the system: the macromolecule (here, DNA) and the ligand (here, MBD2 protein). The macromolecule serves as a reference for aligning trajectory frames and reconstructing the guiding axis of the funnel during simulation (Supplementary Fig. S25B and C). The reference structure included the polar heavy atoms of the central mCpG base pairs, as well as those from three adjacent base pairs upstream and downstream. The ligand can be specified as the center of mass (COM) of a group of atoms. We defined it as the COM of Cα atoms of the MBD2 protein, excluding N- and C-terminal residues. The direction of the funnel was defined by a line connecting two user-defined points, A and B. In our setup, this axis was aligned with the vector between the COMs of the DNA and the MBD2 protein in the initial frame of each simulation. Parameters MINS and MAXS determine the position of the lower and upper walls of the funnel that are perpendicular to the A/B guiding line of the funnel. The Zcc parameters determine a point, where the cylindrical part transforms to the cone part of the funnel. Parameters MINS, MAXS, and Zcc were set individually for each of three systems via manual inspection using PyMol, version 2.5 Schrödinger, LLC [71]. For all systems, the funnel radius (Rcyl) was set to 0.1 nm, and the cone angle (α) was set to 10°. These values were chosen to allow sufficient freedom for protein rotation and translation during unbinding, while still limiting unphysical lateral movement along the DNA axis.

Steered molecular dynamics

SMD simulations were performed using representative structural models of the WT-primary, WT-secondary, and S189A MBD2–DNA complexes [72]. The system setup and force field parameters were consistent with those used in the long-timescale unbiased molecular dynamics (MD) simulations. For each system, 20 independent SMD simulations were run, each lasting 80 ns. Simulations were performed using GROMACS, version 2023 patched with the PLUMED, version 2.9.0 [58, 69, 73, 74]. The pulling coordinate was defined as the distance between the COMs of selected protein and DNA atoms (denoted as "d_com"). The protein COM was calculated using Cα atoms, excluding flexible N- and C-terminal residues, while the DNA COM was computed using the same atoms included in the alignment reference for the funnel potential. The maximum force constant during pulling was set to 2500 kJ mol^{-1} nm^{-2} for WT-primary state and 2000 kJ mol^{-1} nm^{-2} for both WT-secondary state and S189A. A higher force constant was required for WT-primary due to

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

its stronger binding affinity to DNA. We would like to emphasize that SMD simulations were not used for direct binding free energy calculations but rather to generate smooth unbinding trajectories. The restraint force was initially set to zero and gradually increased to its maximum value over the first 0.5 ns. At 0.5 ns, a harmonic potential with the maximal force constant was applied at d_com = 1.6 nm to stabilize the system before initiating the pulling. The protein was then gradually pulled away from the DNA, increasing d_com to 4.3 nm over the course of 80 ns under constant force. The unbinding trajectories of MBD2 from DNA for each of the three systems were selected in two steps. In the first step, we plotted the reaction coordinate (d_com) and the value of the acting funnel potential over time. For each system, we selected three trajectories in which d_com increased steadily while the funnel potential remained relatively low. This approach minimized the likelihood of artifacts caused by the funnel potential and ensured that protein unbinding proceeded smoothly. In the second step, we visually inspected these selected trajectories using PyMOL. For each system, we chose one trajectory where the protein exhibited minimal or no rapid rotational movements during unbinding. These final three trajectories were used as the basis for Umbrella Sampling simulations.

Umbrella sampling

Umbrella sampling simulations were performed for three MBD2–DNA systems: WT-primary, WT-secondary, and S189A. For each system, a single SMD trajectory obtained in a previous step was used. From each SMD trajectory, 35 frames were selected along the unbinding pathway. The first frame corresponded to the fully bound MBD2–DNA state, while the final frame was defined as the point at which all MBD2–DNA contacts were lost. The remaining frames were uniformly distributed between these two endpoints. Each selected frame served as the starting structure for an independent umbrella sampling window. Each window was simulated for 80 ns, resulting in 35 umbrella sampling simulations per system [69, 75]. This simulation length and the number of sampling windows were sufficient to ensure satisfactory histogram overlap (Supplementary Fig. S27). Because biasing potentials were applied in each umbrella sampling simulation, free energy profiles were reconstructed using the binless WHAM reweighting method (all scripts are provided in the Zenodo repository). Bias contributions from all 35 umbrella sampling windows were included in the reweighting procedure. To assess statistical uncertainty, the 80 ns trajectory from each window was divided into 16 consecutive 5-ns intervals (0–5 ns, 5–10 ns,..., 75–80 ns). For each interval, a separate free energy profile was reconstructed by reweighting the corresponding time segment across all 35 independent umbrella sampling replicas. The first 30 ns of data (intervals 0–5 ns through 25–30 ns) were discarded as equilibration [76, 77]. The median profile and standard error of the mean (SEM) were calculated from the remaining ten intervals (30–35 ns through 75–80 ns) to obtain the final free energy estimate and its uncertainty (Fig. 5E). Convergence of the free energy profiles was evaluated by reconstructing profiles using cumulative trajectory data of increasing length: 0–10 ns, 0–20 ns,..., 0–80 ns. To illustrate convergence in the second half of the simulation, profiles corresponding to 0–40 ns, 0–50 ns, 0–60 ns, 0–70 ns, and 0–80 ns are shown in Supplementary Fig. S28.

Experiment

Protein expression and purification

We expressed and purified HsMBD2MBD(S189A) as described previously [78] with only minor changes. In brief, the isolated MBD(S189A) domain (amino acids 150–214) was cloned into a modified pET32a vector [79] with a TEV protease site immediately preceding the domain. We transformed the resulting plasmid into the Rosetta II (DE3) (Invitrogen) *Escherichia coli* strain, which we grew and induced expression as previously. We purified the expressed fusion protein by nickel affinity chromatography, removed the thioredoxin N-terminal tag by TEV protease cleavage, and further purified the domain by gel filtration chromatography (Superdex-200 26/60, Cytiva).

DNA purification

We purchased complementary 17-base pair oligonucleotides (5′-GAGGCGCT(mC)GGCGGCAG-3′) with a single symmetrically methylated site from Integrated DNA Technologies (IDT). The DNA was annealed and purified by ion exchange chromatography (Source 15Q, Cytiva). We purchased a 3′FAM labeled forward oligonucleotide for fluorescence polarization measurements.

Fluorescence polarization (FP) analysis

We buffer-exchanged protein and DNA samples into 10 mM HEPES pH 7.5, 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, and 0.02% sodium azide. The protein was serially diluted in the presence of 10 nM 3′FAM-labeled DNA in a 96-well plate, and fluorescence polarization was measured on a CLARIOstar microplate reader (BMG Labtech). We fit the resulting data to a general equation for two-state binding

$$[mP] = [mP]_{\mathrm{DNA}} + \frac{([D] + K_D + [P]) - \sqrt{([D] + K_D + [P])^2 - 4[P][D]}}{2[D]} \times ([mP]_{\mathrm{sat}} - [mP]_{\mathrm{DNA}})$$

with Prism graphing software (GraphPad). $[D]$ and $[P]$ are the DNA and protein concentration, respectively; $[mP]_{\mathrm{DNA}}$ and $[mP]_{\mathrm{sat}}$ are the polarization for free and protein saturated DNA, respectively. The serial dilutions and measurements were performed in triplicate, and the experiment was repeated three times.

Nuclear magnetic resonance spectroscopy

^{15}N-labeled protein bound to the methylated double-stranded DNA was buffered exchanged into 10 mM $NaPO_4$, pH 6.5, 0.02% sodium azide, 1 mM dithiothreitol, and 10% ^2H_2O and concentrated to approximately 0.3 mM. NMR spectra were collected on a Bruker Avance III 850 MHz instrument at 25 °C, and data were processed and analyzed with NMRPipe [80] and CcpNmr [81], respectively.

Analyses

Native contact analysis

Native versus non-native interactions between protein and DNA are investigated using the native contact concept, introduced by Best *et al.* [82]. The fraction of native contacts Q can

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

be computed as

$$Q(X) = \frac{1}{N}\sum_{(i,j)} \frac{1}{1+\exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]},$$

where the summation is over all N pairs of native contacts (i, j), and $r_{ij}(X)$ and r_{ij}^0 are the distances between atoms i and j in configuration X and in the reference conformation, respectively. As suggested in [82], we used a smooth parameter β of 5 Å^{-1} and a fluctuation factor λ of 1.8.

Similarity index

We adopted the protocol of Leven *et al.* [45]. The residues forming specific contacts with DNA in the recognition mode were identified from the crystal structure, whereas the search mode was represented by positively charged residues forming a patch on the protein surface. To estimate the overlap between the search and recognition states, we defined χ_i for each protein residue forming specific contacts with DNA in the recognition complex according to the following equation:

$$\chi_i = \frac{\sum_j q_j \exp(-a\frac{r_{ij}}{r_c})}{\sum_j \exp(-a\frac{r_{ij}}{r_c})}.$$

In this definition, j corresponds to all protein residues within a cutoff radius of $r_c = 8$ Å from residue i. r_{ij} denotes the distance between residues i and j, q_j represents the point charge associated with residue j, and $a = 5$ is the exponential factor. The total similarity index χ is defined as the mean of all residue-specific χ_i values and spans from -1 to $+1$, with higher χ values corresponding to lower frustration and a higher degree of similarity between the search and recognition states.

RMSD clustering

The RMSD clustering was performed with the gromos [83] method of GROMACS [58] module `gmx cluster` using each 50th frame and a cut-off of 0.25 nm for WT simulations, and each tenth frame and a cut-off of 0.22 nm for the S189A mutation simulations. Cluster centers, meaning representative structures, are the structures with the minimum average RMSD to all other structures within the respective clusters. Skipped frames have been assigned to the cluster of the frame closest in RMSD taken only into account clusters with at least 0.5 % of all data points. More details about the choice of thresholds can be found in the Supplementary Information.

Principal component analysis

Principal component analysis (PCA) has been performed using GROMACS [58]. With `gmx covar` we first calculated the mass-weighted covariance matrix and its eigenvalues and eigenvectors for the fluctuations of C_α (for protein) and P (for DNA) atoms. Then the trajectory is projected onto these eigenvectors using `gmx aneig`. The cosine content of a principal component was calculated with `gmx analyze` to ensure that it does not represent random diffusion [84].

DNA parameters

The DNA conformation, including the geometrical inter and intra bp parameters, DNA grooves, and backbone dihedral angles were calculated with Curves+/Canal [85, 86]. The DNA backbone BI/BII substates are determined by the ε and ζ dihedral angles, where $\varepsilon - \zeta < 0$ means BI configuration and otherwise BII.

