

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/43225870>

# Unique Physical Properties and Interactions of the Domains of Methylated DNA Binding Protein 2

ARTICLE *in* BIOCHEMISTRY · MAY 2010

Impact Factor: 3.02 · DOI: 10.1021/bi9019753 · Source: PubMed

---

CITATIONS

42

READS

58

---

## 8 AUTHORS, INCLUDING:



Lila M Giersch

University of Massachusetts Amherst

234 PUBLICATIONS 11,682 CITATIONS

[SEE PROFILE](#)



Vladimir N Uversky

University of South Florida

653 PUBLICATIONS 33,980 CITATIONS

[SEE PROFILE](#)



Kristopher Charles Hite

Emory University

4 PUBLICATIONS 113 CITATIONS

[SEE PROFILE](#)



Jeffrey C Hansen

Colorado State University

118 PUBLICATIONS 6,907 CITATIONS

[SEE PROFILE](#)



# NIH Public Access

## Author Manuscript

*Biochemistry*. Author manuscript; available in PMC 2011 May 25.

Published in final edited form as:  
*Biochemistry*. 2010 May 25; 49(20): 4395–4410. doi:10.1021/bi9019753.

## Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2)<sup>†</sup>

Rajarshi P. Ghosh<sup>‡</sup>, Tatiana Nikitina<sup>§</sup>, Rachel A. Horowitz-Scherer<sup>§</sup>, Lila M. Giersch<sup>‡,⊥</sup>, Vladimir N. Uversky<sup>||,Θ</sup>, Kristopher Hite<sup>†</sup>, Jeffrey C. Hansen<sup>†</sup>, and Christopher L. Woodcock<sup>§,‡,\*</sup>

<sup>§</sup> Department of Biology, University of Massachusetts, Amherst, MA 01003

<sup>‡</sup> Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, MA 01003

<sup>⊥</sup> Department of Biochemistry and Molecular Biology, and Department of Chemistry, University of Massachusetts, Amherst, MA 01003

<sup>||</sup> Institute for Intrinsically Disordered Protein Research, Center for Computational Biology and Bioinformatics, Department, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202

<sup>Θ</sup> Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

<sup>†</sup> Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523

### Abstract

MeCP2 is a methyl CpG binding protein whose key role is the recognition of epigenetic information encoded in DNA methylation patterns. Mutation or mis-regulation of MeCP2 function leads to Rett syndrome as well as a variety of other Autism Spectrum Disorders. Here, we have analyzed in detail the properties of six individually expressed human MeCP2 domains spanning the entire protein with emphasis on their interactions with each other, with DNA, and with nucleosomal arrays. Each domain contributes uniquely to the structure and function of the full-length protein. MeCP2 is ~60% unstructured, with nine interspersed  $\alpha$ -Molecular Recognition Features ( $\alpha$ -MoRFs), which are polypeptide segments predicted to acquire secondary structure upon forming complexes with binding partners. Large increases in secondary structure content are induced in some of the isolated MeCP2 domains and in the full-length protein by binding to DNA. Interactions between some MeCP2 domains in *cis* and *trans* seen in our assays, likely contribute to the structure and function of the intact protein. We also show that MeCP2 has two functional halves. The N-terminal portion contains the methylated DNA binding domain (MBD) and two highly disordered flanking domains which modulate MBD-mediated DNA binding. One of these flanking domains is also capable of autonomous DNA binding. In contrast, the C-terminal portion of the protein which harbors at least two independent DNA binding domains and a chromatin specific binding domain is largely responsible for mediating nucleosomal array compaction and oligomerization. These findings lead to new mechanistic and biochemical insights regarding the conformational modulations of this intrinsically disordered protein, and its context-dependent *in vivo* roles.

<sup>†</sup>Supported by NIH GM070897 and International Rett Syndrome Foundation (CLW); NIH GM066834 (JCH); NIH GM027616 and OD945 (LMG); NIH LM007688, GM071714, the program of the Russian Academy of Sciences for “Molecular and Cellular Biology”, and the IUPUI Signatures Center Initiative (VNU).

\*Corresponding Author: Dr. Christopher L. Woodcock, Biology Department, University of Massachusetts, Amherst, MA 01003, Tel: 413 545 2825; FAX: 413 455 3243; chris@bio.umass.edu.

## Keywords

MeCP2; Rett syndrome; chromatin; unstructured protein; intrinsically disordered protein

MeCP2 belongs to a family of conserved vertebrate proteins that bind to symmetrically methylated CpG dinucleotides and, at least in some cases, transmit epigenetic signals encoded in DNA methylation (reviewed in 1, 2). The finding that mutations in human MeCP2 result in Rett syndrome (RTT), a debilitating neurodevelopmental disorder (3), and that its misregulation is common in other patients with Autism Spectrum Disorders (4, 5), stimulated a focused effort to determine its function(s) and mechanism(s) of action. It is now clear from work with both humans and mouse models which recapitulate many of the human symptoms, that MeCP2 is required for the development and maintenance of neurons in some regions of the brain (5). However, it has been difficult to determine the molecular event(s) that are affected by MeCP2 deficiency or by RTT-causing mutations which lead to functional deficit(s). A complicating issue is that the function of MeCP2 appears to be largely context dependent, varying with species, tissue and cell type, and presence of binding partners. *Xenopus* MeCP2, for example, is important in neuronal fate decisions in early embryogenesis, a function not seen in mice (6). In humans, MeCP2 deficits have been linked to conditions other than RTT, including cancer (7–10).

Protease-resistance and modeling studies have revealed that MeCP2 is a striking example of an intrinsically unstructured protein containing protease resistant domains having varying degrees of disorder (11). Early work identified a short (~90 residue) ordered region of MeCP2 (between residues 75–164) with the ability to bind methylated DNA (12). This region, named the methylated DNA binding domain (MBD), is highly conserved (only four amino acid differences between *Xenopus* and humans). Its structure has been determined by both NMR and x-ray diffraction (13–15). DNA methylation is typically associated with transcriptional repression, and indeed MeCP2 has been shown to repress methylated genes in artificial systems *in vitro* (16). This transcriptional down-regulation activity was mapped to amino acids 207–310 and the region was accordingly termed the transcriptional repression domain (TRD) (16). The MBD and TRD together comprise ~40% of the 486 residue hMeCP2 sequence (Fig. 1a, upper diagram).

One early proposed mechanism of methylation-dependent repression involves an initial binding of MeCP2 to methylated DNA via the MBD, followed by TRD-mediated recruitment of the Sin3A co-repressor and histone deacetylase complexes (HDACs). The subsequent deacetylation of histones in nucleosomes would render the local chromatin region more refractory to transcription (17). However, it is now clear that the mechanism of repression is often more complex, involving multiple MeCP2 binding partners. For example, in non-neuronal Rat-1 cells, repression of the neuron-specific *NaCh type II* gene by MeCP2 appears to involve at least three direct MeCP2 binding partners (methylated DNA, CoREST, and the histone H3 lysine methyl transferase SUV39H1) as well as indirect contributions from REST/NRSF and HP1 (18). In addition, the TRD of MeCP2 has been shown to be an important recruitment platform for several transcriptional modulators and epigenetic regulators in addition to mSin3A and HDACs. These include Ski, N-COR (19), DNMT1 (20), histone H3K9 methyltransferase (21), PU1 (22), splicing factors (23), BRM (24), RNA (25), and the RNA splicing machinery (26). The identification of numerous complexes that interact with MeCP2 suggests that additional modes of MeCP2 function remain to be discovered and call further attention to its identity as an intrinsically unstructured protein (11), which characteristically has large number of binding partners and multiple functions (27).

Further insight into the complexity of MeCP2 biology has come from recent genome-level studies. These revealed that MeCP2 binding is not confined to chromatin containing methylated DNA (28), and that MeCP2 binding can lead to both repression and up-regulation depending on the gene context (29). Evidence has also been presented that MeCP2 is involved in the maintenance of large-scale chromatin loops, perhaps by physically anchoring loop bases (30). This suggestion is consistent with the ability of MeCP2 to promote nucleosome-nucleosome interactions *in vitro*, a property that is enhanced by, but not dependent on DNA methylation (31–33). These findings establish that MeCP2 is a multifunctional protein, and suggest that the different functions are highly context dependent.

With the exception of the MBD, very little is known about the structural properties of MeCP2 and how they contribute to the functional complexity of the intact protein. Studies of RTT-causing MeCP2 mutations show that the most prominent are a few missense mutations in the MBD that disrupt its structure and affect folding (13, 34). There are, however, RTT-causing mutations throughout the entire molecule (see Rett syndrome database at <http://mecp2.chw.edu.au/mecp2/>), indicating that regions of MeCP2 other than the MBD and TRD contribute to its multiple functions. Indeed, several reports have associated specific functions with individual regions. For example, the N-terminal domain of MeCP2 has been shown to mediate interactions with HP1 needed for transcriptional silencing during myogenic differentiation (35). The short (~45 residue) domain connecting the MBD and TRD has been recently shown to be instrumental in stable MeCP2 binding to chromatin *in vivo* (36), and a study of RTT patients showed that mutations tend to be located in this region (37). Finally, the C-terminal portion of MeCP2 required for chromatin interactions *in vitro* (31) also harbors the Group II WW domain binding motif required for binding to splicing factors (23), and the SPxK DNA-binding motif found in histone H1. The importance of the C-terminal region for MeCP2 function is underscored by the frequent occurrence of C-terminal deletions in RTT patients (37).

In order to better understand the interactions and functions of the different domains of human MeCP2, we have undertaken a systematic study of their properties, focusing on their structure, their interactions, and their DNA and chromatin binding abilities. Our work shows that the different domains are highly diverse in many respects, revealing novel properties and providing new mechanistic insights regarding the overall structure of the protein. The MBD and TRD, with their ability to bind methylated DNA and unmethylated DNA, respectively, are clearly key functional elements. Here we show that the N-terminal domain flanking the MBD modulates the affinity of MBD-DNA binding. Further, the intervening domain (ID) between the MBD and TRD possesses a strong, autonomous methylation-independent DNA-binding activity and also facilitates MBD dependent binding. We also report that some domains show a dramatic acquisition of secondary structure upon DNA binding and, while there are 4 autonomous DNA binding domains in MeCP2, considerable synergism exists in their mode of binding. Further, when bound to DNA, some domains increase the stability of MeCP2. Specific inter-domain interactions are seen both *in cis*, and *in trans*, suggesting that these physical couplings play an important role in MeCP2 structural organization and function. We have also expanded our understanding of MeCP2 as an intrinsically unstructured protein, and show that it has an unusually large number of interspersed Molecular Recognition Features (MoRFs) (38, 39), short regions predicted to acquire structure when complexed with binding partners. The occurrence of several RTT-causing mutations within MoRFs further underscores their importance in MeCP2 function. Taken together, these studies significantly advance our understanding of the molecular basis of the unusual structure of MeCP2, and its relationships to DNA binding and the modulation of chromatin conformation.

## Materials and methods

### Cloning MeCp2 domains and linear combinations of domains

To construct the NTD (residues 1–90), an amplicon extending 100-bp 5' of the NdeI site into the pTYB1 vector sequence and carrying a 3' EcoRI linker (GACCGTGAATT) was generated by PCR from full-length pTYB1-MeCP2 cDNA carrying MeCP2 cDNA between NdeI and EcoRI sites, using the following primer pairs: NTD Forward 5'  
CCGGTTAACCGGGATCTGATCC 3', NTD Reverse 5'  
GTTAGAGAATTCTGTACGGATGATGGAGCGCCGCTG 3'. The forward primer used in this amplification reaction was complementary to a site ~100 bp upstream of the MeCP2 start codon in the pTYB1 vector.

To construct the MBD (residues 75–164), ID (residues 165–210), TRD (residues 207–310), CTD- $\alpha$  (residues 261–330) and CTD- $\beta$  (residues 335–486), amplicons with 5' NdeI and 3' EcoRI linkers and additional hexanucleotide overhangs at each end were engineered using the following primer pairs.

MBD Forward primer 5' CAATGACATATGGAAGCTCTGCCTCCCCAACAGC 3', Reverse primer 5' GTTAGAGAATTCTGCTCCCTCTCCAGTTACCGTGAAG 3' ID Forward primer 5' CAATGAATATGCCCTCCGGCGAGAGCAGAAACC 3', Reverse primer 5' GTTAGAGAATTCCACCTGCACACCCCTCTGACGTGGC 3' TRD Forward primer 5' CAATGACATATGGTGCAGGTGAAAAGGGTCTGGAG 3', Reverse primer 5' GTAGAGAATTCTCCGGTCTTGCCTTGATG 3' CTD- $\alpha$  Forward primer 5' CAATGACATATGCCTCAGGCCATTCCAAGAAACGGG 3', Reverse primer 5' GTTAGAGAATTCTCACCGAGGGTGGACACCAGCAG 3' CTD- $\beta$  Forward primer 5' CAATGACATATGGGACTGAAGACCTGTAAGAGCCTGG 3', Reverse primer 5' GTTAGAGAATTCTGCTAACTCTCGGTACGGCGTC 3'

The domain combinations NTD-MBD (residues 1–164), MBD-ID (residues 75–210) and TRD-CTD (residues 207–486) were engineered using the following primer pairs; NTD-MBD: NTD forward primer, MBD reverse primer; MBD-ID: MBD forward primer, ID reverse primer; TRD-CTD $\alpha$ -CTD $\beta$ : TRD forward primer, CTD- $\beta$  reverse primer. The triple domain NTD-MBD-ID was synthesized using the primers 5'-

CCGGTTAACCGGGATCTGATCCGC - 3' forward and 5'-TTTCAG  
AATTCTGCACACCCCTCTGACGTGGCCGC - 3' as reverse.

