11218–11232 *Nucleic Acids Research, 2014, Vol. 42, No. 17* *Published online 02 September 2014*
doi: 10.1093/nar/gku782

Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for ^{m}CpG/TpG mismatches

Ninad M. Walavalkar[1], Jason M. Cramer[2], William A. Buchwald[1], J. Neel Scarsdale[3] and David C. Williams, Jr[1,*]

[1]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, [2]Department of Biochemistry and Molecular Biology, Virginia Commonwealth University, Richmond, VA 23298, USA and [3]Institute of Structural Biology and Drug Discovery, Center for the Study of Biological Complexity and Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA

Received April 25, 2014; Revised August 12, 2014; Accepted August 18, 2014

ABSTRACT

Unlike other members of the methyl-cytosine binding domain (MBD) family, MBD4 serves as a potent DNA glycosylase in DNA mismatch repair specifically targeting ^{m}CpG/TpG mismatches arising from spontaneous deamination of methyl-cytosine. The protein contains an N-terminal MBD ($MBD4_{MBD}$) and a C-terminal glycosylase domain ($MBD4_{GD}$) separated by a long linker. This arrangement suggests that the $MBD4_{MBD}$ either directly augments enzymatic catalysis by the $MBD4_{GD}$ or targets the protein to regions enriched for ^{m}CpG/TpG mismatches. Here we present structural and dynamic studies of $MBD4_{MBD}$ bound to dsDNA. We show that $MBD4_{MBD}$ binds with a modest preference for ^{m}CpG as compared to mismatch, unmethylated and hydroxymethylated DNA. We find that while $MBD4_{MBD}$ exhibits slow exchange between molecules of DNA (intermolecular exchange), the domain exhibits fast exchange between two sites in the same molecule of dsDNA (intramolecular exchange). Introducing a single-strand defect between binding sites does not greatly reduce the intramolecular exchange rate, consistent with a local hopping mechanism for moving along the DNA. These results support a model in which the $MBD4_{MBD4}$ targets the intact protein to ^{m}CpG islands and promotes scanning by rapidly exchanging between successive ^{m}CpG sites which facilitates repair of nearby ^{m}CpG/TpG mismatches by the glycosylase domain.

INTRODUCTION

DNA methylation involves enzymatic addition of a methyl group at the C5 position of the symmetrically opposed cytosine bases in a double stranded cytosine-guanosine sequence (CpG) and serves as a key epigenetic signal in developmental and tissue type specific regulatory mechanisms such as gene silencing, chromatin modifications and aberrant silencing of tumor suppressor genes in cancer (1). Central to these regulatory functions is a family of proteins that selectively bind symmetrically methylated CG dinucleotide sequences (^{m}CpG) through a common methyl-cytosine binding domain (MBD). The MBD was first described as a ∼70 amino acid region in the MeCP2 protein (2) and subsequently identified by homology in four additional proteins, MBD1–4 (3). MBD4 represents a unique member of the MBD family in that it contains intrinsic enzymatic activity provided by a C-terminal glycosylase domain ($MBD4_{GD}$) in addition to an N-terminal MBD ($MBD4_{MBD}$). The $MBD4_{GD}$ can remove thymine or hydroxymethyluracil in ^{m}CpG/TpG and ^{m}CpG/^{hm}UpG double-stranded mismatches (3–7). The ^{m}CpG/TpG (or ^{m}CpG/^{hm}UpG) mismatch arises from hydrolytic deamination of a methyl-cytosine (or hydroxymethylcytosine) to thymine (or hydroxmethyluracil) and represents one of the more common sources of germ line and somatic DNA point mutations.

The $MBD4_{MBD}$ binds in the major groove of DNA while the $MBD4_{GD}$ binds the minor groove (8) and the two are separated by a long spacer region (∼280 amino acids). Since the two domains bind opposite faces of the DNA, the $MBD4_{MBD}$ could selectively recognize a ^{m}CpG/TpG mismatch and augment the enzymatic activity of the $MBD4_{GD}$ at the same site - either by providing mismatch selectivity, overall binding affinity or stabilizing the transition state. Al-

*To whom correspondence should be addressed. Tel: +1 919 843 9949; Fax: +1 919 966 6718; Email: david_willjr@med.unc.edu

ternatively the $MBD4_{MBD}$ could target the protein to regions enriched for ^{m}CpG sites and allow the $MBD4_{GD}$ to repair mismatches at nearby sites (4). Several pieces of experimental evidence favor the latter model. The $MBD4_{GD}$ markedly distorts the DNA backbone which likely inhibits simultaneous binding with the $MBD4_{MBD}$ (4,8). The isolated $MBD4_{GD}$ maintains enzymatic activity in isolation while addition of free $MBD4_{MBD}$ inhibits activity towards a single ^{m}CpG/TpG mismatch site (9). $MBD4_{GD}$ orthologs in invertebrates frequently lack an associated MBD (Supplementary Table S1), again supporting the observation that the $MBD4_{GD}$ does not require a MBD for function.

In the studies presented here, we determine the solution structure of human $MBD4_{MBD}$ and show that it has a similar overall architecture as exhibited by other MBD domains. We show that $MBD4_{MBD}$ can bind methylated as well as unmethylated, hydroxymethylated and mismatched (^{m}CpG/TpG) DNA with only modest preference for ^{m}CpG (∼5-fold). Hence $MBD4_{MBD}$ does not show a strong preference for methylated DNA, which can be attributed to the reorientation of a critical tyrosine residue. Based on these observations, we hypothesized that dynamic intramolecular exchange by $MBD4_{MBD}$ could contribute to function by allowing MBD4 to scan along CpG rich regions of chromatin. To test this hypothesis, we measured chemical exchange rates to compare intermolecular and intramolecular exchange of $MBD4_{MBD}$ between two DNA binding sites. These studies show that intermolecular exchange of $MBD4_{MBD}$ occurs in the nuclear magnetic resonance (NMR) slow exchange time regime while intramolecular exchange occurs in the NMR fast exchange time regime and demonstrate that $MBD4_{MBD}$ preferentially exchanges along the DNA between sequentially binding sites. They represent the first example reported to date that demonstrates rapid intramolecular exchange for the MBD family of proteins. Together the data support a model in which the $MBD4_{MBD}$ contributes to function by recruiting the protein to ^{m}CpG rich regions and rapidly scanning among the ^{m}CpG sites.

MATERIAL AND METHODS

Protein expression and purification

The $MBD4_{MBD}$ (amino acids 80–148) was cloned and expressed with a hexahistidine tag and a thioredoxin fusion in a modified pET32a vector (10). The expression vector was transformed into the BL21(DE3) *Escherichia coli* strain, grown at 37°C and induced with 1 mM isopropyl-β-d-thiogalactopyranoside at an A_{600} ∼0.8. Induced bacteria were harvested and lysed with the B-PER reagent (Thermo Scientific). The soluble fraction was passed over a nickel-sepharose column and protein was eluted with a step gradient of imidazole. The $cMBD2_{MBD}$ was expressed and purified as described previously. (11) For NMR analyses, uniform double (^{13}C, ^{15}N) and triple (^{13}C, ^{15}N, ^{2}H) labeled protein samples were generated by standard techniques and the thioredoxin and hexahistidine fusion tags were removed by thrombin cleavage overnight at room temperature. The labeled protein was further purified by gel filtration over a Superdex-75 column (GE Healthcare) followed by reverse phase chromatography over a SOURCE-15RPC column (GE Healthcare). For SPR analysis, the fusion protein was purified over a nickel-sepharose column followed by ion exchange chromatography over a MonoS 10/100 GL (GE Healthcare) and size exclusion chromatography. The Y109F mutation was introduced using the QuickChange® site-directed mutagenesis kit (Agilent) following the manufacturer's protocol. The final proteins for all experiments were > 95% pure as estimated by SDS-PAGE analysis.

DNA purification

Complimentary DNA oligonucleotides were purchased from integrated DNA technologies. Forward and reverse oligonucleotides were dissolved in standard buffer (20 mM Tris pH 8.0), mixed in equimolar concentrations, incubated at >90°C for 10 min and cooled slowly to anneal. Subsequently, dsDNA was purified by ion exchange chromatography on MonoQ 10/100 column (GE Healthcare). 3′ biotinylated forward oligonucleotides (purchased from integrated DNA technologies) were mixed with regular unlabeled complimentary reverse oligonucleotides, annealed and further purified using MonoQ 10/100 column for surface plasmon resonance binding studies.

Surface plasmon resonance

Protein and DNA samples were prepared in standard buffer (10 mM HEPES pH 6.5, 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT). Binding affinities of MBD domain with 3′ biotinylated DNA variants were determined using a NLC sensor chip on ProteOn™ XPR36 (Bio-Rad). Biotinylated dsDNAs were immobilized to the ligand channels of NLC chip using biotin-streptavidin chemistry until the final response units were in the range of ∼100 RU, control channels were blocked without linking DNA. Various concentrations of $MBD4_{MBD}$ were passed over the analyte channels (at a flow rate of 30 μl/min) in running buffer (10 mM HEPES pH 7.4, 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.1%BSA, 0.005% polysorbate 20). Data analysis, plotting and curve fitting were performed with pro Fit software (QuantumSoft).

NMR spectroscopy

Purified protein was combined with 10% excess purified dsDNA and buffer exchanged into 10 mM $NaPO_4$ pH 6.5, 1 mM dithiothreitol, 10% $^{2}H_2O$ and 0.02% sodium azide and concentrated to 0.2–1 mM. NMR spectra from standard experiments for resonance assignments, distance and torsional angle restraints were collected on a Bruker Avance III 700 MHz instrument at 25°C. Data were processed using NMRPipe (12) and analyzed with CcpNmr (13). Residual dipolar couplings were measured for complexes containing ^{2}H,^{13}C,^{15}N labeled protein using standard IPAP experiments and samples aligned by adding ∼12 mg/ml pf1 bacteriophage (Asla Biotech, Ltd.) (14,15).