Hydrogen bonds

Hydrogen bonds have been determined based on geometric criteria. The angle between hydrogen, donor and acceptor atom of a hydrogen bond deviated no more than 42° from linearity and the distance between hydrogen and acceptor atom has to be lower than 2.2 Å.

Contact area

The contact area between protein and DNA has been determined based on the solvent accessible surface are (SASA) [87] of the protein (A_{prot}), the DNA (A_{DNA}) and the protein–DNA complex (A_{complex}) as

$$A_{\text{prot-DNA}} = \frac{1}{2}\left(A_{\text{prot}} + A_{\text{DNA}} - A_{\text{complex}}\right).$$

Results

The specific complex exhibits high stability

The highest specificity and affinity of MBD2 for mCpG, compared to other MBD proteins, suggests the formation of a highly stable MBD2–mCpG complex. To evaluate the stability of the MBD2 bound specifically to mCpG, we carried out eight 1 μs MD simulations (R1–R8) starting from the X-ray structure (PDB ID: 7MWK), named recognition simulations. Figure 1A illustrates the MBD2 protein sequence along with the DNA sequence containing mCpG used in this study. RMSD analysis of the recognition simulations reveals that the MBD2–mCpG complex maintains a stable conformation with an average RMSD of 1.87 ± 0.35 Å (for full time series see Supplementary Fig. S1A). Figure 1B depicts the cartoon model alignment of the medoid structure calculated from all recognition simulations with respect to the MD-equilibrated model based on the crystal structure (hereafter, "reference model"), demonstrating high structural similarity with an RMSD of 1.35 Å between the two structures (see Supplementary Table S2). Two key subdomains of the MBD2 DNA-binding domain, here referred to as the head and tail domain (Fig. 1B), establish unique hydrogen bonds with mCpG DNA. Figure 1C–E illustrate the hydrogen bond interactions and their occupancies calculated from the recognition simulations. In the head domain, the two residues R188 and S189 form strong hydrogen bonds with the base of G_{-1} and the backbone of mC_{+1}, respectively (Fig. 1D,E). The formation of the S189–mC_{+1} hydrogen bond also facilitates the moderate interaction of R195 with mC_{+1}. In the tail domain, the key base-specific interaction, illustrated in Fig. 1C, is between R166 and G_{+1}, the complementary base of the methylated cytosine mC_{+1}. Furthermore, tail domain residues L170 and S171 jointly contribute strong hydrogen bonds to the backbone of G_{+1} (Fig. 1C,D). This is also supported by the water-mediated interaction of Y178 with mC_{-1} (illustrated in Supplementary Fig. S2A,B). Outside of mCpG (Fig. 1D), the only moderate base-specific interaction occurs between K174 and G_{+3}. Residue R209 also forms a strong hydrogen bond interaction with the backbone of G_{+3}. These detailed, strong interactions—clearly maintained throughout our recognition simulations—are crucial for stabilizing the specific MBD2–mCpG complex.

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

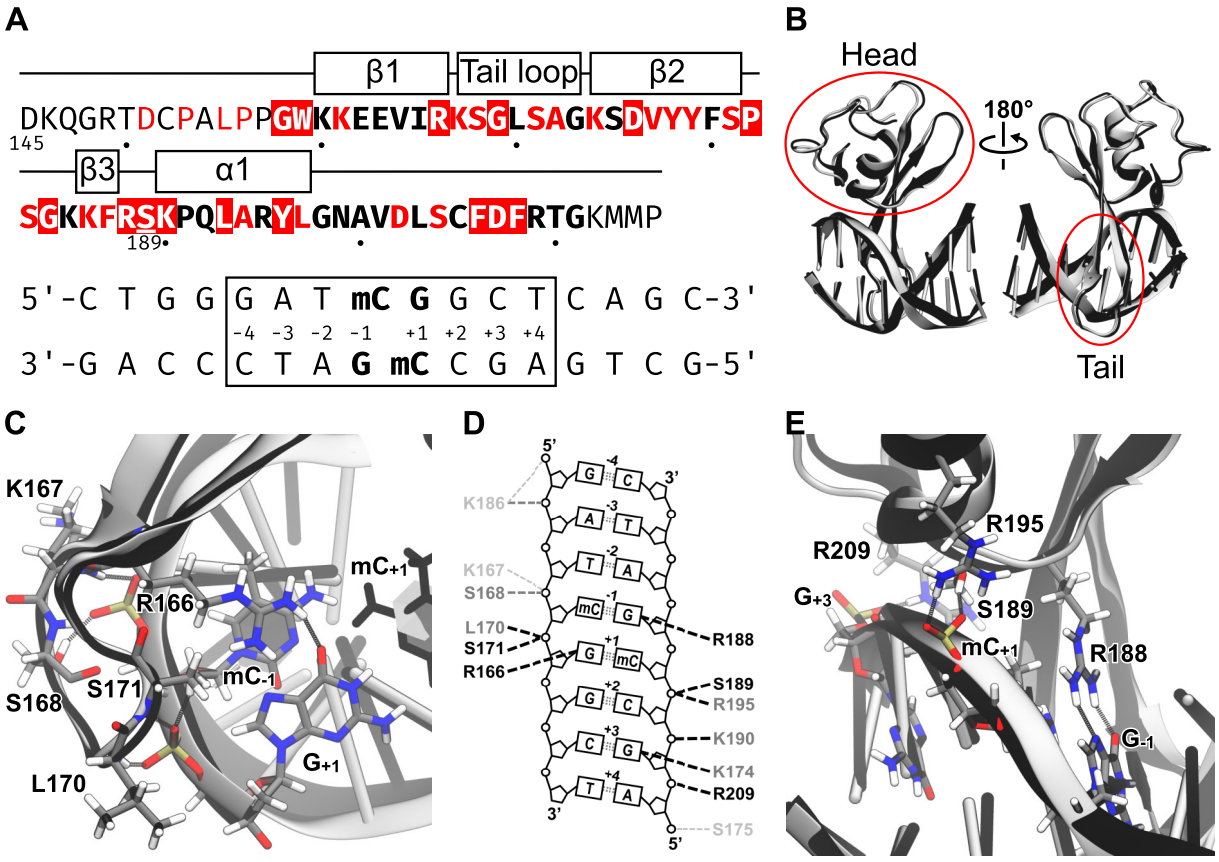


Figure 1. (**A**) Top: Amino acid sequence of MBD2 with the secondary structure indicated on top; highly conserved residues among the MBD proteins are indicated by a red background, mostly conserved residues are written in red and the core sequence for all analyses is bold. Additionally, S189 is highlighted. Dots help in identifying the residue numbers by marking the tens place. Bottom: DNA sequence of the recognition and interrogation simulations with numbering scheme and the mCpG dinucleotide highlighted in bold. Analyzed eight central basepairs are highlighted by a box. (**B**) Medoid structure (black) and reference model (gray) of the recognition simulations show high structural similarity with an RMSD of 1.35 Å. The head and tail domain of MBD2 are indicated. (**C**) Cartoon representation of the medoid structure's tail domain of MBD2 bound to DNA (black), with hydrogen bond interactions indicated. For comparison, the reference model is shown in gray. (**D**) Hydrogen bond interactions between MBD2 and DNA observed in the recognition simulations during the 300–1000 ns interval. The color of each residue indicates the occupancy (strength) of its corresponding hydrogen bond: light gray for [0.25, 0.50) (weak), dark gray for [0.50, 0.75) (moderate, and black for [0.75, 1.00] (strong). The thickness and color of each connecting line represent the relative error: black for [0.00, 0.25], dark gray for (0.25, 0.50], and light gray for (0.50, 0.75]; otherwise, a hydrogen bond is not shown. A square bracket indicates the inclusion of a value, while a round bracket indicates the exclusion, so that the intervals do not overlap. (**E**) Cartoon representation of the medoid structure's head domain of MBD2 bound to DNA (black), with hydrogen bond interactions indicated. For comparison, the reference model is shown in gray.

MBD2 being at the target site does not guarantee recognition

The first deposited PDB structure of the MBD2–mCpG complex was resolved using NMR spectroscopy [31]. However, owing to the low proton density observed in the corresponding NMR study, the DNA in this structure was modeled in part on the B-DNA conformation. As shown in Supplementary Fig. S3, a comparison of the NMR-resolved structures with the X-ray structure reveals differences in the positioning of the head and tail domains as well as the DNA structure relative to the X-ray structure. To assess the influence of these structural variations on the conformational stability of the MBD2–mCpG complex, we performed eight additional MD replicas (I1–I8), initiated from the NMR structure (PDB ID: 2KY8), named interrogation simulations. In contrast to the recognition simulations, RMSD analysis of the interrogation simulations indicates lower stability, with higher RMSD values relative to the reference model (for time series see Supplementary Fig. S1B). However, three out of the eight interrogation MD replicas, i.e. I2, I3, and I7 converged to the reference model based on the X-ray structure (Supplementary Fig. S1B, while the remaining simulations failed to achieve the specific conformation within the 1 μs simulation time span and also remained far from the initial NMR structure, indicating that they were still in a search state (Supplementary Fig. S1B,C). These simulations indicate that mere localization at the target site is not sufficient for formation of the recognition complex; rather, a final transition occurring on a timescale exceeding 1 μs is required to reach the full recognition state.

Fast target localization precedes slow recognition complex formation

After confirming the stability of the MBD2–mCpG recognition complex in recognition simulations and showing that this state is not readily reached from the NMR-based starting structure, we probed target-site search by positioning MBD2 one base pair downstream from its recognition site. This setup enabled us to assess target localization, specific

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

mCpG binding, and the molecular determinants of the transition. We therefore generated a non-native initial complex (Supplementary Fig. S4A) and performed long-timescale MD simulations (seven replicas; see Materials and methods), hereafter referred to as search simulations (S1–S7). Initially, we utilized the similarity index (χ), introduced by Leven *et al.* [45] (see Materials and methods), to quantify the difficulty for a DNA-binding protein in transitioning from its search mode to the recognition mode. The similarity index χ reflects the overlap between the search and recognition states, where higher χ values correspond to lower frustration, stronger similarity, and a reduced free energy barrier, whereas lower χ values indicate the opposite. For the MBD2–mCpG complex, our analysis yielded $\chi = 0.35$, indicating rather similar MBD2's conformation between these states. Our search simulations reveal that the transition of MBD2 from its initial configuration to a position localized on the mCpG dinucleotide occurs rapidly within a sub-microsecond timescale, as illustrated in Fig. 2A by the RMSD evolution from the initial structure, which is also supported by the rotation angle evolution of MBD2 around the DNA (Supplementary Fig. S5). This rapid sliding motion, however, may also be attributed to a DNA conformational shift from the canonical B-DNA form to a protein-induced adapted conformation. Figure 2B shows the RMSD profiles from the search simulations with respect to the reference model. Notably, four search simulations (Fig. 2B, replicas S1, S2, S6, and S7) successfully adopt a structure in agreement with the recognition state, characterized by an RMSD of approximately 2 Å. The observed rotational angle $\Delta\theta_1 \approx 30^\circ$ (Supplementary Fig. S5) suggests an approximately 1 bp rotational movement along the DNA in these simulations. In contrast to the rapid sliding motion, the transition from the localized state to the native recognition state occurs on a timescale of several microseconds, as indicated by the non-highlighted regions of replicas S1, S2, S6, and S7 in Fig. 2B. This suggests that localization of the protein at the specific DNA sequence does not necessarily result in immediate formation of the specific conformation. This is in agreement with the moderate search-to-recognition transition barrier indicated by the similarity index.