Following PCR amplification the amplicons were double digested with EcoRI and NdeI and cloned into double digested pTYB1 vector following standard ligation procedure.

Full-length wildtype MeCP2 and R294X were prepared as described. (31)

To synthesize MBD-tetraCys, a modified pTYB1 expression vector was constructed where the cDNA corresponding to the tetraCys sequence. AEAHRWCCPGCCKTF  
(GTTAGAGAATTC GCTGCTCATCGTTGGTGTCTGGTTGTTGTA  
AAACTTTT CTCGAG GATTGA, (underlined bases constitute the hexanucleotide extension to facilitate restriction digestion, italicized bases represent the EcoR I and Xho I site) was inserted in frame between the Eco RI and Xho I sites in the polylinker preceding the intein tag (the *Sce* VMA intein/chitin binding domain) such that insertion of the MBD amplicon between the NdeI and EcoRI sites generated MBD domain-tetraCys-intein-CBD tag fusion. The tetraCys peptide with flanking HRW and KTF tripeptides was chosen because HRWCCPGCCKTF was shown earlier to have a better quantum yield than the core CCPGCC peptide on binding to FlAsH reagent (40).

## Protein purification

Isoform 1 of human MeCP2 (WT and R294X) as well as the individual domains and linked domain constructs were purified using the IMPACT system (New England Biolabs) as described (31, 34). For NTD and CTD- $\beta$ , proteins were applied to heparin HP columns in 100 mM NaCl and eluted using salt steps from 0.1 to 1.0 M NaCl with increments of 0.1 M. Salt fractions containing the pure proteins were pooled and, if required, concentrated using Centricon concentrators (Amicon Inc).

## DNA and NA preparation

Methylated and unmethylated 45-bp segment of promoter IV of the mouse brain-derived neurotrophic factor (BDNF) gene were prepared as described (34). 601-12 DNA was purified, methylated and reconstituted into saturated/undersaturated nucleosomal arrays as described (31).

## Electrophoretic mobility shift assay (EMSA)

To analyze DNA/nucleosomal array (NA) binding efficiency and methylation specificity of the MBD, ID, MBD-ID, NTD-MBD, NTD-MBD-ID and TRD-CTD $\alpha$ -CTD $\beta$  constructs, unmethylated or methylated target DNA/NA (200ng), was mixed and incubated with various amounts of polypeptide in binding buffer (100 mM NaCl, 10 mM Tris, 0.025% NP-40, 0.25 mM EDTA [pH 7.4]) at room temperature for 30 min. Two fold unmethylated competitor DNA or mononucleosome (400 ng) was included in experiments using DNA and NA as substrates respectively. Electrophoresis was performed on prechilled 1% agarose type IV gels, which were run at 85 V for 4 h at 4°C in TAE (40 mM Tris, 24 mM acetic acid, 0.5 mM EDTA [pH 8.3]) buffer.

To compare the DNA/NA binding efficiency of the rest of the MeCP2 fragments, methylated 601-12 DNA/NA was incubated with various amounts of protein (NTD, MBD, ID, TRD, CTD- $\alpha$ , CTD- $\beta$ , TRD-CTD $\alpha$ -CTD $\beta$ ) in the same binding buffer in absence of competitor and electrophoresed as mentioned above. Gels were stained by ethidium bromide, photographed with the Kodak Gel 200 system, and analyzed using ImageJ. For each EMSA experiment two to three trials were performed.

## Electron Microscopy (EM)

Sample fixation, grid preparation, and darkfield EM imaging were as described (31).

## Circular Dichroism (CD)

For each CD related experiment two to four trials were performed. CD spectra of domains and their DNA complexes were acquired and analyzed as described (43). Estimates of secondary structure were calculated using CONTINLL and reference set 7 on Dichroweb (41, 42), which, in addition to a set of structured proteins (43, 44), also contains five denatured proteins, aimed at moderating any possible structural bias of CONTINLL. Estimates of secondary structure derived using CDSSTR closely resembled those from CONTINLL. We showed earlier (34) that estimates of secondary structure in MBD and full-length MeCP2 derived using CONTILL and LIN-COMB is almost identical (45). To further probe the consistency and reproducibility of structural estimates derived using CONTINLL, two to four independent data acquisitions for each individual fragment were deconvolved.

To determine the nature of the spatial packing between different domains of MeCP2 we used a fragment complementation approach. For each pairwise comparison (NTD+MBD, MBD+ID, MBD+TRD, MBD+CTD- $\beta$ , NTD+ID, ID+TRD, ID+CTD- $\beta$ , NTD+TRD) CD data were acquired separately for the two domains and also for their mixture, keeping the concentrations

constant (46). Each mixture was incubated for 15 minutes at room temperature prior to acquisition of data. The CD spectrum for each pair mix (A+B) was then subtracted from the spectrum obtained by addition of the individual spectra ((A)+(B)) of the constituent domains, and the difference at each wavelength was expressed as a percent of the sum of the individual spectra  $[(\text{(A)+(B)}) - (\text{A+B})]/(\text{(A)+(B)})$ .

### Fluorescence Spectroscopy

For thermal unfolding studies, fluorescence emission spectra of MBD and polypeptides containing the MBD together with contiguous domains were collected and analyzed, and Tm values derived as described (34). Experiments with domains alone included a re-cooling step to verify reversibility. DNA-containing samples used a 1:2 ratio of protein: DNA. Thermal melting reversibility cannot be assessed for DNA protein complexes, and the Tm values are therefore denoted as ‘apparent’. Data shown in Table 4 are averages of three independent sets of data.

### Solvent accessibility of Trp104 using acrylamide quenching

Two or three independent fluorescence quenching measurements for each polypeptide were performed on a PTI QM1 spectrofluorometer over a 95 nm window from 305 to 400 nm using 2 nm emission and excitation slits with an integration time of 0.3 s and 0.5 nm steps. 4M acrylamide stock solution was prepared in buffer containing 10 mM Tris, pH 7.4, 100 mM NaCl, and 0.25 mM Na2EDTA. Fluorescence excitation was carried out at 295 nm. At this wavelength there is no inner filter effect due to acrylamide (47). In agreement with this, at 295 nm and at 324 nm the maximum concentration of acrylamide used in the assay (310 μM) had negligible absorbance. For each acrylamide concentration, solvent-only spectra were subtracted from the solvent + protein data. Examination of the solvent-only spectra showed that emission intensity was not affected by acrylamide concentration. Fluorescence quenching was assessed by the addition of varying amounts of 4.0M acrylamide stock solution to 2.5 μM protein in a final volume of 600 μl. For each acrylamide concentration a separate reaction mixture was prepared and incubated ~30 minutes prior to data acquisition to ensure attainment of equilibrium. There was no spontaneous quenching of Trp104 during the ~3 min data acquisition period. Overlays of spectra of the same reaction mixture after 30 minutes showed no visible change. Fluorescence quenching data were analyzed by using the Stern-Volmer equation:

$$F_0/F = 1 + K_{sv} * Q \quad (\text{Eq 1})$$

, where F0 and F are the initial (in absence of quencher) and final (in presence of quencher) fluorescence intensities, Q is the quencher concentration and  $K_{sv}$  is the effective quenching constant. Fit of this equation to the raw data ( $F_0/F$ ) yielded the  $K_{sv}$  and a y intercept of 1.

### Fluorescence labeling of tetraCys-MBD and anisotropy

MBD-tetraCys stock solution was incubated with 10 mM DTT for 6 hr at 4°C following which it was dialyzed extensively against 50 mM Tris-HCl (pH 7.8), 100 mM NaCl, 2.5 mM TCEP, 0.1 mM EDT and 1 mM EDTA. ~50 μM MBD-tetraCys was then incubated overnight at room temperature with 2 molar equivalents of FlAsH EDT2 (LumioGreen™ labeling reagent, Invitrogen) and dialyzed extensively against the same buffer in the dark at 4°C. A two-fold molar excess of FlAsH label for purified tetraCys fusions has been shown to be sufficient for efficient labeling (48, 49). FlAsH-EDT alone has negligible fluorescence whereas the tetraCys peptide-bound FlAsH undergoes a boost in quantum yield in excess of  $10^4$  fold (50), indicating that the observed fluorescence originated from the bound reagent. Fluorescence anisotropy measurements were performed on a PTI QM1 spectrofluorometer equipped with an excitation

and emission polarizer, using 8 nm emission and excitation slits with an integration time of 1sec. The excitation wavelength ( $\lambda_{ex}$ ) was 500nm and the emission scan used a window of 10nm (525–535nm). A constant amount of labeled MBD (100nM) was mixed with various amounts of other domains covering a range of 10nM to 10 $\mu$ M, and incubated for 5 minutes prior to data acquisition. Fluorescence anisotropy was calculated using:

$$r = (I_{vv} - GI_{vh}) / (I_{vv} + 2GI_{vh}) \quad (\text{Eq 2})$$

, where r is the fluorescence anisotropy of FlAsH labeled MBD-tetraCys,  $I_{vv}$  and  $I_{vh}$  are the fluorescence intensities collected with a vertically oriented excitation polarizer and vertically ( $I_{vv}$ ) and horizontally ( $I_{vh}$ ) oriented emission polarizer (47). G is the correction factor for the difference in sensitivity of the detection system for vertically and horizontally polarized light and expressed as:

$G = I_{hv}/I_{hh}$ , where  $I_{hv}$  and  $I_{hh}$  are the fluorescence intensities collected with a horizontally oriented excitation polarizer and vertically ( $I_{hv}$ ) and horizontally ( $I_{hh}$ ) oriented emission polarizer. G factor correction was done for each data acquisition cycle keeping the machine settings identical. The basal anisotropy of FlAsH-labeled tetraCys MBD varied from ~0.1–0.12 between experiments. The anisotropies of complexes were normalized by dividing the anisotropies at each input concentration by the anisotropy of the tetraCys-MBD fusion in the respective experiment. Plots of normalized anisotropy versus increasing complementary unlabeled protein fragment (TRD, ID) concentration were fit to a four parameter logistic binding model using Psi plot (51). The goodness of fit for TRD and ID were  $r^{TRD}=0.998$  and  $r^{ID}=0.992$  respectively. The dissociation constant  $y=D+((A-D)/(1+(X/C)^B))$ , where Y is the normalized anisotropy, d is the anisotropy at infinite concentration of the complementary MeCP2 domain, A is the anisotropy at zero concentration of the complementary MeCP2 domain, x is the concentration of the different MeCP2 domains, and C is the inflection point on the fitting curve which is equivalent to the dissociation constant.

### DNA binding affinities of MeCP2 fragments

The blunt ended fluorescein labeled 22bp duplex with a single symmetrically methylated CpG was synthesized by annealing complementary single strands of an HPLC-purified 22-bp DNA segment of mouse BDNF promoter IV strand 1 5'-/56-FAM/CCCTATAA/Me-dC/GGAATTCTATAATG - 3' strand 2 5'-CATTATGAATTC/Me-dC/GTTATAGGG.

Fluorescence anisotropy measurements were performed on a PTI QM1 spectrofluorometer equipped with an excitation and emission polarizer, using 20 nm emission and 26nm excitation slits with an integration time of 8 sec. The  $\lambda_{ex}$  used was 480 nm and the emission was collected over a window of 4nm (518–521nm). A constant amount of labeled DNA (250pM) was mixed with various amounts of MBD, MBD-ID, NTD-MBD, and TRD-CTD $\alpha$ -CTD $\beta$  covering a range of 100pM to 60nM and 1nM of the same DNA was mixed with various amounts of NTD, ID, TRD, CTD- $\alpha$  and CTD- $\beta$  covering a range of 1nM to 600nM and incubated for 10 min prior to data acquisition. Anisotropy was calculated at 520nm using Eq 2. Anisotropy values were normalized using the equation  $r_{norm} = (r_n - r_0) / (r_{max} - r_0)$  where  $r_0$  is the raw anisotropy at 0 protein input,  $r_{max}$  is the raw anisotropy at maximum protein input,  $r_n$  is the raw anisotropy at each protein concentration and  $r_{norm}$  is the corresponding normalized anisotropy. The global dissociation constant of domain DNA interaction was obtained from least square fits of plots of normalized fluorescence anisotropy versus protein concentration. For comparing the binding affinity of NTD and CTD- $\beta$  to the other domains (e.g. ID) anisotropy values were normalized using the equation  $r_{norm} = r_n/r_0$ .

## Sedimentation velocity

Sedimentation velocity experiments were performed with Beckman Optima XL-I analytical ultracentrifuge using absorbance optics. 208-12 nucleosomal arrays reconstituted on methylated DNA were mixed with the appropriate MeCP2 construct in 50mM NaCl, 10mM Hepes, 0.25mM EDTA and sedimented at a velocity of 18000 rpm and temperature of 20±0.1 C. The sample absorbance was 0.7 A260 and the molar ratio of protein to nucleosomal array ranged from 1 to 4. The data were analyzed (52) using the Ultrascan data analysis program. Plots of the boundary fractions against their corresponding  $S_{20,w}$  values yielded the integral distribution of sedimentation coefficients. Each experiment was repeated two to three times.

## Compositional profiling

To gain insight into the relationships between sequence and disorder, amino acid composition of MeCP2 was analyzed using an approach recently developed for intrinsically disordered proteins (53). To this end, the fractional difference in composition between MeCP2 (or a set of disordered proteins from the DisProt database, (54) and a set of ordered proteins was calculated for each amino acid residue. The fractional difference was calculated as (CX-Corder)/Corder, where CX is the content of a given amino acid in a given protein (or protein set), and Corder is the corresponding content in a set of ordered proteins and plotted for each amino acid. In corresponding plots, the amino acids were arranged from the most order-promoting to the most disorder-promoting according to the amino acid distribution in DisProt database (54). For each amino acid residue, this analysis provides two important parameters, the sign of a corresponding bar and its height. The sign of a bar reflects enrichment or depletion of a given residue in a query sequence as compared to the content of this same residue in a set of ordered proteins. The height of the bar corresponds to the value of the fractional difference, where large bars correspond to the residues whose content in the query sequence is the most different from their content in the set of ordered proteins.