Structure calculations

The solution structure was determined for $MBD4_{MBD}$ bound to a dsDNA fragment with a central symmetrically methylated ^{m}CpG dinucleotide. Initially, a complete

NMR dataset was collected for $MBD4_{MBD}$ bound to a 17 bp fragment derived from the $p16^{INK4a}$ promoter as described previously (16). These data revealed significant line-broadening for residues at the protein–DNA interface which suggested dynamic exchange between binding modes. To limit intramolecular exchange, a second dataset was collected for $MBD4_{MBD}$ bound to a 10 bp fragment of ds-DNA (GGAT^{m}CGGCTC) identical to that in the solution structure of cMBD2 (11). This dataset showed a significant reduction of line-broadening and was used for all subsequent structure calculations and analyses. The structure was calculated by standard simulated annealing techniques using the Xplor-NIH software package (17) and minimized against a target function that included Nuclear Overhauser Effect (NOE)-derived interproton distances, torsion angles and residual dipolar couplings restraints; a quartic van der Waals repulsion term for the non-bonded contacts (18); a torsion angle data base potential of mean force (19) and a radius of gyration restraint to ensure optimal packing (20). Backbone torsion angle restraints were derived from chemical shift indexing as implemented by TALOS-N software (21,22). $^{3}J_{N\text{-}C}$ and $^{3}J_{CO\text{-}C}$ coupling constants were measured to determine a limited number of sidechain torsion angle restraints. Hydrogen bond distance and angle restraints were incorporated for regions of secondary structure as predicted by TALOS-N and confirmed by characteristic NOE crosspeak patterns. Based on chemical shift changes observed for Arg^{97} $^{15}N\epsilon$ (as described previously (11,16)), hydrogen bond distance and angle restraints between Arg^{97} $H\epsilon$/HH21 and Asp^{107} $O_{\delta 1}/O_{\delta 2}$, respectively, were incorporated in structure calculations.

The DNA assignments and NOE restraints were derived from double filtered (^{13}C,^{15}N) homonuclear NOE experiments and by comparison with previous spectra and assignments for cMBD2 bound to the same DNA (11). Assignments of the key 5-methyl-cytosine H5 protons were confirmed by the presence of strong NOEs between Thy^{204} H6 and both Thy^{204} H5 and $^{m}Cyt^{205}$ H5. In addition to the NOE restraints, hydrogen bond distance and planarity restraints as well as B-form DNA backbone torsion angle restraints were incorporated into structure calculations. Intermolecular protein–DNA NOEs were derived from an isotope-filtered 3D ^{13}C HMQC-NOESY spectrum. As was described previously (11), hydrogen bond distance and angular restraints were incorporated between Arg^{97}/Arg^{109} NH_2 and Gua^{206}/Gua^{216} O6 and N7, respectively, in the final simulated annealing calculations.

Molecular dynamics simulations

A B-form methylated DNA structure comprising the 10 bp sequence with three additional cytosine bases on each end for added stability was generated using the 3D-DART webserver (23) and Visual Molecular Dynamics software (24). The DNA and preliminary $MBD4_{MBD}$ structures were docked and solvated in a box with 10 Å of TIP3P water surrounding the complex and 75 mM NaCl. All dynamics simulations were carried out using NAMD 2.9 (25) and the CHARMM27 (26) force field. The system was equilibrated with two rounds of preparative constant number, volume, and temperature (NVT) simulations. In the first round, 5000 steps of minimization were followed by 30 ps of dynamics ps at 300 K with all atoms of the protein–DNA complex held rigid. In the second round, the same protocol was used except that the rigid restraints on the protein–DNA complex were replaced by a harmonic restraint (5 kcal/mol/Å) applied to the backbone atoms of the protein–DNA complex. Finally a 1 ns preparative constant number, pressure, and temperature (NPT) simulation was carried out on the system using a Langevin piston barostat (27,28) with a target pressure of 1.01325 bar, a decay period of 100 fs and a piston temperature of 300 K followed by 40 ns of unrestrained NPT dynamics.

Nz-exchange spectroscopy

Two separate samples comprising ^{2}H,^{15}N labeled $MBD4_{MBD}$ at a final concentrations of 200 μM and 333 μM were combined with a mixture of wild-type and inverted 10 bp methylated DNA at final molar ratios of 1:2:2 (protein:wild-type DNA:inverted DNA). A series of 2D ^{1}H-^{15}N TROSY based N_z-exchange spectra (29) were collected with total exchange delays of 11.9, 14.3, 16.8, 21.8, 29.3, 36.8, 49.3, 61.8, 111.8 and 211.8 ms. The data were processed as pseudo-3D spectra using a Lorentz-to-Gaussian window function in both dimensions and the spectra fit using an automated lineshape fitting algorithm within the NMRPipe software (12). The intensities for the auto and exchange crosspeaks were fit to four coupled equations describing chemical exchange in the slow exchange limit (30) using pro Fit software (Quantum Soft). For comparison, two samples of ^{1}H,^{15}N labeled $cMBD2_{MBD}$ at final concentrations of 185 μM and 370 μM were combined with a mixture of wild-type and inverted 10 bp methylated DNA at final molar ratios of 1:2:2 (protein:wild-type: inverted DNA). A series of 2D ^{1}H-^{15}N N_z-exchange spectra (non-TROSY versions) were collected and analyzed in a similar manner.

RESULTS

Solution structure of $MBD4_{MBD}$ bound to methylated DNA

We determined the solution structure of the methyl binding domain from human MBD4 (amino acids 80–148) bound to a 10 bp DNA fragment containing a central symmetrically methylated ^{m}CpG dinucleotide. This 10 bp DNA was derived from the chicken ρ globin promoter and was used previously in structural studies of $cMBD2_{MBD}$ (11) (Table 1). The structure was well determined with an overall root mean square deviation (RMSD) from the mean for the backbone of the $MBD4_{MBD} = 0.7 \pm 0.1$ Å and for the complex $= 1.2 \pm 0.3$ Å (Table 2, Figure 1a). The structure comprises a 3-strand β-sheet that extends down and across the major groove and includes many of the DNA contacts (Figure 1b). The loop connecting the first two β-strands shows conformational heterogeneity within the ensemble of structures, which is similar to that seen for MBD3 (16). An α-helix follows the β-sheet which is similar in length to the same helix in MeCP2 and longer than for MBD1, cMBD2 and MBD3 (Figure 1c). The length of this helix reflects a four amino acid insertion common to both MeCP2 and MBD4, which results in a larger hydrophobic core.

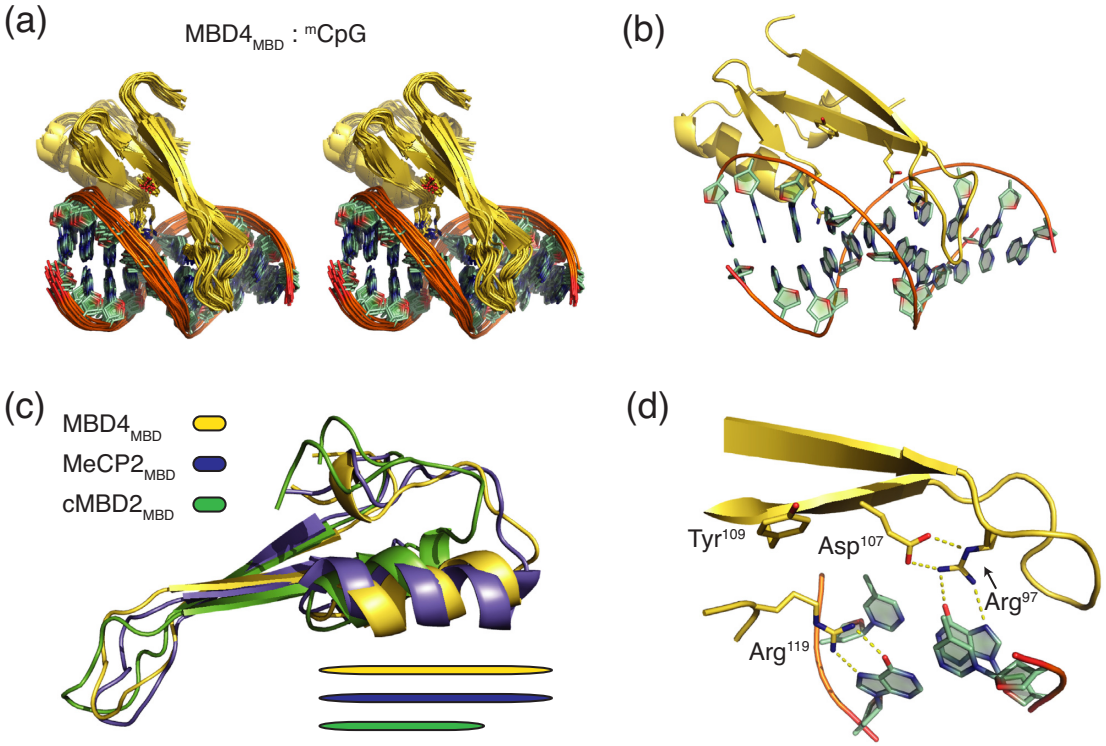

Figure 1. Solution structure of $MBD4_{MBD}$ bound to methylated DNA. (**a**) A best-fit superimposition stereo cartoon diagram depicts the twenty calculated structures of the $MBD4_{MBD}$ (yellow) and DNA (green) with key interacting protein residues shown as stick diagrams. (**b**) A cartoon diagram depicts the lowest energy structure of $MBD4_{MBD}$ (yellow) and DNA (green) with key interacting protein residues shown as stick diagrams. (**c**) A best-fit superimposition cartoon diagram of the $MBD4_{MBD}$ (yellow), $MeCP2_{MBD}$ (blue) and $cMBD2_{MBD}$ (green) shows that the C-terminal α-helix is longer in MeCP2 and MBD4 as compared to cMBD2. Horizontal bars below the diagram highlight the relative lengths of this α-helix in each. (**d**) An expanded mixed rendering diagram shows critical interactions involved in DNA binding with key residues shown as sticks and potential hydrogen bonds highlighted with yellow dashed lines. Note that Tyr^{109} points away from DNA, which differs from other MBD proteins.