An additional binding state emerges from the search simulations

Despite successful formation of the recognition complex in replicas S1, S2, S6, and S7, simulations S3–S5 fail to reach this state, instead stabilizing at an RMSD of $\sim$ 4 Å relative to the reference model (Fig. 2B). To determine whether the MBD2–mCpG complex in these simulations form distinct, well-defined conformations different from the recognition complex, we performed pairwise RMSD clustering across all trajectories from the search, recognition, and interrogation simulations. Additionally, we analyzed the evolution of native contacts (Q) in all simulations with respect to the native complex, i.e. the reference model. Figure 2C displays the native contact profiles mapped onto RMSD-based clusters with distinct regions highlighted by different colors.

Cluster analysis reveals two stable conformations

Our cluster analysis identifies two predominant conformational states associated with stable binding: the primary state, represented by cluster C_1^{wt}, which corresponds to the native recognition complex (depicted in green in Fig. 2C), and the secondary state, represented by cluster C_2^{wt}, which reflects an alternative binding conformation (depicted in orange). The corresponding pairwise RMSD distribution (Supplementary Fig. S6A) exhibit a valley, supporting separated states. A PCA based on Cα and P atom fluctuations provides additional support for the separation into these two states (Supplementary Fig. S7). As shown in Fig. 2C, left, the recognition simulations remain entirely within the primary specific state; with a fraction of native contacts Q close to 1 and RMSD clustering showing a single cluster C_1^{wt}, corresponding to the specific complex. However, only two of the interrogation simulations (I2 and I7) achieve a conformation resembling the specific complex (Fig. 2C, middle). In I3, although the simulation ultimately falls within cluster C_1^{wt}, the native contacts are not completely formed, with $Q \simeq 0.5$. This indicates that the specific conformation can only be achieved when both clustering and native contact conditions are fulfilled. The four replicas I4, I5, I6, and I8 in the interrogation simulations adopt conformations closer to cluster C_2^{wt}. In the search simulations, as noted above, replicas S1, S2, S6, and S7 adopt the primary state, which resembles the specific complex, as indicated in Fig. 2C, right by their assignment to cluster C_1^{wt} and a native contact value (Q) close to one. However, the search replicas S3–S5 predominantly adopt the distinct secondary state, represented by cluster C_2^{wt} (Fig. 2C, right). Notably, the secondary state can persist over multiple microseconds (replicas S3 and S5), thus emphasizing its stable nature.

The secondary state adopts a distinct complex conformation

Here, we examine the characteristics that differentiate the secondary state from the primary state. Figure 3A depicts the representative structures of the primary (green) and secondary states (orange) in cartoon illustration with respect to the reference model (gray). The representative primary state closely resembles the reference model (RMSD of 1.8 Å), whereas the secondary state shows a larger deviation (RMSD of 3.6 Å); the two states differ by 4.1 Å (Supplementary Table S3). The key difference between secondary and primary states (or crystal structure) lies in the distinct positioning of the protein relative to the DNA, primarily due to a shift in MBD2's head domain accompanied by a minor displacement of the tail domain, resulting in a reconfigured protein–DNA interface. In contrast to the primary state, where one base-pair sliding occurs mainly through pure rotation around a defined axis, sliding in the secondary state involves coupled rotation and conformational changes caused by rearrangement of the MBD2 head domain and DNA. Supplementary Figures S8 and S9 show two representative examples (S7 and S5) of such sliding along DNA toward the primary and secondary states, respectively. Interestingly, in S5, MBD2 transiently slides even beyond one base-pair before subsequently relaxing back to the stable secondary state (Supplementary Fig. S9).

To examine the difference between these observed bistable equilibrium in more detail, we calculated the hydrogen bond interaction map for the primary and secondary states, as shown in Fig. 3B. A comparison of the hydrogen bond interactions of the primary state (Fig. 3B, left) and those of the recognition simulations (Fig. 1D) shows a high degree of similarity. The hydrogen bond maps of the primary and secondary state indicate that the three base-specific interactions crucial for the selective recognition of mCpG DNA—R166, K174, and R188—are also retained in the secondary state. However,

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

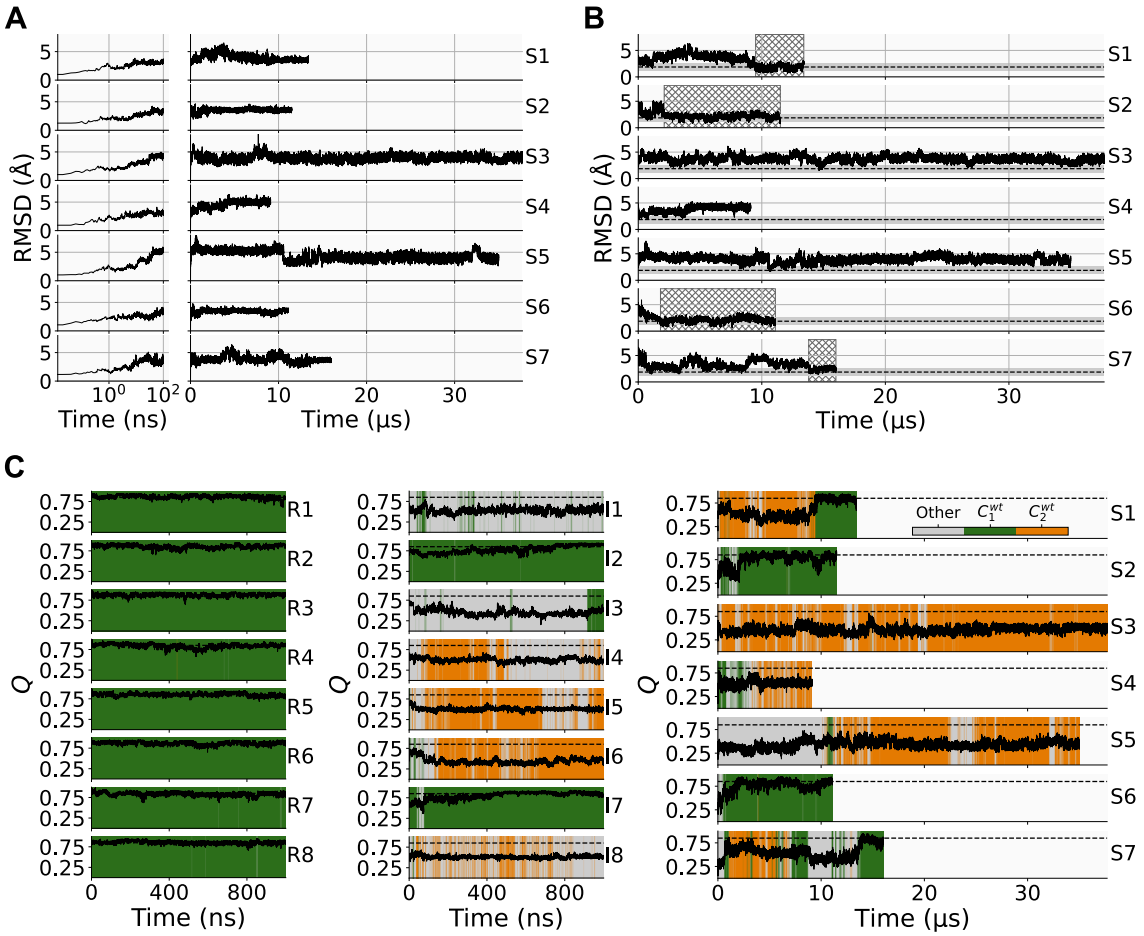


Figure 2. (**A**) RMSD of the search simulations with respect to the first frame. The first 100 ns are plotted additionally on a log-scale. (**B**) RMSD of the search simulations with respect to the reference model. The average RMSD of 1.87 Å (300 – 1000 ns) of the recognition simulations is highlighted as a dashed line, while the gray area around it additionally indicates the fluctuation around that average as two standard deviations. Parts of the trajectory that are within that range are highlighted by a hatched area. (**C**) Clustering based on pairwise RMSD with a threshold of 0.25 nm over time for all recognition, interrogation, and search simulations is shown in green, orange and gray for clusters C_1^{wt}, C_2^{wt} and all others, respectively. The fraction of native contacts Q with respect to the reference model over time is shown as solid black line, while the average fraction of native contacts of the specific complex from the recognition simulations is shown as a black dashed line. Note the different time axis scales.

a slight shift in hydrogen bond interactions is observed toward the downstream region of the DNA, resulting in altered interactions of residues K174 and R188, which establish additional moderate contacts with G_{+2} and T_{-2}, respectively. A key loss of interaction contributing to the formation of the secondary state is the disruption of the hydrogen bond between S189 and mC_{+1}, which remains strong in the primary state (Fig. 3B). The absence of this DNA backbone-mediated interaction is a key factor in the displacement of the MBD2 head domain within the DNA major groove, resulting in formation of the secondary state. Figure 3C depicts the MBD2 head domain in both states, highlighting its conformational shift and the positioning of residue S189 within the DNA-binding interface. As a result of this conformational shift the R209–G_{+3} and K190–C_{+2} hydrogen bonds in the primary state, are rearranged into R209–C_{+2} and K190–mC_{+1} interactions in the secondary state (Fig. 3B,C), respectively, meaning they shift one nucleotide downstream. Additionally, the conformational change of the MBD2 tail domain in the secondary state, compared to the primary state, disrupts the moderate interactions of residues K167, L170, and S171, which are observed only in the primary state (Fig. 3B and structurally shown in Supplementary Fig. S10A,B). Consequently, contact area analysis reveals a larger buried surface at the protein–DNA interface in the primary state than in the secondary state (Supplementary Fig. S11A), consistent with the stronger hydrogen bond interactions observed in the primary state. The secondary state's distinct conformation with respect to the primary state also significantly alters Y178's positioning relative to the mCpG bases. As shown in Fig. 3D and Supplementary Fig. S12 (distance distribution), residue Y178 forms a strong water-mediated hydrogen bond interaction with mC_{-1} in the primary state, whereas in the secondary state this interaction is largely absent, as supported by the distance distributions between the Y178-OH atom and the N4 atom of mC_{-1} and mC_{+1} in Supplementary Fig. S12. The overall MBD2–mCpG contact maps (Supplementary Fig. S13) show distinct contact patterns for the primary and secondary states relative to the recognition simulations, while the primary and recognition conformations differ only subtly.