## Disorder, $\alpha$ -MoRF prediction, and modeling

Disorder predictions for MeCP2 were made using PONDR® VLXT (53). Potential interaction sites, molecular recognition features (MoRFs) that gain functionality upon a disorder-to-order transition induced by binding to a partner, were identified by the  $\alpha$ -MoRF predictor which detects short ( $\leq 20$  residue) stretches within long regions of disorder with the potential for helical structure acquisition upon binding (38; 39). The algorithm utilizes a stacked architecture, where PONDR® VLXT is used to identify short predictions of order within long predictions of disorder, and then a second level predictor determines whether the order prediction is likely to be a binding site based on attributes of both the predicted ordered region and the predicted disordered region surrounding it. (

The UCSF Chimera software ([www.cgl.ucsf.edu](http://www.cgl.ucsf.edu)) was used to visualize the MoRF residues within the structure of a MBD-methylated DNA complex (PDB 3c2i) (15).

## Results

### MeCP2 domain nomenclature

In this study, we establish the structural and functional properties of 6 domains of hMeCP2 isoform 1 both individually, and as contiguous fusions. With the exception of the MBD, these regions do not constitute independently evolving structural units. However, in addition to their protease resistance, there is strong evidence, discussed below, that they have unique structural and functional properties and in that context, are considered domains. Table 1 lists the salient features of the MeCP2 polypeptides we prepared, and Figure 1a (upper panel) shows their locations within the parent protein. Flanking the MBD are the N-terminal and Intervening

domains, termed NTD and ID respectively. The long C-terminal domain (CTD) includes a highly protease-sensitive segment, which, when cleaved, results in two fragments (11) denoted CTD- $\alpha$  and CTD- $\beta$ .

### Most MeCP2 domains are extensively disordered

The predicted distribution of structured and unstructured domains in MeCP2 is well illustrated by the output of PONDR (Predictor of Naturally Disordered Regions) (53, 55) (Fig. 1a, lower panel). A portion of the MBD contains a region predicted to adopt a stable secondary structure, and is the only region for which structure is known at the atomic level (13, 15). Other short segments of predicted order occur throughout the protein and are found in all the domains except the short ID (Fig. 1a). This alternating pattern of disorder and order is also predicted by the FoldIndex algorithm (11). Unstructured proteins tend to have an amino acid composition that favors structure-disrupting residues (53, 54, 56), and MeCP2 represents an extreme case of this skewing, exceeding that of the unstructured proteins in the DisProt (54) data base (Fig. 1b).

To assess the inherent secondary structure content of individual domains and determine whether the distribution of secondary structure agrees with the predicted disorder/order map of MeCP2, two to four independent circular dichroism (CD) spectra were recorded for each domain (Fig. 2a). Of the six domains, the MBD was the only one showing a characteristic positive band in its CD spectrum at  $\sim$ 197 nm, indicative of significant ordered secondary structure (Fig. 2a). The others had a negative band in this region, indicating extensive disorder, with the NTD and TRD being the most strongly disordered. Estimates of the different types of secondary structure were obtained using CONTINLL (34, 41, 42) deconvolution (see Materials and Methods for details). Deconvolution produced highly reproducible estimates of secondary structure for each domain (Table 2).

In agreement with previous findings (11, 13, 34), the MBD is  $\sim$ 60% structured, with  $\sim$ 45%  $\beta$ -sheet/turn, and  $\sim$ 15%  $\alpha$ -helix. The proportion of predicted unstructured sequence for the NTD, ID, and CTD ranged from  $\sim$ 62% to  $\sim$ 78% (Table 2), consistent with the net 60% unstructured sequence in intact MeCP2 (11, 34). The amount of disorder in each of the domains determined by CD (Table 2) is close to that predicted by the PONDR (Fig. 1a) and FoldIndex (11) algorithms.

### MeCP2 domains differ in their ability to bind DNA and chromatin

Native electrophoretic mobility shift assays (EMSA) provide a qualitative estimate of the DNA and chromatin interaction properties of MeCP2 (31, 32), allowing the exploration of a wide range of DNA: protein ratios. The distribution of shifted species provides some indication of the nature of the interactions involved. We previously reported strikingly large mobility shifts when full-length MeCP2 interacts with DNA and nucleosomal arrays (NAs) (31). Here, we dissect the extent of these mobility shifts domain-by-domain. Substrates consisted of tandem ( $n=12$ ) arrays of a 207 bp sequence containing the '601' nucleosome positioning sequence (57), either as naked DNA, or after reconstitution with core histones to yield 12-mer nucleosomal arrays (NAs). Input ratios are expressed as moles of protein per nucleosome or 207 bp DNA. Since each 207 bp fragment contains 18 methylatable CpGs, the highest molar ratio of peptides used here in the absence of competitor is approximately equivalent to one polypeptide per two methyl CpGs. The higher input ratios of peptides in EMSA and EM experiments were used to simulate a situation of local enrichment of peptides as may occur in cases such as the MeCP2-regulated BDNF (Brain Derived Neurotropic Factor) promoter III which contains a region of closely spaced CpGs (58). Also, we have found that closely spaced methylated CpGs favor cooperative DNA binding of MeCP2 (RPG and CLW, in preparation).

Full length MeCP2 induces pronounced shift with both DNA and NAs even at low molar inputs of the protein. These shifts are enhanced in case of methylated substrates (Fig. 2b far right). Among the individual domains, the NTD and CTD- $\beta$  stand out as inducing only very minor mobility shifts in DNA, whereas the ID, TRD, and CTD- $\alpha$  fragment all induce marked shifts (Fig. 3a). The shifts induced by these three domains are methylation-independent (Figure 3c), accounting for the substantial methylation-independent binding observed with intact MeCP2 (Fig 3b) (11, 32, 33). Further support for DNA binding by the ID comes from the finding that a R188E MeCP2 mutant has a significantly lowered gel mobility compared to wild type (data not shown). A slightly different fragment containing the TRD (residues 198–305) also has been shown to bind DNA (11). In general, the shifts with DNA and NAs are qualitatively similar. However, the CTD- $\beta$  is a clear exception, inducing a moderate but reproducible shift with NAs but not with naked DNA (Fig. 3a, c). This result is consistent with earlier findings that deletion of 192 residues of the C-terminal portion of MeCP2 results in deficient NA compaction and oligomerization (31), and suggests that the CTD- $\beta$  contains a unique histone binding region (s).

### The DNA and NA binding properties of the MBD and CTD are modulated by their flanking domains

To determine if there were inter-domain effects, we compared the DNA and chromatin interactions of several constructs comprising multiple contiguous MeCP2 domains. (Note that throughout, we use hyphens to denote constructs encompassing adjacent MeCP2 domains yielding information on DNA and chromatin interactions in *cis*, and the '+' symbol to denote different MeCP2 domains combined in solution and providing information on *trans* interactions). The TRD-CTD fusion comprising the C-terminal 280 residues of MeCP2 promotes pronounced shifts with both DNA (Fig. 3a, far right) and NAs (Fig. 3d, far right). This is consistent with DNA binding by TRD and CTD- $\alpha$ , and chromatin binding by the CTD- $\beta$  (Fig. 3a, d). The gel shifts seen with the TRD-CTD construct reflect the additive binding effect of the constituent domains, and is consistent with the concerted binding by these domains that would be required for chromatin condensation and/or oligomerization (also see Figs. 4 and 5).

For the MBD-containing constructs, it was important to compare gel shifts obtained with methylated and unmethylated DNA or chromatin substrates. For these experiments we used a two-fold excess of unmethylated 207 base pair DNA or mononucleosome competitor to enhance methylation-dependent effects (31, 34). With this level of competitor, higher protein:DNA ratios are needed to observe significant mobility shifts. As a control, the MBD alone shows a reproducible methylation-dependent enhancement in mobility shift with methylated DNA (Fig. 3b) and chromatin (Fig. 3d far left). The NTD-MBD construct showed methylation independent and dependent binding both to DNA and chromatin, but, surprisingly, the observed mobility shifts were much larger than expected from a mere sum of the moderate shift induced by MBD and zero shift induced by NTD (Fig. 3a, b and d, e). This synergistic shift enhancement suggests a strong structural and/or functional coupling between NTD and MBD eliciting a binding mode uncharacteristic of either of the individual domains. This phenomenon cannot be explained by a simple increase in electrostatic shielding since the net charge of the MBD (5 at pH 7.4) is higher than that of NTD-MBD (4 at pH 7.4) in the reaction buffer. Thus, while the MBD may be the minimal domain necessary to recognize methylated DNA, MBD function is clearly impacted by the adjacent non-DNA binding NTD in a way that is likely to enhance its DNA-binding affinity (see later). The EMSA patterns obtained with MBD-ID were also different than MBD, with significant smearing (Fig. 3b). This likely reflects formation of non-specific higher order complexes via cross-linking by the MBD-ID fragment, with its two independent DNA-binding regions. The NTD-MBD-ID construct interacted with DNA and chromatin much like NTD -MBD, although the gel shifts were slightly more pronounced with

the longer construct (Fig. 3b, e). It should be stressed that methylation-enhanced shifts occur only with MBD-containing polypeptides. Examples of this are shown in Fig. 3c where, for ID and the TRD-CTD fusion (TRD-CTD $\alpha$ -CTD $\beta$ ), the methylation state of the DNA has no effect on the induced shift.

### Contiguous fusions of certain MeCP2 domains induce condensation of NAs

Full length MeCP2 is a potent chromatin architectural protein, inducing extensive compaction and self-association of nucleosomal arrays (31, 32). To determine which domains of MeCP2 were important for this phenomenon, we first investigated changes in sedimentation velocity of defined nucleosomal arrays (NAs), which provide a sensitive and quantitative assessment of their state of compaction (59). We prepared methylated 601-12 NAs and measured the influence of individual domains and multi-domain constructs on their sedimentation properties. MeCP2 domains and constructs comprising multiple domains, shown by EMSA to interact with chromatin, were mixed with methylated 601-12 NAs, and diffusion-corrected sedimentation coefficient distributions were obtained by analysis of sedimentation boundaries using van Holde-Weischet method which is particularly well suited for polydisperse systems such as population of nucleosomal arrays with multiple compaction states (52). Results from two to three independent experiments show that at the ionic strength used in these experiments (50 mM NaCl), NAs alone give a sedimentation profile typical of a slightly folded conformation, with a nearly homogenous population between boundary fractions of ~20% to ~70%, and an average  $s_{20,w}$  value of ~32 ( $\pm 2$ ) S (Fig. 4). Full-length MeCP2 at a 1:1 input ratio induced a dramatic increase to ~61 ( $\pm 2$ ) S (Fig. 4), similar to our earlier observation in the 5S 208-12 NA system (32). Higher input ratios of full-length protein result in the formation of rapidly sedimenting complexes due to MeCP2-mediated self-association of NAs (data not shown).

For the MBD and constructs that included its flanking domains, the sedimentation coefficient distributions obtained at an input ratio of two polypeptides per nucleosome are presented in Fig. 4. The MBD alone increased the sedimentation coefficient by only ~2 ( $\pm 1$ ) S, indicating binding but little or no array compaction (binding of two MBD molecules to each nucleosomal unit in an array results in a ~10% increase in mass for the complex and would be expected to increase the array sedimentation coefficient by a few S units). The NTD-MBD construct, which produced a prominent electrophoretic mobility shift (Fig. 3d), caused an increase in sedimentation coefficient of ~4.5 ( $\pm 0.5$ ) S, whereas MBD-ID resulted in a sedimentation coefficient increase of ~6 ( $\pm 1$ ) S. These results show that each of these fragments binds to nucleosomal arrays but induces only small increases in array compaction. A more substantial increase in compaction of ~11 ( $\pm 1$ ) S was seen with the NTD-MBD-ID construct (Fig. 4), and the most striking result was obtained with TRD-CTD. Binding of the latter caused an increase of ~26 ( $\pm 3$ ) S in the homogeneous segment of the population indicating a level of NA compaction similar to that caused by full-length MeCP2 albeit at double the protein input. Binding of the TRD-CTD fragment also produced a significant fraction of heterogeneous self-associated arrays (Fig. 4, boundary fraction >60%), as did full-length MeCP2. These results suggest that the TRD-CTD fragment may be able to recapitulate the chromatin condensing functions of the full-length protein. At a four-fold molar input of domains and domain fusions, the relative differences in levels of compaction of NAs remained largely the same, although as expected the absolute sedimentation values increased (data not shown).

### Different domains of MeCP2 induce distinctive changes in NA morphology

Observation by electron microscopy (EM) of the compaction state of individual NAs provides a direct measure of the impact of protein binding on array morphology and complements the EMSA and sedimentation studies. From the images it is possible to compare compaction effects quantitatively, and examine the initial changes at the nucleosome and linker DNA level that

lead to overall increases in condensation. For these purposes, it is useful to employ “subsaturated” NAs where linker remains visible during the initial stages of compaction (31). Hence, in this study, we reconstituted methylated 601 DNA templates with 6–8 (rather than 12) nucleosomes, exposed them to the defined fragments of MeCP2 at various input ratios, and imaged the resulting complexes using darkfield EM.

