Cytosine methylation is a well-characterized epigenetic mark and occurs at both CG and non-CG sites in DNA. Both methylated CG (mCG)- and mCH (H = A, C, or T)-containing DNAs, especially mCAC-containing DNAs, are recognized by methyl-CpG-binding protein 2 (MeCP2) to regulate gene expression in neuron development. However, the molecular mechanism involved in the binding of methyl-CpG-binding domain (MBD) of MeCP2 to these different DNA motifs is unclear. Here, we systematically characterized the DNA-binding selectivities of the MBD domains in MeCP2 and MBD1–4 with isothermal titration calorimetry–based binding assays, mutagenesis studies, and X-ray crystallography. We found that the MBD domains of MeCP2 and MBD1–4 bind mCG-containing DNAs independently of the sequence identity outside the mCG dinucleotide. Moreover, some MBD domains bound to both methylated and unmethylated CA dinucleotide–containing DNAs, with a preference for the CAC sequence motif. We also found that the MBD domains bind to mCA or nonmethylated CA DNA by recognizing the complementary TG dinucleotide, which is consistent with an overlooked ligand of MeCP2, i.e. the matrix/scaffold attachment regions (MARs/SARs) with a consensus sequence of 5′-GGTGT-3′ that was identified in early 1990s. Our results also explain why MeCP2 exhibits similar binding affinity to both mCA- and hmCA-containing dsDNAs. In summary, our results expand the genome-wide activity of MBD-containing proteins in gene regulation.

Cytosine methylation occurring prevalently at CG dinucleotide sites, with about 70% of CG sites being subject to methylation (m)³ in the human genome (1). Nevertheless, cytosine methylation is also present at CH (H = A, T, or C) sites (2, 3), and non-CG methylation (mCH) accounts for about 25% of the total cytosine methylation in both embryonic stem cells and neurons, contributing to transcriptional repression and imprinting, similar to CG methylation (4–6). Non-CG methylation occurs in virtually all human tissues and is associated with repression of development-related genes during differentiation of adult stem cells (7).

mCG-mediated transcriptional repression is through its binding to a family of proteins containing the MBD domain, a specific methyl-CpG-binding domain of about 70 residues. 11 MBD domains have been identified in mammals, including MeCP2, MBD1–6, SETDB1/2, and BAZ2A/B.

In both mouse and human neurons, mCH is mainly located in chromatin regions of low CG density, which is established and maintained by DNMT3A (2, 4). Among the three CH dinucleotides, CA is the major target for cytosine methylation (2, 4, 8, 9). A flurry of recent studies demonstrate that MeCP2, a protein involved in neuron development whose mutations are linked to Rett syndrome and other neurological diseases (10, 11), interacts with mCH sites, particularly the mCA sites, in neurons, implying that the MeCP2–mCA interaction plays a key role in regulation of gene expression in normal neuron development (4, 12–14). MeCP2 mainly represses long genes (>100 kb) with high mCA density that are primarily expressed in brain (13). EMSA analysis indicates that MeCP2 binds to mCA as tightly as to mCG DNA and that MeCP2 prefers mCA over mCT and mCC (13, 14). Hydroxylation of mCG into hmCG (hmC is 5-hydroxymethylcytosine) significantly reduces its binding affinity to MeCP2, whereas hydroxylation of mCA into hmCA does not affect its binding to MeCP2 (14).

Recent progress in understanding the physiological role of mCA recognition by MeCP2 motivated us to carry out systematic analysis of mCG and mCH binding to the MBD domains of human MeCP2 and MBD1–4 by using ITC and crystallography. We found that the MBD domains of MeCP2 and MBD1–4 bound to mCG DNAs independent of the sequence identity outside the mCG dinucleotide, and the MBD domains of both MeCP2 and MBD1/2/4 could bind to mCA DNAs with a preference for the mCAC sequence motif. We next determined the crystal structures of the MBD2 MBD domain in complex with several different DNA ligands, including mCG, mCAT, mCAC,
and unmodified CAC dsDNAs. We found that the MBD domain of MBDB2 recognizes the mCA or CA via binding to their complementary TG dinucleotide and explained why the MBD domains favor the mCAC motif. Taken together, our results presented here imply that the unmethylated CA (or TG) DNAs also serve as the binding sites for MeCP2 and other MBD proteins, and also provide a foundation to study how the TG dinucleotide–binding ability of some MBD proteins, including that of MeCP2, impacts their genome-wide distributions and associated gene expression regulation.