Table 1. Oligonucleotide sequences

Oligonucleotide	Sequence
^{m}CpG (17 bp)	5′-GAGGCGC T^{m}CGG CGGCAG-3′ 3′-CTCCGCG AG^{m}CC GCCGTC-5′
CpG (17 bp)	5′-GAGGCGC TCGG CGGCAG-3′ 3′-CTCCGCG AGCC GCCGTC-5′
^{hm}CpG (17 bp)	5′-GAGGCGC T^{hm}CGG CGGCAG-3′ 3′-CTCCGCG AG^{hm}CC GCCGTC-5′
^{m}CpG/TpG (17 bp)	5′-GAGGCGC T^{m}CGG CGGCAG-3′ 3′-CTCCGCG AGTC GCCGTC-5′
^{m}CpG (10 bp)	5′-GGA T^{m}CGG CTC-3′ 3′-CCT AG^{m}CC GAG-5′
Inverted (10 bp)	5′-GGA C^{m}CGA CTC-3′ 3′-CCT GG^{m}CT GAG-5′
Tandem (30 bp)	5′-CACGGA T^{m}CGG CT CCCC CGAG T^{m}CGG TCCCGC-3′ 3′-GTGCCT AG^{m}CC GA GGGG GCTC AG^{m}CC AGGGCG-5′
Tandem nicked (30 bp)	5′-CACGGA T^{m}CGG CT CCCC CGAG T^{m}CGG TCCCGC-3′ 3′-GTGCCT AG^{m}CC GA GG - - GCTC AG^{m}CC AGGGCG-5′
Tandem (20 bp)	5′-GGA T^{m}CGG CTC GGA C^{m}CGA CTC-3′ 3′-CCT AG^{m}CC GAG CCT GG^{m}CT GAG-5′

Only a relatively few intermolecular NOEs between protein and DNA could be measured which is consistent with an interface dominated by solvent mediated interactions and few base specific contacts (Figure 1d). The base specific interactions primarily involve two arginine side chains (Arg^{97} and Arg^{119}) that establish bidentate hydrogen bonds with the symmetrically related bases of the CpG dinucleotide. A highly conserved aspartate (Asp^{107}) makes side chain hydrogen bonds and stabilizes Arg^{97} in the appropriate conformation. Unlike other MBD proteins, a critical tyrosine residue (Tyr^{109}) no longer points towards the methyl group of mCyt, instead this residue points towards the phosphate backbone and solvent. The latter observation is confirmed both by NOEs detected between Tyr^{109}

Table 2. NMR and refinement statistics

	Protein	Nucleic acid
NMR distance and dihedral constraints		
Distance restraints		
Total NOE	487	115
Intraresidue	105	68
Inter-residue	382	47
Sequential ($\|i-j\|=1$)	149	30
Non-sequential ($\|i-j\|>1$)	233	17
Hydrogen bonds	18	37
Hydrogen bonds protein–nucleic acid	4	
Protein–nucleic acid intermolecular	12	
Total dihedral angle restraints		
Protein		
ψ	48	
φ	47	
χ^1	15	
Nucleic acid		
Backbone		120
Sugar pucker		20
RDC Q% (number of constraints)		
NH	1.0 ± 0.4 (37)	
Structure statistics		
Violations (mean and s.d. for the complex)		
Distance constraints (Å)	0.035 ± 0.004	
Dihedral angle constraints (°)	0.34 ± 0.04	
Maximum dihedral angle violation (°)	2.2	
Maximum distance constraint violation (Å)	0.39	
Deviations from idealized geometry		
Bond lengths (Å)	0.0027 ± 0.0002	
Bond angles (°)	0.546 ± 0.008	
Impropers (°)	0.32 ± 0.01	
Average pairwise RMS deviation[a] (Å)		
Protein		
Heavy	1.3 ± 0.2	
Backbone	0.7 ± 0.1	
DNA		
Heavy		0.3 ± 0.1
Backbone		0.5 ± 0.1
Complex		
Heavy	1.4 ± 0.3	
Backbone	1.2 ± 0.3	
Ramachandran plot summary[a]		
Most favored regions	94.0%	
Additionally allowed regions	5.9%	
Generously allowed regions	0.1%	
Disallowed regions	0.0%	

[a]Pairwise RMS deviation and Ramachandran plot summary was calculated among 20 refined structures for structured residues (amino acids 86–140 of $MBD4_{MBD}$ and base pairs 202–212 of DNA).

and Val^{93}/Lys^{95} as well as J-coupling measurements for the χ^1 torsion angle of Tyr^{109} ($J_{C'C\gamma} \approx 0$ Hz). As we previously described in the solution structure of cMBD2:dsDNA, base specific intermolecular NOEs between $MBD4_{MBD}$ and DNA indicate that the protein binds predominantly in a single orientation on DNA. Distinct NOEs between Arg^{97} and $mCyt^{205}$ and between Arg^{119} and $mCyt^{215}$ can only be satisfied by a single orientation on the DNA. In addition, we detected intermolecular NOEs between Thr^{102} and both Gua^{206} and Gua^{207} that confirm this orientation.

Molecular dynamics simulations of the $MBD4_{MBD}$:dsDNA complex

To further characterize the $MBD4_{MBD}$:DNA interface and better understand the orientation preference of Tyr^{109}, we carried out molecular dynamics simulations of the $MBD4_{MBD}$:dsDNA complex. The two key arginine residues Arg^{97} and Arg^{119} established persistent bidentate hydrogen bonds with the symmetrically related guanosine bases of the CpG dinucleotide Gua^{206} and Gua^{216}, respectively (Figure 2a), which is consistent with the experimental solution structure. Likewise the side chain of Asp^{107} also forms persistent bidentate hydrogen bonds with Arg^{97} (Figure 2a). Plots of distances between hydrogen bond donors and acceptors for these interactions show that these hydrogen bonds are established early and broken only intermittently during the simulation (Figure 2c). In addition, Tyr^{109} maintains a *gauche(+)* χ^1 torsion angle throughout the simulation as opposed to the *gauche(−)* χ^1 torsion angle in MeCP2 (Figure 2b). As can be seen in Figure 2b, Tyr^{109} fits within a hydrophobic pocket formed by the side chains of Val^{93}, Lys^{95}, Ile^{111} and Lys^{117}. The conformation of these side chains is stabilized by interactions between the lysine ϵ-

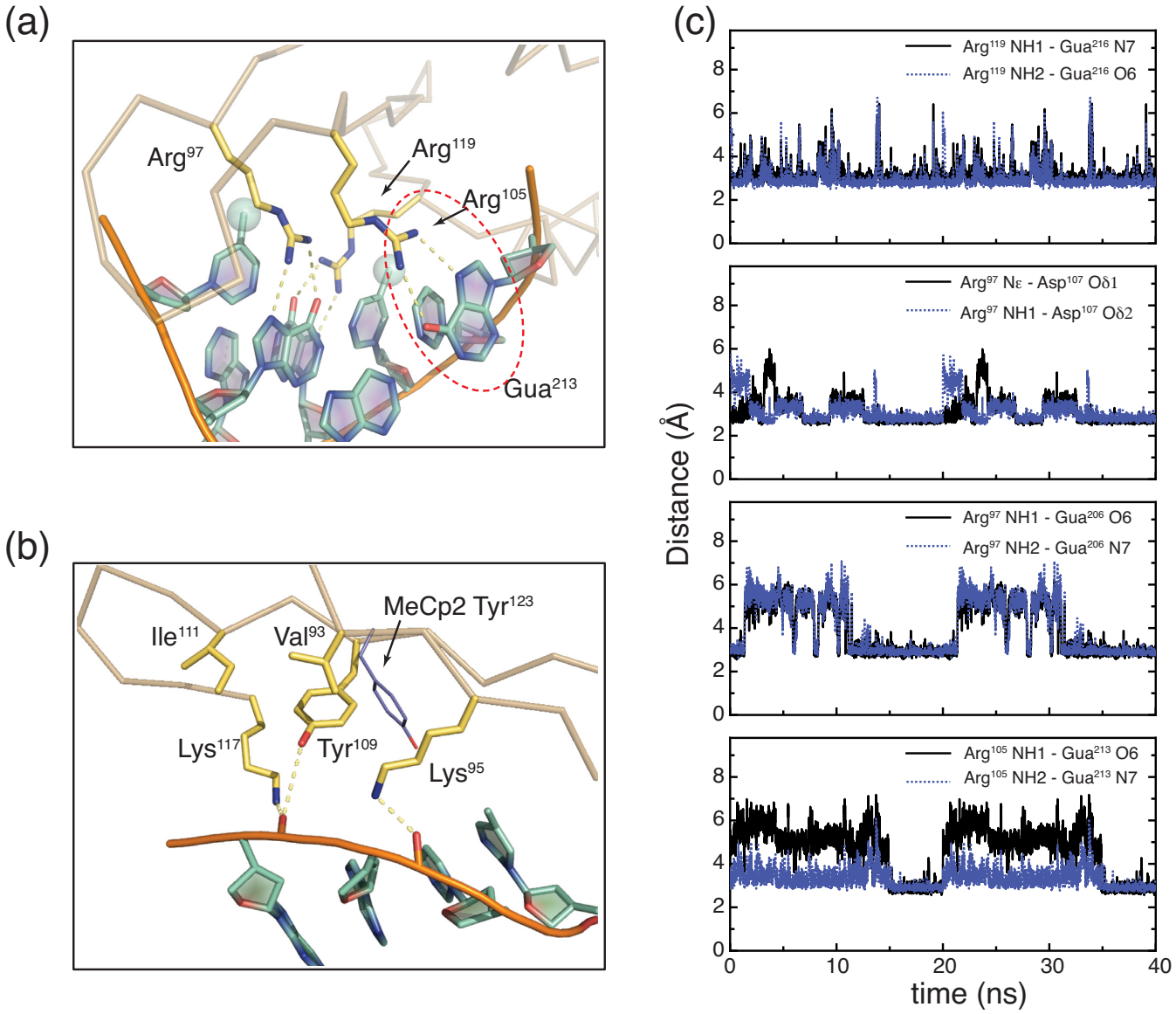

Figure 2. Molecular dynamics simulation highlights critical $MBD4_{MBD}$–DNA interactions. (**a-b**) Expanded mixed rendering diagrams from the final frame of a 40 ns molecular dynamics simulation show key interactions with DNA. Basic specific interactions between arginine residues and DNA are highlighted in (a) including the potentially orienting interaction between Arg^{105} and Gua^{213}. Methyl groups of ^{m}Cyt are depicted as translucent spheres while potential hydrogen bonds are shown as yellow dashed lines. Residues that stabilize the *gauche(+)* χ^1 torsion angle of Tyr^{109} are highlighted in (b) with potential hydrogen bonds to the phosphate backbone of DNA shown as yellow dashed lines. For comparison the orientation of the equivalent tyrosine in MeCP2 (Tyr^{123}) is shown in blue. (**c**) Distances between specific heavy atoms are plotted every 10 ps of the 40 ns molecular dynamics simulation. These distances are shown for potential bidentate hydrogen bonds between Arg^{119} and Gua^{216}, Arg^{97} and Asp^{107}, Arg^{97} and Gua^{206} and Arg^{105} and Gua^{213}.

amino groups and the phosphate backbone of DNA as well as between the tyrosine hydroxyl and the phosphate backbone of DNA.