Moreover, our analysis reveals a significant conformational change in the DNA when comparing the primary and secondary state. In particular, we observe a significant change in the backbone BI/BII conformation, a prominent structural

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

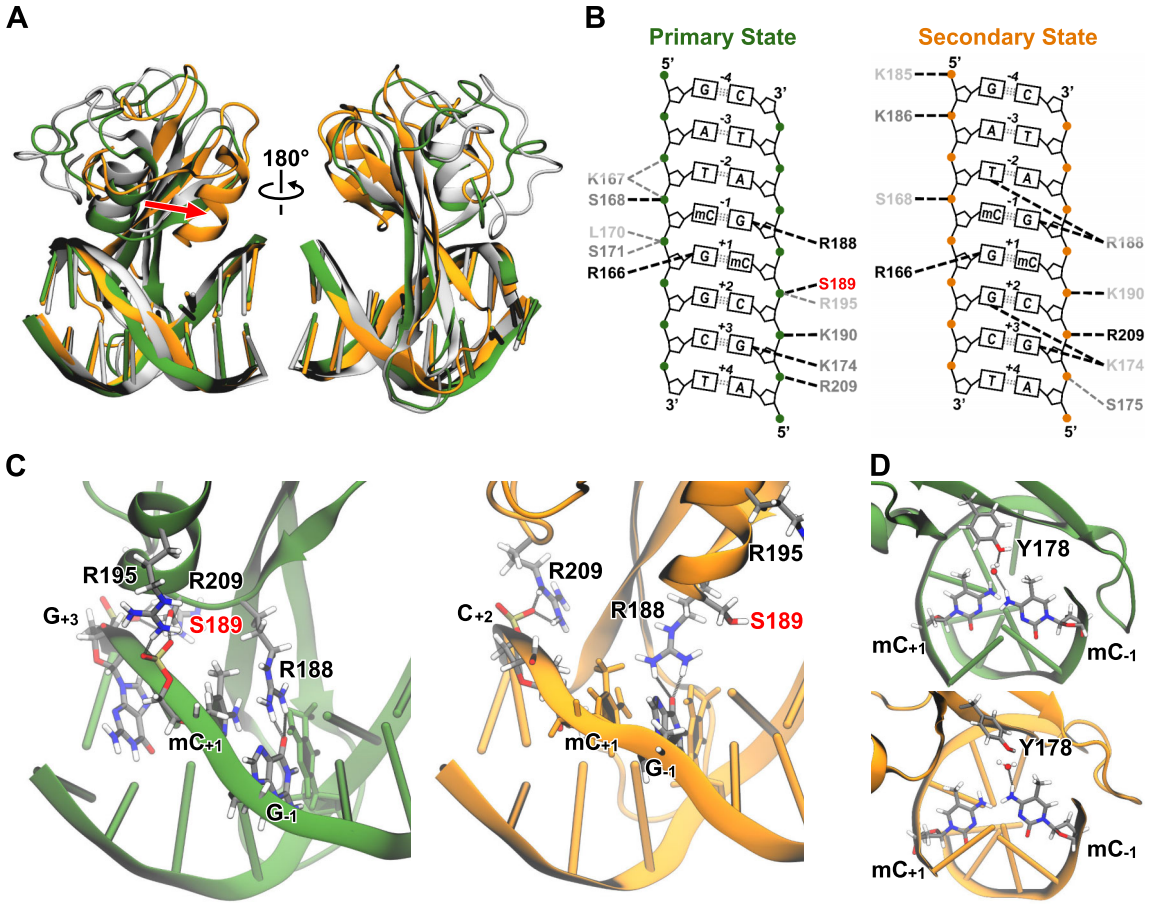


Figure 3. (**A**) Cartoon representation of the cluster centers (green and orange) corresponding to the clustering shown in Fig. 2C and the reference model (gray). A red arrow indicates the shift of the head domain's helix $\alpha 1$ for cluster C_2^{wt} (orange) with respect to the recognition complex. (**B**) Hydrogen bond interactions between protein and DNA for the two main clusters, which we identify as primary (green) and secondary (orange) states. The color of the residue corresponds to the occupancy (strength) of the corresponding hydrogen bond: light gray for [0.25, 0.50) (weak), dark gray for [0.50, 0.75) (moderate), and black for [0.75, 1.00] (strong). The thickness and color of the line on the other hand represents the relative error of that hydrogen bond: black for [0.00, 0.25], dark gray for (0.25, 0.50], and light gray for (0.50, 0.75]; otherwise, a hydrogen bond is not shown. A square bracket indicates the inclusion of a value, while a round bracket indicates the exclusion, so that the intervals do not overlap. (**C**) Cartoon representation of the head domain of MBD2 bound to DNA showing the hydrogen bond interactions for the two cluster centers C_1^{wt} (left, green) and C_2^{wt} (right, orange), corresponding to the primary and secondary states, respectively. (**D**) The absence or presence of the water-mediated hydrogen bonds between Y178 and the base of mC_{-1} in clusters C_1^{wt} (top, green) and C_2^{wt} (bottom, orange), respectively.

parameter that critically influences DNA flexibility and function [88–94]. As shown for G_{-1} in Fig. 4A and for mC_{-1} and mC_{+1} in Supplementary Fig. S14, the secondary state exhibits distinct, well-defined BI/BII backbone conformations, whereas the primary state remains nearly identical to the specific complex (dashed lines). Figure 4B depicts the backbone of G_{-1} in BI and BII conformations, showing that this transition affects not only the base position but also the neighboring mC_{+1}, as reflected by the phosphate-group positioning. In addition, the DNA geometric parameters, including inter- and intra-base-pair as well as groove parameters (Supplementary Figs S15–S17), reveal considerable conformational changes in the DNA in the secondary state compared to the primary state, whereas the primary state displays a geometry closely resembling to that of the specific recognition complex. This is particularly evident in the mCpG dinucleotide and, to some extent, in the nucleotides directly flanking mCpG, as reflected in the intra-base-pair parameters such as opening, propeller, x-, and y-displacement (Supplementary Fig. S15), as well as inter-base-pair parameters such as shift and twist (Supplementary Fig. S16). Notably, the distinct DNA conformation observed in the secondary state, compared to the primary state, extends not only at the base-pair level but also to the backbone, as evidenced by the minor and major groove parameters (Supplementary Fig. S17).

To further validate the secondary state as an alternative sequence-specific bound conformation, we performed additional search simulations using a different DNA sequence containing the native mCpG sites at positions -53 and -50 relative to the transcription start site of the γ-globin gene promoter in adult erythroid cells, where MBD2 functions as a cofactor in gene silencing [29]. The new DNA sequence, labeled DNA 2, together with the previously used sequence (DNA 1), is shown in Fig. 4C. Similar to the previous search simulations, in the new simulations with DNA 2 (9 replicas, each 10 μs), MBD2 was positioned one base pair downstream relative to the mCpG site at position -50 (Fig. 4C, shown in bold). Here, RMSD clustering of the new search simulations (Supplementary Fig. S18A) also reveals two states corresponding to the primary and secondary states observed with DNA 1 (RMSD distances can be found in Supplementary Table S4), with the S189–mC_{+1} interaction re-emerging as

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

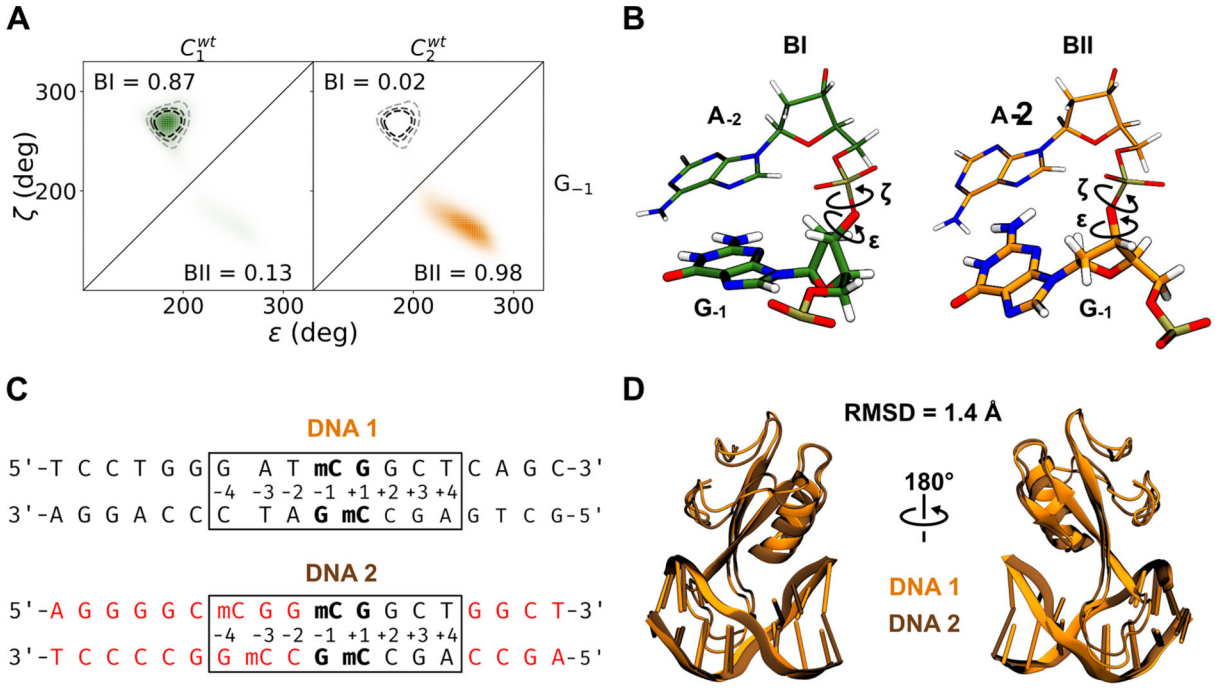


Figure 4. A) Distribution of backbone angles ε and ζ representing the backbone conformation, which is either BI or BII, is shown for the guanine G_{-1} of clusters C_1^{wt} (left, green, primary state) and C_2^{wt} (right, orange, secondary state) as well as for the specific complex (dashed gray lines). (**B**) Representative structures showing the G_{-1} backbone conformation, which is mainly BI for the primary state (left, green) and mainly BII for the secondary state (right, orange). Backbone angles ε and ζ, which determine the backbone conformation, are indicated. (**C**) Comparison of the two DNA sequences used for search simulations, DNA 1 (top) and DNA 2 (bottom). The differences in sequence are shown in red for DNA 2. The central eight basepairs are highlighted by a box and the mCpG dinucleotide in bold. (**D**) Cartoon representation of the cluster centers for the secondary state of search simulations performed on DNA 1 (orange) and DNA 2 (brown) show high structural agrrement.

the key macro-switch. Notably, five out of nine new search replicas converge to the secondary state conformation, which closely recapitulates the secondary state observed in the previous simulations with DNA 1, with an RMSD of 1.4, Å between them, as shown by the superimposed structures in Fig. 4D (for the primary state see Supplementary Fig. S18B,C).

