DISSERTATION

EXPOLRING INDUCED SECONDARY STRUCTURE AND UNMETHYLATED DNA BINDING DOMAINS OF METHYL CpG BINDING PROTEIN 2 (MeCP2)

Submitted by

Kristopher Charles Hite

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2011

Doctoral Committee:

Advisor: Jeffrey C. Hansen

Robert W. Woody
Eric D. Ross
Donald L. Mykles

ABSTRACT

EXPLORING INDUCED SECONDARY STRUCTURE AND UNMETHYLATED DNA BINDING DOMAINS OF METHYL CpG BINDING PROTEIN 2 (MeCP2)

Our understanding of Methyl CpG binding protein 2 (MeCP2) structure and function has changed and expanded considerably over the last two decades. Mutations along the entirety of the human MeCP2 gene product lead to a disease state - Rett syndrome. The clinical connection of this protein has continued to drive intense research into the nature of MeCP2 structure and function. There is now considerable and corroborated evidence that proves MeCP2 is an archetypical intrinsically disordered protein acting as a global ATP independent chromatin architectural protein. The ubiquity of MeCP2 in vertebrate neuronal nuclei has only recently been realized and has focused my investigations. Results from my work demonstrate a clear relationship between predicted α-molecular recognition features and inducible α- helical structure. From these data I suggest that inducible α-helices and maintained intrinsic disorder participate in binding the pool the twenty reported MeCP2 binding partners. In addition to structural studies I have identified two non-specific unmethylated DNA binding domains unreported in the literature at the onset of my work. I have also shown that MeCP2 acquires some secondary structural stability when bound to DNA and relatively little additional stability when bound to methylated DNA. The results presented here improve the fine resolution functional understanding of MeCP2 by observing isolated fragments of MeCP2 using both structural and functional methods. This approach is significant in and of itself as,

like the large disordered subset of all eukaryotic proteins, the full-length MeCP2 molecule has proven impossible to crystallize thus far. Therefore narrowing the amino acid residues responsible for DNA binding activity or any other measureable functionality in a solution state is valuable.

ACKNOWLEDGMENTS

Primarily I would like to express gratitude to Dr. Jeffrey C. Hansen who helped me realize my independence as a scientist. I began working with Dr. Hansen in the summer of 2006 and remained in his lab for 5 years. Not only did he help me learn how to solve problems on my own but he also guided me in the art of scientific writing. For this opportunity I am forever grateful. I also had an extraordinarily encouraging graduate committee that steered me on course to a successful defense. My graduate committee consisted of Dr. Robert Woody, Dr. Donald Mykles and Dr. Eric Ross. Not only were they fully engaged with my progress for five long years, but they took great care in editing this dissertation with a fine-toothed comb. I thank each committee member for their continued encouragement, ideas for experimental improvement, and positive attitudes in sending me out into the world with open eyes.

Throughout my six years in the MRB building at Colorado State I had the great pleasure of becoming friends with inspiring individuals. Dr. Kevin Flynn was my first lab mentor as I rotated through Dr. James Bamburg's Lab. While working with Kevin I met Dr. Chi Pak who would later become my house-mate and good friend.

As I entered the Hansen lab I became friends with my long-term lab-mates. Dr. Xu Lu was my first mentor in the Hansen lab. I was fascinated with his ability fix cars and to use the internet to save money on all kinds of things. Christine Krause and I shared space and many stories of family and music. An honest-to-

goodness giant named Dr. Michael Resch was working in the Hansen lab when I started. We instantly hit it off with our shared love of high ceilings, hiking off the beaten path, and fringe politics.

I am forever indebted to Dr. Heather Szerlong for her bravery in dog-sitting for me in my third year of graduate school. Anyone who has met my yellow Labrador retriever Ben knows he is more than a handful. When Heather unwittingly agreed to watch him while I was away one summer, Ben decided it was a good idea to swallow a gym-sock the day before I returned. Heather rushed Ben to the vet-emergency room allowing him to be sitting with me now as I type! Thanks again Heather.

My most fond memory in Colorado is of watching leonid meteors dart across the night sky above Mirror Lake in Rocky Mountain National Park on a camping trip with biochemistry buddies. This came as part of an annual pilgrimage across the high country from the west side of Rocky to the departmental retreat at Pingree Park in late August. Mike Resch, Heather Szerlong, Teri McClain, Jim Bogenberger and myself slept on a huge granite slab under the stars sharing good food and good stories. This was the closest I came to feeling like I had family in Colorado.

I would also like to thank Dr. Steve McBryant, Dr. Uma Muthurajan, Dr. Anna Kalashnikova, Troy Sorenson, and Dr. Mary Porter-Goff for their professional comradery and friendship. This community of people enlivened my

time in Fort Collins, CO and made this period of my life transformative beyond my wildest expectations.

Finally I would like to thank my family. My mother Peg and my father Ron have given me unconditional and life-long support in my pursuit of education. It brought tears of happiness to my eyes to be with them for my graduation.

TABLE OF CONTENTS

CHAPTER 1

Literature Review

1.1. Introduction: An Historical Perspective of MeCP2 Structure and Function

MeCP2 (methyl CpG-Binding Protein 2) is a 53 kDa nuclear protein named for its ability to bind methylated DNA (1,2). In addition to preferentially binding methylated DNA, it was found that MeCP2 facilitated transcriptional repression in an in vitro transcription assay using native MeCP2 from rat brain nuclear extracts (3). These initial studies set a precedent for MeCP2 function to be predicated on methylated-DNA interactions. In this chapter I will discuss the heuristic evolution of the view of MeCP2 from a single-function protein to a multifunctional nuclear protein that directly affects chromatin architecture and is involved with actively transcribed and silenced regions of chromatin (4,5).

A nuclear protein that preferentially bound methylated DNA *in vitro* without apparent regard for a consensus binding sequence was first described in 1989 and named methyl CpG-binding protein (MeCP) (6,7). This was accomplished using synthetic double stranded DNA oligomers methylated by bacterial methyl transferases. Competition of the protein from synthetic DNA probes was observed only when the competitor DNA was methylated (6). These assays selectively pulled down a protein in complexes from mouse brain, spleen, kidney, rat liver, and rabbit liver extracts. In each extract, MeCP preferentially bound the methylated DNA templates without sequence specificity. Further experiments revealed that MeCP was two distinguishable proteins: MeCP1 and MeCP2 (1), with MeCP1 requiring at least 12 symmetrically methylated CpGs and MeCP2 able to bind a single methylated CpG pair. MeCP2 was reported to be 100 times more abundant in adult somatic nuclei than MeCP1 (1). When forming the initial hypothesis of MeCP2 function, it was proposed that MeCP2 normally binds methylated DNA in the context of chromatin, leading to long-term transcriptional repression. This hypothesis was corroborated by results showing that native MeCP2 purified from rat brain extracts was released upon micrococcol nuclease digestion of a

methylated DNA probe but not present on a non-methylated DNA probe (1). The cDNA sequence for MeCP2 was also determined in this study by first deriving amino acid sequences of digested native MeCP2 protein from rat brain extract, and subsequently designing oligonucleotide primers to amplify the mRNA by the polymerase chain reaction. More evidence was provided when transiently transfected recombinant genes coding MeCP2 fused to the LacZ gene were expressed in mouse cell cultures, and similar localization to centromeric heterochomatin was observed when compared to endogenous MeCP2 stained with anti-MeCP2 antibody. The inability of the MeCP2-LacZ fusion protein to localize to centromeric hereochomatin in methyltransfererase-deficient mouse cells was provided as evidence that MeCP2 required a methylated chromatin substrate for binding (2).

The early driving hypothesis in MeCP2 research was methylated DNA binding preference. This hypothesis gave MeCP2 a name and also led to the isolation of the first functional domain: the methyl CpG-binding domain (MBD) (8). By truncating either end of full length MeCP2 and probing for methylated DNA binding preference, Nan et al. (8) demonstrated that the MBD encompassed residues 78-163. This study was conducted using murine MeCP2 in mouse cell cultures. MeCP2 homologs in other mammals have not been fully characterized. Specifically, preferences for DNA methylation by MeCP2 homologs have not been thoroughly addressed in the current literature. The preference for methylated over unmethylated DNA for the *Xenopus* MeCP2 homolog is 20 fold, whereas the preference of mouse (9) and human (10) MeCP2 for methylated compared to unmethylated DNA is 2 to 3 fold.