We previously showed that Lys^{32} of $cMBD2_{MBD}$ makes base specific interactions with Gua^{107} which contributes to the binding affinity and preferred orientation on an $^{m}CpGG$ trinucleotide (11). Since $MBD4_{MBD}$ binds in a similar orientation on this same DNA sequence, we hypothesized that the equivalent residue, Arg^{105}, would make similar base specific interactions. During the molecular dynamics simulation, the side chain of Arg^{105} reaches across the major groove to form intermittent bidentate hydrogen bonds with Gua^{213} (Figure 2b). This base specific interaction differs from that observed $cMBD2_{MBD}$ in that the guanosine is on the opposite strand and one base removed from the CpG dinucleotide while the base specific interaction identified in cMBD2 involved Gua^{107}, which immediately follows the CpG dinucleotide. The interaction between Arg^{105} and Gua^{213} helps explain the preferred orientation on DNA since Ade^{203}, which has only one hydrogen bond acceptor, occupies the equivalent position to Gua^{213} if $MBD4_{MBD}$ were to bind in the reverse orientation. A different and palindromic DNA sequence was used in a recently deposited crystal structure of human $MBD4_{MBD}$ such that

Arg^{105} hydrogen bonds to O4 of thymidine immediately following the CpG dinucleotide (PDB ID: 4LG7). In contrast, lysine occupies the equivalent position in a crystal structure of mouse MBD4$_{MBD}$ which interacts with the phosphate backbone instead of making a base specific interaction (31).

Comparison with other MBDs

Structural information is now available for most of the MBD family of proteins (MeCP2, MBD1–4). The crystal structure of MeCP2$_{MBD}$ (32) and the NMR structures of MBD1$_{MBD}$ (33), MBD2$_{MBD}$ (11) and MBD3$_{MBD}$ (16) have previously been reported. More recently the crystal structures of mouse MBD4$_{MBD}$ bound to methylated, hydroxymethylated and mismatch DNA were reported (31) and a crystal structure of human MBD4$_{MBD}$ bound to methylated DNA was deposited in the RCSB (PDB ID: 4LG7) by the Structural Genomics Consortium Toronto. All of the MBD structures show similar folds and demonstrate that the key DNA contacting residues are largely conserved and form a similar DNA recognition interface (Figure 1). The backbone RMSD between the solution structure of MBD4$_{MBD}$ and other MBDs is as follows: MeCP2 - 1.7 Å; MBD1 - 2.0 Å; cMBD2 - 2.5 Å; MBD3 - 2.3 Å; human MBD4$_{MBD}$ crystal structure - 1.5 Å; mouse MBD4$_{MBD}$ (excluding a small C-terminal helix) bound to ^{m}CpG - 1.8 Å, ^{hm}CpG - 1.8 Å, ^{m}CpG/TpG mismatch - 1.8 Å; and mouse MBD4$_{MBD}$ bound to both a ^{m}CpG/TpG mismatch and a non-specific complex - 1.6 Å. The MBD family can be divided into two groups (MBD1/2/3 and MeCP2/MBD4) reflecting the four amino acid insertion described above. Because of this insertion, both MBD4$_{MBD}$ and MeCP2$_{MBD}$ contain a longer C-terminal α-helix that provides a larger hydrophobic core. Consistent with this observation, both MeCP2 (34) and MBD4 adopt a regular folded structure in isolation as opposed to MBD2 and MBD3 which undergo a disorder to order transition upon binding DNA (11,16). 2D ^{15}N-HSQC spectra of the isolated MBD4$_{MBD}$ (not bound to DNA) contains sharp well-dispersed peaks indicative of a folded domain while that of cMBD2 does not (Supplementary Figure S1).

A few changes at the protein–DNA interface potentially explain binding affinity and methylation selectivity differences among the MBDs. Both mouse and human crystal structures of MBD4$_{MBD}$ as well as our solution structure of MBD4 show the same reorientation of Tyr^{109,} which plays a critical role in methylation specific DNA binding for MBD2 and MBD3. As described by Otani *et al.* (31), this change in Tyr^{109} conformation opens up a large solvent accessible cavity at the protein–DNA interface. In addition, this change removes one of the key interactions that drives binding selectivity for methylated DNA which indicates that MBD4$_{MBD}$ would not show the same level of methylation selectivity as cMBD2$_{MBD}$ and mutating this Tyr^{109} would not reduce binding affinity and selectivity as seen with other MBDs. We test both of these hypotheses in the binding analysis reported below.

One of the more notable differences between the crystal structures of mouse MBD4$_{MBD}$ reported by Otani *et al.* (31) and the solution structure reported here is that a small C-terminal helical region adopts an extended conformation and swaps positions with symmetry related molecules in three of the crystal structures (PDB IDs: 3VXX, 3VXV and 3VYB). This same region adopts the more typical compact monomeric fold in the solution structure reported here as well as in the fourth mouse MBD4$_{MBD}$(PDB ID: 3VYQ) and human MBD4$_{MBD}$ crystal structures. To confirm the relative orientation of this region, we fit the residual dipolar couplings measured for MBD4$_{MBD}$ in solution to the crystal structure of mouse MBD4$_{MBD}$ bound to methylated DNA using singular value decomposition as implemented by PALES software (35). If we included the entire dataset, the quality of fit was poor ($Q = 78.9\%$) but if we excluded values from the C-terminal helical region, the fit markedly improves($Q = 26.7\%$). Hence residual dipolar coupling measurements confirm the monomeric fold in solution which supports the interpretation by Otani *et al.* (31) that apparent dimerization of mouse MBD4$_{MBD}$ reflects a crystal lattice induced artifact.

MBD4$_{MBD}$ preferentially binds methylated over mismatch, hydroxymethylated and unmethylated DNA

The binding affinity of MBD4$_{MBD}$ for modified and unmodified 17 bp oligonucleotides (Table 1) was determined by surface plasmon resonance (Figure 3) and the results are summarized in Table 3. Although the overall affinity for DNA is relatively weak, MBD4$_{MBD}$ shows a modest preference for methylated DNA (K_D ∼6.4 μM) over ^{m}CpG/TpG mismatch (K_D ∼11.5 μM), hydroxymethylated (K_D ∼14.2 μM) and unmethylated DNA (K_D ∼17.2 μM). These findings are similar to those reported by Hashimoto *et al.* (36) for human MBD4$_{MBD}$ and suggest that the reorientation of Tyr^{109} reduces the overall affinity and methylation selectivity. Hence MBD4$_{MBD}$ can bind a variety of modified CpGs with comparable affinity and only modest selectivity for ^{m}CpG. Based on these observations, we hypothesized that a Y109F mutation in MBD4$_{MBD}$ would not affect binding affinity to the same degree as observed for other MBDs (16,37,38). As shown in Figure 3 and Table 3, MBD4$_{MBD(Y109F)}$ binds methlylated DNA with similar overall affinity (K_D ∼7.9 μM) and relative selectivity over ^{m}CpG/TpG mismatch (K_D ∼14.0 μM), hydroxymethylated (K_D ∼16.0 μM) and unmethylated (K_D ∼19.6 μM) DNA. These findings contrast with the large reduction in binding affinity and reduction in methylation selectivity associated with the same modification in cMBD2 and the differences in binding specificity between MBD3 and MBD2 (16,37,38).

We previously showed that the ^{15}N, ^{1}H chemical shifts of select residues in the MBD reflected a distribution between CpG specific and non-specific binding modes of MBD2 and MBD3. Hence we compared the chemical shifts of the same residues in MBD4$_{MBD}$. As described previously for MBD2, the ^{1}Hε chemical shift of Arg^{97} is shifted far downfield to ∼9.6 ppm when bound to ^{m}CpG (Supplementary Figure S2), which is consistent with stabilization of the side chain hydrogen bond with Asp^{107}. When MBD4$_{MBD}$ binds unmethylated DNA, this peak shifts ∼0.8 ppm upfield to ∼8.8 ppm and when bound to hydroxymethylated DNA, the same peak falls between the two extrema at ∼9.1 ppm. Likewise, the ^{15}N chemical shift of Gly^{100} is

Table 3. DNA binding affinity

	K_D (μM)	R_{max}	X^2 (10^{-4})
$MBD4_{MBD}$-^{m}CpG	6.4 ± 1.5	0.74	34
$MBD4_{MBD}$-TpG	11.5 ± 1.2	0.59	2.3
$MBD4_{MBD}$-CpG	17.2 ± 2.0	0.91	3.5
$MBD4_{MBD}$-^{hm}CpG	14.2 ± 1.9	0.72	3.9
$MBD4_{MBD(Y109F)}$-^{m}CpG	7.9 ± 1.6	0.63	14
$MBD4_{MBD(Y109F)}$-TpG	14.0 ± 1.4	0.51	1.0
$MBD4_{MBD(Y109F)}$-CpG	19.6 ± 3.5	0.76	3.5
$MBD4_{MBD(Y109F)}$-^{hm}CpG	16.0 ± 3.0	0.59	3.4