**Results and discussion**

**MBD domains of MeCP2 and MBD1–4 bind to mCG DNA independent of the sequence outside the mCG dinucleotide**

The methyl-CG binding ability and sequence selectivity of MBD domains have been studied extensively. For instance, MeCP2 has been reported to prefer some A/T nucleotides surrounding the fully methylated CG dinucleotide (15). On the basis of the SELEX selection assay, the MBD domain of MBD1 has been shown to preferentially recognize mCG within the TCGCA and TGCGCA sequence contexts (16). By surface plasmon resonance (SPR) and structural analysis, the MBD domain of cMBD2 (chicken MBD2) was reported to preferentially recognize the mCGG sequence (17). The MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22). The MBD4 protein contains a T/G or U/G mismatch-specific glycosylase domain in addition to the MBD domain of MBD3, which was initially found to lack mCpG-binding ability (18–20), has been reported to display preferential binding to 5hmC by EMSAs (21) or preferential binding to mCG by residual dipolar coupling analysis (22).

In both crystal structures the base-specific protein–DNA interactions are largely confined to the mCG dinucleotide motif (Fig. 1, A–C). We did not observe any base-specific interaction between protein and methylated DNA outside the mCG dinucleotide. The two MBD2–mCG complex structures could be well superimposed with a root mean square deviation of 0.66 Å over aligned backbone Cα atoms (Fig. 1D). Different from the published cMBD2–mCG structure (17), we found that Lys-174 of human MBDB2 did not interact with the guanine following the mCG dinucleotide, explaining why MBDB2 does not display sequence selectivity other than the mCG dinucleotide (Fig. 1, E–G). Although the human MBDB2 MBD domain is 95% identical to that of cMBD2, our affinities were slightly stronger than those of cMBD2 (17). Based on the complex structures, we could not establish a causal link between the few differing sequence positions and the observed difference in affinity because these different amino acids do not play an obvious role in binding. Thus, we propose that the binding discrepancies for

### Table 1

<table>
<thead>
<tr>
<th>DNA Sequences</th>
<th>MeCP2 (Kd in 80-164)</th>
<th>MBD2 (Kd in 143-220)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GCCACGmCGGTGCGC-3’</td>
<td>1.5±0.2</td>
<td>0.7±0.1</td>
</tr>
<tr>
<td>5’-GCCACGmCGGTGCGC-3’</td>
<td>0.6±0.1</td>
<td>0.1±0.0</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>DNA Sequences</th>
<th>MBD1 (Kd in 1-77)</th>
<th>MBD3 (Kd in 1-71)</th>
<th>MBD4 (Kd in 55-152)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GCCAACCGTGGGCGG-3’</td>
<td>10±1</td>
<td>13±1</td>
<td></td>
</tr>
</tbody>
</table>

**Structure of MBD domain in complex with mCA DNA**

[Supplementary figure and tables]

**Table 1**

<table>
<thead>
<tr>
<th>DNA Sequences</th>
<th>MeCP2 (Kd in 80-164)</th>
<th>MBD2 (Kd in 143-220)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GCCACGmCGGTGCGC-3’</td>
<td>1.5±0.2</td>
<td>0.7±0.1</td>
</tr>
<tr>
<td>5’-GCCACGmCGGTGCGC-3’</td>
<td>0.6±0.1</td>
<td>0.1±0.0</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>DNA Sequences</th>
<th>MBD1 (Kd in 1-77)</th>
<th>MBD3 (Kd in 1-71)</th>
<th>MBD4 (Kd in 55-152)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GCCAACCGTGGGCGG-3’</td>
<td>10±1</td>
<td>13±1</td>
<td></td>
</tr>
</tbody>
</table>
our human MBD2 and reported cMBD2 may result from the different experimental techniques, i.e. ITC versus SPR (17). The sequence-independent binding of these MBD domains is not only consistent with our binding results, but is also in line with crystal structures of MeCP2 and MBD4 in complex with the mCG DNA solved by others, which reveal that the MBD domains of MeCP2 and MBD4 barely make contact with any bases other than the CG dinucleotide (Fig. S2) (24, 26, 27). Taken together, the MBD domains of MeCP2 and MBD1–4 display no sequence selectivity outside the mCG dinucleotide.