A solution structure of the isolated MBD was solved by NMR spectroscopy (11). The structured core appears as a wedge made up of a three-stranded anti-parallel β-sheet on one side with an α-helix on the C-terminal side. β-strands 1 and 2 are

connected by a disordered loop of 5 residues with one positively charged and two polar residues. The central wedge-shaped fold encompasses residues 103-145 of the MBD and is flanked by 26 residues on the N-terminal side and 19 residues on the C-terminal side having no detectable secondary structure. A high-resolution X-ray crystal structure of the MBD bound to DNA was published showing that the hydration state of the methyl group on the 5 carbon of cytosine stabilizes the MBD-DNA interface (12). The hydrophilic interaction between the MBD and stabilizing water molecules countered the previously proposed hypothesis that hydrophobic interactions stabilize the interaction. Of note, while the atomic structures have provided much detail about the MBD wedge motif, they do not explain the need for the flanking unstructured regions in recognizing methylated DNA.

I recently participated in a collaborative investigation in which I expressed and purified recombinant human MeCP2 protein isoform e2, and methylated and unmethylated 198 base pair 5S DNA, and reconstituted MeCP2-DNA complexes. I sent these reagents to collaborators who used hydrogen-deuterium exchange (HD/X) to precisely monitor which amino acids in the MBD were stabilized when bound to unmethylated and methylated DNA (13). Results indicate that residues in the amino terminal of MBD, but not part of the ordered “wedge” structure, gain increased solvent protection when bound to methylated versus unmethylated DNA (see appendix II). This result suggests that residues outside the reported NMR and crystal structures are essential for the observed preference of the MBD for methylated DNA.

The second MeCP2 domain to be characterized was the transcription repression domain (TRD). Using an in vitro ß-actin transcription assay, different regions of MeCP2 were fused to the Gal 4 DNA-binding domain. Results showed that residues 205-310 were required for transcriptional silencing, defining the TRD (3). Several mechanisms for

the transcriptional repression by the TRD have been proposed. These mechanisms were reviewed by Zlatanova (14). Perhaps the most compelling hypothesis for transcriptional repression is based on the observation that the TRD binds the corepressor mSin3A, which is thought to recruit histone deacetylases (15-17). In this model, MeCP2 indirectly causes changes in chromatin architecture through mediating post-translational modifications of the histone tails. This model is not universally applicable to MeCP2 function as histone deacetylase-independent transcriptional repression has been observed (18). A lack of global histone tail modifications in MeCP2 null mice also argues against this molecular model (19). Since last reviewed, transcriptional repression by MeCP2 has been connected to cancer-causing gene expression patterns. Repression of tumor-associated genes by the TRD of MeCP2 has been observed in a chromatin immunoprecipitation assay (20).

The identification and characterization of the MBD and TRD led to a more refined model in which MeCP2 functions as a methyl DNA specific proximal gene silencer that recruits co-repressors and HDACs (1,3,21). In this regard, MeCP2 also binds to other transcriptional silencing factors besides mSin3a, including N-CoR, and c-Ski (22).

When the link between mutations in MeCP2 and RTT was established in 1999, research on MeCP2 increased dramatically and new studies were undertaken relating to how the protein functions on a global genomic scale (23). The first implication that MeCP2 has an HDAC-independent role in coordinating global chromatin architecture came from in vitro studies in 2003, which demonstrated that MeCP2 could directly compact chromatin without DNA methylation, ATP, or other proteins such as mSin3A (24). This study opened the door for understanding MeCP2 as a complex multifunctional nuclear protein with a prominent role in regulating global chromatin architecture. Since

this observation, several other functions have been suggested for the protein. Some of these additional functions are depicted in Figure 1.1.

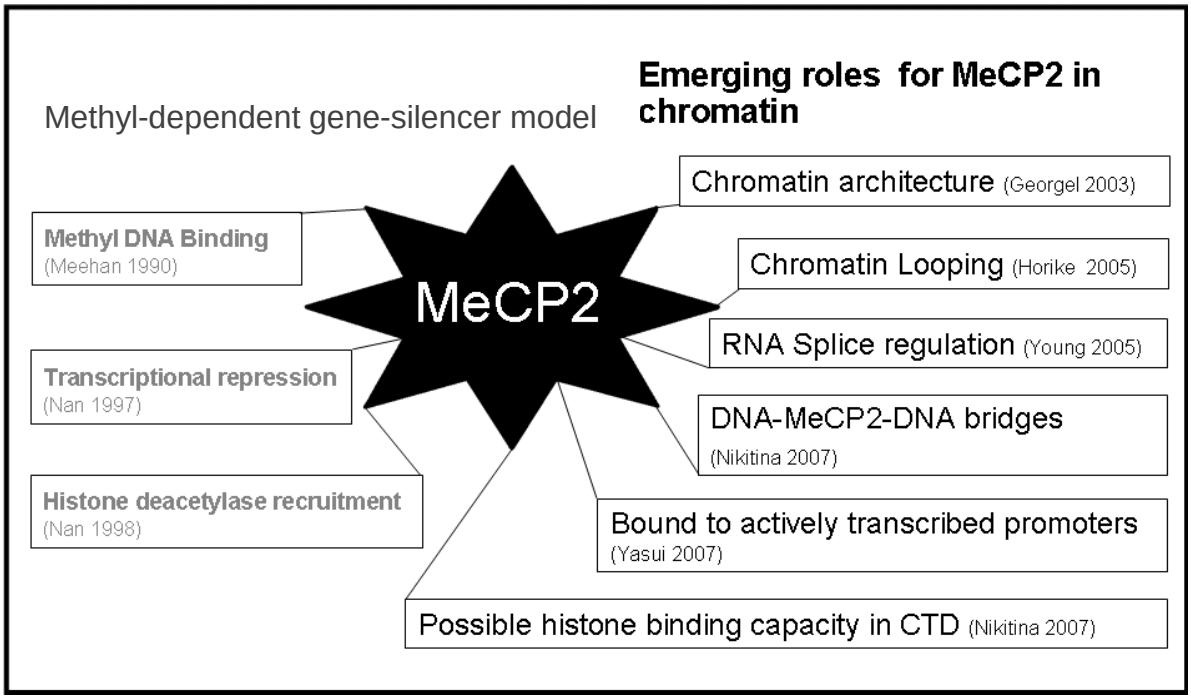

Figure 1.1. Chronological representation of MeCP2 functional understanding. Progression of traditional understanding of MeCP2 as a methyl DNA-dependent proximal gene silencer is shown in the time line. Studies implicating MeCP2 as having additional functions are listed at the right.

In addition to the proposed roles in transcriptional repression and modulation of chromatin structure, there is a link between MeCP2 function and mRNA splicing. Using coimmunoprecipitation from HeLa cell extracts, MeCP2 was shown to interact with YB-1 (25). The YB-1 protein is a highly conserved component of messenger ribonucleoprotein particles (mRNPs) and functions as the main mRNA-packaging protein. The interaction between MeCP2 and YB-1 requires the presence of RNA, as coimmunoprecipitation in HeLa cell extracts treated with RNase failed to pull down YB-1 with MeCP2. It is unclear whether the RNA bridges MeCP2 and YB-1 or stabilizes a protein-protein interaction between the two. In the same study, it was also observed that MeCP2 affects the splicing of reporter mini-genes, and a functional MBD was not required for the YB-1 interaction or splice regulation. These findings, in conjunction with the observation that there are aberrant alternative splicing patterns in a mouse model of RTT (25), implies that MeCP2 has a previously uncharacterized function as a splice-site regulator. Previously reported evidence that MeCP2 directly binds RNA with high affinity (0-10 nM) (26) is of renewed interest in light of these more recent observations. Considering Rett syndrome as a neuronal-specific phenotype caused by mutations in MeCP2, it is worth testing predictions made by the model of MeCP2 having a crucial role in splice-site regulation and RNA-binding. Predictions include neuronal-specific aberrant splice variants of those genes already observed to have dysregulated expression in MeCP2 null mouse models (27).

New data about the MeCP2 gene itself has also been reported. The identification of a MeCP2 splice variant added an extra dimension to understanding how and where MeCP2 functions. This new splice variant was labeled e1 isoform and differs from the previously characterized e2 isoform only in the segments at their extreme amino termini. The e1 isoform has a 21-residue segment with an acidic pI of 4.25, while

the e2 isoform N-terminal segment is 9 residues and has an basic pI of 9.5 (28) (see Fig. 1.2). These distinct structural differences, together with the differential distribution of the e1 and e2 isoforms between the dorsal thalamus and hypothalamus in developing post-natal mouse brains (29), suggest that there are important undiscovered differences between the functions of the two isoforms.