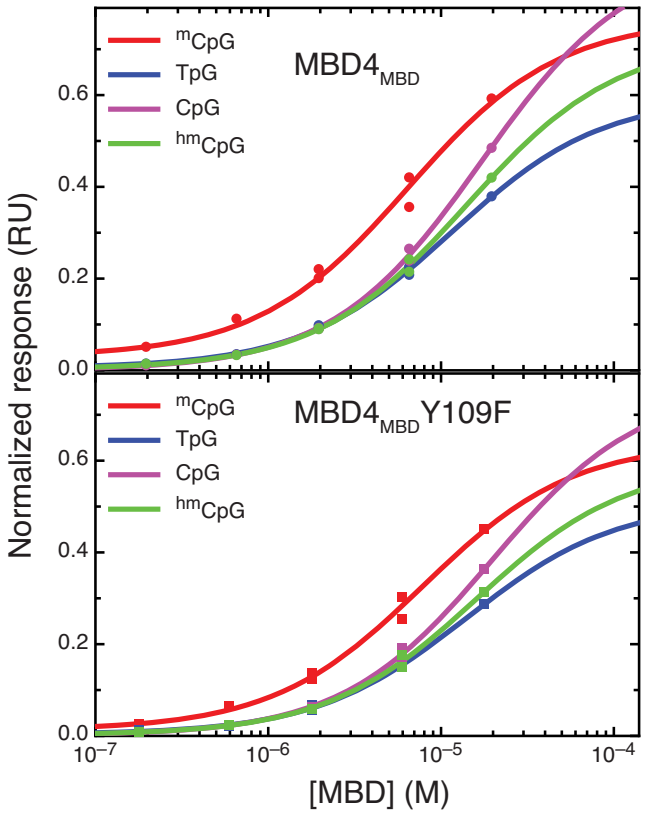

Figure 3. $MBD4_{MBD}$ shows a modest binding preference for methylated DNA. Steady state response curves are shown for surface plasmon resonance analyses of wild-type $MBD4_{MBD}$ (upper panel) and mutant $MBD4_{MBD(Y109F)}$ (lower panel) binding to methylated (green), mismatch (blue), unmethylated (purple) and hydroxymethylated (green) DNA. Both wild-type and mutant $MBD4_{MBD}$ show a similar modest preference for methylated DNA.

shifted far upfield when bound to ^{m}CpG (Supplementary Figure S2) and is shifted downfield by ∼1 ppm when bound to unmethylated and hydroxymethylated DNA. Interestingly the peak for Gly^{100} ^{15}N-^{1}H is shifted upfield while the peak for Arg^{97} $^{15}N\epsilon$-$^{1}H\epsilon$ is broadened and not observed when bound to ^{m}CpG/TpG mismatch. Together these findings support the relative binding affinities for the different modified substrates and indicate that, like MBD3 and MBD2, $MBD4_{MBD}$ distributes between CpG specific and non-specific binding modes when bound to DNA.

Intermolecular and intramolecular exchange of $MBD4_{MBD}$ between methylated sites

We previously established that exchanging the bases immediately flanking the mCpG causes the $MBD2_{MBD}$ to reverse orientation on the DNA (11). In doing so, we noted that binding to the inverted sequence led to chemical shift changes in select backbone amide (^{15}N, ^{1}H) resonances. 2D ^{15}N-^{1}H HSQC spectra of ^{15}N-$MBD4_{MBD}$ also show distinct chemical shifts for selected reporter residues, Arg^{105} and Phe^{106}, (Figure 4) when bound to DNA with the central four bases in the inverted orientation (Table 1). This observation allows us measure chemical exchange between these closely related ^{m}CpG binding sites by N_z-exchange NMR spectroscopy (29). When bound to a mixture of the wild-type and inverted sequences (1:2:2 molar ratio of $MBD4_{MBD}$, wild-type and inverted DNA), two distinct peaks were observed for these same reporter residues (Figure 4a and b) which indicate that $MBD4_{MBD}$ exchanges slowly on the NMR timescale between two DNA molecules.

To measure the intermolecular exchange rate, we collected ^{1}H-^{15}N TROSY based N_z-exchange spectra (29) with exchange delays ranging from 11.9–211.8 ms. As can be seen in Figure 4b exchange crosspeaks, AB and BA (which represent exchange from wild-type to inverted or inverted to wild-type sequences, respectively) buildup with increasing delay. The intensities of the auto and exchange peaks were fit to four coupled equations describing chemical exchange in the slow exchange limit (Figure 4c) (30). These data show that at 200 μM protein and 800 μM total DNA intermolecular exchange of $MBD4_{MBD}$ occurs with a mean lifetime of ∼40 ms. Increasing the protein and total DNA concentration to 333 μM and 1.33 mM, respectively, shortens the mean lifetime to ∼23 ms as would be expected for intermolecular exchange.

To test whether $MBD4_{MBD}$ exchanges more rapidly between two binding sites in the same DNA molecule, we bound $MBD4_{MBD}$ to a 30 bp oligonucleotide containing both wild-type and inverted mCpG sequences separated by 12 bp (Table 1) at a protein concentration of 200 μM and total ^{m}CpG binding site concentration of 800 μM. A 2D TROSY ^{15}N-HSQC of this complex no longer showed distinct peaks for the two bound states (Figure 5b). Instead the spectrum contains only a single peak for Arg^{105} located at approximately 62% of the distance between the peaks for same residue in the wild-type and inverted DNA complexes. Assuming the chemical shifts for $MBD4_{MBD}$ bound to either the wild-type or inverted sequences do not change

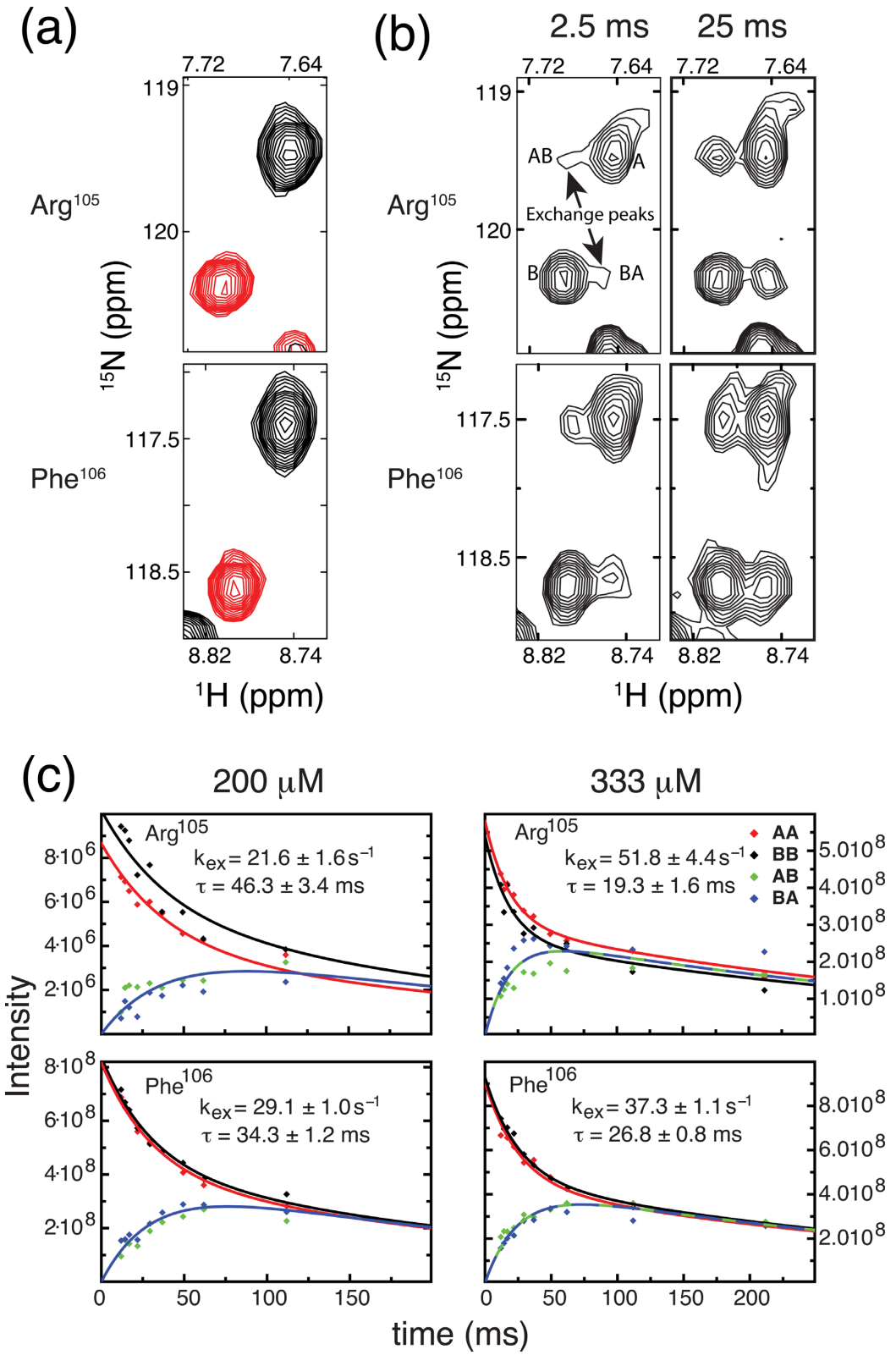

Figure 4. $MBD4_{MBD}$ exchanges between molecules of DNA on the slow NMR timescale. (**a**) 2D ^{1}H-^{15}N HSQC spectra of $MBD4_{MBD}$ bound to wild-type (black) and inverted (red) DNA sequences show that the chemical shifts for Arg^{105} and Phe^{106} differ between the two complexes. (**b**) 2D ^{1}H-^{15}N N_z-exchange spectra of $MBD4_{MBD}$ bound to a 1:1 mixture of wild-type and inverted (10 bp) sequences shows two distinct peaks for Arg^{105} and Phe^{106} which represent between binding to the different DNA sequences. The exchange crosspeaks, AB and BA, increase in intensity with increasing exchange delay (2.5 and 25 ms) and represent intermolecular exchange between DNA molecules. (**c**) The intensities for auto (AA and BB) and exchange (AB and BA) crosspeaks for Arg^{105} (upper panels) and Phe^{106} (lower panels) were fit to four coupled equations to determine the rate of intermolecular exchange at 200 μM (left panels) and 333 μM (right panels) $MBD4_{MBD}$ concentration.