In the absence of protein, the NAs were fully extended (Fig. 5a). Little change in conformation was seen with either MBD or NTD alone (Fig. 5b, c), consistent with the EMSA data on saturated NAs (Fig. 3). All the other individual domains induced partial clustering of nucleosomes within arrays (Fig. 5d–g), leading to significant ( $p < 0.001$ ) reductions in array diameter (diameter of the smallest circle that fully encloses the array; Fig. 5k). The three contiguous constructs examined (MBD-ID, NTD-MBD, and TRD-CTD) also showed nucleosome clustering and array compaction (Fig. 5h–j), with decreased array diameters (Fig. 5k) that parallel the increased sedimentation coefficients seen in Fig. 4. Consistent with its potency in inducing gel shifts and increasing sedimentation velocity, the TRD-CTD was the most effective at compaction, with a mean diameter significantly ( $p < 0.0001$ ) smaller than all of the other MeCP2 fragments. In addition to compacted individual arrays, the TRD-CTD construct induced self-association of arrays (Fig. 5j, far right panel), consistent with our observation of a large fraction of heterogeneous rapidly sedimenting material (Fig. 4). The full-length protein induces extensive array oligomerization under these conditions, (32) precluding accurate measurements of array compaction.

### The MBD is structurally coupled to other MeCP2 domains

In a previous study focusing on the properties of the MBD, we used fluorescence spectroscopy to monitor the solvent exposure (47) of the single tryptophan at position 104, and reported that domains flanking the MBD provided solvent protection to W104 (34). Here, we examine this effect in more detail, using the collisional quenching agent acrylamide to avoid complications due to the differing fragment sizes, and hence tumbling rates. The extent of quenching of the W104 by acrylamide provides a direct measurement of its solvent accessibility, and allows the identification of flanking domains that provide protection from quenching in cis. The results of two to three independent measurements for each fragment (Fig. 6) show that acrylamide most effectively quenches the fluorescence signal of tryptophan in the MBD alone (quenching constant  $K_{sv}$   $8.4 \text{ M}^{-1}$ ), and least effectively in full-length MeCP2 ( $K_{sv}$   $3.9 \text{ M}^{-1}$ ). Intermediate levels of quenching ( $K_{sv}$   $7.3 \text{ M}^{-1}$  and  $6.8 \text{ M}^{-1}$ ) were obtained for NTD-MBD and MBD-ID, respectively. Tryptophan fluorescence from the truncation RTT mutant R294X, which is approximately equivalent to NTD-MBD-ID-TRD showed greater protection from quenching than other constructs ( $K_{sv}$   $4.9 \text{ M}^{-1}$ ) but was less effective than the full-length protein. These results indicate that within the full-length protein, the structure of the MBD is influenced by associations with other domains that lead to the shielding of W104 from solvent exposure. All the MeCP2 domains contribute to this shielding (particularly NTD, ID, and TRD), suggesting that inter-domain coupling occurs within the overall structure of the full-length protein. In this respect, it is interesting to note that an NMR study of a region of the chicken homolog of MeCP2 approximately equivalent to the human MBD-ID construct, suggested that it may act as a platform for interaction with other regions of MeCP2, or the binding of other proteins (60).

### Inter-domain coupling occurs in trans and affects secondary structure

Inter-domain interactions that are crucial to the stability of the native state of the protein are often strong enough to be sustained in *trans* in mixtures of protein fragments (61). Fragment complementation approaches to define tertiary organization of domains are particularly helpful for proteins refractory to crystallization as is the case with MeCP2. To test for such associations between MeCP2 domains, we prepared a construct in which a tetra-cysteine motif was fused

to the MBD and labeled with the FlAsH reagent (50–62) at a level that produced a robust emission (Fig. 7A insert). The labeled MBD was titrated with a second domain and evaluated using fluorescence anisotropy for inter-domain interactions that cause it to tumble more slowly. The results indicated that the ID and TRD were able to bind the MBD in solution, with dissociation constants of ~4  $\mu$ M and ~2.5  $\mu$ M respectively (Fig. 7a). In contrast, no interactions were observed between MBD and NTD. A weak effect was seen with the CTD- $\beta$ , but only at high molar inputs.

We also investigated inter-domain associations in *trans* using CD, which can detect changes in secondary structure of one or both components in a mixture of the two (63). For each pairwise comparison, data were acquired separately for the individual domains and also for their mixture. With the NTD+MBD, MBD+ID, and MBD+TRD mixtures, we observed a clear difference between the summed individual CD spectra and the spectrum of an equimolar mixture (Fig. 7b, c), indicating that an interaction between domains led to change(s) in secondary structure. The failure to observe an interaction between the MBD and NTD with anisotropy may be due to the location of the tetra-Cys moiety at the largely unstructured C-terminus of the MBD. Also, binding by NTD distal to the C-terminus may not result in changes in rotational freedom of the label and thus not affect anisotropy.

The CD approach also allows domain interactions not involving the MBD to be investigated. None of the domain combinations that lacked the MBD gave any evidence of interaction. Examples are shown for the NTD + ID and ID + CTD- $\beta$  pairs (Fig. 7c). Thus, it appears that the role of the MBD is not solely as a methylation-dependent DNA binding domain, but it also acts as a structural ‘core’ of the protein, participating in multiple inter-domain interactions (see Discussion).

### Impact of DNA binding on secondary structure

Intrinsically disordered proteins often undergo binding-induced increases in secondary structure content. Given the multiple DNA-binding domains and large degree of intrinsic disorder spread throughout MeCP2, we were interested to determine whether DNA binding leads to structural alterations. For DNA-binding experiments, we selected a 45 bp segment of the brain-derived neurotropic factor (BDNF) III promoter that has a single CpG unit and is known to be a target for *in vivo* MeCP2 binding (58–64). A similar stretch of BDNF DNA was used in the MeCP2 complex for which the x-ray structure has been determined (15). The ID and TRD, when mixed with an equimolar amount of the 45 bp DNA substrate, resulted in striking changes in far-UV CD profiles, independent of the DNA methylation state (Fig. 2b, c). These changes result from significant increases in secondary structure content (from ~38% to ~59% for ID; from ~22% to ~30% for the TRD (Table 3)). For ID, the acquired structure is approximately equally divided between  $\alpha$ -helix and  $\beta$ -strand, while for the TRD; the increase is in the  $\beta$ -strand component. As previously reported, in the presence of DNA, the MBD shows no change in secondary structure with unmethylated DNA, but acquires a methylation-dependent ~6% increase in  $\alpha$ -helix (34).

### Domains of MeCP2 differ in their affinity for DNA, and contribution to thermal stabilization upon DNA binding

To measure affinity between DNA and different MeCP2 domains, we used a 22 bp segment of the BDNF promoter containing a single centrally located methylated CpG dinucleotide and a fluorescein label at one end. DNA binding by a given protein fragment reduces the DNA tumbling rate in solution and can be measured by following changes in steady-state anisotropy of the fluorescein label. Fig. 8a–c shows the changes in anisotropy as a function of concentration of MeCP2 domains, and constructs with linked domains. Amongst MBD and MBD containing contiguous domain fusions, NTD-MBD showed the highest DNA binding

affinity ( $K_d$  0.8 nM), 10-fold higher than the MBD alone which showed the weakest binding ( $K_d$  8.5 nM) (Fig. 8a, c). MBD-ID ( $K_d$  1.4 nM) bound to DNA with 6 fold higher affinity than MBD and ~50-fold higher affinity than ID alone ( $K_d$  75nM) (Fig. 8a, b). This clearly shows that while the ID contains an autonomous DNA binding domain, it also facilitates MBD-mediated binding. Unlike ID, NTD does not bind to DNA as an isolated domain (Fig. 8c), but when coupled to MBD markedly enhances its binding affinity, consistent with our EMSA data.

The constituent domains of TRD-CTD fusions showed considerable variability in their DNA binding affinity. Unlike TRD and CTD- $\alpha$ , which bound with affinities of 20 nM and 96nM respectively (Fig 8b), CTD- $\beta$  induced small increases in fluorescence anisotropy and only at very high protein concentrations (~600nM) (Fig. 8c). Interestingly the TRD-CTD fusion construct bound DNA more strongly ( $K_d$  3.6 nM) than its constituent domains suggesting that, in addition to their autonomous binding capacities, coupling between the constituent domains results in emergent binding properties. In summary, in increasing order of DNA binding affinity, the domains of MeCP2 can be arranged as NTD, CTD- $\beta$ , CTD- $\alpha$ , ID, TRD, and MBD.

Since DNA binding to MeCP2 also confers thermal stability to the protein (34), it was of interest to determine the contributions of the different domains to the overall stability. To assess stability, we monitored the fluorescence emission of the single tryptophan in MeCP2 (W104 in the MBD) over the temperature range from 10°C to 85°C. We first compared the melting profiles of the MBD alone with those of the longer constructs containing the MBD. As previously reported, the MBD has an apparent  $T_m$  (50% melt temperature) of ~45 °C (34), and similar values were obtained with MBD-ID and MeCP2<sup>1–294</sup> which includes most of the TRD. Unexpectedly, the  $T_m$  of NTD-MBD domain fusion was lower than MBD (Table 4), suggesting that the NTD holds the MBD in a relatively destabilized state (thus lowering the  $T_m$ ).

Each protein fragment was then mixed with methylated or unmethylated DNA consisting of the 45 bp segment of the BDNF promoter with a single centrally located methylatable site, and thermal melting profiles obtained and analyzed (Table 4). Two important patterns emerge. First, unmethylated DNA acts as a stabilizing agent, consistently inducing increases in  $T_m$ . Here, the MBD alone stands out in inducing only a small stabilization of only ~2 °C, whereas additional domains generate a much greater enhancement in thermal stability. Second, when the DNA is methylated, there is a consistent increase of ~ 8 °C in the  $T_m$  of all constructs. Interestingly, the stabilizing effect of methylated DNA on the NTD-MBD-ID fragment exceeds that of the full-length protein. The very similar apparent thermal stabilities of MeCP2<sup>1–294</sup> and full-length MeCP2<sup>1–486</sup> indicate that the C-terminal domain of the molecule does not contribute to the overall thermal stability of the protein when bound to DNA.

## Discussion

Due to its high degree of disorder, the overall structure of full-length MeCP2 is not readily amenable to structure determination by x-ray crystallography or NMR. However, our domain-by-domain dissection of MeCP2 has provided important insights into the physical and functional properties of this unique unstructured protein. An overarching aspect of MeCP2 biology is the role that disorder plays in supporting its multiple functions. Importantly, the structure content of the full length protein is largely similar to the sum of the weighted average structure of its domains and the small difference can be attributed to changes in secondary structure resulting from interdomain associations largely involving the MBD in the context of the full length protein. As further discussed below, the MBD appears to be the central hub for MeCP2 tertiary structure, forming contacts with the NTD, ID, and the TRD.

### The large number of MoRFs may account for the functional and structural versatility of MeCP2

The functions of intrinsically disordered proteins are often coupled to the acquisition of structure upon binding to a partner (65–68). In this respect, it is significant that MeCP2 gains secondary structure and undergoes striking thermal stabilization upon binding to DNA (Table 4).  $\alpha$ -MoRF predictors (38, 39) (see materials and methods) predict nine  $\alpha$ -MoRFs in MeCP2. This is an unusually large number for a protein of this size (68), but is consistent with its predicted distribution of order and disorder. There is at least one predicted MoRF in each MeCP2 domain except the ID. Some of these sites may be involved in long-range intra-protein contacts within MeCP2, promoting the *cis* and *trans* domain interactions documented above.

Two MeCP2 point mutations, R133C and A140V located in the MBD resident  $\alpha$ -MoRF 133–150, have been shown to impair MeCP2's interactions with the ATRX protein leading to improper nuclear localization of ATRX, a phenomenon implicated in ATRX ( $\alpha$ -thalassemia/mental retardation X-linked) syndrome (69). In the x-ray structure of the MBD-DNA complex (15) the two discontinuous MoRFs, 87–104 and 133–150 (Fig 1), form a continuous surface (Fig 9). Both  $\alpha$ -MoRFs are thus maximally accessible and constitute potential interaction surfaces for MeCP2 domains or protein partners. Strikingly, several RTT-causing mutations in the MBD are proximal to these candidate interaction surfaces, explaining the deleterious effects of these mutations on MeCP2 function (Fig. 9).

### MeCP2 harbors multiple autonomous binding sites that affect the overall interactions of the protein with DNA and chromatin

Although MeCP2 was originally identified as a protein that binds specifically to methylated DNA, it has been shown since that MeCP2 can also bind to unmethylated DNA, albeit less efficiently (11, 28, 31, 32). In this study we have shown that of the 4 autonomous DNA binding domains, ID and TRD acquire significant secondary structure upon binding to DNA, a phenomenon that has also been reported for key proteins such as Jun, Fos, GCN4 and histone H1 (70–72). MBD is solely responsible for methylation specific binding (Fig. 3b). The autonomous non-specific DNA binding capabilities of the ID, TRD and CTD- $\alpha$  suggest that full-length MeCP2 is poised to make unusually extensive contacts with DNA which coupled with their highly disordered nature and ability to undergo binding-induced structural changes, suggests that the simultaneous or selective engagement of these domains will promote considerable functional variability.

The slight but consistent enhancement of shift seen with nucleosomal arrays when incubated with CTD- $\beta$ , over DNA with which there was virtually no shift (Fig. 3d), can be attributed to a putative histone-binding region(s) in the C terminal half of MeCP2. This is in agreement with earlier work suggesting a role of the C-terminal domain in specific binding to chromatin, most likely to histone H3 (32, 73, 74).

### Properties of multi-domain fragments of MeCP2 reveal structural and functional synergism between domains

Upon incubation with NTD-MBD both DNA and NA undergoes striking enhancement of electrophoretic shift compared to the minor shift with MBD and zero shift with NTD. Furthermore a 10-fold higher binding affinity of NTD-MBD to methylated DNA compared to MBD alone (Fig. 3b, e and 8a, b), suggest that conformational coupling between these domains possibly through their MoRFs results in a synergistic increase in DNA binding efficiency, and/or methylation specificity. Furthermore, thermal unfolding suggests that the NTD holds the MBD in a relatively destabilized state which may be more potent in DNA binding as evident from the higher affinity of NTD-MBD for methylated DNA than of MBD. This is also evident

from the fact that upon binding to DNA, NTD-MBD undergoes a more pronounced thermal stabilization than the MBD.