It has been reported that the MBD domains recognize the duplex mCG dinucleotide through two highly conserved arginine “fingers” (Fig. 1B, Fig. S2, B and D) (26, 27). Each of the two arginine fingers recognizes one mCG dinucleotide from the duplex mCG DNA and forms a stair motif (28). This stair-shaped motif is usually bound together by means of three kinds of interactions: bidentate hydrogen bonds between the arginine side chain and the guanine base; cation–π interactions between the guanidinium group of the same arginine side chain and the 5-methylcytosine (5-mC) 5’ to the guanine; and the nucleobase stacking interactions between the two bases in the mCG dinucleotide (Fig. 1, B and C, and Fig. S2). Cytosine methylation enlarges the binding interface and enhances cation–π interactions between 5-methylcytosine and arginine (28). The stair-shaped motif is also found in other protein–DNA complexes and usually consists of an arginine residue interacting with consecutive bases (pyrimidine followed by guanine) (29, 30). Therefore, we propose that the two arginine and the two symmetrically related mCG steps would be the structural determinants in the specific interactions between the MBD domains and mCG DNA.

**MBD domains of MeCP2 as well as MBD1/2/4 bind to mCA DNA with a preference for mCAC sequence motif**

As the MBD domain of MeCP2 recognizes mCA DNA in addition to mCG DNA (4, 12–14), we also measured the binding affinities of the MBD domains of MBD1–4 in addition to...
MeCP2 to different non-CG DNAs by ITC (Fig. 2, A and B, Fig. S1, and Tables 1 and 2). We found that the MBD domains of MeCP2 and MBD1/2/4 bound to mCA DNA, albeit weaker than to mCpG DNA in general, and the MBD domain of MBD3 exhibited only weak binding ability to mCA (Tables 1 and 2 and Fig. S1). We found that Tyr-178 of MBD2 formed a water-mediated hydrogen bond with mCG DNA in the MBD2 complex structures (Fig. 1B), and this interaction is also conserved in the MeCP2–mCpG DNA structure (Fig. S2B) (26). This conserved tyrosine residue has been proposed to be critical for mCG binding (26, 27, 31), but it is substituted with phenylalanine (Phe-34) in MBD3 (Fig. 2C), which cannot form a hydrogen bond as tyrosine does in MBD2 and MeCP2 (26, 27, 31). As a result, MBD3 is a weaker mCG binder, and an even weaker binder to mCA DNA (Table 2 and Fig. S1). Our ITC binding results also revealed that MBD domains bind to mCT and mCC DNAs only weakly (Tables 1 and 2 and Fig. S1), consistent with the earlier report that the MBD domain of MeCP2 binds to mCT and mCC DNAs as weakly as unmethylated CG DNA (13).

Motif analysis of the genome-wide CH methylation identifies that CH methylation prominently occurs in the context of trinucleotide mCAG in neuron cells (4, 8, 32, 33). Interestingly, our ITC results also revealed that the MBD domains of MeCP2 and MBD1/2/4 preferred mCAG over other mCAH (H = T, G and A) DNA (Fig. 2, A and B, Fig. S1, and Tables 1 and 2), in line with the observation that the preferential binding of MeCP2 to mCAC is critical for cerebral gene expression in the brain (32). Taken together, the MBD domains of MeCP2 and MBD1/2/4 exhibited binding abilities to mCA DNAs with a preference for the mCAC sequence motif.

Figure 2. MBD domain proteins possess CAC-binding ability. A, ITC-binding curves for the MBD domain of MBD2 and its mutants with different dsDNAs. B, ITC-binding curves for the MBD domain of MeCP2 and its mutants with different dsDNAs. NB, no detectable binding. C, sequence alignment of the MBD domains of human MBD2 (NP_003918.1), MBD1 (NP_001191065.1), MBD3 (NP_001268382.1), MBD4 (NP_001263199.1), and MeCP2 (NG_007107.2). The secondary structures of MBD2 and MeCP2 are indicated at the top and bottom of the sequences, respectively. The mCG dinucleotide-interacting residues of MBD2 and MeCP2 are labeled.
determined the crystal structure of the MBD domain of MBD2 in complex with mCAT DNA at a resolution of 2.05 Å (Fig. 3, A–C, and Table S1). In the MBD2–mCAT complex structure, the MBD domain of MBD2 adopted a canonical MBD-fold, with a C-terminal α-helix packed against the three-stranded β-sheet. The β-sheet was inserted into the major groove of mCA DNA and interacted with the mCA dinucleotide extensively (Fig. 3A).