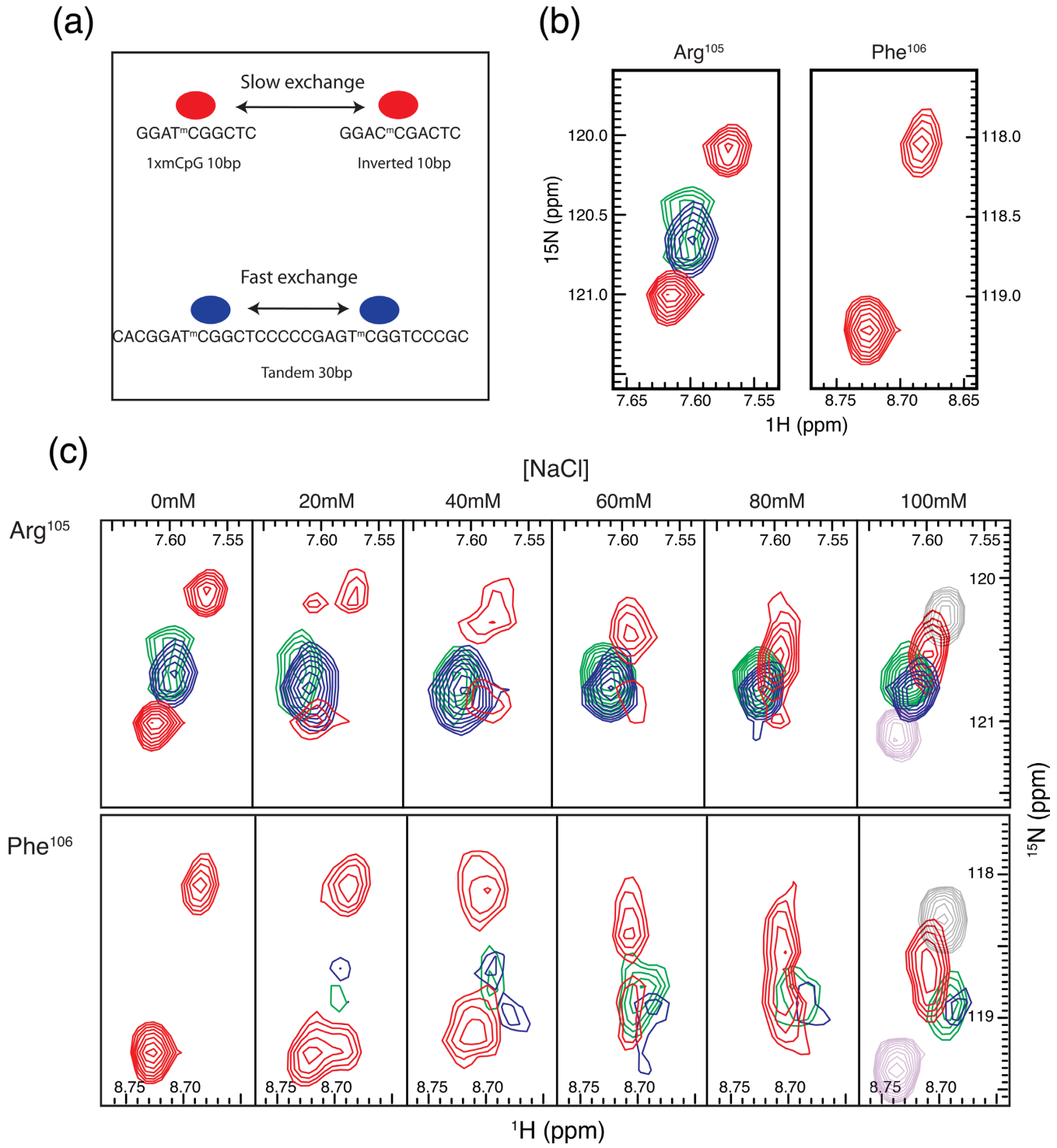

Figure 5. $MBD4_{MBD}$ exchanges between methylated sites in the same molecule of DNA on the fast NMR timescale. (**a**) A diagram depicting slow intermolecular and fast intramolecular exchange by $MBD4_{MBD}$. (**b**) 2D ^{1}H-^{15}N TROSY HSQC spectra of $MBD4_{MBD}$ bound to methylated wild-type and inverted (10 bp) DNA (red), tandem (30 bp) DNA (blue) and tandem nicked (30 bp) DNA (green) show that $MBD4_{MBD}$ more rapidly exchanges between sites in the same molecule of DNA. The spectra show only a single crosspeak for Arg^{105} (left panel) when bound to DNA containing both wild-type and inverted sites, which is consistent with fast intramolecular exchange. A similar comparison for Phe^{106} (right panel) shows marked broadening of the intramolecular crosspeaks consistent with intermediate exchange on the NMR timescale which likely reflects line-broadening from additional internal dynamic motions. Incorporating a single-strand defect does significantly alter rapid intramolecular exchange. (**c**) In contrast, increasing NaCl concentration accelerates both intermolecular and intramolecular exchange. The two separate peaks reflective of intermolecular exchange (red) for both Arg^{105} (upper panels) and Phe^{106} (lower panels) coalesce into a single peak with increasing NaCl concentration. This change indicates that the intermolecular exchange rate has increased from the slow to fast NMR timescale. Likewise the crosspeaks reflective of intramolecular exchange (blue and green) sharpen with increasing NaCl concentration, which is consistent with accelerated intramolecular exchange. For comparison, 2D ^{1}H-^{15}N TROSY HSQC spectra of $MBD4_{MBD}$ bound to methylated wild-type (gray) and inverted (purple) DNA are shown at 100 mM NaCl.

within the context of the 30 bp DNA, these findings indicate rapid exchange between the two binding sites. Hence the position of the crosspeak reflects a weighted average for the two binding modes, which suggests that $MBD4_{MBD}$ slightly favors binding to the inverted sequence, spending ~62% of time on that site. The crosspeak for Phe^{106} is not observed in the 30 bp complex likely reflecting additional internal dynamic motions leading to significant line broadening. Together these results show that $MBD4_{MBD}$ exchanges more efficiently between binding sites in the same molecule consistent with rapid facilitated diffusion along the DNA.

Rapid intramolecular exchange does not require continuous double-stranded DNA

To test whether a defect in the DNA can impact rapid intramolecular exchange by $MBD4_{MBD}$, we introduced a two-base stretch of single-stranded DNA between the ^{m}CpG binding sites (Table 1). A 2D TROSY ^{15}N-HSQC of $MBD4_{MBD}$ bound to the 30 bp nicked tandem DNA (Figure 5b, green) shows similar fast exchange peaks as seen for the 30 bp tandem DNA complex. Adding a small defect in double stranded DNA does not impede rapid intramolecular exchange of $MBD4_{MBD}$. This data indicate that $MBD4_{MBD}$ does not necessarily maintain continuous contact with the major groove of dsDNA during translocation, consistent with the local hopping mechanism of facilitated diffusion (39–43).

Increasing salt concentration accelerates both intermolecular and intramolecular exchange by $MBD4_{MBD}$

Intermolecular exchange rates for protein–DNA complexes often show a strong dependence on NaCl concentration (44). Therefore we evaluated the effect of increasing NaCl concentrations on exchange rates for $MBD4_{MBD}$ when bound to (i) 1:1 mixture of wild-type and inverted DNA, (ii) 30 bp tandem DNA and (iii) 30 bp nicked tandem DNA. At very low salt concentrations, Arg^{105} (Figure 5c, top panel) shows slow intermolecular exchange and rapid intramolecular exchange as described above. With increasing concentrations of NaCl, the intramolecular peaks for Arg^{105} become sharper (blue and green) indicating an increase in exchange rate. Concomitantly the intermolecular exchange peaks for Arg^{105} (red) undergo profound changes such that the two peaks observed at low salt concentrations (slow exchange) merge into one peak (fast exchange) at an average chemical shift. Likewise, Phe^{106} (Figure 5c, bottom panel) shows slow intermolecular exchange and marked line broadening indicative of intermediate intramolecular exchange at low salt concentrations. With increasing concentrations of NaCl the intramolecular exchange peak for Phe^{106} sharpens and becomes detectable (blue and green) consistent with a transition from intermediate to fast exchange on the NMR timescale. At 100 mM NaCl concentration, the crosspeaks for the individual wild-type (gray, Figure 5c) and inverted DNA (purple, Figure 5c) complexes remain well separated while the intermolecular exchange peaks (red) collapse into a single peak indicating fast exchange. Importantly, these changes in intramolecular exchange kinetics are similar whether the DNA has a 2 bp defect or not (Figure 5c, green).

Although at 100 mM NaCl both exchange rates fall within the fast exchange time regime, the linewidths indicate that the intramolecular exchange rate remains much faster than the intermolecular exchange rate. As can be seen in Figure 5c and Supplementary Figure S3, the intermolecular exchange peaks (red) are much broader, especially in the ^{15}N dimension, than the intramolecular exchange peaks (blue and green). The ^{15}N linewidths at half-height for intramolecular exchange are 16.7 Hz (Arg^{105}) and 25.0 Hz (Phe^{106}) and for intermolecular exchange are 25.5 Hz (Arg^{105}) and 34.2 Hz (Phe^{106}) (Supplementary Figure S3), even though the molecular weight of $MBD4_{MBD}$ bound to the 30 bp tandem DNA (26 551 g/mol) is nearly double that of $MBD4_{MBD}$ bound to either the wild-type or inverted DNA sequences (14 159 g/mol).

For comparison, we measured intermolecular exchange kinetics for $cMBD2_{MBD}$ binding to the same DNA (Figure 6a). Consistent with an increased binding affinity, the mean lifetime for intermolecular exchange of $cMBD2_{MBD}$ is somewhat longer at similar protein concentrations (τ ~66 ms, 185 μM protein; τ ~34 ms, 370 μM protein). Spectra of $cMBD2_{MBD}$ bound to 20 and 30 bp tandem ^{m}CpG sequences (Figure 6b) show that intramolecular exchange changes from fast to slow exchange with a small increase in binding site separation. Hence $cMBD2_{MBD}$ does not exchange between binding sites as efficiently as $MBD4_{MBD}$.