Taken together, our results suggest that the secondary state represents an alternative conformational equilibrium distinct from the primary state while preserving the key structural features required for for the formation of a stable, specific MBD2–mCpG complex. Its recurrence in a second, biologically relevant DNA sequence further substantiates this alternative specific bound conformation.

Validating the secondary state via S189A mutation

S189A mutation shifts the MBD2–mCpG complex toward the secondary state conformation

As demonstrated in the previous section, a critical distinction between the primary and secondary state is the absence of the S189–mC_{+1} hydrogen bond interaction in the secondary state. To determine whether the S189A mutation might shift the binding equilibrium toward the secondary state, we performed eight independent 3 μs MD simulations of the specific complex bearing the S189A mutation based on the 7MWK crystal structure (*S189A simulations*). The fraction of native contacts Q in the S189A simulations (Supplementary Fig. S19A) reveals that the protein–DNA complex gradually deviates from the native specific configuration over the course of the simulation in nearly all trajectories. Performing RMSD clustering on the entire set of S189A simulations, following the same approach as in the wild-type (wt)-simulations, reveals three distinct clusters, as illustrated in Supplementary Fig. S19A. The cluster shown in dark blue (cluster C_1^{mu}) corresponds to a conformation similar to the specific complex (or primary state structure). Figure 5A shows that the cluster center of cluster C_1^{mu} closely resembles the primary state structure (cluster center C_1^{wt}, green) with an RMSD of 1.1 Å. It can be mainly found in the beginning of all S189A replicas. Six out of the eight S189A simulation replicas—M1, M3, M4, and M6–M8 (for cluster time series see Supplementary Fig. S19B)—converge toward cluster C_2^{mu}) which closely resembles the conformation and binding interactions of the secondary state cluster C_2^{wt} (orange) observed in the wt-simulations (Fig. 5B), with an RMSD of 2.2 Å This convergence from the primary toward the secondary state is further supported by the native contact fractions of the S189A replicas relative to both states (Supplementary Fig. S19B); specifically, the native contact fraction decreases over time relative to the primary state while increasing relative to the secondary state. Together, this indicates that the protein–DNA contacts in cluster C_2^{mu} closely align with those observed in the secondary-state complex. The other two replicas, M2 and M5, adopt an intermediate conformation C_3^{mu}, where the head domain positioning on DNA resembles C_2^{mu}, while the tail domain remains similar to cluster C_1^{mu}, see Supplementary Fig. S20B. However, the DNA structure in C_3^{mu} adopts a conformation similar to that of cluster C_2^{mu}, as evidenced by DNA geometry parameters (Supplementary Figs S21–S23) as well as the BI/BII conformation (Supplementary Fig. S24) Taken together, our simulation results demonstrate that the S189A

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

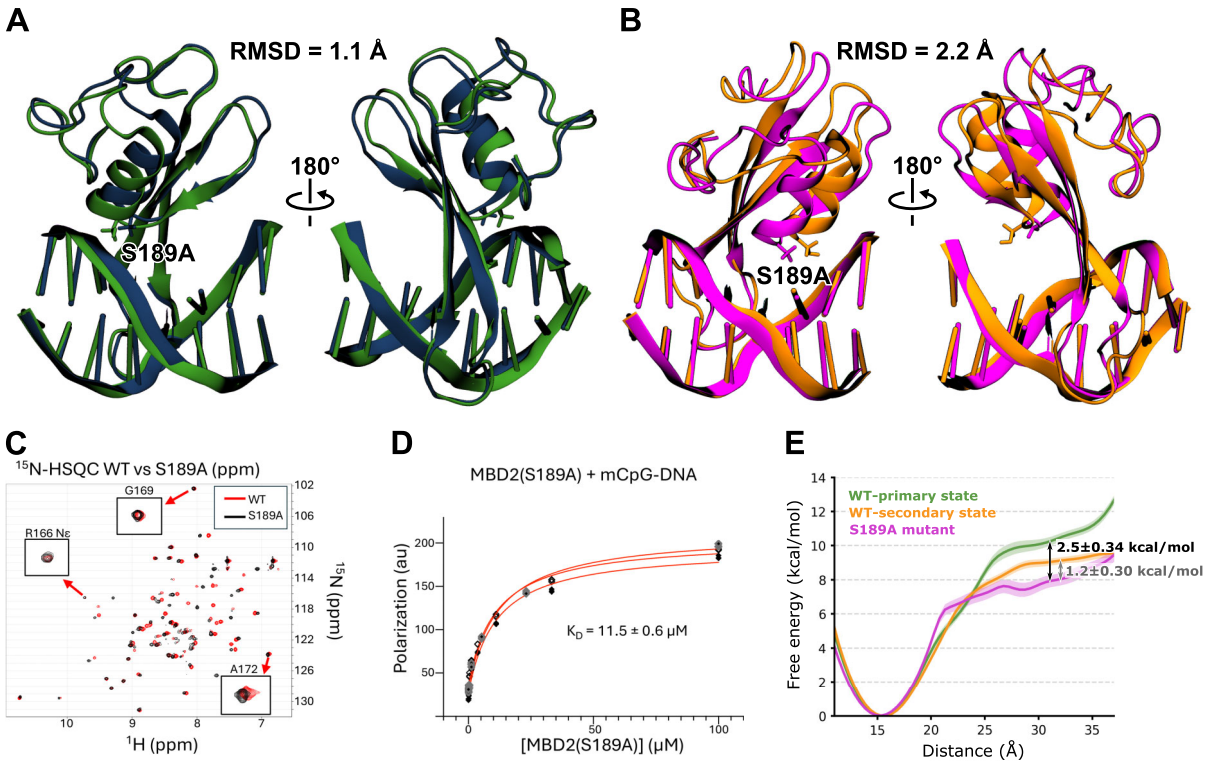


Figure 5. Methylation selectivity of MBD2(S189A). (**A**) Comparison of cluster center C_1^{wt} (green) obtained from the wild-type simulations and the cluster center C_1^{mu} (blue) obtained from the S189A mutation simulations. (**B**) Comparison of cluster center C_2^{wt} (orange) obtained from the wild-type simulations and cluster center C_2^{mu} (pink) obtained from the S189A mutation simulations. The mutation is highlighted. (**C**) An overlay of 2D ^{15}N-HSQC spectra is shown for MBD2–MBD wild-type and S189A. Three sub-spectra show expanded views of reporter resonances (R168 Nϵ, G169, and A172) that show large chemical shift changes between methylation-specific and nonspecific binding modes. (**D**) The binding affinity of MBD2–MBD(S189A) for methylated DNA was measured by fluorescence polarization analysis. The fluorescent polarization intensities and corresponding fits are depicted for each of the three replicates ($N = 3$). (**E**) Binding free energy profiles from Umbrella Sampling simulations for the wt-primary state, wt-secondary state, and the S189A mutant. Bold lines indicate the statistical mean, and shaded areas represent the standard error of the mean (SEM). Arrows show free energy differences between wt-primary and S189A, and between wt-secondary and S189A (in kcal/mol).

mutation shifts the MBD2–mCpG conformation toward the secondary state observed in the wt-simulations.

S189A mutation retains MBD2 selectivity for mCpG but reduces binding affinity

To assess whether the S189A mutation affects MBD2's selectivity for mCpG, we performed a ^{15}N HSQC NMR experiment. Previous studies have identified three cross-peaks in MBD2, corresponding to highly conserved residues G169, R166, and A172, whose chemical shifts reflect mCpG specific binding by MBD2. These three cross-peaks exhibit large chemical shift changes between nonspecific and methylation-specific binding modes [18, 36, 95, 96]. For example, the cross-peak for ^{1}Hϵ of R166 shifts over 1 ppm downfield when bound to methylated DNA compared to unmethylated DNA. Our ^{15}N HSQC data (Fig. 5C) indicate that the cross-peaks corresponding to these residues, i.e. G169, R166, and A172, remain unchanged upon the S189A mutation, demonstrating that this mutation does not alter MBD2's selectivity for mCpG. In contrast, many of the other resonances show chemical shift changes with the S189A mutation, consistent with a change in the relative orientation between the MBD2 and DNA. Hence, the mutant protein maintains methylation-specific binding even though its orientation with respect to DNA likely changes. Next, to quantify the effect of the S189A mutation on binding affinity, we performed a fluorescence polarization (FP) experiment measuring MBD2–S189A binding to mCpG DNA. As shown in Fig. 5D, we measured a dissociation constant of $K_D = (11.5 \pm 0.6)\,\mu$M, indicating a ten-fold reduction in binding affinity compared to *wt*-MBD2, which has a dissociation constant of $K_D = (0.72 \pm 0.06)\,\mu$M as we have previously determined using the same DNA and experimental conditions [78]. Therefore, the binding free energy difference as a result of the S189A mutation is (1.67 ± 0.51) kcal/mol, consistent with the expected energetic contribution of a single strong hydrogen bond interaction [97, 98]. We further sought to evaluate the binding free energy of the MBD2–mCpG complex for both wt (primary and secondary states) and S189A variants using advanced enhanced sampling simulation techniques (see Materials and methods). Using steered molecular dynamics (SMD) with a funnel potential to guide the dissociation pathway, combined with with umbrella sampling (details depicted in Supplementary Fig. S25), we calculated a binding free energy difference of (2.4 ± 0.34) kcal/mol between the wt-primary state complex and the S189A mutant (Fig. 5E). A similar calculation yielded a binding free energy difference of (1.2 ± 0.43) kcal/mol between the wt-secondary state complex and the S189A mutant. These results indicate that the binding free energy difference between the wt-secondary state and the S189A mutant is approximately half that of the difference between wt-primary state and the S189A mutant, highlighting the greater conformational similarity of the secondary state to S189A compared to the primary state. Consistent with these findings, the bind-

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

ing free energy measured by fluorescence polarization (FP) falls between the two computed free energy bounds (Fig. 5E), demonstrating strong agreement between simulation and experiment. Taken together, while our FP measurements and biased MD simulations demonstrate a reduced binding affinity upon the S189A mutation, NMR experiments reveal that this mutation does not affect MBD2's selectivity for mCpG.