In the MBD2–mCAT complex structure, Arg-166 formed two hydrogen bonds with the guanine base and simultaneously formed cation–π interactions with the pyrimidine ring of thymine in the TG dinucleotide, completing an R/TG stair interaction motif (Fig. 3, B and C). Despite the same positively charged binding groove and the similar Arg-166 binding pattern between the MBD2–mCAT and other available MBD–mCG structures (Figs. 1B and 3B and Fig. S3, A and B) (15, 18–20, 26), there are significant differences between the mCA and mCG recognition. Different from the second mC–G pair recognition by Arg-188 in the MBD2–mCG complex, Arg-188, the other arginine finger, did not interact with the adenine of mCA dinucleotide, because both the side chain of Arg-188 and the 6-NH$_2$ group of adenine function as hydrogen bond donors and could not form a hydrogen bond with each other. Instead, the side chain of Arg-188 was pushed away from the interaction interface, resulting in the loss of the cation–π interactions between Arg-188 and 5-mC (Fig. 3D and Fig. S4, A and B). The 5-mC did form a water-mediated hydrogen bond with Asp-176 and a C–H–π hydrogen bond with the main chain carbonyl oxygen of Arg-188 (Fig. 3, B and C) (34).

The arginine finger Arg-166 forms a salt bridge with the conserved residue Asp-176, as observed in the MBD–mCG complex structures (Fig. 3C). Because Arg-166 was fixed by Asp-176 with two intramolecular hydrogen bonds, and Arg-188 had more flexibility, Arg-166 was used to recognize the TG dinucle-
otide; otherwise, if the fixed Arg-166 recognized the complementary CA dinucleotide, then the adenine would form close contacts with Arg-166 because both are hydrogen bond donors. Consistently, our mutagenesis binding results revealed that mutating Arg-166 to alanine severely diminished its binding to mCA, whereas mutating Arg-188 to alanine just reduced its binding to mCA by about 4-fold, highlighting that Arg-166 is essential for the binding of MBD2 to mCA DNA (Fig. 2A and Fig. S1).

Interestingly, in the MeCP2–mCG DNA structures, Arg-133 (corresponding to Arg-188 in MBD2) also formed a hydrogen bond with Glu-137, in addition to the conserved salt bridge interactions between Arg-111 and Asp-121 (corresponding to Arg-166 and Asp-176 in MBD2, respectively) (Fig. 2C and Fig. S2, A and B). To investigate how MeCP2 recognizes mCA DNA, we also mutated Arg-111 and Arg-133 to alanine, and found that R111A disrupted the mCA DNA binding, whereas the R133A still retained modest mCA DNA binding (Fig. 2B and Fig. S1), implying that MeCP2 adopts a binding mode similar to that of MBD2 in binding mCA DNA.

Our structure also explained why mCC and mCT DNAs displayed significantly reduced binding affinities toward the MBD domains (Tables 1 and 2 and Fig. S1), because Arg-166 could not form cation–π interactions with the purine ring of adenine or guanine as it does with methylcytosine or thymine (Fig. 3, F and F). This binding mode also explained why MeCP2 exhibits similar binding affinities to both mCA and hmCA (14), because its MBD domain recognized the mCA mainly through its complementary sequence TG, a mimic of mCG, regardless of the modification status of CA.

**Molecular basis for the preferential mCAC binding by the MBD domain**

To further address why the MBD domains of MeCP2 and MBD1/2/4 prefer mCAC over other mCAH (H = A, T, and G) DNAs, we also determined the structures of the MBD2 MBD domain in complex with two different mCAC DNAs, respectively (Fig. 4, A–C, Fig. S3, C and D, and Table S1). The only difference between these two mCAC DNA sequences is that a