Based on the observation that neurological defects in a mouse model of RTT can be reversed by re-expression of the wild-type protein, understanding the molecular mechanisms of native MeCP2 and how these mechanisms are malfunctioning in RTT has potential clinical application (30). This observation is significant in establishing RTT as a neurodevelopmental rather than neurodegenerative disorder, and is consistent with a model for MeCP2 as a positive and negative regulator of transcription, a gene-specific splicing factor, and a chromatin architectural protein.

Since the time of this initial “rescue” study a clinical research group has had success in rescuing dysfunctional synaptic activity in MeCP2 knock-out mice. They achieved this rescue by injecting exogenous brain-derived neurotrophic factor (BDNF) into the brainstem nucleus *tractus solitarius* site in MeCP2-null mice. They hypothesize that application of exogenous BDNF may be a viable treatment for Rett patients (31).

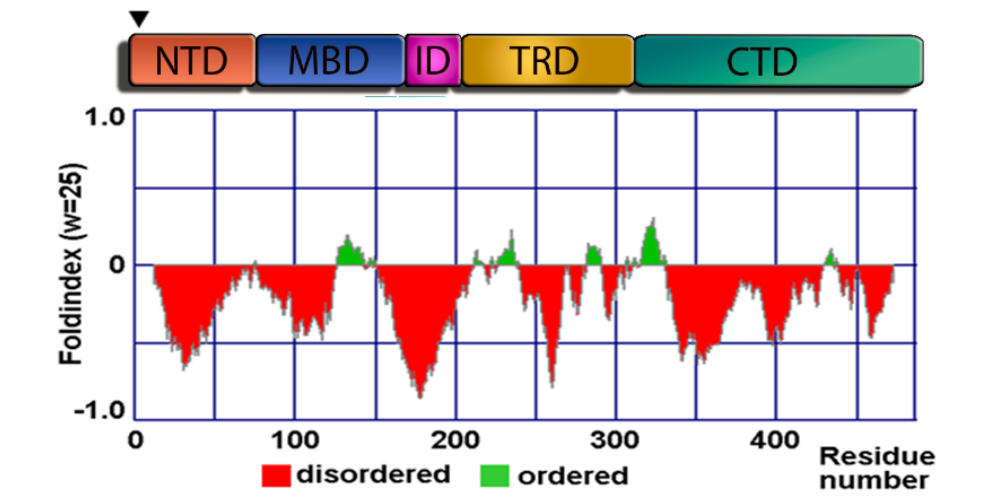

Figure 1.2 Schematic overlay of MeCP2 Foldindex and domain organization. The full length protein sequence of MeCP2 was entered into the fold index algorithm (http://bioportal.weizmann.ac.il/fldbin/findex) with a window setting of 25 and a step value of 1. Note that under these setting the first and last 25 amino acids are not represented in the final plot. The biochemically defined domains are presented to scale above the fold index plot. (▼) N-terminal isoform junction region. Neither exon 1 or exon 2 are represented in this fold index graph due to the window setting of 25.

1.2 The Structural and Intrinsically Disordered Domains of MeCP2

Protein function is inexorably linked to structure; therefore I will first consider recent advances in understanding the fundamental biochemistry of MeCP2. In particular, recent solution biophysical and protease digestion experiments have established that native MeCP2 is an intrinsically disordered protein composed of at least six distinct domains (32) (Fig 2). Trypsin digestion of MeCP2 has the potential to occur throughout the length of the polypeptide chain as there are approximately ninety potential digestion sites dispersed evenly throughout the human MeCP2 sequence. The rate of fragment appearance is related to trypsin accessibility to the tertiary structure. N-terminal sequence analysis of kinetically stable tryptic bands identified six distinct MeCP2 domains (32). Listed from amino to carboxy termini, these are the NTD, MBD, ID, TRD, CTD-α, and CTD-β (Fig. 1.2). Of note, the observed trypsin cleavage sites mapped to the boundaries of the two well characterized functional domains (the MBD and TRD). There were also additional trypsin cleavage sites within both the MBD and TRD (32). The N-terminal domain (NTD) shares amino acid composition similarity with the HMGA2 protein and was rapidly digested to completion once released from the neighboring MBD domain. Likewise, the second HMG-like domain now called the intervening domain (ID) was not detected as a kinetically stable band. The TRD sequence was recovered from the trypsin-digestion experiment indicating that the availability of trypsin cleavage sites is more restricted in this domain. The experimental approach allowed the CTD to be divided into a CTD-α (residues 310-354) and a CTD-β (residues 355-486) (32). The CTD-β contains two identifiable sequence motifs: seven consecutive histidines between residues 366-72 and a group 2 WW binding protein motif within a larger proline-rich region at residues 381-393. The His-rich motif in MeCP2 is highly conserved between

species. The WW binding protein motif (residues 384-387) has been demonstrated to interact with splicing and transcription factors (33,34). Interestingly, the proline-rich region in the CTD-β (residues 355-486) has been shown to bind HMGB1 (35). At the level of chromatin structure, the CTD is required for MeCP2-mediated chromatin compaction. Nikitina et. al. showed that the R294X MeCP2 mutant was able to bind naked DNA as well as wild-type but could not condense chromatin into higher order structures (36). This result is consistent with the early studies of Chandler, who showed that a region of the MeCP2 CTD contributed to the footprint of MeCP2 on nucleosomes (37).

While delineating the structural domains of MeCP2 was an important advance, understanding how these domains are organized into a functional tertiary structure is essential for understanding the normal cellular functions of the protein, and why certain mutations lead to a RTT phenotype. The next section discusses recent studies indicating that MeCP2 has an atypical tertiary structure permeated with an unusual amount of disorder.

1.3 MeCP2 Tertiary Structure

Due to the lack of NMR or X-ray crystal structure data, characterization of the tertiary structure of full-length MeCP2 has been accomplished by solution biochemical and biophysical methods. Circular dichroism (CD) of recombinant human MeCP2 showed that full-length protein was approximately ~35% β-strand/turn, 5% α-helix and almost 60% unstructured. CD further indicated that the isolated MBD fragment (residues 78-168) was ~10% α-helix, 51% β -strand/turn, and 38% unstructured (32), levels that approximated the amount of secondary structure seen by NMR (11). A recombinant

fragment of MeCP2 comprising the TRD (residues 198-305) was 85% unstructured according to CD. When studied by analytical ultracentrifugation, MeCP2 behaved as a monomer over a wide range of ionic conditions and molar concentrations, had an unusually low sedimentation coefficient (2.2 S), and a correspondingly high frictional coefficient ratio (f/f_o = 2.4). These results were in agreement with sucrose gradient results obtained at a single set of experimental conditions (38). Importantly, the CD data and the high f/f_o value indicate that MeCP2 has a coil-like tertiary structure similar to that of a partially denatured protein (32). Based on these observations it is not surprising that MeCP2 has been reported to have certain anomalous physicochemical properties. For example, the 53 kDa MeCP2 monomer yields an apparent molecular mass of 500 kDa according to gel filtration, and migrates at an apparent mass of 75-80 kDa on SDS gels (38).

The CD and sedimentation results, together with the anomalous behavior in gel matrices, indicate that the MeCP2 tertiary structure possesses the features of an intrinsically disordered protein (32). The concept of intrinsic or native disorder in proteins has recently gained much attention (39-44). Prediction algorithms indicate that there is preponderance of intrinsic disorder (ID) in proteins such as transcription factors (45,46). It has been hypothesized that the presence of ID permits transient, low-affinity protein-protein and protein-nucleic acid interactions (41,45). In the case of MeCP2, the location of order and disorder can be reproducibly predicted by several programs, e.g., PONDR (47) and FoldIndex (48). The FoldIndex prediction plot for MeCP2 is shown in Figure 1.2 aligned with the domains of MeCP2. Unlike the core and linker histones, which are predicted to have disorder at their terminal domains, MeCP2 is predicted to have short stretches of order interspersed between long stretches of internal disorder over the length of the entire peptide chain. Of note, many of the predicted order/disorder

boundaries coincide with the boundaries of the structural domains identified by proteolysis. Also, the known functional domains (i.e., the MBD and TRD) are predicted to be significantly disordered, consistent with the experimental data. Taken together, all available evidence suggests that ID is a key determinant of MeCP2 tertiary structure, and is likely to be an important feature of all six MeCP2 structural domains (Figs. 2, 3).