DISCUSSION

MBD4 occupies a unique niche within the MBD family of proteins. It is the only member of this family that incorporates enzymatic activity within the same protein, contributes directly to DNA mismatch repair and does not specifically recruit histone deacetylase activity. The observation that the $MBD4_{GD}$ can recognize and repair mismatch in isolation raises questions about the functional role of the $MBD4_{MBD}$. Two alternative, yet not mutually exclusive, models of $MBD4_{MBD}$ can be proposed: (i) the $MBD4_{MBD}$ bind to the same site as the MBD_{GD} and thereby augment sequence specificity or enzymatic activity; or (ii) the $MBD4_{MBD}$ could target the protein to regions enriched for ^{m}CpG and by virtue of the long intervening linker allow the $MBD4_{GD}$ to identify and repair nearby mismatches. Our recent work on the dynamic distribution of MBD3 on DNA (16) suggests that a similar dynamic behavior of $MBD4_{MBD}$ would augment the scanning mechanism inherent in the second model of $MBD4_{MBD}$ function. Hence we studied the structure and dynamics of this protein on DNA.

The basic structural motifs and DNA binding interface of $MBD4_{MBD}$ are very similar to other MBDs, yet local structural differences modify binding specificity and affinity. Changing a single tyrosine to phenylalanine at the DNA interface in $MBD3_{MBD}$ dramatically reduces methylation selectivity and overall binding affinity. This same tyrosine in $MBD4_{MBD}$ (Tyr^{109}) changes orientation with respect to the DNA which correlates with reduced selectivity for ^{m}CpG. Binding analyses by surface plasmon resonance reveals a relatively low overall affinity for methylated DNA and only a modest preference for symmetrically methylated CpG over ^{m}CpG/TpG mismatch, hydroxymethlyated and unmethylated DNA. Consistent with the reorientation of

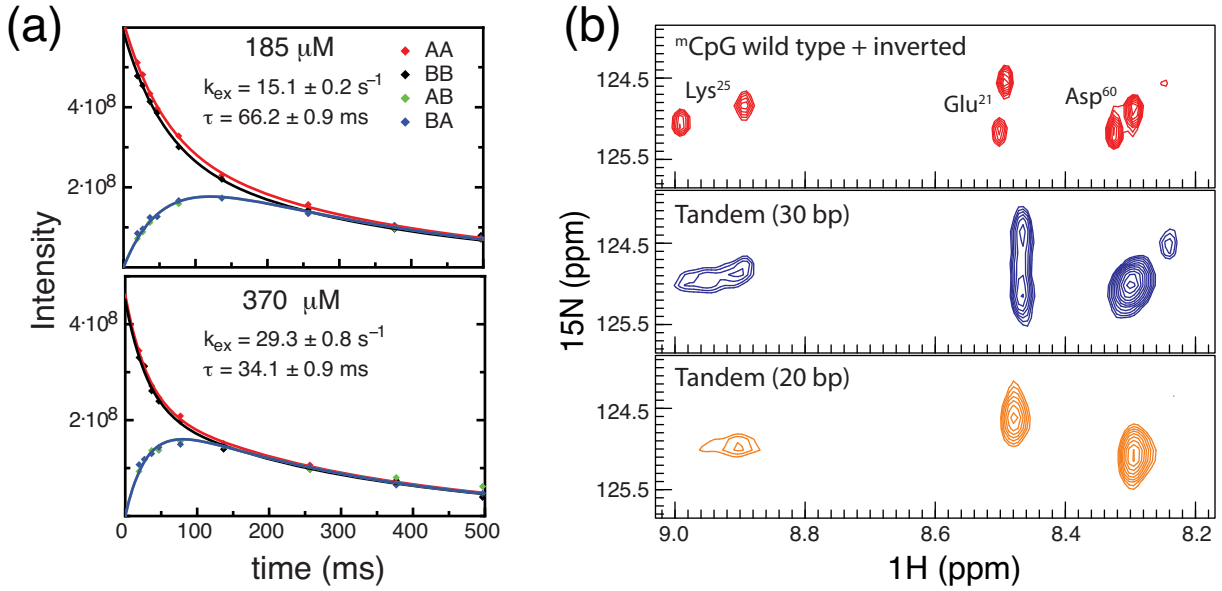

Figure 6. cMBD2$_{MBD}$ does not exchange between binding sites as efficiently as MBD4$_{MBD}$. (**a**) The intensities for auto (AA and BB) and exchange (AB and BA) crosspeaks for Glu^{21} from N_z-exchange spectra were fit to four coupled equations to determine the rate of intermolecular exchange at 185 μM (upper panel) and 370 μM (lower panel) cMBD2$_{MBD}$ concentration. (**b**) 2D ^{1}H-^{15}N TROSY HSQC spectra of cMBD2$_{MBD}$ bound to methylated wild-type and inverted (10 bp) DNA (red, upper panel), tandem (30 bp) DNA (blue, middle panel) and tandem (20 bp) DNA (orange, lower panel) show that intramolecular exchange for cMBD2$_{MBD}$ remains in the slow to intermediate NMR timescale when bound to the tandem (30 bp) DNA. If the sites are 4 bp closer together as in the tandem (20 bp), then cMBD2$_{MBD}$ exchanges between sites on the fast NMR timescale.

Tyr^{109} away from DNA, the Y109F mutation does not appreciably alter binding affinity and selectivity. The same mutation in cMBD2 reduced binding affinity by ∼50-fold. The structural rearrangement of Tyr^{109} places MBD4$_{MBD}$ between the highly ^{m}CpG selective MBD2$_{MBD}$ and the minimally ^{m}CpG selective MBD3$_{MBD}$.

As demonstrated by the crystal structures of mouse, MBD4$_{MBD}$ bound to a variety of modified substrates (31), a solvent filled protein–DNA interface allows MBD4$_{MBD}$ to adapt to different binding sites. We hypothesized that this adaptability and limited selectivity would promote efficient exchange between ^{m}CpG along the same DNA molecule. This hypothesis arises from the well-established model of facilitated diffusion by which DNA binding proteins use one-dimensional diffusion along the DNA to more rapidly identify a specific binding site (39). In Figure 7, a diagram depicting facilitated diffusion shows the different modes of exchange between DNA binding sites which include: (i) simple 3D diffusion through bulk solvent, (ii) jumping between nearby strands of DNA; (iii) sliding along the DNA between sites; and (iv) local hopping which does not required continuous contact with the DNA.

To study intermolecular and intramolecular exchange for MBD4$_{MBD}$, we took advantage of a previous observation that inverting the central 4 bp of the target DNA leads to chemical shift changes for select residues (11). Based on this observation, we used NMR spectroscopy to measure exchange between binding sites as described for other transcription factors (29,45–47). Following the peaks for these residues allowed us to measure the mean lifetime for intermolecular exchange (τ ∼40 ms at 200 μM protein) of MBD4$_{MBD}$. In contrast, exchange between two binding sites in the same molecule of dsDNA occurs on the fast exchange timescale which indicates that the mean lifetime is much less than the difference in chemical shift for the two states ($\tau << \Delta\nu$) (30). Hence, the mean lifetime can be estimated to be < 2 ms, which is at least a 20-fold in-

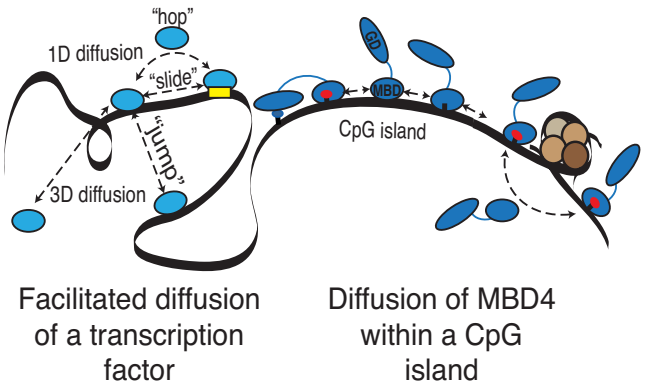

Figure 7. A diagram depicting a model of MBD4$_{MBD}$ function in which rapid intramolecular exchange allows MBD4 to efficiently scan CpG islands for ^{m}CpG/TpG mismatches. The left side of the diagram shows the different modes of searching for binding sites by transcription factors. The transcription factor (light blue circle) can search for binding sites by diffusion through bulk solvent (3D diffusion) or by different modes of facilitated diffusion (1D diffusion) including long-range jumps, local hops and sliding. The right side of the diagram shows the MBD4$_{MBD}$ (blue circle) exchanging between methylated (red ovals) and unmethylated CpG sites (black marks). Intramolecular exchange involves local hops that allow MBD4 to navigate around obstacles. The tethered MBD4$_{GD}$ (blue oval) identifies and repairs nearby ^{m}CpG/TpG mismatches (blue oval). In this manner, the MBD4$_{MBD}$ facilitates identification of ^{m}CpG/TpG mismatches within regions of increased CpG content.

crease in rate for intramolecular as compared to intermolecular exchange. Introducing a small stretch of single stranded DNA does not significantly reduce the intramolecular exchange rate. However, raising the NaCl concentration does increase both intermolecular and intramolecular exchange rates. Together, these data show that $MBD4_{MBD}$ efficiently exchanges between binding sites along the DNA and suggest that a local hopping mechanism contributes to the exchange process. It would be interesting to test whether efficient intramolecular exchange by MBD4 can be detected on a longer length scale using alternative approaches such as single molecule fluorescence measurements.