---

**Figure 4. Structural basis for the recognition of mCAC and CAC DNA by the MBD domain of MBD2.** A, overall structure of the MBD domain of MBD2 in complex with an mCAC DNA in a schematic representation. The protein is shown in blue, and the DNA ligand is shown in green except the A4–T4 (gray), G5–C5 (gray), T6–A6’ (yellow), and G7–mC7’ (red) bp. The mCAC-interacting residues in MBD2 are shown as stick models. B, specific recognition of the mCAC trinucleotide by MBD2. The interacting residues and DNA bases are shown in the same mode as in A, C, schematic diagram of the detailed interactions between MBD2 and mCAC DNA. The direct hydrogen bonds and water-mediated hydrogen bonds are indicated by solid and dashed red arrows, respectively. The stacking interaction between Arg-166 and T6 is indicated by a gray arrow. D, overall structure of the MBD domain of MBD2 in complex with a CAC DNA in a schematic representation. The protein and DNA are shown the same as observed in A, E, schematic diagram of the detailed interactions between MBD2 and CAC DNA, with the intermolecular interactions indicated in the same way as shown in C, A, B, and D, E, hydrogen bonds formed between protein residues and DNA bases are marked as black dashed lines, and gray dashed lines represent hydrogen bonds between bp.
Structure of MBD domain in complex with mCA DNA

Figure 5. Structural basis for preferential recognition of mCAC DNA by the MBD domain of MBD2. A, superposition of the MBD2–mCAT (blue) and MBD2–mCAC (orange) structures. B, superposition of the MBD2–mCAC (orange) and MBD2–CAC (green) structures. Hydrogen bonds formed between protein residues and DNA bases are marked as black dashed lines, and gray dashed lines represent hydrogen bonds between bp.

Cytosine methylation of the CA dinucleotide is not essential for the binding of MBD domains

The structural revelation that the MBD domain of MBD2 bound to the mCA DNA by specifically recognizing the complementary TG dinucleotide prompted us to investigate whether MBD2 was also able to recognize the unmethylated CA (or TG) DNA. Our binding results indeed revealed that the MBD domains of MBD2, MBD4, and MeCP2 could bind to the unmethylated CA DNA, albeit weaker than to mCA DNA (Fig. 2, A and B, Fig. S1, and Tables 1 and 2), presumably due to the lack of the C–H–O hydrogen bond between the 5-methyl group of the 5-mC and the main chain carbonyl oxygen of Arg-188 in MBD2. To illustrate the structural basis of the recognition of unmethylated CA DNA by the MBD domains, we determined the complex structure of the MBD2 MBD domain bound to a CAC-containing DNA (Fig. 4, D–F, and Table S1). The MBD2–CAC complex structure confirmed our hypothesis that the only difference between the MBD2–mCAC and MBD2–CAC structures is the loss of the C–H–O hydrogen bond between the cytosine of the CA dinucleotide and the main chain carbonyl oxygen of Arg-188 (Figs. 4, C and F, and 5B, and Fig. S4, C and D).

Although the MBD domain has been long established as a methyl-CG–binding domain (35), surprisingly, back to 1991 it has been reported that the chicken attachment region-binding protein (ARBP) protein, which was later found to be the MeCP2 homolog in chicken (36), recognizes the matrix/scaffold attachment regions (MARs/SARs) through a consensus sequence of 5′-GGTGT-3′ with flanking AT-rich sequences (37, 38), and this recognition depends on the MBD domain and a central 5′-GGTGT-3′ sequence (36, 37). Mutation of the central three nucleotides GTG of 5′-GGTGT-3′ motif either abolishes or diminishes its binding to ARBP (or MeCP2) (37). The GTG sequence corresponds to the CAC sequence in the complementary strand of the DNA duplex. Furthermore, by re-assessing the previously published DNA binding database generated from the protein-binding microarray (PBM) assay, a technology developed to characterize DNA-binding sequence specificities of proteins, including transcription factors, in a high-throughput manner, we found that the MBD domain of MeCP2 selectively bound to unmethylated CA/TG sequence (Fig. 6A) (39, 40). Hence, these observations together with our findings presented here demonstrated that the binding of MBD domains, such as those of MeCP2 and MBD2, to mCA DNAs, is through the recognition of the complementary TG dinucleotide, and cytosine methylation of the CA dinucleotide is not essential for the binding of MBD domains.

The ability of some MBD domains recognizing both mCG and TG DNA is analogous to those of some other transcription factors (41), such as KLF4 (Krüppel-like factor 4) and Kaiso (42–44). Nevertheless, unlike KLF4 and Kaiso that bind to both mCG and TG DNA located within some specific sequences (42–45), the MBD domains recognize mCG or GTG DNA without additional sequence selectivity. Compared with the KLF4–TG and Kaiso–TG complex structures, we found that, apart from the water-mediated interaction between Lys-178 and DNA, MBD2 utilizes the conserved arginine residue and acidic amino acid to recognize the TG dinucleotide (Fig. 6, B–D). The TG motif binding by MBD domains also reminds us of another DNA sequence motif, i.e. the GT box motif, a GGGTGGG-like sequence (46). The GT box is predominantly found in the proximal promoter regions or the more distal regulatory regions of mammalian genes with its CG-rich sequence unmethylated (also called GC box) (46). The GT and GC boxes together function as the recruiting elements for the Sp (speci-
protein residues and the top or the bottom CG pairs are marked as respectively.