Several conceptual models of MeCP2 tertiary structure can be imagined to fit the data. In one model, several β-sheet/turn structural motifs are interspersed along the length of the MeCP2 amino acid chain, connected by disordered regions, and with no inter-motif interactions. This could be viewed as an "inchworm" model. However, electron micrographs of MeCP2 do not support this model (36). Another model is based on the CD finding that beta strands/turns are the predominant type of classical secondary structure in MeCP2. In this model, MeCP2 strand-forming regions that are separated by tens or possibly hundreds of amino acids apart on the linear polypeptide chain are connected by beta sheets or other forms of beta structures. For example, one can picture a small half beta barrel with large regions of intrinsic disorder making up loops on either side of the beta sheet stacks. In Chapter 4 of this dissertation I propose an alternative model to explaining how MeCP2 may orchestrate chromatin architecture. This model is consistent with the data I will present in Chapters 2 and 3. I discuss this model in depth in section 4.2.1 "Alternative hypothesis and future directions."

Although the new studies represent significant progress, a better understanding of the unusual tertiary structure of MeCP2 will be necessary to decipher the molecular links between MeCP2 domain organization, the multifunctionality of the protein, and the cellular pathogenesis of RTT.

1.4 Evidence for Global Genomic Functions of MeCP2

A recent genomics paper has raised questions about the current paradigms of how MeCP2 acts at the cellular level. In the first study a ChIP-to-chip approach was used. Yasui et. al. have performed an overall epigenomic binding analysis of MeCP2 (5). Two important aspects of MeCP2 function were revealed by this work. First, MeCP2 is not always associated with transcriptionally repressed genes and instead is often associated with actively transcribed genes. Second, the majority (59%) of double-stranded DNA binding sites for MeCP2 within the genome are thousands of base pairs away from intragenic regions, let alone methylated promoters. These new results suggest that the fundamental question of what MeCP2 is doing in the nucleus remains to be answered. Clearly, the finding that a majority of MeCP2 molecules bind actively transcribed promoters indicates that this protein has additional functions that are not explained by the traditional proximal gene silencer model.

This ChIP-to-chip analysis consisted of scanning 26.3 Mb of imprinted and non-imprinted chromosomal loci of known or suspected genes targeted by MeCP2. A custom high-density gene chip or oligonucleotide (oligo) micro-array was constructed with 50-mer oligos attached with a step of 32 base pairs per 50-mer such that each successive oligo contained 18 bp in the 5' end of the oligo matching the last 18 bp of the 3' end of the previous oligo. Repetitive sequences were removed prior to gene chip construction so only unique sequences were assayed for binding in the selected chromosomal regions. Human neurons from the SH-SY5Y cell line were used as the source of chromatin for immunoprecipitation of MeCP2. The SH-SY5Y cell line was selected based on its ability to doubly express MeCP2 during differentiation (49). The high affinity IgY specific antibody used in this study recognized the C-terminus of MeCP2. The c-terminus is identical in both e1 and e2 MeCP2 splice variants, so the antibody did

not distinguish between isoforms. This yielded an averaged binding along the assayed chromosomal cross-section. Further understanding of the difference between the two isoforms would come with a similar epigenomic analysis using isoform-specific antibodies.

In the CHiP to chip experiments, MeCP2 was observed to occupy many active promoters and bind mostly to non-methylated sites along intergenic spaces (5). Of those intergenic binding sites 58.4% were 10 kb or more away from transcription start sites or transcription end sites. Interestingly, the majority of the MeCP2 intragenic binding sites were intronic, which is consistent with a potential functional role for MeCP2 in pre-mRNA splicing (25). The other main conclusion gathered by the epigenomic ChIP-to-chip analysis was that the majority (62.6%) of MeCP2-bound promoters are actively expressed genes including, for example, the immediate early response gene, JUNB (5).

Perhaps the crux of this ChIP-to-Chip analysis, and the most contradictory to the initial paradigm of MeCP2 function, was the finding that MeCP2 is not concentrated at densely methylated promoters. To determine which of the assayed promoters had the highest methylation levels, genome-wide promoter methylation analysis by methylated DNA immunoprecipitation (MeDIP) was performed on a microarray containing 24,275 presumed human promoters (5). This methylation-seeking immunoprecipitation technique was applied to identically differentiated SH-SY5Y cells. Comparing the results from the promoter methylation-dependent immunoprecipitation assay to those from the MeCP2-binding assay revealed that only 2.2% of the top 4,062 promoters with highest measurable levels of methylation were bound by MeCP2. The finding that MeCP2 is not concentrated at densely methylated promoters contradicts the original paradigm of MeCP2 function. The hypothesis that MeCP2 promoter binding is coupled to transcriptional repression is further compromised by the finding that with certain

promoter occupancy, expression levels of the target protein actually increase. The observation that the RNASEH2A gene expression decreases with MeCP2 deficiency is another specific example implicating MeCP2 as a regulator rather than strictly a repressor of targeted genes associated with neuronal development (5).

There are caveats to the applicability of results generated by in vitro ChIP to chip techniques to an *in vivo* model of RTT. Using neuroblastoma cells as a source of chromatin may not reflect tissue-specific chromatin binding profiles for MeCP2 in normal neuronal tissue, and low-affinity DNA binding events may be over-exaggerated on the gene chip results. However, the model of MeCP2 as being associated with actively transcribed genes and a majority of non-methylated CpG sequences is reinforced by more recent literature. Gene expression patterns studied using the hypothalamus of mice reveal that nearly 85% of genes bound by MeCP2 are actively transcribed (4). This study also adds to the list of MeCP2 interaction partners by demonstrating that CREB1 protein is pulled down in chromatin immunoprecipitation with MeCP2. The specificity of the interaction is further established by MeCP2 interacting with CREB1 at activated promoters but not at repressed targets.

Recent *in vivo* work also corroborates the conclusions of the CHiP-to-chip epigenomic MeCP2 binding analysis. MeCP2-knockout mice and RTT patients do not aberrantly express genes regulated by promoter methylation (50,51). These results, together with the evidence that MeCP2 binds a majority of actively transcribing promoters, argue for MeCP2 having a role as both a positive and negative regulator of transcription (4,5). This evidence further supports the view that MeCP2 recognizes something in addition to, or other than, methylation for certain cellular functions, and suggests a model in which MeCP2 functions as an architectural chromatin protein, and

both a positive and negative regulator of transcription rather than a gene-specific, methylation-directed silencer.

As mentioned previously, in vitro studies are also consistent with a role for MeCP2 in regulating global chromatin architecture independent of methylation status. Genomic double-stranded DNA is packed in chromatin by wrapping around octamers of basic histone proteins which form nucleosomes (52). Nucleosomes are interspersed at semi-regular intervals along the DNA strand, creating arrays of nucleosomes termed chromatin. MeCP2 has been shown to bind to chromatin fibers and directly compact them into folded and oligomeric structures when bound to model nucleosomal arrays (10,24,36). The arrays used in these studies were unmethylated, implicating a role for MeCP2 in modulating chromatin architecture independent of methylation status. Though it is unequivocal that MeCP2 preferentially binds methylated DNA, and even more specifically the methylated linker DNA of mononucleosomes and nucleosomal arrays (53), the fact remains that MeCP2 is able to significantly alter higher order chromatin structure independent of methylation remains a clearly documented property of the protein, and is likely due in part to the presence of multiple MeCP2 DNA/chromatin binding domains besides the MBD (see above).

These observations have since been further corroborated by work from the lab of Dr. Adrian Bird. They recently reported that MeCP2 is expressed at nearly one molecule of MeCP2 per every two histone octamers in neuronal tissue (54). Though their own data demonstrate genome-wide distribution of MeCP2 in neuronal tissue, this group of researchers still maintains that MeCP2 “tracks” methylation. I would argue that though MeCP2 may concentrate slightly at sites of genome methylation, the vast majority of MeCP2 binding sites in neuronal nuclei are methylation-independent. They do concede that their data suggest a genome-wide role for MeCP2.

> "Our data argue that MeCP2 may not act as a gene-specific transcriptional repressor in neurons, but might instead dampen transcriptional noise genome-wide in a DNA methylation-dependent manner" (54).

The abundance of MeCP2 in adult somatic nuclei (1-5 $x10^5$ molecules per nucleus) implies that there is a global genomic role for MeCP2. The *in vitro* studies of chromatin compaction are currently the best explanation for the genomic occupancy of MeCP2. Even though MeCP2 may function as a global chromatin architectural protein, no change in histone modification profiles were observed compared to wild-type in MeCP2-null mice (19). This finding can be reconciled by considering that MeCP2 can modify genomic chromatin architecture directly and independently of histone modifier proteins (24,36). Considering that specific genes such as BDNF and TRKB are either down-or-up regulated respectively in the MeCP2-knockout mouse model, this leaves open the prospect of unidentified factors directing MeCP2 localization (55).