Discussion

In this study, we elucidated the atomic-level dynamics and stability of MBD2 binding to methylated CpG (mCpG) sites using extensive microsecond-scale MD simulations, validated by NMR spectroscopy and fluorescence polarization assays.

Initially, we assessed the stability of the MBD2–mCpG complex using its experimentally determined X-ray (PDB ID: 7MWK) and NMR (PDB ID: 2KY8 [31]) structures. Using the X-ray structure as the initial conformation for MD simulations, we observe that the MBD2–mCpG complex remains fully stable, maintaining both its overall conformation and key base-specific interactions. This stability is mediated by the key MBD2 residues R166 and R188, which act as a clamp to lock MBD2 onto the mCpG site by interacting with the bases of opposing methylated cytosines, while Y178 reinforces the structure through a strong water-mediated hydrogen bond interaction with the base of methylcytosine mC_{-1}. These residues govern both the selectivity and affinity of MBD2 for mCpG [22]. Additionally, DNA methylation is a key determinant of the high stability and specificity of MBD2–mCpG interactions [99].

In contrast to the recognition simulations (based on the X-ray structure) the MBD2–mCpG complex in the interrogation simulations (based on the NMR structure) appears unstable. This instability arises from the fact that the DNA in the NMR structure is partially modeled on canonical B-DNA due to low proton density observed during structure determination [31]. Notably, despite the presence of key protein–DNA direct hydrogen bonds in the NMR structure, only three out of eight MD replicas (with a 1 μs simulation time per replica) resemble the specific state observed in the X-ray structure, underscoring the significance of the indirect (shape) readout mechanism [100–102] for stabilizing the specific complex. Figure 4A illustrates that a unique DNA conformation is necessary for the formation of the specific MBD2–mCpG complex. To adopt such unique DNA conformation required for proper recognition, MBD2—as shown in the NMR based interrogation simulations—must overcome an energetic barrier, which, based on the calculated similarity index [45], is of moderate magnitude.

Consistent with these findings, our search simulations, for which MBD2 was initially positioned one base pair downstream of the mCpG site, reveal a moderate energetic barrier to recognition. In four out of seven replicas, MBD2 reaches the specific state observed in the X-ray structure on a timescale of several microseconds. Strikingly, even from a nonspecific B-DNA starting structure and with MBD2 positioned one base pair away from its specific target site, fully unbiased MD simulations spontaneously drive MBD2 to its native recognition complex, demonstrating the remarkable capability of MD simulations to recapitulate precise, functionally relevant protein–DNA recognition.

In three independent search simulations, MBD2 formed a stable alternative complex with mCpG, structurally distinct from the primary recognition state and remarkably persistent over multiple microseconds—2 up to about 35 μs—highlighting its intrinsically stable nature. To uncover the structural determinants of the secondary state, we analyzed possible macro-switches distinguishing it from the primary binding mode. Our data indicate that the essential arginines R166 and R188 maintain their base-specific recognition of mCpG, suggesting a preserved core of selective interactions. Disruption of either residue has been experimentally shown to substantially impair MBD2's affinity for methylated CpG [17, 31]. In addition, the critical contact between K174 and a flanking mCpG base persists in both conformational states. Y178, which has also been associated with mCpG selectivity in MBD2, forms a water-mediated hydrogen bond in the primary state, whereas in the secondary state it interacts primarily through van der Waals interactions. Therefore, both states retain a core set of key interactions that robustly support MBD2's binding selectivity and functional engagement with mCpG.

The key macro-switch differentiating the primary and secondary state is the interaction of S189 with the backbone of methylated cytosine mC_{+1}. We observe that the absence of this interaction shifts the MBD2–mCpG conformation toward the secondary state. To investigate the precise role of S189 in mediating selective mCpG recognition by MBD2, we conducted ^{15}N HSQC NMR spectroscopy on the MBD2–mCpG complex containing the S189A mutation and directly compared the resulting spectra with those of the wild-type complex [78]. Importantly, our NMR analysis reveals that the S189A mutation does not compromise the selective recognition of mCpG by MBD2 (Fig. 5C). However, both our fluorescence polarization experiments and biased MD simulations consistently show that the S189A mutation reduces MBD2's binding affinity for mCpG (Fig. 5D,E). Similar muations in MBD1 (S45A) and MeCP2 (S134A) have been shown to decrease their affinity for mCpG by approximately half [32, 33]. The S134C mutation in MeCP2 also reduces binding affinity for mCpG without affecting selectivity, and is even clinically significant, as it is associated with Rett syndrome, underscoring the biological importance of this serine residue in maintaining proper epigenetic regulation [103]. We suggest that, analogous to S189A in MBD2, the S134C mutation could drive MeCP2 (whose conformation closely resembles that of MBD2, see Supplementary Fig. S26) into a binding conformation that diverges from its canonical state, potentially carrying greater functional consequences.

From our MD simulations of the MBD2–S189A variant, we observe that the S189A mutation indeed shifts the MBD2–mCpG complex toward a conformation characteristic of the secondary state. This finding strongly suggests that the secondary state represents a valid and selective recognition state, albeit with reduced binding affinity relative to the primary state.

To further establish the secondary state as a genuine alternative sequence-specific bound conformation, we conducted additional search simulations using a native DNA sequence. In line with previous results, bistability is recapitulated, with the secondary state closely resembling that observed with the initial DNA sequence.

Such bistable binding behavior in DNA-binding proteins were initially identified through single-molecule spectroscopy by Liu *et al.*, demonstrating that the DNA-binding domain of the *Arabidopsis* WRKY protein undergoes bistable oscil-

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

lations upon specific target engagement [104]. Their findings further revealed that the dwell time, defined as the period the system resides in each oscillatory state, varies with temperature [104]. Therefore, as X-ray crystallography preferentially captures well-ordered conformational states, the observation of only the primary state—being more ordered and less dynamic than the secondary state—may reflect this inherent crystallographic bias and may also explain why the complete structure could not be fully resolved in the NMR ensemble, owing to the dynamic nature of the MBD2–mCpG complex [31, 105, 106].

It is also noteworthy that, since the S189–mC_{+1} interaction serves as a critical macro-switch distinguishing the two stable states observed with MBD2 bound at the target site, external factors could plausibly shift the equilibrium between these states by modulating the strength of this interaction. For instance, ions have been shown to modulate structural transitions in DNA-binding proteins [107]. Specifically, changes in ion concentrations regulate electrostatic interactions, particularly with the DNA backbone, which are critical for protein–DNA recognition [107–109]. Thus, variations in ionic strength may influence the S189–mC_{+1} backbone interaction, thereby increasing the likelihood of transitions between the two distinct stable states.

Moreover, MBD2 binding to mCpG, in either the primary or secondary state, induces distinct DNA conformations. Our simulations reveal that disruption of the S189–mC_{+1} interaction leads to a distinct BI/BII backbone conformation at the mCpG site, which plays a critical role in determining the overall DNA geometry [88–94]. Notably, MBD2 has been shown to retain high-affinity binding to mCpG across diverse DNA geometries—including both straight and bent conformations—suggesting that bistable recognition may balance conformational adaptability with high-affinity DNA binding. It has been shown that MBD2 becomes trapped in mCpG-rich genomic regions, yet can still rapidly exchange between closely spaced mCpG sites [22]. In this context, the presence of an alternative bound state may provide a structural basis for maintaining specific mCpG recognition while enabling local conformational flexibility that facilitates exchange between neighboring methylated sites. Specifically, this bistable state mechanism may provide MBD2 with a strategic advantage, ensuring precise mCpG binding while also enabling dynamic participation in assembling larger complexes, such as MBD2–NuRD.

Conclusions

As a pivotal epigenetic regulator, MBD2 serves as a key reader of 5mC, characterized by its high selectivity and tight binding to methylated CpG dinucleotides. To gain a deeper atomic-level understanding of the MBD2–mCpG complex, we performed extensive molecular dynamics simulations, both in its specific bound state and during transitions from nonspecific to specific binding. In summary, our study reveals the following key insights: (i) MBD2 adopts a bistable binding equilibrium with mCpG, a mechanism that may enable it to maintain high-affinity binding while remaining flexible for interactions across different molecular contexts and assemblies. (ii) By favoring a conformation similar to the secondary state, the S189A mutation maintains MBD2's selectivity for mCpG but compromises its binding affinity. (iii) Our observation of bistable binding in MBD2 with two different DNA sequences, along with single-molecule evidence from the WRKY domain (Liu *et al.*), supports the notion that bound-state conformational equilibria may be a fundamental feature of sequence-specific protein–DNA recognition.

Acknowledgements

The authors thank the International Max Planck Research School for Biology and Computation (IMPRS-BAC), and the Bavarian Equal Opportunities Sponsorship—Realisierung von Chancengleichheit von Frauen in Forschung und Lehre (FFL)—for supporting S.V. The authors gratefully acknowledge the computing time made available to them on the high-performance computer "Lise" at the NHR center NHR@ZIB. This center is jointly supported by the Federal Ministry of Education and Research and the state governments participating in the NHR (www.nhr-verein.de). The authors gratefully acknowledge the computing time provided by them on the high-performance computers Barnard and Romeo at NHR@TUD, which is funded by the Federal Ministry of Education and Research and the state governments participating on the basis of the resolutions of the GWK for the national high performance computing at universities (www.nhr-verein.de/unsere-partner). We thank Dustin Vivod for helpful discussions.

Author contributions: Senta Volkenandt (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Julia Belyaeva (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal]), Torry Li (Formal analysis [equal], Investigation [equal], Validation [equal]), Matthias Elgeti (Supervision [equal], Validation [equal], Writing—review & editing [equal]), Ralf Metzler (Conceptualization [equal] Validation [equal] Writing—review & editing [equal]), David C. Williams, Jr. (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—review & editing [equal]), and Mahdi Bagherpoor Helabad (Conceptualization [lead], Data curation [lead], Formal analysis [equal], Investigation [lead], Methodology [equal], Software [equal], Supervision [lead], Validation [lead], Visualization [equal], Writing—original draft [lead], Writing—review & editing [lead]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

The authors declare no competing interests.

Funding

This work was funded by the the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 514664767, subproject B07 to M.E., project number IM141/1-3 to P.I., project number 318763901, suproject B10

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

to R.M., and National Institute of Diabetes, Digestive, and Kidney Diseases [R01DK115563 to D.W.]. Funding to pay the Open Access publication charges for this article was provided by Martin-Luther-Universität Halle-Wittenberg Universitäts- und Landesbibliothek Sachsen-Anhalt Zeitschriftenstelle.