Two distinct mechanisms for coupled binding and folding of unstructured proteins have recently been proposed: folding upon binding, and conformational selection (75). For MeCP2, both mechanisms may be operating: while the acquisition of structure by the ID and TRD domains upon binding to DNA is definitely a case of binding-induced folding, inter-domain interactions within MeCP2 could select for conformations favorable for DNA binding, as in the NTD-induced enhancement of the DNA binding affinity of MBD. In addition, the unstructured regions of MeCP2 will populate an ensemble of conformations from which those that favor binding by specific partner proteins may be selected.

Another case of synergistic binding is seen with TRD-CTD, which binds to DNA with 6-fold higher affinity than TRD and 30-fold higher affinity than CTD- $\alpha$ . It is also the most potent in terms of inducing mobility shifts of DNA and NAs (Fig. 3a, c, d). Furthermore, the remarkable compaction and self-association of NAs induced by the TRD-CTD fusion suggests contributions from both the DNA and histone binding sites in this fragment.

The apparent synergism between certain domains in DNA and chromatin binding is consistent with the conformational coupling between MBD and other domains of MeCP2 detected in *cis* and *trans*. Long-range interactions between interspersed structured segments may result in a loose folded structure with intrinsically disordered domains extending from one or two structural ‘hubs’. Such a structure is likely to promote simultaneous interaction with multiple partners conferring considerable functional flexibility to MeCP2.

### **The two halves of MeCP2 involved in DNA and chromatin binding**

The DNA binding properties of the MBD are strongly modulated by the flanking domains, suggesting that the NTD-MBD-ID region constitutes a functional entity with both methylation-dependent and independent DNA binding abilities. The TRD-CTD also functions as an independent unit with chromatin compacting and oligomerizing properties. This suggests that MeCP2 is effectively organized into an N-terminal functional unit composed of the NTD, MBD and ID and a C-terminal unit composed of the TRD, CTD- $\alpha$  and CTD- $\beta$ .

While both the TRD (207–310) and CTD- $\alpha$  (261–330) fragments bind DNA efficiently, DNA-induced structure acquisition is limited to the TRD polypeptide. This suggests that residues 207–260 within the TRD house DNA binding activity while residues 310–335 harbor the DNA binding activity seen in CTD- $\alpha$ . In support of the former, deletion mutagenesis of the isolated TRD suggests its DNA binding region lies between residues 245–270 (data not shown). While the extent of NA compaction induced by TRD-CTD approaches that seen with the wild type protein, a higher ratio of the TRD-CTD polypeptide is needed to see an equivalent effect. This would be expected if the NTD-MBD-ID mediated DNA binding acts in concert with TRD-CTD mediated DNA binding and also contributes to inter- and intra-nucleosomal associations of NAs.

### **In vitro functions of MeCP2 domains strongly correlate with their function in vivo**

Our results indicating the importance of MeCP2 domains other than the MBD and TRD are supported by *in vivo* data. For example, the heterochromatin-associated HP1 protein has been shown to interact with the NTD (35), perhaps contributing to the co-localization of MeCP2 and HP1 in pericentromeric heterochromatin. However, while the NTD appears to be necessary for the pericentromeric localization of MeCP2, it is evidently not sufficient, but also requires the MBD (76). This is in agreement with our data that NTD largely plays a role in regulating MBD mediated binding. Important roles for the ID have also been established. The ID has been

shown to be a universal component of the MeCP2 fragments required for interaction with the co-repressors mSin3A (77), N-CoR and Ski (18), H3 methyl transferase (20), p20, a putative Xenopus protease (78), as well as the DNA methyltransferase Dnmt1 (19). A fluorescence recovery after photobleaching study (36) has recently demonstrated that deletion of either the ID or the TRD markedly increases MeCP2 intranuclear mobility. This suggests that the DNA induced disorder-to-order transition of the ID and the TRD reported here significantly stabilizes MeCP2-chromatin complexes, perhaps providing a larger window for downstream repressor recruitment or secondary interactions required for the structural modulation of chromatin.

Our finding that full-length MeCP2 and TRD-CTD induce clustering of nucleosomes with looping out of non-nucleosomal DNA (Fig. 5) (31) is consistent with the growing *in vivo* evidence for a role of MeCP2 in stabilizing large chromatin loops (30, 79, 80). The presence of independent DNA and chromatin-binding domains in multiple regions of MeCP2, allowing a single MeCP2 molecule to bind two or more regions of chromatin, would contribute to the stabilization of a loop base. Interestingly, in this regard, the MBD-ID fragment has been reported to contain a MAR binding site (23). A role in loop maintenance would not be possible with the MBD alone with its very limited compaction ability, but may require the additional DNA-binding properties of the ID and the chromatin binding properties of the CTD. The pathological effects of C-terminal truncations of MeCP2 both in RTT patients and a mouse model (17) may be understood in this context.

On the basis of the work presented here, a new picture of MeCP2 biology is emerging in which its intrinsically disordered nature is a key property. The novel properties of MeCP2 elucidated here support a global structure for MeCP2 that can engage in a wide range of potential binding events and conformational changes, promoting functional outcomes which will vary according to the specific context of the gene locus, DNA methylation density, and availability of binding partners.

## Acknowledgments

We thank Dr. Stephen Eyles for assistance with analytical ultracentrifugation.

## Abbreviations

MeCP2	Methyl CpG binding protein 2
RTT	Rett syndrome
PONDR	Predictor of naturally disordered regions
FlAsH	fluorescein arsenical hairpin
NA	nucleosomal array
MoRF	Molecular recognition feature

## References

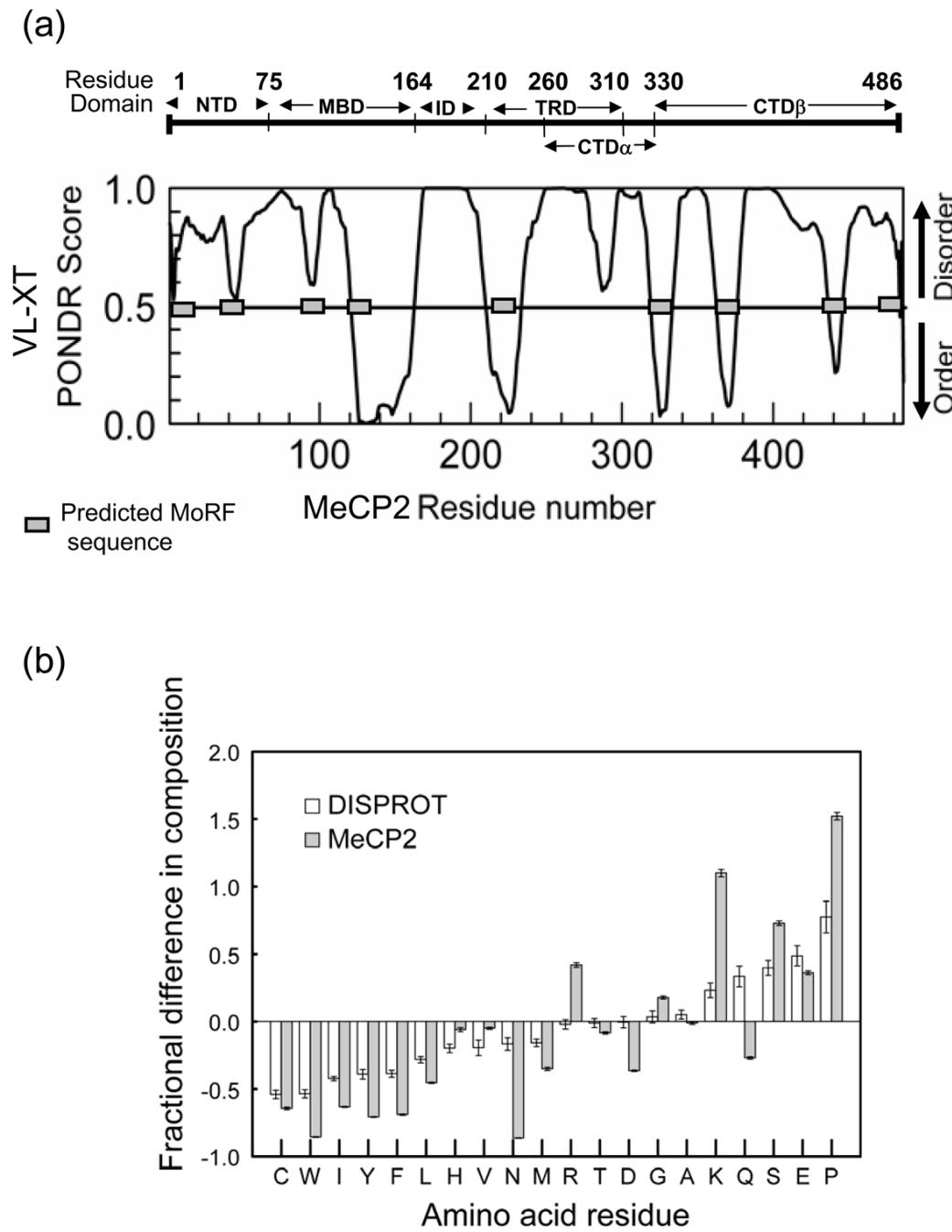
1. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006;31:91–97.
2. Dhasarthy J, Wade PA. The MBD protein family – reading an epigenetic mark? *Mut Res* 2008;647:39–43. [PubMed: 18692077]
3. Amir RE, den Veyber IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MeCP2, encoding methyl-CpG-binding protein 2. *Nat Genet* 1999;23:185–188. [PubMed: 10508514]

4. LaSalle JM, Hogart A, Thatcher KN. Rett syndrome: a Rosetta stone for understanding the molecular pathogenesis of autism. *Int Rev Neurobiol* 2005;71:131–165. [PubMed: 16512349]
5. Moretti P, Zoghbi HY. MeCP2 dysfunction in Rett syndrome and related disorders. *Curr Opin Genet Dev* 2006;16:276–281. [PubMed: 16647848]
6. Kishi N, Macklis JD. MeCP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol Cell Neurosci* 2004;27:306–321. [PubMed: 15519245]
7. Bernard D, Gil J, Dumont P, Rizzo S, Monte D, Quatannens B, Hudson D, Visakorpi T, Fuks F, de Launoit Y. The methyl-CpG-binding protein MeCP2 is required for prostate cancer cell growth. *Oncogene* 2006;25:1358–1366. [PubMed: 16331274]
8. Chatagnon A, Bougel S, Perriaud L, Lachuer J, Benhattar J, Dante R. Specific association between the methyl-CpG-binding domain protein 2 and the hypermethylated region of the human telomerase reverse transcriptase promoter in cancer cells. *Carcinogenesis* 2009;30:28–34. [PubMed: 18952593]
9. Ballestar E, Esteller M. Methyl-CpG-binding proteins in cancer: blaming the DNA methylation messenger. *Biochem Cell Biol* 2005;83:374–384. [PubMed: 15959563]
10. Ballestar E, Paz MF, Valle L, Wei S, Fraga MF, Espada J, Cigudosa JC, Huang TH, Esteller M. Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J* 2003;22:6335–6345. [PubMed: 14633992]
11. Adams VH, McBryant SJ, Wade PA, Woodcock CL, Hansen JC. Intrinsic disorder and autonomous region function in the multifunctional nuclear protein, MeCP2. *J Biol Chem* 2007;282:15057–15064. [PubMed: 17371874]
12. Nan X, Meehan RR, Bird A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nuc Acids Res* 1993;21:4886–4892.
13. Wakefield RID, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow PN. The solution structure of the region from MeCP2 that binds to methylated DNA. *J Mol Biol* 1999;291:1055–1065. [PubMed: 10518942]
14. Ohki I, Shimotake N, Fujita Nn, Jee JG, Ikegami T, Nakao M, Shirakawa M. Solution structure of the methyl-CpG binding region of human MBD1 in complex with methylated DNA. *Cell* 2001;105:487–497. [PubMed: 11371345]
15. Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MW. MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell* 2008;29:525–531. [PubMed: 18313390]
16. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 1997;88:471–481. [PubMed: 9038338]
17. Shahbazian M, Young J, Yuva-Paylor L, Spencer C, Antalffy B, Noebels J, Armstrong D, Paylor R, Zoghbi H. Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* 2002;35:243–254. [PubMed: 12160743]
18. Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, Rosenfeld MG. Corepressor-dependent silencing of chromosomal regions encoding neuronal gene. *Science* 2002;298:1747–1752. [PubMed: 12399542]
19. Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S. The Ski protein family is required for MeCP2-mediated transcriptional repression. *J Biol Chem* 2001;276:34115–34121. [PubMed: 11441023]
20. Kimura H, Shiota K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyl transferase, Dnmt1. *J Biol Chem* 2003;288:4806–4812. [PubMed: 12473678]
21. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Bird AP. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* 2003;278:4035–4040. [PubMed: 12427740]
22. Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Oikawa T. Direct association between PU.1 and MeCP2 that recruits mSin3A-HDAC complex for PU.1-mediated transcriptional repression. *Oncogene* 2003;22:8688–8698. [PubMed: 14647463]
23. Buschdorf JP, Stratling WH. A WW region binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J Mol Med* 2004;82:135–143. [PubMed: 14618241]
24. Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L, Craig JM, Jones PL, Sif S, El-Osta A. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* 2005;37:254–264. [PubMed: 15696166]