1.5 Potential Clinical Correlation between MeCP2 Domains and Rett Syndrome:

Missense and nonsense mutations that cause RTT are found in all six structural domains of MeCP2 (56). This argues that all domains are required for proper function of the protein. Toward this end, recent studies have attempted to correlate the location of the mutation with specific facets of RTT pathogenesis. Though the majority of RTT cases are caused by either single point mutations or truncation mutations in MeCP2, the diagnosis of RTT is defined via clinical observation, not genetically (57).

Perhaps the most influential factor in disease severity is the type of X-chromosome inactivation (XCI). Females with favorably skewed XCI may have very mild

learning disabilities while a spectrum of severity correlates with increasingly unfavorable XCI to autism and on to later onset RTT (57). Some attempts have been made to correlate MeCP2 mutations to disease severity by comparing data in the international RTT database to clinically reported severity. Two common RTT mutations in MeCP2, the R168X and T158M mutations, were chosen for a study in which the degree of XCI skewing and direction were observed and correlated with RTT severity (58). The study reports that there is a statistically significant correlation between unfavorable XCI and measured clinical severity. Though the correlation appears to be statistically significant, it does not appear very robust. One MeCP2 mutation that has been directly implicated to have elevated severity regardless of XCI skew is the R270X mutant. This truncation mutant decreases life expectancy in RTT patients compared to all other recorded mutations (59). Recently Neul et. al have established that individuals with the R168X mutation are more severely affected than those with R294X and other late carboxy-terminal truncating mutations (60) highlighting the importance of domains that are C-terminal to the MBD. These researchers established that different point mutations affect the severity of the three main pathological problems associated with RTT: loss of language, walking, and hand use. The correlation between domain mutation and disease phenotype is depicted in Figure 1.3. The most severe mutation observed was the R168X. Individuals with this truncation mutation lose the ability to walk, use their hands properly, and more frequently lose their entire vocabulary. The truncation mutations in the carboxyl-terminal are less severe. These patients have a higher probability of walking and retaining vocabulary, although they exhibit other significant problems. The R306C point mutation only affects language skills. Taken together, the clinical studies and the distribution of mutations in the RTT database suggest that all six MeCP2 structural domains shown in Figure 1.2 must function together to mediate the normal cellular actions of MeCP2.

The fundamental question of how MeCP2 functions in the nucleus of nearly all vertebrate tissues should be approached in the future from the new structural and functional perspectives discussed above. Specifically, at the molecular level, what is the relationship between intrinsic disorder, MeCP2 domain organization and protein function? How do different mutations in the highly disordered MeCP2 tertiary structure cause different neurodevelopmental RTT symptoms? How does MeCP2 affect chromatin architecture in vivo? Does MeCP2 have to bind to currently unknown proteins that modulate its capacity to regulate transcription? What, exactly, is the role of MeCP2 in pre-mRNA splicing? Does MeCP2 directly bind to RNA molecules? Are there cellular functions of MeCP2 that have yet to be discovered? The new data requires that I step back and ask these and many other related questions.

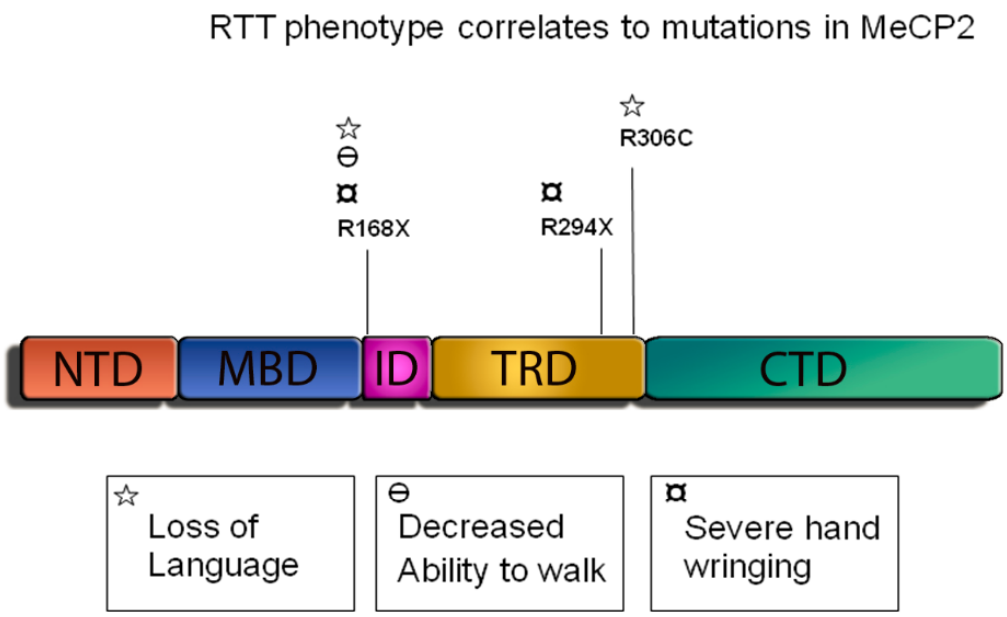

Figure 1.3 Schematic depicting MeCP2 with contingent domains labeled. Two common RTT truncation mutations (R168X and R294X) and a RTT point mutation (R306C) are indicated along the peptide. The phenotypic pathology or combinations are correlated to each mutation with symbols.

CHAPTER 2

Induced α-Helix Formation in Methyl CpG Binding Protein 2 and its Isolated Domains

2.1 Introduction

The ability of full-length Methyl CpG binding protein 2 (MeCP2) and its isolated domains to gain α-helical structure in 2, 2, 2—trifluoroethanol (TFE) has been determined by deconvolution of circular dichroism (CD) data. The isolated MeCP2 domains studied include the N-terminal domain (NTD), the methyl DNA binding domain (MBD), the intervening domain (ID), the transcription repression domain (TRD), and the C-terminal domain (CTD). Full-length MeCP2, which is 68% unstructured and 7% α-helical in 0% TFE, becomes 55% α-helical in 70% TFE while remaining 24% unstructured. The NTD, MBD, TRD, and CTD acquired α-helical structure in 70% TFE, but not the ID. Thus, only those domains that contained predicted α-molecular recognition features (α -MoRFs) were induced to become more α -helical. In quantitative terms, the NTD, MBD, and CTD gained ~25-30% α-helix in 70% TFE while remaining ~35% unstructured. This is in contrast to the TRD, which became almost completely α-helical in 70% TFE. Based on these results we discuss the potential relationships between coil to helix transitions, intrinsic disorder, and MeCP2 structure and function.

MeCP2 is a nuclear protein that has the ability to selectively recognize methylated DNA (2). It is particularly abundant in the nuclei of neuronal cells, where its stoichiometry approaches one MeCP2 per nucleosome (54). MeCP2 is mainly distributed in both promoter and intergenic regions *in vivo* (5). While originally described as a DNA methylation-dependent gene repressor (16), more recently it has been shown that MeCP2 both activates and represses specific gene transcription (4). In the process, MeCP2 interacts with unmethylated and methylated DNA, nucleosomes (61,62) and chromatin (63), RNA and RNA splicing machinery (34,64-66), and thus far nearly 20 different protein partners (16,22,67-81). In addition to defining the structure/function

relationships that apply to MeCP2 in the normal physiological state, there is intense interest in MeCP2 because of its role in the neurodevelopmental disease, Rett Syndrome (RTT). Specifically, many different missense and nonsense mutations located throughout the length of the MeCP2 polypeptide chain are associated with the onset of RTT (23,82-84).

MeCP2 is 486 residues in length and has a mass of 53 kDa. It is monomeric when free in solution (85). MeCP2 is digested into a number of limiting peptides by trypsin (85). Subsequent N-terminal sequencing of these peptides yielded the linear domain organization of the protein. The N-terminal domain (NTD) encompasses residues 1-78, the methyl DNA binding domain (MBD) residues 79-167, the intervening domain (ID) residues 168-205, the transcription repression domain (TRD) residues 206-299, and the C-terminal domain (CTD) residues 300-486 (85). Importantly, the MBD (8), ID (86), and TRD (3) have previously been mapped to essentially the same locations using various functional assays (e.g. DNA binding, transcriptional repression), and nonsense mutations that delete the C-terminal ~200 residues are common in RTT (87). Thus, one obtains the same linear MeCP2 domain organization whether derived from structural or functional sources.