Data availability

The data that support the findings of this study are openly available at Zenodo at https://doi.org/10.5281/zenodo.15757877. NMR assignments used in this work were previously published in another paper under the BMRB entry 16936.

References

1. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. *Science* 2018;361:1336–40. https://doi.org/10.1126/science.aat6806
2. Kanwal R, Gupta S. Epigenetic modifications in cancer. *Clin Genet* 2012;81:303–11. https://doi.org/10.1111/j.1399-0004.2011.01809.x
3. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology* 2013;38:23–38. https://doi.org/10.1038/npp.2012.112
4. Schübeler D. Function and information content of DNA methylation. *Nature* 2015;517:321–6. https://doi.org/10.1038/nature14192
5. Bagherpoor Helabad M, Kanaan N, Imhof P. Base flip in DNA studied by molecular dynamics simulations of differently-oxidized forms of methyl-cytosine. *Int J Mol Sci* 2014;15:11799–816. https://doi.org/10.3390/ijms150711799
6. Jiang S, Guo Y. Epigenetic clock: DNA methylation in aging. *Stem Cells Int* 2020;2020:1047896. https://doi.org/10.1155/2020/1047896
7. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. *Trends Genet* 2012;28:33–42. https://doi.org/10.1016/j.tig.2011.09.004
8. Smith ZD, Hetzel S, Meissner A. DNA methylation in mammalian development and disease. *Nat Rev Genet* 2025;26:7–30. https://doi.org/10.1038/s41576-024-00760-8
9. Pasque V, Jullien J, Miyamoto K *et al.* Epigenetic factors influencing resistance to nuclear reprogramming. *Trends Genet* 2011;27:516–25. https://doi.org/10.1016/j.tig.2011.08.002
10. Blake LE, Roux J, Hernando-Herraez I *et al.* A comparison of gene expression and DNA methylation patterns across tissues and species. *Genome Res* 2020;30:250–62. https://doi.org/10.1101/gr.254904.119
11. Hu S, Wan J, Su Y *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* 2013;2:e00726. https://doi.org/10.7554/eLife.00726
12. Bogdanović O, Veenstra GJC. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* 2009;118:549–65. https://doi.org/10.1007/s00412-009-0221-9
13. Du Q, Luu PL, Stirzaker C *et al.* Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* 2015;7:1051–73. https://doi.org/10.2217/epi.15.39
14. Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl CpG-binding proteins. *Genet Res* 1998;72:59–72. https://doi.org/10.1017/S0016672398533307
15. Liu K, Xu C, Lei M *et al.* Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNAJ. *Biol Chem* 2018;293:7344–54. https://doi.org/10.1074/jbc.RA118.001785
16. Liu K, Lei M, Wu Z *et al.* Structural analyses reveal that MBD3 is a methylated CG binder. *FEBS J* 2019;286:3240–54. https://doi.org/10.1111/febs.14850
17. Leighton G, Williams DC, Jr. The methyl-CpG–Binding domain 2 and 3 proteins and formation of the nucleosome remodeling and deacetylase complex. *J Mol Biol* 2020;432:1624–39. https://doi.org/10.1016/j.jmb.2019.10.007
18. Cramer JM, Scarsdale JN, Walavalkar NM *et al.* Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNAJ. *Biol Chem* 2014;289:1294–302. https://doi.org/10.1074/jbc.M113.512236
19. Zhang Y, Ng HH, Erdjument-Bromage H *et al.* Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* 1999;13:1924–35. https://doi.org/10.1101/gad.13.15.1924
20. Hainer SJ, McCannell KN, Yu J *et al.* DNA methylation directs genomic localization of MBD2 and MBD3 in embryonic stem cells. *eLife* 2016;5:e21964. https://doi.org/10.7554/eLife.21964
21. Wood KH, Zhou Z. Emerging molecular and biological functions of MBD2, a reader of DNA methylation. *Front Genet* 2016;7:93. https://doi.org/10.3389/fgene.2016.00093
22. Leighton GO, Irvin EM, Kaur P *et al.* Densely methylated DNA traps methyl-CpG–binding domain protein 2 but permits free diffusion by methyl-CpG–binding domain protein 3. *J Biol Chem* 2022;298:1–14. https://doi.org/10.1016/j.jbc.2022.102428
23. Mian OY, Wang SZ, Zhu SZ *et al.* Methyl-binding domain protein 2–dependent proliferation and survival of breast cancer cells. *Mol Cancer Res* 2011;9:1152–62. https://doi.org/10.1158/1541-7786.MCR-11-0252
24. Lei Q, Zhang W. Methyl CpG binding domain protein 2 (MBD2) in inflammation. *Chin Med J* 2022;135:2880–2. https://doi.org/10.1097/CM9.0000000000002482
25. Lax E, Do Carmo S, Enuka Y *et al.* Methyl-CpG binding domain 2 (MBD2) is an epigenetic regulator of autism-risk genes and cognition. *Transl Psychiatry* 2023;13:259. https://doi.org/10.1038/s41398-023-02561-9
26. Baubec T, Ivánek R, Lienert F *et al.* Methylation-dependent and-independent genomic targeting principles of the MBD protein family. *Cell* 2013;153:480–92. https://doi.org/10.1016/j.cell.2013.03.011
27. Günther K, Rust M, Leers J *et al.* Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res* 2013;41:3010–21. https://doi.org/10.1093/nar/gkt035
28. Yu X, Azzo A, Bilinovich SM *et al.* Disruption of the MBD2-NuRD complex but not MBD3-NuRD induces high level HbF expression in human adult erythroid cells. *Haematologica* 2019;104:2361. https://doi.org/10.3324/haematol.2018.210963
29. Shang S, Li X, Azzo A *et al.* MBD2a-NuRD binds to the methylated γ-globin gene promoter and uniquely forms a complex required for silencing of HbF expression. *Proc Natl Acad Sci USA* 2023;120:e2302254120. https://doi.org/10.1073/pnas.2302254120
30. Millard CJ, Fairall L, Ragan TJ *et al.* The topology of chromatin-binding domains in the NuRD deacetylase complex. *Nucleic Acids Res* 2020;48:12972–82. https://doi.org/10.1093/nar/gkaa1121
31. Scarsdale JN, Webb HD, Ginder GD *et al.* Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res* 2011;39:6741–52. https://doi.org/10.1093/nar/gkr262
32. Ohki I, Shimotake N, Fujita N *et al.* Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J* 1999;18:6653–61. https://doi.org/10.1093/emboj/18.23.6653
33. Free A, Wakefield RI, Smith BO *et al.* DNA recognition by the methyl-CpG binding domain of MeCP2. *J Biol Chem* 2001;276:3353–60. https://doi.org/10.1074/jbc.M007224200

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

34. Cheadle JP, Gill H, Fleming N *et al.* Long-read sequence analysis of the MeCP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. *Hum Mol Genet* 2000;9:1119–29. https://doi.org/10.1093/hmg/9.7.1119
35. Kucukkal TG, Yang Y, Uvarov O *et al.* Impact of Rett syndrome mutations on MeCP2 MBD stability. *Biochemistry* 2015;54:6357–68. https://doi.org/10.1021/acs.biochem.5b00790
36. Walavalkar NM, Cramer JM, Buchwald WA *et al.* Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic Acids Res* 2014;42:11218–32. https://doi.org/10.1093/nar/gku782
37. Pan H, Bilinovich SM, Kaur P *et al.* CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. *Nucleic Acids Res* 2017;45:9164–77. https://doi.org/10.1093/nar/gkx548
38. Bigman LS, Greenblatt HM, Levy Y. What are the molecular requirements for protein sliding along DNA?. *J Phys Chem B* 2021;125:3119–31. https://doi.org/10.1021/acs.jpcb.1c00757
39. Blainey PC, Luo G, Kou S *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* 2009;16:1224–9. https://doi.org/10.1038/nsmb.1716
40. Marklund E, van Oosten B, Mao G *et al.* DNA surface exploration and operator bypassing during target search. *Nature* 2020;583:858–61. https://doi.org/10.1038/s41586-020-2413-7
41. Cencini M, Pigolotti S. Energetic funnel facilitates facilitated diffusion. *Nucleic Acids Res* 2018;46:558–67. https://doi.org/10.1093/nar/gkx1220
42. Tempestini A, Monico C, Gardini L *et al.* Sliding of a single lac repressor protein along DNA is tuned by DNA sequence and molecular switching. *Nucleic Acids Res* 2018;46:5001–11. https://doi.org/10.1093/nar/gky208
43. Ahmadi A, Rosnes I, Blicher P *et al.* Breaking the speed limit with multimode fast scanning of DNA by Endonuclease V. *Nat Commun* 2018;9:5381. https://doi.org/10.1038/s41467-018-07797-4
44. Arbel-Goren R, McKeithen-Mead SA, Voglmaier D *et al.* Target search by an imported conjugative DNA element for a unique integration site along a bacterial chromosome during horizontal gene transfer. *Nucleic Acids Res* 2023;51:3116–29. https://doi.org/10.1093/nar/gkad068
45. Leven I, Levy Y. Quantifying the two-state facilitated diffusion model of protein–DNA interactions. *Nucleic Acids Res* 2019;47:5530–8. https://doi.org/10.1093/nar/gkz308
46. Marcovitz A, Levy Y. Frustration in protein–DNA binding influences conformational switching and target search kinetics. *Proc Natl Acad Sci USA* 2011;108:17957–62. https://doi.org/10.1073/pnas.1109594108
47. Pulkkinen O, Metzler R. Distance matters: the impact of gene proximity in bacterial gene regulation. *Phys Rev Lett* 2013;110:198101. https://doi.org/10.1103/PhysRevLett.110.198101
48. Hedström L, Metzler R, Lizana L. Enhancer-insulator pairing reveals heterogeneous dynamics in long-distance 3D gene regulation. *PRX Life* 2024;2:033008. https://doi.org/10.1103/PRXLife.2.033008
49. Marklund EG, Mahmutovic A, Berg OG *et al.* Transcription-factor binding and sliding on DNA studied using micro-and macroscopic models. *Proc Natl Acad Sci USA* 2013;110:19796–801. https://doi.org/10.1073/pnas.1307905110
50. Terakawa T, Takada S. p53 dynamics upon response element recognition explored by molecular simulations. *Sci Rep* 2015;5:1–10. https://doi.org/10.1038/srep17107
51. Bhattacherjee A, Krepel D, Levy Y. Coarse-grained models for studying protein diffusion along DNA. Wiley Interdiscip. *Rev Comput Mol Sci* 2016;6:515–31. https://doi.org/10.1002/wcms.1262
52. Chu X, Muñoz V. Roles of conformational disorder and downhill folding in modulating protein–DNA recognition. *Phys Chem Chem Phys* 2017;19:28527–39. https://doi.org/10.1039/C7CP04380E
53. Wieczór M, Czub J. How proteins bind to DNA: target discrimination and dynamic sequence search by the telomeric protein TRF1. *Nucleic Acids Res* 2017;45:7643–54. https://doi.org/10.1093/nar/gkx534
54. Zacharias M. Atomic resolution insight into Sac7d protein binding to DNA and associated global changes by molecular dynamics simulations. *Angew Chem Int Ed* 2019;58:5967–72. https://doi.org/10.1002/anie.201900935
55. Tian J, Wang L, Da LT. Atomic resolution of short-range sliding dynamics of thymine DNA glycosylase along DNA minor-groove for lesion recognition. *Nucleic Acids Res* 2021;49:1278–93. https://doi.org/10.1093/nar/gkaa1252
56. Dai L, Xu Y, Du Z *et al.* Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA. *Proc Natl Acad Sci USA* 2021;118:e2102621118. https://doi.org/10.1073/pnas.2102621118
57. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12. https://doi.org/10.1002/jcc.20084
58. Abraham MJ, Murtola T, Schulz R *et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1:19–25. https://doi.org/10.1016/j.softx.2015.06.001
59. Maier JA, Martinez C, Kasavajhala K *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;11:3696–713. https://doi.org/10.1021/acs.jctc.5b00255
60. Ivani I, Dans PD, Noy A *et al.* Parmbsc1: a refined force field for DNA simulations. *Nat Methods* 2016;13:55–8. https://doi.org/10.1038/nmeth.3658
61. Liebl K, Zacharias M. How methyl-sugar interactions determine DNA structure and flexibility. *Nucleic Acids Res* 2019;47:1132–40. https://doi.org/10.1093/nar/gky1237
62. Carvalho AT, Gouveia L, Kanna CR *et al.* Understanding the structural and dynamic consequences of DNA epigenetic modifications: Computational insights into cytosine methylation and hydroxymethylation. *Epigenetics* 2014;9:1604–12. https://doi.org/10.4161/15592294.2014.988043
63. Jorgensen WL, Chandrasekhar J, Madura JD *et al.* Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35. https://doi.org/10.1063/1.445869
64. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007;126:014101. https://doi.org/10.1063/1.2408420
65. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 1981;52:7182–90. https://doi.org/10.1063/1.328693
66. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92. https://doi.org/10.1063/1.464397
67. Essmann U, Perera L, Berkowitz ML *et al.* A smooth particle mesh Ewald method. *J Chem Phys* 1995;103:8577–93. https://doi.org/10.1063/1.470117
68. Hess B, Bekker H, Berendsen HJ *et al.* LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997;18:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12⟨1463::AID-JCC4⟩3.0.CO;2-H
69. PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* 2019;16:670–3. https://doi.org/10.1038/s41592-019-0506-8
70. Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci USA* 2013;110:6358–63. https://doi.org/10.1073/pnas.1303186110