MBD2 (aa 143–220) expression constructs as the template, (Agilent Technologies) using the MeCP2 (aa 80–164) and were obtained by QuickChange site-directed mutagenesis MeCP2 (R111A and R133A) and MBD2 (R166A and R188A)

GST-tagged fusion proteins. The MBD domain mutants of ics Consortium) expression vector to generate N-terminal were subcloned into the pET28-GST-LIC (Structural Genomics Consortium) expression vector to generate N-terminal His-tagged fusion proteins, pET28-MHL (Structural Genomics Consortium) expression 55–152) fragments of human genes were subcloned into the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific). For each sample, we measured at least three times to get an average concentration. ITC measurements were carried out at the concentrations of MBD domain proteins and DNA ligands ranging from 20 to 60 μM and from 0.5 to 1 mM, respectively. The assays were performed using MicroCal ITC or ITCA200 (GE Healthcare) at 25 °C. Regarding the ITC titrations, for most samples, we did just once; for the other samples, we did more than once until we found optimal experimental condi-
tions, mainly protein/DNA concentrations, which gave nice ITC curves with significant heat change so that we could calculate the $K_d$ reliably. We just used the best curves for each and every binding pair to calculate $K_d$, and the standard errors are the fitting errors from the best ITC titration curves of each binding pair. All the ITC curves with the corresponding thermodynamic parameters are shown in Fig. S1. To determine the $K_d$ values, the data were fitted using the ITC data analysis module of Origin 7.0 (MicroCal Inc.) with the one-site binding model.

**Crystallization**

The purified proteins were mixed at a 1:1 molar ratio with different DNA ligands followed by incubation on ice for 30 min. The protein/DNA reaction mixtures were crystallized using the sitting drop vapor diffusion method at 18 °C by mixing 0.5 μl of the complex samples with 0.5 μl of the reservoir solution. Finally, we successfully obtained the complex crystals for MBD2 (aa 143–220) with the respective DNA ligands. The detailed crystallization conditions for each MBD–DNA complex are summarized in Table S1.

**Data collection and structure determination**

The native crystals were soaked in the crystallization solution plus a final concentration of 15% glycerol and frozen by immersion in liquid nitrogen. Diffraction data were collected at synchrotron or rotating anode X-ray sources under cooling to 100 K, processed with XDS (50), and merged with SCALA or AIMLESS (51). Structures were solved by molecular replacement with Phaser (52) using coordinates from PDB entries 3QM G and 2KY8 (for MBD2–CmCGG) or unpublished models (for remaining MBD2 structures) as required. The MBD2–AmCAT complex was used as a starting model for the nearly isomorphous triclinic MBD2–AmCAC complex structure, which in turn was used as a starting model for the MBD2–ACAC complex. In these cases, molecular replacement search was not needed, and POINTLESS (51) analysis and initial refinement were controlled by a DIMP LE (ccp4.github.io/dimple/) script.4 ARP/WARP (53) was used for electron density map improvement and COOT (54) for interactive model building. Restrained model refinement was performed with PHENIX.REFINE (55), REFMAC (56), and AUTOBUSTER (Cambridge, United Kingdom, Global Phasing Ltd.). MOLPROBITY (57) and PARVATI server (58) were used for analysis of model geometry and atomic anisotropic displacement parameters, respectively. PDB_EXTRAC (59) and IOTBX.CIF (60) were used for the compilation of data collection and refinement statistics summarized in Table S1.

Coordinates and structure factors for the structures of the MBD domains in complex with respective DNA ligands, have been deposited into Protein Data Bank (PDB) under the accession codes 6C1A, 6C1U, 6C1T, 6C1V, 6CNP and 6CNQ.

---

4 Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.
Structure of MBD domain in complex with mCA DNA


Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B., and Baig, A. (2002) The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding
protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res.* **30**, 2911–2919 CrossRef Medline