25. Jeffery L, Nakielny S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J Biol Chem* 2004;279:49479–49487. [PubMed: 15342650]
26. Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S, Zoghbi HY. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci USA* 2005;102:17551–17558. [PubMed: 16251272]
27. Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 2008;18:756–764. [PubMed: 18952168]
28. Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP, Thatcher KN, Farnham PJ, LaSalle JM. Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc Natl Acad Sci U S A* 2007;104:19416–19421. [PubMed: 18042715]
29. Chahrour M, Jung SJ, Shaw C, Zhou X, Wong STC, Qin J, Zoghbi HY. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 2008;320:1224–1229. [PubMed: 18511691]
30. Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* 2005;37:31–40. [PubMed: 15608638]
31. Nikitina TN, Shi X, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Woodcock CL. Multiple interactions between MeCP and chromatin. *Mol Cell Biol* 2007a;27:864–877. [PubMed: 17101771]
32. Georgel PT, Horowitz-Scherer RA, Adkins N, Woodcock CL, Wade PA, Hansen JC. Chromatin compaction by human MeCP2 – assembly of novel secondary structures in the absence of DNA methylation. *J Biol Chem* 2003;278:32181–32188. [PubMed: 12788925]
33. Chadwick LH, Wade PA. MeCP2 in Rett syndrome: transcriptional repressor or chromatin architectural protein. *Curr Opin Genet Dev* 2007;17:1–5.
34. Ghosh RP, Horowitz-Scherer RA, Nikitina T, Giersch LM, Woodcock CL. Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J Biol Chem* 2008;283:20523–20534. [PubMed: 18499664]
35. Agarwal N, Hardt T, Brero A, Nowak D, Rothbauer U, Becker A, Leonhardt H, Cardoso MC. MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nuc Acids Res* 2007;35:5402–5408.
36. Kumar A, Kamboj S, Malone BM, Kudo S, Twiss JL, Czymbek K, LaSalle JM, Schanen NC. Analysis of protein regions and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J Cell Sci* 2008;121:1128–1137. [PubMed: 18334558]
37. Laccone F, Huppke P, Hanefeld F, Meins M. Mutation spectrum in patients with Rett syndrome in the German population: evidence of hot spot regions. *Hum Mut* 2001;17:183–190. [PubMed: 11241840]
38. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 2005;44:12454–12470. [PubMed: 16156658]
39. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). *J Mol Biol* 2006;362:1043–1059. [PubMed: 16935303]
40. Martin BR, Giepmans BN, Adams SR, Tsien RY. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* 2005;10:1308–1314. [PubMed: 16155565]
41. Lobley A, Whitmore L, Wallace BA. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 2002;18:211–212. [PubMed: 11836237]
42. Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nuc Acids Res* 2004;32:W668–673.
43. Provencher SW, Glöckner J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 1981;20:33–37. [PubMed: 7470476]

44. Pancoska P, Bitto E, Janota V, Urbanova M, Gupta VP, Keiderling TA. Comparison of and limits of accuracy for statistical analyses of vibrational and electronic circular dichroism spectra in terms of correlations to and predictions of protein secondary structure. *Prot Sci* 1995;4:1384–1401.
45. Perczel A, Park K, Fasman GD. Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Anal. Biochem* 1992;203:83–91.
46. Greenfield NJ. Circular dichroism analysis for protein-protein interactions. *Methods Mol Biol* 2004;261:55–78. [PubMed: 15064449]
47. Lakowicz JR. Principles of fluorescence spectroscopy. 3. Springer; New York: 2006.
48. Luedtke NW, Dexter RJ, Fried DB, Schepartz A. Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nat Chem Biol* 2007;3:779–784. [PubMed: 17982447]
49. Krishnan B, Giersch LM. Cross-strand split tetra-Cys motifs as structure sensors in a beta-sheet protein. *Chem Biol* 2008;15:1104–1115. [PubMed: 18940670]
50. Griffin BA, Adams RS, Tsien YR. Specific covalent labeling of recombinant protein molecules inside living cells. *Science* 1998;281:269–272. [PubMed: 9657724]
51. Findlay JW, Dillard RE. Appropriate calibration curve fitting in ligand binding assays. *AAPS J* 2007;9:E260–E267. [PubMed: 17907767]
52. Van Holde KE, Weischet WO. Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers* 1978;17:1387–1403.
53. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 2001;42:38–48. [PubMed: 11093259]
54. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. DisProt: the database of disordered Proteins. *Nuc Acids Res* 2007;35:D786–93.
55. Iakoucheva LM, Dunker KA. Order, disorder, and flexibility: prediction from protein sequence. *Structure* 2003;11:1316–1317. [PubMed: 14604521]
56. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 2007;19:211. [PubMed: 17578581]
57. Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 1998;276:19–42. [PubMed: 9514715]
58. Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* 2003;302:885–889. [PubMed: 14593183]
59. Hansen JC, Lebowitz J, Demeler B. Analytical ultracentrifugation of complex macromolecular systems. *Biochemistry* 1994;33:13155–13163. [PubMed: 7947722]
60. Heitmann B, Maurer T, Weitzel JM, Stratling W, Kalbitzer HR. Solution structure of the matrix attachment region-binding domain of chicken MeCP2. *Eur J Biochem* 2003;270:3263–3270. [PubMed: 12869202]
61. Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* 2007;7:2833–2842. [PubMed: 17640003]
62. Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, Llopis J, Tsien RY. New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J Am Chem Soc* 2002;124:6063–6076. [PubMed: 12022841]
63. Shuman CF, Jiji R, Kerfeldt KS, Linse S. Reconstitution of calmodulin from domains and subdomains: influence of target peptide. *J Mol Biol* 2006;358:870–881. [PubMed: 16530223]
64. Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* 2003;302:890–893. [PubMed: 14593184]
65. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 2005;579:3346–3354. [PubMed: 15943980]
66. Cortese MS, Uversky VN, Dunker AK. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol* 2008;98:85–106. [PubMed: 18619997]

67. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9, Suppl 2008;1:S1.
68. Mészáros B, Tompa P, Simon I, Dosztányi Z. Molecular principles of the interactions of disordered proteins. *J Mol Biol* 2007;372:549–561. [PubMed: 17681540]
69. Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X, Kriaucionis S, Bird A. Interaction between chromatin proteins MeCP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc Natl Acad Sci U S A* 2007;104:2709–2714. [PubMed: 17296936]
70. Patel L, Abate C, Curran T. Altered protein conformation on DNA binding by Fos and Jun. *Nature* 1990;347:572–575. [PubMed: 2120592]
71. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. Folding transition in the DNA-binding region of GCN4 on specific binding to DNA. *Nature* 1990;347:575–578. [PubMed: 2145515]
72. Roque A, Iloro I, Ponte I, Luis J, Arrondo R, Suau P. DNA induced secondary structure of the carboxy terminal region histone H1. *J Biol Chem* 2005;280:32141–32147. [PubMed: 16006555]
73. Chandler SP, Guschin D, Landsberger N, Wolffe AP. The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry* 1999;38:7008–7018. [PubMed: 10353812]
74. Nikitina TN, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Grigoryev SA, Woodcock CL. MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J Biol Chem* 2007;282:28237–28245. [PubMed: 17660293]
75. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol* 2009;19:31–38. [PubMed: 19157855]
76. Marchi M, Guarda A, Bergo A, Landsberger N, Kilstrup-Nielsen C, Ratto GM, Costa M. Spatio-temporal dynamics and localization of MeCP2 and pathological mutants in living cells. *Epigenetics* 2007;2:187–197. [PubMed: 17965612]
77. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998;393:386–389. [PubMed: 9620804]
78. Carro S, Bergo A, Mengoni M, Bachi A, Badaracco G, Kilstrup-Nielsen C, Landsberger N. A novel protein, Xenopus p20, influences the stability of MeCP2 through direct interaction. *J Biol Chem* 2004;279:25623–25631. [PubMed: 15056664]
79. Weitzel JM, Buhrmester H, Strätling WH. Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. *Mol Cell Biol* 2007;17:5656–66. [PubMed: 9271441]
80. Eivazova ER, Gavrilov A, Pirozhkova I, Petrov A, Iarovaia OV, Razin SV, Lipinski M, Vassetzky YS. Interaction in vivo between the two Matrix attachment regions flanking a single chromatin loop. *J Mol Biol* 2008;386:929–37. [PubMed: 19118562]

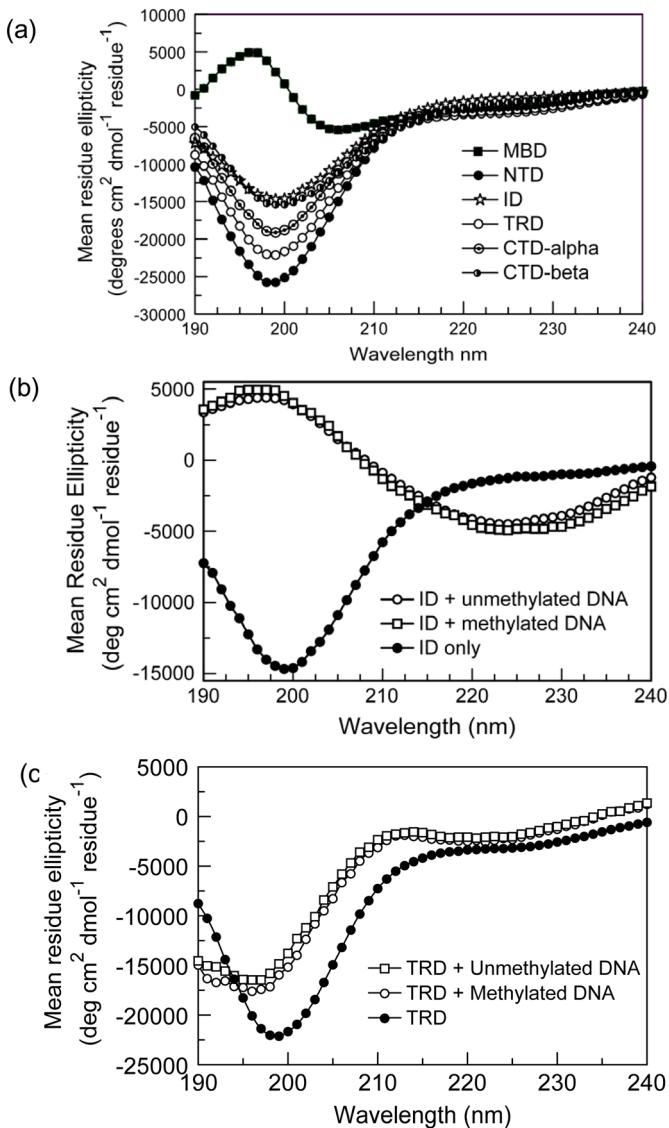


**Figure 1. Organization of MeCP2 and relation to disorder predictions**

(a) Upper panel – map of MeCP2 showing the six major domains identified by partial proteolysis (11). The graph shows the order-disorder score of MeCP2 predicted by PONDR VLXT, a neural network predictor of native disorder (39). Grey bars denote predicted molecular recognition features (MoRFs) – see discussion.

(b) The amino acid composition of MeCP2 is characteristic of a highly unstructured protein. Bar chart (filled bars) shows differences in amino acid composition between MeCP2 and the average composition of a set of ordered proteins for each amino acid. Positive values and negative values correspond to greater and lesser abundance of an amino acid in MeCP2 compared to ordered proteins.

Clear bars show the differences in average composition for each amino acid between disordered proteins from the DisProt database (54) and the same set of ordered proteins. The amino acid residues are arranged in an increasing order of disorder promoting potential (54). For explanation regarding calculation of fractional difference in composition see materials and methods.



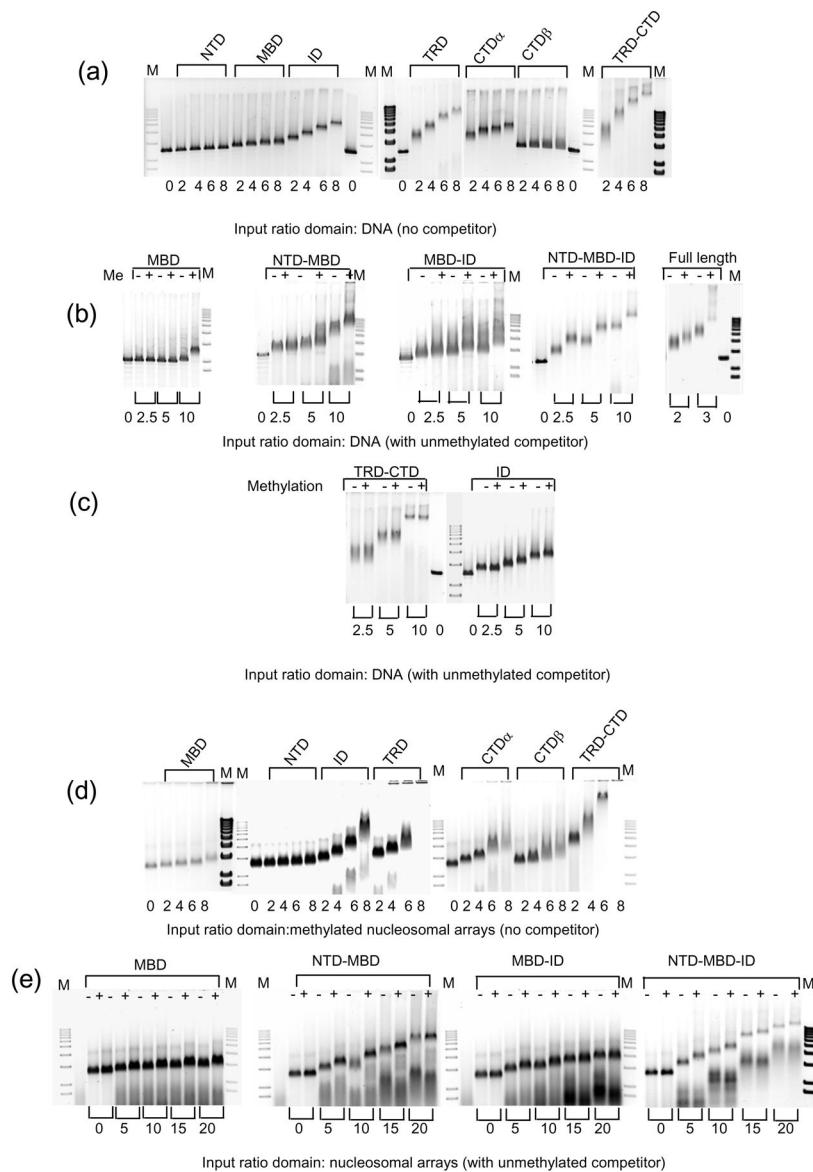
**Figure 2. Circular dichroism spectra of MeCP2 domains reveal marked differences in secondary structure content**

CD spectra are representative of two - four separate acquisitions.