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

71. . The PyMOL molecular graphics system. *Version* 2015;1:8.
72. Grubmüller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 1996;271:997–9. https://doi.org/10.1126/science.271.5251.997
73. Van Der Spoel D, Lindahl E, Hess B *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26:1701–18. https://doi.org/10.1002/jcc.20291
74. Tribello GA, Bonomi M, Branduardi D *et al.* PLUMED 2: new feathers for an old bird. *Comput Phys Commun* 2014;185:604–13. https://doi.org/10.1016/j.cpc.2013.09.018
75. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys* 1977;23:187–99. https://doi.org/10.1016/0021-9991(77)90121-8
76. Kumar S, Rosenberg JM, Bouzida D *et al.* The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 1992;13:1011–21. https://doi.org/10.1002/jcc.540130812
77. Souaille M, Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun* 2001;135:40–57. https://doi.org/10.1016/S0010-4655(00)00215-0
78. Desai MA, Webb HD, Sinanan LM *et al.* An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res* 2015;43:3100–13. https://doi.org/10.1093/nar/gkv168
79. Cai M, Williams DC, Wang G *et al.* Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIAGlucose and the cytoplasmic domain of the glucose transporter IICBGlucose of the *Escherichia coli* glucose phosphotransferase system. *J Biol Chem* 2003;278:25191–206. https://doi.org/10.1074/jbc.M302677200
80. Delaglio F, Grzesiek S, Vuister GW *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–93. https://doi.org/10.1007/BF00197809
81. Vranken WF, Boucher W, Stevens TJ *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 2005;59:687–96. https://doi.org/10.1002/prot.20449
82. Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 2013;110:17874–9. https://doi.org/10.1073/pnas.1311599110
83. Daura X, Gademann K, Jaun B *et al.* Peptide folding: when simulation meets experiment. *Angew Chem Int Ed* 1999;38:236–40. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2⟨236::AID-ANIE236⟩3.0.CO;2-M
84. Hess B. Convergence of sampling in protein simulations. *Phys Rev E* 2002;65:031910. https://doi.org/10.1103/PhysRevE.65.031910
85. Lavery R, Moakher M, Maddocks JH *et al.* Conformational analysis of nucleic acids revisited: curves. *Nucleic Acids Res* 2009;37:5917–29. https://doi.org/10.1093/nar/gkp608
86. Blanchet C, Pasi M, Zakrzewska K *et al.* CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res* 2011;39:W68–73. https://doi.org/10.1093/nar/gkr316
87. Eisenhaber F, Lijnzaad P, Argos P *et al.* The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J Comput Chem* 1995;16:273–84. https://doi.org/10.1002/jcc.540160303
88. Djuranovic D, Hartmann B. DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers* 2004;73:356–68. https://doi.org/10.1002/bip.10528
89. Heddi B, Foloppe N, Bouchemal N *et al.* Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J Am Chem Soc* 2006;128:9170–7. https://doi.org/10.1021/ja061686j
90. Tian Y, Kayatta M, Shultis K *et al.* 31P NMR investigation of backbone dynamics in DNA binding sites. *J Phys Chem B* 2009;113:2596–603. https://doi.org/10.1021/jp711203m
91. Westwood M, Ljunggren K, Boyd B *et al.* Single-base lesions and mismatches alter the backbone conformational dynamics in DNA. *Biochemistry* 2021;60:873–85. https://doi.org/10.1021/acs.biochem.0c00784
92. Westwood M, Pilarski A, Johnson C *et al.* Backbone conformational equilibrium in mismatched DNA correlates with enzyme activity. *Biochemistry* 2023;62:2816–27. https://doi.org/10.1021/acs.biochem.3c00230
93. Savelyev A, MacKerell AD, Jr. Differential deformability of the DNA minor groove and altered BI/BII backbone conformational equilibrium by the monovalent ions Li+, Na+, K+, and Rb+ via water-mediated hydrogen bonding. *J Chem Theory Comput* 2015;11:4473–85. https://doi.org/10.1021/acs.jctc.5b00508
94. Precechtelová J, Novák P, Munzarová ML *et al.* Phosphorus chemical shifts in a nucleic acid backbone from combined molecular dynamics and density functional calculations. *J Am Chem Soc* 2010;132:17139–48. https://doi.org/10.1021/ja104564g
95. Cramer JM, Cramer JM, Pohlmann D *et al.* Methylation specific targeting of a chromatin remodeling complex from sponges to humans. *Sci Rep* 2017;7:40674. https://doi.org/10.1038/srep40674
96. Sperlazza MJ, Bilinovich SM, Sinanan LM *et al.* Structural basis of MeCP2 distribution on non-CpG methylated and hydroxymethylated DNAJ. *Mol Biol* 2017;429:1581–94. https://doi.org/10.1016/j.jmb.2017.04.009
97. Lee HR, Helquist SA, Kool ET *et al.* Importance of hydrogen bonding for efficiency and specificity of the human mitochondrial DNA polymerase. *J Biol Chem* 2008;283:14402–10. https://doi.org/10.1074/jbc.M705007200
98. Gorfe AA, Jelesarov I. Energetics of sequence-specific protein–DNA association: computational analysis of integrase Tn916 binding to its target DNA. *Biochemistry* 2003;42:11568–76. https://doi.org/10.1021/bi026937p
99. Ginder GD, Williams DC, Jr. Readers of DNA methylation, the MBD family as potential therapeutic targets. *Pharmacol Ther* 2018;184:98–111. https://doi.org/10.1016/j.pharmthera.2017.11.002
100. Rohs R, Jin X, West SM *et al.* Origins of specificity in protein–DNA recognition. *Annu Rev Biochem* 2010;79:233–69. https://doi.org/10.1146/annurev-biochem-060408-091030
101. Samee MAH, Bruneau BG, Pollard KS. A de novo shape motif discovery algorithm reveals preferences of transcription factors for DNA shape beyond sequence motifs. *Cell Syst* 2019;8:27–42. https://doi.org/10.1016/j.cels.2018.12.001
102. , Mishra SK, Bhattacherjee A. Role of shape deformation of DNA-binding sites in regulating the efficiency and specificity in their recognition by DNA-binding proteins. *JACS Au* 2024;4:2640–55. https://doi.org/10.1021/jacsau.4c00393
103. Yang Y, Kucukkal TG, Li J *et al.* Binding analysis of methyl-CpG binding domain of MeCP2 and Rett syndrome mutations. *ACS Chem Biol* 2016;11:2706–15. https://doi.org/10.1021/acschembio.6b00450
104. Liu W, Li J, Xu Y *et al.* Complete mapping of DNA–protein interactions at the single-molecule level. *Adv Sci* 2021;8:2101383. https://doi.org/10.1002/advs.202101383
105. Tilton RF, Jr, Dewan JC, Petsko GA. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K. *Biochemistry* 1992;31:2469–81. https://doi.org/10.1021/bi00124a006

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

106. Fraser JS, Van Den Bedem H, Samelson AJ *et al*. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc Natl Acad Sci USA* 2011;108:16247–52. https://doi.org/10.1073/pnas.1111325108
107. Xu C, Lu Y, Wu Y *et al*. Sodium ion-induced structural transition on the surface of a DNA-interacting protein. *Adv Sci* 2024;11:2401838. https://doi.org/10.1002/advs.202401838
108. Pasi M, Maddocks JH, Lavery R. Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res* 2015;43:2412–23. https://doi.org/10.1093/nar/gkv080
109. Yu B, Pettitt BM, Iwahara J. Dynamics of ionic interactions at protein–nucleic acid interfaces. *Acc Chem Res* 2020;53:1802–10. https://doi.org/10.1021/acs.accounts.0c00212

Downloaded from academic.oup.com/nar/article/54/19/gkag952/8937190 by Universitaetsbibliothek Potsdam user on 21 October 2026

Received: July 2, 2025. **Revised:** August 5, 2026. **Accepted:** August 22, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.