Small changes in the protein–DNA binding interface allow $MBD4_{MBD}$ to exchange more efficiently between successive binding sites along the same DNA molecule. This intramolecular exchange appears to involve a local hopping mechanism, which would allow $MBD4_{MBD}$ to avoid obstacles such as other DNA binding proteins on the DNA. This attribute is consistent with a model of $MBD4_{MBD}$ function in which the MBD targets the protein to regions of increased ^{m}CpG density and allows the glycosylase domain to scan nearby sites for ^{m}CpG/TpG mismatches. The disorder prediction algorithm PONDR® VLXT (48,49) indicates that the majority of the ∼280 amino acid linker separating the $MBD4_{MBD}$ and $MBD4_{GD}$ is disordered in solution (Supplementary Figure S4). The average end-to-end length of a 280 amino acid unstructured polypeptide is approximately 190 Å (as calculated from $(C_{\infty}nl^2)^{1/2}$ where C_{∞} is the limiting characteristic ratio (9.27), n is the number of residues and l the average Cα-Cα distance (3.8 Å)) (50,51). With this long linker, the two domains can span at least 50 bp (3.4 Å rise in B-DNA), which will contain many CpG dinucleotides in a CpG island. This arrangement allows the two domains to occupy distinct sites in the DNA such that the MBD could scan along the same molecule of DNA or hop between neighboring strands and thereby function to keep the protein in regions of increased mCpG density when the MBD_{GD} transiently diffuses off of the DNA.

In Figure 7, a diagram of this model depicts the $MBD4_{MBD}$ and $MBD4_{GD}$ as blue circle and oval, respectively, separated by a long linker. The $MBD4_{MBD}$ binds to DNA and exchanges among the different methylated (red circles), unmethylated (no circle) and ^{m}CpG/TpG mismatch (blue circle) sites. The exchange process is depicted as a combination of sliding local hopping events that avoids a protein obstacle while the $MBD4_{GD}$ identifies and repairs a nearby ^{m}CpG/TpG mismatch. This model helps explain why the $MBD4_{MBD}$ appears to impede (not augment) enzymatic activity when using a small synthetic oligonucleotide with a single T-G mismatch (4). The $MBD4_{MBD}$ competes with $MBD4_{GD}$ when there is only a single binding site in the substrate. In contrast, full-length MBD4 shows increased activity when a large nucleosome substrate with a single T-G mismatch is methylated and in the presence of competing unmethylated DNA (52). Hence the $MBD4_{MBD}$ helps localize the enzyme to methylated regions when there is a large amount of competing unmethylated DNA but does not augment activity on isolated mismatches. Since a T-G mismatch arises from spontaneous deamination of methyl-cytosine, preferential localization to methylated regions would increase the likelihood that MBD4 would rapidly identify a newly formed mismatch. The ultimate test of this model will require comparing the rate at which spontaneous C to T point mutations arise and the location of these mutations in cell lines or whole animals with and without MBD4 or with a truncated MBD4 that lacks the $MBD4_{MBD}$.

In summary, the studies presented here show how a subtle structural rearrangement, one that is not readily identified by primary sequence analysis alone, can lead to functional differences and specialization within the MBD family of proteins

ACCESSION NUMBERS

The coordinates and NMR restraints for the $MBD4_{MBD}$–dsDNA complex have been deposited in the Protein Data Bank (PDB ID: 2moe); the NMR assignments have been deposited in the Biological Magnetic Resonance Bank (BMRB accession: 19939).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank the University of North Carolina at Chapel Hill and the Research Computing group for providing computational resources and support that have contributed to these research results. The VCU Massey Cancer Center Biological Macromolecule Shared Resource and the Structural Biology Core provided services, instrumentation and products for this research project. The University of North Carolina at Chapel Hill Center for Structural Biology provided instrumentation for this research project.

FUNDING

National Institutes of Health (NIH) [R01-GM098624 to D.C.W.]. NCI Cancer Center Core Support (in part) for VCU Massey Cancer Center Biological Macromolecule Shared Resource and the Structural Biology Core [Grant P30 CA016059]. Funding for open access charge: National Institutes of Health (NIH) [R01-GM098624].
Conflict of interest statement. None declared.

REFERENCES

1. Hendrich,B. and Tweedie,S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.*, **19**, 269–277.
2. Nan,X., Meehan,R.R. and Bird,A. (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.*, **21**, 4886–4892.
3. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
4. Wu,P., Qiu,C., Sohail,A., Zhang,X., Bhagwat,A.S. and Cheng,X. (2003) Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *J. Biol. Chem.*, **278**, 5285–5291.
5. Hendrich,B., Hardeland,U., Ng,H.H., Jiricny,J. and Bird,A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, **401**, 301–304.

6. Sjolund,A.B., Senejani,A.G. and Sweasy,J.B. (2013) MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles. *Mutat. Res.*, **743–744**, 12–25.
7. Hashimoto,H., Zhang,X. and Cheng,X. (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*, **40**, 8276–8284.
8. Manvilla,B.A., Maiti,A., Begley,M.C., Toth,E.A. and Drohat,A.C. (2012) Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *J. Mol. Biol.*, **420**, 164–175.
9. Aziz,M.A., Schupp,J.E. and Kinsella,T.J. (2009) Modulation of the activity of methyl binding domain protein 4 (MBD4/MED1) while processing iododeoxyuridine generated DNA mispairs. *Cancer Biol. Ther.*, **8**, 1156–1163.
10. Cai,M., Williams,D.C., Wang,G., Lee,B.R., Peterkofsky,A. and Clore,G.M. (2003) 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.*, **278**, 25191–25206.
11. Scarsdale,J.N., Webb,H.D., Ginder,G.D. and Williams,D.C. Jr. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.*, **39**, 6741–6752.
12. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
13. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
14. Clore,G.M., Starich,M.R. and Gronenborn,A.M. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc.*, **120**, 10571–10572.
15. Hansen,M.R., Hanson,P. and Pardi,A. (2000) Filamentous bacteriophage for aligning RNA, DNA and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol.*, **317**, 220–240.
16. Cramer,J.M., Scarsdale,J.N., Walavalkar,N.M., Buchwald,W.A., Ginder,G.D. and Williams,D.C. Jr. (2014) Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J. Biol. Chem.*, **289**, 1294–1302.
17. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
18. Clore,G.M. and Gronenborn,A.M. (1998) New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5891–5898.
19. Clore,G.M. and Kuszewski,J. (2002) Chi(1) rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J. Am. Chem. Soc.*, **124**, 2866–2867.
20. Kuszewski,J., Gronenborn,A.M. and Clore,G.M. (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.*, **121**, 2337–2338.
21. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
22. Shen,Y. and Bax,A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, **56**, 227–241.
23. Van Dijk,M. and Bonvin,A.M. (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.*, **37**, W235–W239.
24. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38, 27–28.
25. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
26. MacKerell,A.D. Jr, Banavali,N. and Foloppe,N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, **56**, 257–265.
27. Feller,S.E., Zhang,Y., Pastor,R.W. and Brooks,B.R. (1995) Constant pressure molecular dynamics simulation: the langevin piston method. *J. Chem. Phys.*, **103**, 4613–4621.
28. Martyna,G.J., Tobias,D.J. and Klein,M.L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, **101**, 4177–4189.
29. Sahu,D., Clore,G.M. and Iwahara,J. (2007) TROSY-based z-exchange spectroscopy: application to the determination of the activation energy for intermolecular protein translocation between specific sites on different DNA molecules. *J. Am. Chem. Soc.*, **129** 13232–13237.
30. Rule,G. and Hitchens,T. (2006) Fundamentals of Protein NMR Spectroscopy. Springer, Dordrecht, The Netherlands.
31. Otani,J., Arita,K., Kato,T., Kinoshita,M., Kimura,H., Suetake,I., Tajima,S., Ariyoshi,M. and Shirakawa,M. (2013) Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.*, **288**, 6351–6362.
32. Ho,K.L., McNae,I.W., Schmiedeberg,L., Klose,R.J., Bird,A.P. and Walkinshaw,M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell*, **29**, 525–531.
33. Ohki,I., Shimotake,N., Fujita,N., Jee,J., Ikegami,T., Nakao,M. and Shirakawa,M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.
34. Wakefield,R.I., Smith,B.O., Nan,X., Free,A., Soteriou,A., Uhrin,D., Bird,A.P. and Barlow,P.N. (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.*, **291**, 1055–1065.
35. Zweckstetter,M. and Bax,A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.*, **122**, 3791–3792.
36. Hashimoto,H., Liu,Y., Upadhyay,A.K., Chang,Y., Howerton,S.B., Vertino,P.M., Zhang,X. and Cheng,X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.*, **40**, 4841–4849.
37. Saito,M. and Ishikawa,F. (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.*, **277**, 35434–35439.
38. Fraga,M.F., Ballestar,E., Montoya,G., Taysavang,P., Wade,P.A. and Esteller,M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.*, **31**, 1765–1774.
39. Tafvizi,A., Mirny,L.A. and van Oijen,A.M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem*, **12**, 1481–1489.
40. Bonnet,I., Biebricher,A., Porte,P.L., Loverdo,C., Benichou,O., Voituriez,R., Escude,C., Wende,W., Pingoud,A. and Desbiolles,P. (2008) Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.*, **36**, 4118–4127.
41. Loverdo,C., Benichou,O., Voituriez,R., Biebricher,A., Bonnet,I. and Desbiolles,P. (2009) Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.*, **102**, 188101-1-188101-4.
42. Gorman,J., Plys,A.J., Visnapuu,M.L., Alani,E. and Greene,E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.
43. Gorman,J., Wang,F., Redding,S., Plys,A.J., Fazio,T., Wind,S., Alani,E.E. and Greene,E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3074–E3083.
44. Iwahara,J. and Clore,G.M. (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
45. Iwahara,J. and Clore,G.M. (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J. Am. Chem. Soc.*, **128**, 404–405.
46. Iwahara,J., Zweckstetter,M. and Clore,G.M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15062–15067.
47. Doucleff,M. and Clore,G.M. (2008) Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13871–13876.

48. Romero,Obradovic and Dunker,K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform. Ser. Workshop Genome Inform.*, **8**, 110–124.
49. Li,X., Romero,P., Rani,M., Dunker,A.K. and Obradovic,Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.*, **10**, 30–40.
50. Cantor,C.R. and Schimmel,P.R. (1980) Biophysical Chemistry. W. H. Freeman, San Francisco, CA, USA.
51. Williams,D.C. Jr, Cai,M. and Clore,G.M. (2004) Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. *J. Biol. Chem.*, **279**, 1449–1457.
52. Ishibashi,T., So,K., Cupples,C.G. and Ausio,J. (2008) MBD4-mediated glycosylase activity on a chromatin template is enhanced by acetylation. *Mol. Cell. Biol.*, **28**, 4734–4744.