(a) Compared to the 195nm peak indicative of  $\beta$ -sheet structure within the MBD (black squares), all the other domains show a negative band in the 195nm–198nm region indicative of disorder. NTD (black circles), TRD (white circles) and CTD- $\alpha$  (black rhombi) have lower structure content than ID (stars) and CTD- $\beta$  (half filled circles) (see Table 1 for quantitation).

(b) Addition of DNA (methylated as well as unmethylated) to the ID induces changes typical of the formation of  $\beta$ -structure, namely a marked increase in positive ellipticity at 195nm and negative in the 220–225nm range.

(c) Addition of DNA (methylated as well as unmethylated) to the TRD results in an increase in order irrespective of the methylation status of the DNA.



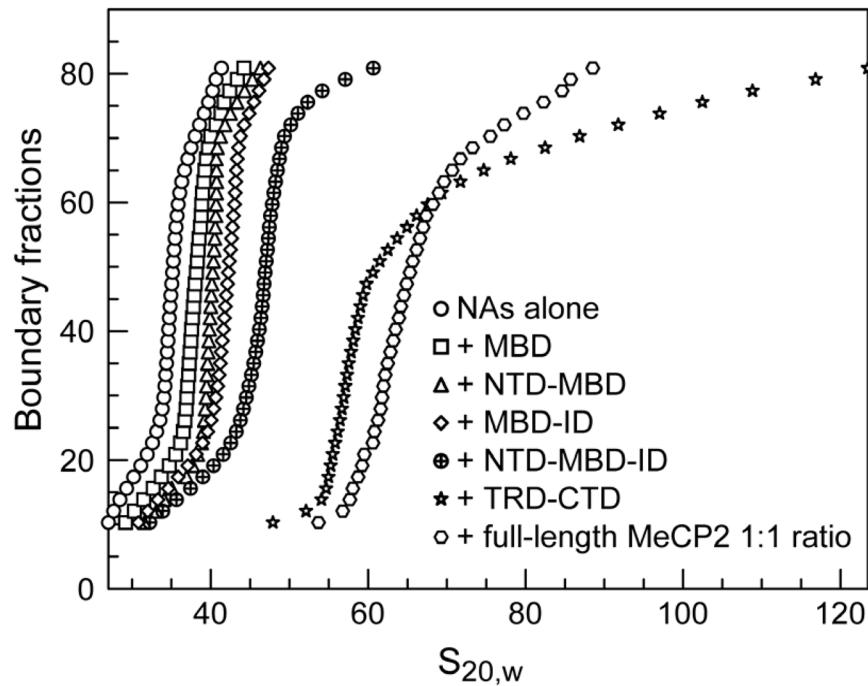
**Figure 3. MeCP2 domains induce electrophoretic mobility shifts upon addition to DNA or chromatin**

Gel images for each experiment are representative of two to three separate trials.

(a) Interaction between individual MeCP2 domains and DNA. Domains (NTD, MBD, ID, TRD, CTD- $\alpha$ , CTD- $\beta$ , TRD-CTD $\alpha$ -CTD $\beta$ ) were incubated with methylated 601-12 DNA at molar input ratios of 0 to 8, and the products are displayed on 1% agarose gels. The ID, TRD and CTD- $\alpha$  induce substantial retardation of the DNA. In contrast, the MBD shows only minor shifts and the NTD appears to have virtually no interaction with DNA.

(b) To examine methylation specificity, MBD and constructs that include its flanking domains were incubated with unmethylated (-) or methylated (+) DNA in the presence of two-fold excess of 208-1 DNA competitor at molar input ratios of 0 to 10. A distinct methylation-dependent enhancement of the gel shifts is seen in all constructs containing the MBD. Of particular interest is the large shift shown by the NTD-MBD construct, which suggests a synergism between these two domains. Full length MeCP2 produces pronounced gel shift at much lower input than the MBD containing contiguous domain fusions.

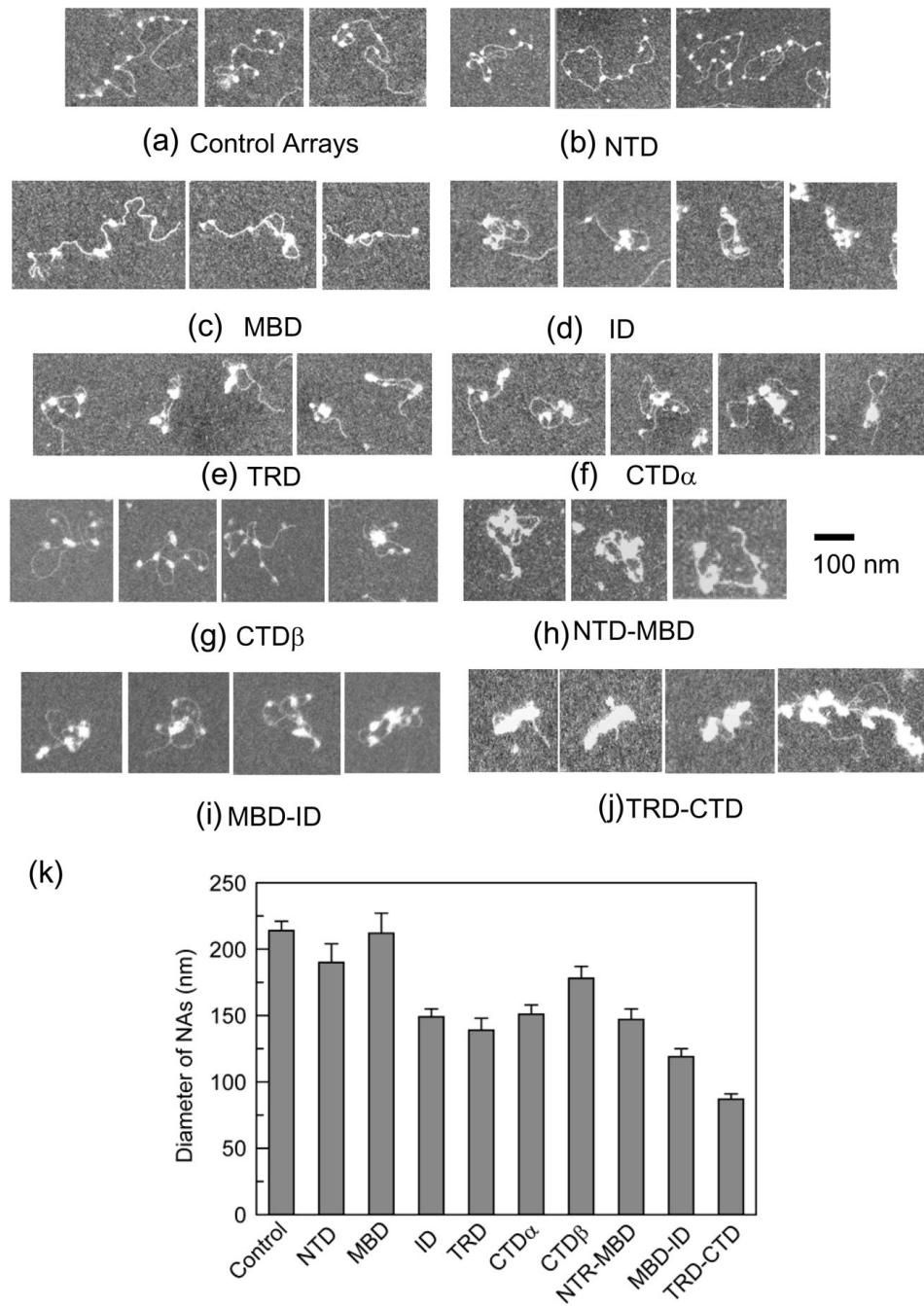
- (c) The ID and TRD-CTD polypeptides produce strong shifts, but there is no methylation-dependent enhancement.
- (d, e) as (a, b) but with 601-12 nucleosomal arrays (NAs) as substrate and 208-1 mononucleosomes as competitor. With the exception of CTD- $\beta$  which induces a moderate but consistent mobility shift with chromatin but not with naked DNA, the patterns of electrophoretic shift with DNA and NAs are similar.  
M denotes molecular weight marker lanes.



**Figure 4. Sedimentation velocity reveals differences in the ability of MeCP2 domains to compact nucleosomal arrays**

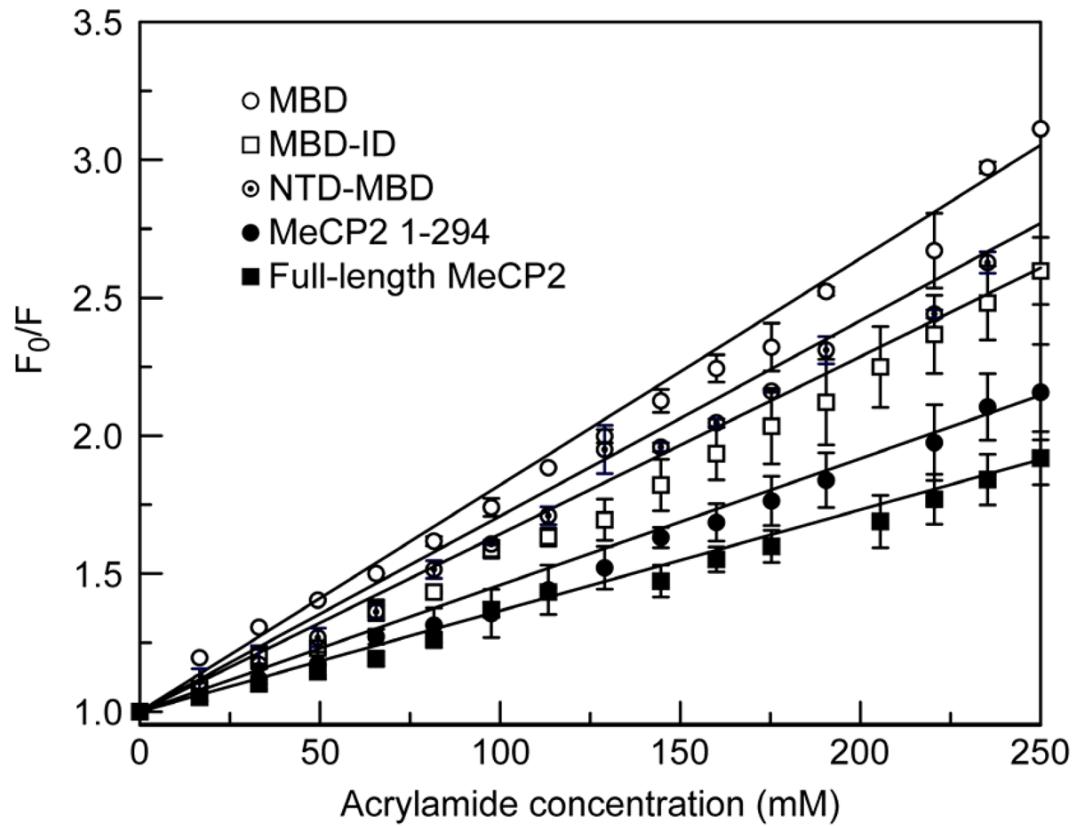
Methylated 601-12 nucleosomal arrays were incubated with a two-fold molar input of MeCP2 constructs in 50mM NaCl, 10mM Hepes, 0.25 mM EDTA and analyzed by sedimentation velocity. NAs alone (circles) MBD (squares), NTD-MBD (triangles), MBD-ID (diamonds), NTD -MBD-ID (circles with cross), TRD-CTD (stars) and full-length MeCP2 (hexagons). For characterization of the full-length MeCP2 an equimolar input of protein was used since a two-fold input causes extensive self-association and oligomerization.

Data were consistent over two to three separate trials for each experiment.