Native MeCP2 is ~60-70% unstructured as shown by circular dichroism (CD) (85,88) and has a frictional coefficient indicative of a random coil as demonstrated by analytical ultracentrifugation (85). Hydrogen/deuterium exchange (H/DX) experiments have revealed that the entire MeCP2 polypeptide chain exhibits very rapid H/DX except for the MBD, and even the MBD shows faster H/DX than a typical globular protein (13). Moreover, while DNA binding locally stabilized the MBD fold, the remainder of the protein remained largely unstructured (13). The biophysical studies indicate that MeCP2

is an intrinsically disordered protein when free in solution and when bound to unmethylated and methylated DNA. Intrinsically disordered proteins lack secondary structure elements throughout much or all of their polypeptide chain, yet are functional (89-95). Consistent with the biophysical data, both the FoldIndex and PONDR algorithms predict that MeCP2 has long stretches of disordered sequence interspersed with short regions of secondary structure (85,96). MeCP2 is also predicted to have nine α-molecular recognition features (α-MoRFs), which are segments of an intrinsically disordered protein that become α-helical upon binding to other macromolecules (97,98). The predicted α-MoRFs are distributed in each MeCP2 domain except the ID (63).

In the present studies our working hypothesis was that full-length MeCP2 and those isolated domains that contain α-MoRFs would gain α-helix content in the presence 2-2-2-trifluoroethanol (TFE), which stabilizes intramolecular hydrogen bonds in proteins (99-102). TFE has been shown to induce α-helical structure in intrinsically disordered proteins such as the linker histones (103,104), measles virus nucleoprotein (105) and the carboxyl-terminal domain of heat-shock factor 1 (106). To test our hypothesis, we have used CD to determine the fraction of MeCP2 and its domains that are α-helical, β-strand/turn, and unstructured as a function of TFE concentration. Our results indicate that full-length MeCP2 and all isolated domains except the ID become significantly more α-helical in TFE. Thus, there was a direct correlation between the presence of predicted α-MoRFs and TFE-induced acquisition of α-helical structure. The TRD became almost completely α-helical in 70% TFE, while the NTD, MBD, ID, and CTD remained partially unstructured under these conditions. Based on these results we discuss ways in which coil-to-helix transitions and intrinsic disorder may contribute to MeCP2 structure and function.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 Protein expression and purification. Histone variant H5 was purified from native chromatin derived from mature chicken erythrocytes as described previously (107). Full-length MeCP2, NTD, MBD, ID, and TRD domains were expressed and purified using the Intein Mediated Purification with an Affinity Chitin-binding Tag (IMPACT) system (New England Biolabs). The CTD was expressed using a modified pET28a vector plasmid (Novagen). The MBD and TRD constructs contained an added sequence, EFLEGSSC, on their C-terminal ends as a result of previously described cloning methods (85,96). The ID was cloned without this vestigial sequence and corresponded to the DNA sequence that codes for amino acids 168-205 from wild-type MeCP2 using the following primers (ID5') 5'-GGA GCC CCC ATA TGC GAG AGC AGA AAC CAC CTA AGA AGC C-3'; and (ID3') 5'-CAT AGG CTC TTC GGC ACT CTG ACG TGG CCG CCT TGG GTC TC-3'. Primers were designed to amplify DNA fragments cleavable by restriction enzymes Nde1 on the 5' side and Sap1 on the 3' side. The NTD construct was amplified to express wild type MeCP2 amino acids 1-78 using the following primers; (NTD5') 5'-GAC ATA TGG TAG CTG GGA TGT TAG GGC TCA GGG AAG-3'; and (NTD3') 5'- CAG AAT TCA GAA GCT TCC GGC ACA GCC GGG GC-3'. The NTD insert was amplified from the wild type MeCP2 template such that it was cleaved by NdeI and EcoR1, gel-purified, and ligated into correspondingly digested and purified pTYB1 vector plasmid. The NTD also expresses with an additional non-native eight amino acids, EFLEGSSC as a result of the NdeI and EcoR1 cloning strategy.

E. coli BL21RP+ was used as host bacteria for expression by pTYB1 plasmid. A transformed bacterial colony was selected and grown in lysogeny broth to an optical density of 0.5 absorbance units (590 nm) at 37ºC. Translation was induced with 0.4 mM

isopropyl-β-D-thiogalactopyranoside (IPTG) at 18℃ for 3 hours before harvest. Bacteria were centrifuged in an Avanti J-26 XPI preparative centrifuge (Beckman Coulter) in a JLA 8.100 rotor at 5,000 g for 10 minutes. Pellets were resuspended in wash buffer (25 mM Tris, pH 7.5/100 mM NaCl) and centrifuged again under the previous conditions. Bacterial pellets were resuspended in column buffer (25 mM Tris–HCl pH 8.0, 500 mM NaCl) with 0.1% Triton X-100, 0.2 mM PMSF, and Protease Inhibitor Cocktail Set II (Calbiochem) added. Re-suspended bacteria were subjected to two rounds of sonication, 90s each with a Branson Sonifier 450. A large tip set to 50% duty cycle with a power output of 7 was employed. Sonicated lysate was poured into Oakridge tubes and centrifuged at 21,000 *g* for 25 min in a JA-17 rotor (Beckman). Clear supernatant was mixed with chitin beads (New England Biolabs) previously equilibrated in column buffer and the mixture was incubated at 4ºC overnight. Errant contaminating bacterial DNA-MeCP2complexes were washed from the samples from the chitin beads with five column volumes column buffer, followed by an equal volume column buffer at 900 mM NaCl final concentration. Chitin beads were re-equilibrated with an additional 5 column volumes of 500 mM NaCl column buffer. Column buffer with 50 mM DTT was passed over the column such that 1 cm buffer remained between the meniscus and top of the column bed in a 10 cm Kontes FlexColumn (Fischer). The column was left for 48-72 hours which allowed intein mediated cleavage. Column buffer was used to elute protein from the chitin column. The eluant was diluted with column buffer with no salt from 500 mM to 100 mM NaCl and loaded onto a HiTrap Heparin HP column (GE healthcare). Proteins were eluted via step gradient from 100 mM NaCl to 1M NaCl buffer from the heparin column. A gradient using 100mM NaCl steps in 25 mm Tris, pH 7.5, 10% glycerol background buffer was employed. Peak fractions were loaded on a 12% SDS polyacrylamide gel to analyze for purity and/or degradation, then pooled and dialyzed into 10 mM Tris pH 7.5.(108)

The CTD was expressed using a modified pET28a vector plasmid (Novagen) with an N-terminal 6-histidine tag cleavable by PreScission protease (GE Healthcare). Modified vector was kindly supplied by Dr. Wayne Lilyestrom. The CTD fragment corresponds to residues 310-486. The CTD construct was amplified with an Nde1 site on the 5' end and a BamH1 site on the 3' end with a stop codon included directly 3' to the BamH1 site. The forward primers used to amplify these constructs were (CTD5') 5'-GAC ATC CAT ATG GAG ACG GTC AGC ATC GAG G-3'; and (TRD-CTD5') 5'-GAC ATC CAT ATG GAG GGT GTG CAG GTG AAA AGG G-3'. The reverse primer sequence is (MeCP2 486 3') 5'-CTG GGA TCC CTA GCT AAC TCT CTC GGT CAC G-3'. Constructs were expressed and purified using Ni-NTA agarose beads (Qiagen) as previously described (109).

2.2.2 Circular dichroism.

Stock solutions of the respective proteins were prepared at ~500 µg/ml concentration and stored in glass test tubes. Concentrations were determined by bicinchoninic acid assay (Pierce BCA™ protein assay kit, Thermo Scientific) conducted on a 96-well microplate and measured in a Bio-rad Model 680 microplate reader at 560 nm. Peptide concentration accuracy was verified by total amino acid analysis (Biophysics Core, Department of Biomolecular Structure, University of Colorado Denver Anschutz Medical Campus). Dilutions were prepared in glass test tubes to a final volume of 220 µL such that the final protein concentration of 0.12 mg/ml in 1 mM Na phosphate, 0.2 mM Na_3EDTA with either 0, 20, 50, or 70% volumetric concentrations of 99.8% trifluoroethanol (Acros organics, Fisher). The final pH was adjusted to 7.4 when necessary. All buffers were purified though a 0.2 µM filter (Metricel® Pall Corporation)

and proteins were dialyzed extensively into their respective buffers using 3,500 MWCO dialysis tubing (Spectrum labs).

CD spectra were recorded from 260-190 nm on a Jasco-720 spectropolarimeter. Samples were transferred to a cuvette with a 1 mm path length and cooled to 20ºC. Spectra were collected at a bandwidth of 1 nm with a scanning rate of 10 nm/s in continuous scanning mode and a response time of 16 s. From raw data, collected spectra were buffer-subtracted and converted from millidegrees to molar ellipticity using the following equation.

$$[\theta] = {\theta obs \times M}/{10 \times l \times C}$$

Where [θ] is the mean residue ellipticity (degrees cm^2 $dmol^{-1}$ $residue^{-1}$), θ_{obs} is the ellipticity measured in millidegrees, *M* is the protein mean residue molecular weight, *l* is the optical path length of the cuvette in cm, and *C* is the concentration of the protein in mg/ml. Ellipticity data were analyzed using Spectra Manager software version 1.53.01 (Jasco Corporation) and saved as text files. Data files were then deconvoluted against the SDP48 basis set using CDpro software including CONTINLL, SELCON3 and CDSTRR methods (110). Estimates of percent secondary structure were averaged among the three methods used and standard deviations calculated.

2.3 RESULTS

2.3.1 Histone H5 and full-length MeCP2 gain α-helical content in TFE.

As a control we first examined the effect of TFE on histone H5, which has been previously studied by CD (103). Representative CD curves for H5 are shown in Figure 2.1A. Visual inspection of the curves showed a trend toward α-helix formation with increasing TFE, as indicated by the disappearance of the trough at ~200-205 nm and an increase in the peaks at ~195 and 220 nm. To quantitate these results, the CD curves were analyzed with CDPro (Experimental Procedures; (110)) and the % calculated secondary structure plotted against TFE concentration (Fig. 2.1B). Results indicate that there was a dose-dependent interconversion of unstructured content to α-helical content with increasing TFE concentration, while the amount of β-strand/turn stayed essentially constant. Specifically, in 0% TFE, H5 was 68% unstructured, 9% α-helix, and 23% β-strand/turn, while in 70% TFE H5 was 31% unstructured, 38% α-helix, and 31% β-strand/turn. The data obtained with H5 reproduce the results of earlier studies (103), validating the experimental results obtained below with MeCP2 and its domains.

The CD curves for full-length MeCP2 in TFE are shown in Figure 2.2A and a plot of % secondary structure against TFE concentration is shown in Figure 2.2B. As with H5, MeCP2 showed a dose-dependent increase in α-helix with increasing TFE concentration. In 0% TFE MeCP2 was 68% unstructured and 7% α-helix. This changed to 24% unstructured and 55% α-helix in 70% TFE. The β-stand/turn component was ~25% in all TFE concentrations. Thus, for full-length MeCP2, ~40% of the unstructured residues were converted to α-helix in 70% TFE while the amount of β-strand/turn remained essentially constant.

2.3.2 Induction of α-helix in isolated MeCP2 domains.

To determine how the gain in secondary structure is distributed along the length of the MeCP2 sequence, we next characterized the effects of TFE on isolated MeCP2 domains. A schematic illustration of the MeCP2 domain organization is shown in Chapter 1 Figure 1.2 and 1.3. A plot of % secondary structure against TFE concentration for the NTD is shown in Figure 2.3. The α-helix content was 5% in 0% TFE and 34% in 70% TFE. Under these same conditions the unstructured content decreased from ~65% to ~34%. The β-strand/turn content stayed constant at ~30%. Thus, for the NTD, ~25% of the initial unstructured residues were converted to α-helix in 70% TFE

Our analysis of the MBD is shown in Figure 2.4. In 0% TFE, the MBD was ~32% unstructured, ~10% α-helix, and ~54% β-strand/turn. By 70% TFE, the values were ~29% unstructured content, ~35% α-helix, and ~37% β-strand/turn. The data in Figure 2.5 indicate that for the MBD some of the β-strand/turn was converted to α-helix, while the % unstructured residues stayed constant.

Results obtained for the ID are shown in Figure 2.5. The data indicate that there was no significant change in secondary structure content in TFE for this domain; the % unstructured, α-helical and β-strand/turn content remained constant at ~32%, ~12%, and ~55%, respectively. The lack of α-helix formation is notable given that the ID is the only domain that is not predicted to contain an α-MoRF (see Discussion).

Characterization of the TRD is shown in Figure 2.6. This domain showed fundamentally different behavior compared to the others. In 0% TFE the unstructured content was 81%, α-helix was 7%, and β-strand/turn was 12%. This changed to 5% unstructured content, 89% α-helix and 7% β-strand/turn in 70% TFE. Thus, despite the

initial high degree of disorder, the TRD was induced to almost completely form α-helix in 70% TFE.

Lastly, we analyzed the CTD. Plots of % secondary structure against TFE concentration for the intact CTD peptide are presented in Figure 2.7. In 0% TFE, the unstructured content was 70%, α-helix ~5% and β-strand/turn 25%. This changed to 36% unstructured content, 32% α-helix and 34% β-strand/turn in 70% TFE. Thus, for the CTD, ~30% of the initial unstructured residues were converted to α-helix in 70% TFE.

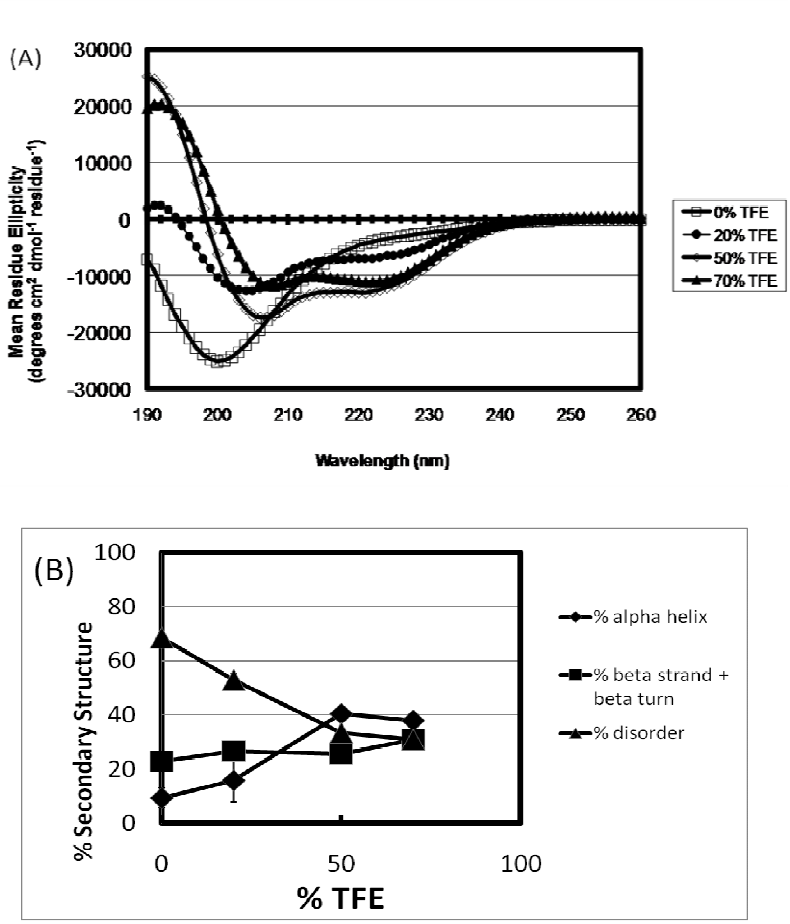

Figure 2.1. Histone H5 (*Gallus gallus*) gains α-helical content in a dose-dependent manner with increased TFE concentration according to circular dichroism. (a) Circular dichroism spectra of histone H5 measured in 0% (□), 20% (●), 50% (◊), and 70% (▲) TFE revealed conversion to alpha helical secondary structure by disappearance of the trough between 200-205 nm and heightened peaks at 195 and 220 nm. (b) Percent secondary structure is plotted against % TFE after deconvolution of CD data from panel (a) using CDpro software (see experimental procedures). The % disordered (▲) decreased from 68% to 31% in 0% and 70% TFE respectively. The % α-helix (♦) increased from 9% to 38% in 0% and 70 TFE respectively, while % β strand/turn (■) remained relatively constant at ~30%.