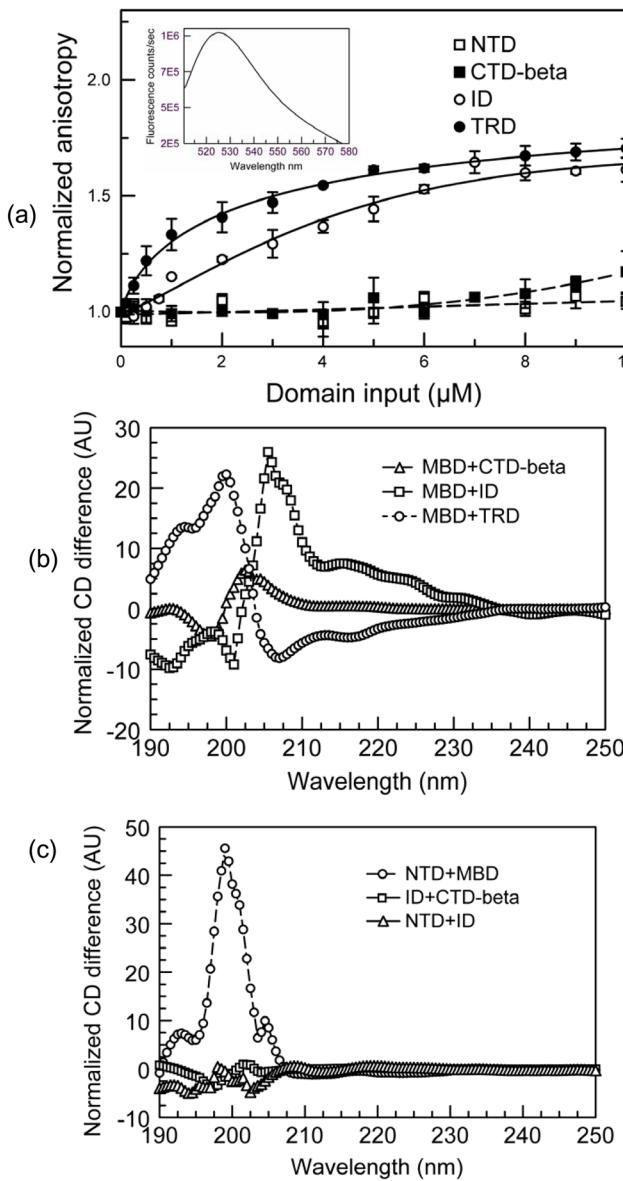


**Figure 5. Direct EM observation reveals differences in conformational changes induced in undersaturated nucleosomal arrays by MeCP2 domains**

Subsaturated NAs were mixed with different MeCP2 fragments at input ratios of 8 molecules of protein per 208 bp DNA, fixed, and imaged using darkfield EM. (a-i) Representative images of NAs showing the range of conformational changes from none for the MBD and NTD, to extensive compaction and self-association for the TRD-CTD fusion. (k) Mean array diameters with standard errors.



**Figure 6. Interactions between the MBD and flanking domains revealed by tryptophan accessibility**  
Fluorescence quenching by acrylamide analyzed using Stern Volmer plots shows that the fluorescence of Trp 104 in the MBD is differentially accessible depending on the flanking domains present. MBD only (circles), NTD-MBD (stars), MBD-ID (diamonds), MeCP2 1-294 (circles with cross), full-length MeCP2 (squares). Plots are linear up to ~250 mM acrylamide. Error bars represent standard errors of mean.

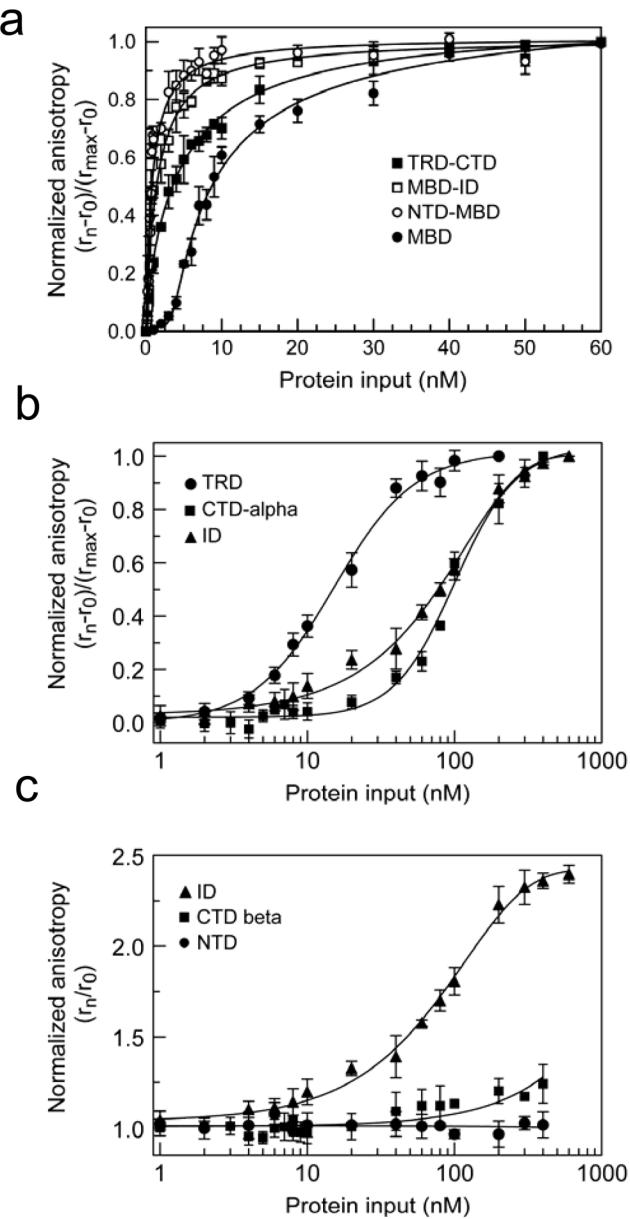


**Figure 7. In trans interactions between MeCP2 domains revealed by fluorescence anisotropy and CD**

(a) Fluorescence anisotropy of fluorescently-labeled MBD upon mixing with other MeCP2 domain constructs. In the presence of ID (open circles) and TRD (filled circles), the MBD shows marked increase in anisotropy whereas addition of NTD (open squares) caused no change in anisotropy. CTD $\beta$  (filled squares) gave a small increase at higher input ratios. Error bars denote standard errors of mean. Insert shows the robust emission spectrum of the MBD-tetraCys bound FlAsH complex.

(b, c) Fragment complementation was also detected by using CD to monitor interactions between domain pairs caused by changes in secondary structure. For each pairwise comparison (A:B), data were acquired separately for the two different domains (A) and (B) and also for their mixture (A+B). Plots show the difference spectra at each wavelength expressed as a percent of the spectrum obtained by addition of the individual spectra  $[(A)+(B)] - [(A+B)] / [(A)+(B)]$ .

- (b) MBD+ID (squares) and MBD+TRD (circles) spectra show strong differences from the composite spectrum of the individual domains while the MBD + CTD- $\beta$  pair shows only a minor change.
- (c) The NTD + MBD pair (circles) show a distinct difference at  $\sim$ 198nm while NTD+ID (triangles) and ID + CTD- $\beta$  (squares) show no differences.



**Figure 8. Quantitation of DNA binding affinity of MeCP2 fragments**

(a) Normalized fluorescence anisotropy,  $r_{norm} = (r_n - r_0)/(r_{max} - r_0)$ , of a 5'-fluorescein labeled 22bp fragment of BDNF promoter DNA with a single methylated CpG was measured in the presence of increasing concentrations of MeCP2 fragments: MBD (black squares), NTD-MBD (white circles), MBD-ID (black circles) and TRD-CTD (white squares). Error bars denote standard errors of mean. X axis (protein concentration) and Y axis (normalized fluorescence anisotropy) are linear normal.

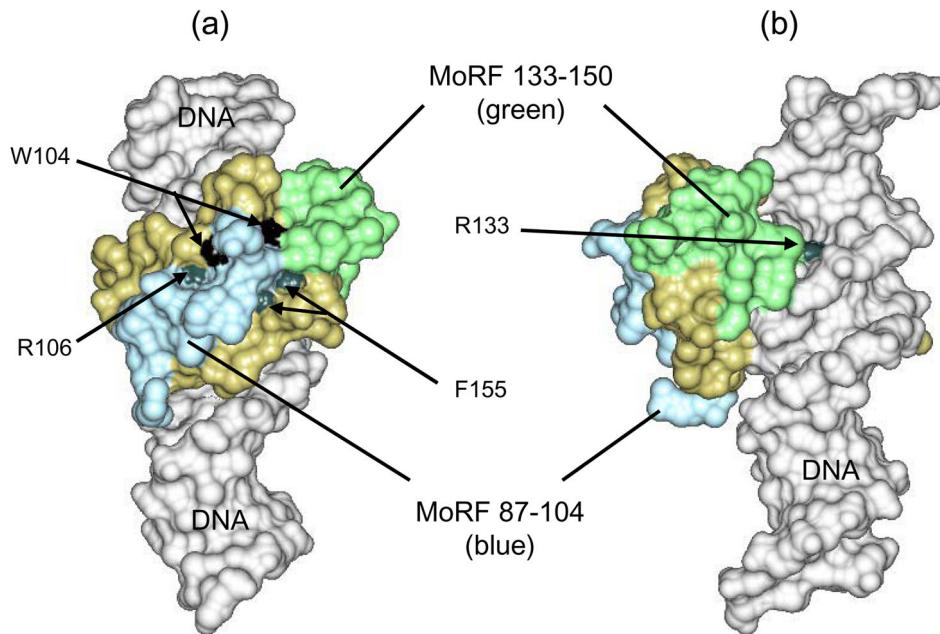
(b) as in (a) but with different MeCP2 domains: ID (black triangle), TRD (black circle), CTD- $\alpha$  (black squares). Error bars denote standard errors of mean. X axis (protein concentration) is log decimal and Y axis (normalized fluorescence anisotropy) is linear normal.

(c) Normalized fluorescence anisotropy,  $r_{norm} = (r_n/r_0)$ , of the same DNA substrate as in (a) and (b) in the presence of increasing concentrations of NTD (black circle), ID (black triangle), CTD- $\beta$  (black square).

ID is included both in 8b and 8c to provide a reference scale for the two different types of normalizations used in 8b and 8c.

Error bars denote standard errors of mean.

$r_0$  = raw anisotropy at 0 protein input,  $r_{max}$  = raw anisotropy at maximum protein input,  $r_n$  = raw anisotropy at each protein concentration and  $r_{norm}$  is the corresponding normalized anisotropy.



**Figure 9. Structure of MBD bound to DNA suggests that MoRFs flank interaction surfaces**  
Model of the MBD (tan, light blue and green) of hMeCP2 bound to 20bp of BDNF promoter DNA (gray), PDB file 3c2i (6). The MBD α-MoRFs are located in residues 87–104 (light blue) and 133–150 (green). The two MoRFs form a contiguous surface that is predominantly hydrophilic, winding across the MBD opposite the DNA interaction surface. Arrows point to the solvent accessible surface area of Trp104 (black), and the surface exposed regions (blue) of Arg106, Arg 133, and Phe155 where Rett syndrome-causing point mutations result in significant changes in the local surface properties (13). These all contribute to a MoRF surface, suggesting a role in inter- and intra-protein interactions related to MoRF disorder-to-order transitions.

**Table 1**

Domains of human MeCP2 used in this study

Nomenclature used here	Alternate name	Polypeptides prepared for this study	Number of residues
N-terminal domain (NTD)	HMGD1(11)	1–90	90
Methylated DNA binding domain (MBD) (34)		75–164	90
Intermediate domain (ID) (36)	HMGD2 (11)	165–210	46
Transcriptional repression domain (TRD) (16)		207–310	104
C-terminal domain (CTD)		261–330	70
CTD- $\alpha$			
CTD- $\beta$		335–486	156
NTD-MBD		1–164	164
MBD-ID		75–210	136
NTD-MBD-ID		1–208	208
TRD-CTD		207–486	280
MeCP2 <sup>1–294</sup>		1–294	294

Secondary structure content of MeCP2 domains based on CONTINLL deconvolution of CD data

**Table 2**

Domain	% ordered secondary structure (standard error)			% unstructured (standard error)	
	$\alpha$ -helix	$\beta$ -strand	$\beta$ -turn	total	
NTD	9 (0.3)	4 (0.3)	9 (0)	22	78 (0.6)
MBD(34)	15 (0.3)	27 (0.3)	18 (0.3)	60	40 (0.3)
ID	9 (0.9)	16 (0.33)	13 (0.6)	38	62 (0.6)
TRD	8 (0.3)	3 (0.8)	11 (0.4)	22	78 (0.5)
CTD- $\alpha$	7 (0.5)	12 (1.0)	12 (0)	31	69 (0.5)
CTD- $\beta$	9 (0.5)	16 (0)	15 (0)	40	60 (0.5)

**Table 3**

Changes in secondary structure of MeCP2 domains upon DNA binding.

Domain	% ordered secondary structure	
	Without DNA	With DNA
Full-length MeCP2	35 <sup>l</sup>	42 <sup>l</sup>
NTD	22	23
MBD	60 <sup>l</sup>	66 <sup>l</sup>
ID	38	59
TRD	22	30
CTD- $\alpha$	31	30

<sup>l</sup>From (34)

**Table 4**

## Thermal stability of MeCP2

MeCP2 polypeptide	T <sub>m</sub> alone (°C)	T <sub>m</sub> (apparent) with unmethylated DNA (°C)	Δ T <sub>m</sub> (unmet)	T <sub>m</sub> (apparent) with methylated DNA (°C)	Δ T <sub>m</sub> (met)
MBD <sup>75-164</sup>	44.9 (±0.1)	46.6 (±0.6)	1.7	54.3 (±0.3)	9.4
NTD-MBD <sup>1-164</sup>	40.4 (±0.6)	46.3 (±0.1)	5.9	54.5 (±0.3)	14
MBD-ID <sup>75-210</sup>	46.7 (±0.2)	55.7 (±0.3)	9	63.8 (±0.2)	17.1
NTD-MBD-ID <sup>1-208</sup>	43.1 (±0.1)	58.6 (±0.4)	15.5	67.1 (±0.2)	24
MeCP2 <sup>1-294</sup>	45.3 (±0.1)	56.3 (±0.3)	11	64.6 (±0.2)	19.3
Full-length MeCP2 <sup>1-486</sup>	44.5 (±0.2)	55.7 (±0.5)	11.2	63.1 (±0.4)	18.6

ΔT<sub>m</sub> – increase in melting temperature upon DNA bindingT<sub>m</sub> values in the presence of DNA are denoted 'apparent' since reversibility cannot be tested.