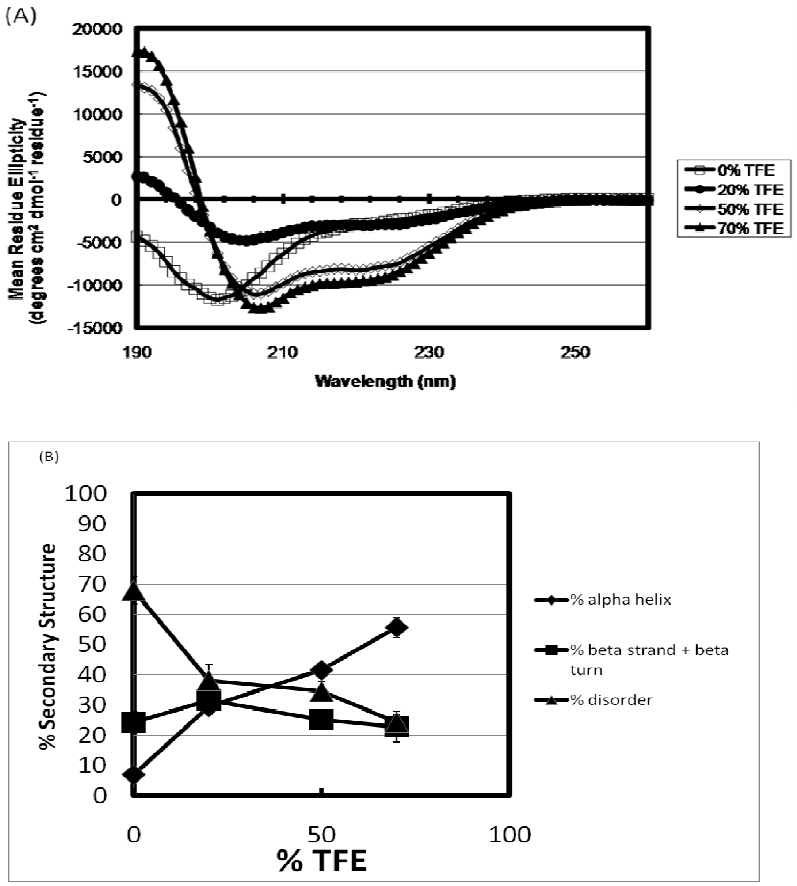

Figure 2.2. Full-length wild-type MeCP2 isoform e2 (*Homo sapiens*) gains α-helical content in a dose-dependent manner with increased TFE concentration according to circular dichroism. (a) Circular dichroism spectra of MeCP2 measured in 0% (□), 20% (●), 50% (◊), and 70% (▲) TFE reveal conversion to alpha helical secondary structure by disappearance of the trough between 200-205 nm and heightened peaks at 195 and 220 nm. (b) Percent secondary structure is plotted against % TFE after deconvolution of CD data from panel (a) using CDpro software (see experimental procedures). Similarly to H5, wild type MeCP2 has decreased % disordered (▲) from 68% to 24% in 0% and 70% TFE respectively. The % α-helix (♦) increased from 7% to 55% in 0% and 70 TFE respectively, while % β strand/turn (■) remained relatively constant at ~25%.

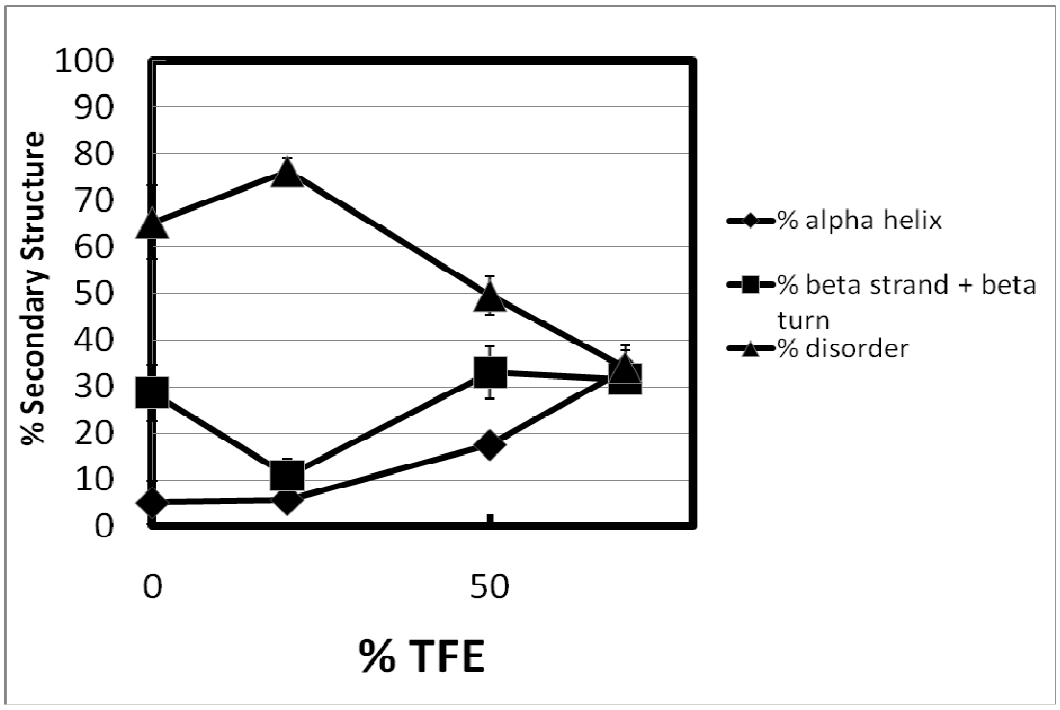

Figure 2.3. The N-terminal domain (NTD) of MeCP2 gains alpha-helical structure while losing disorder in increasing TFE concentration. Percent secondary structure is plotted against % TFE after deconvolution of CD data from using CDpro software (see experimental procedures). The % disordered (▲) decreased from 65% to 34% in 0% and 70% TFE respectively, while % α-helix (♦) increased from 5% to 34% , and % β strand/turn (■) remained constant at ~30%.

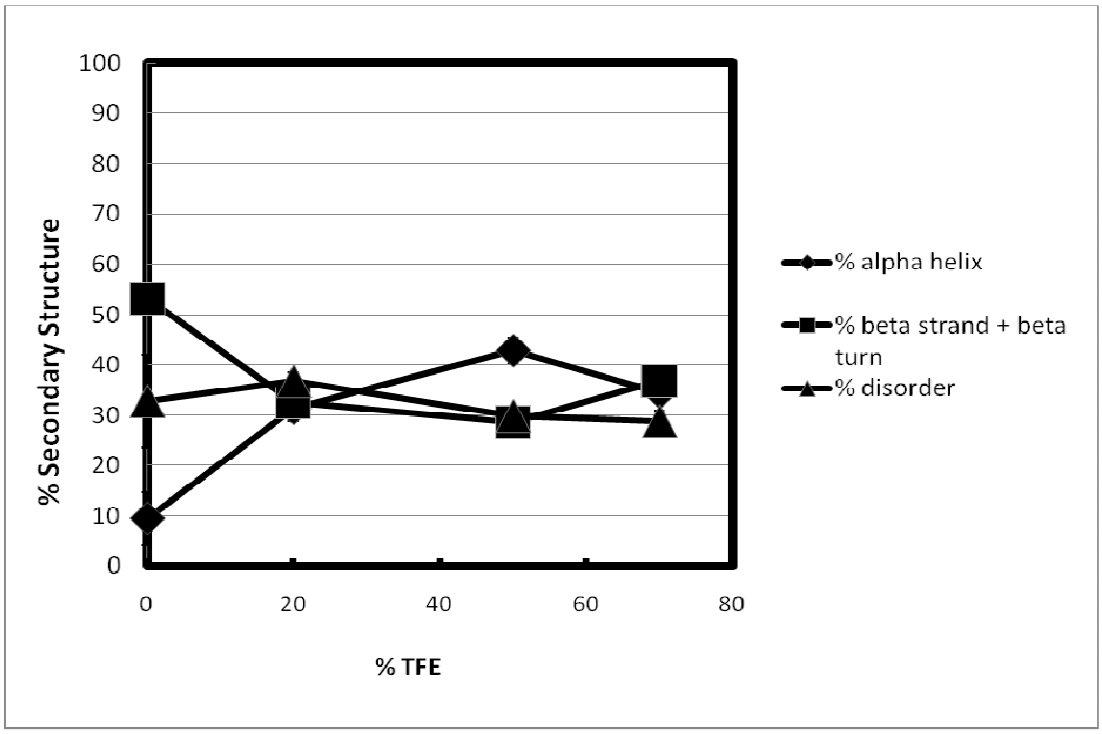

Figure 2.4. The Methyl CpG Binding Domain (MBD) of MeCP2 gains alpha-helical structure at the expense of beta strand/turn content in increasing TFE concentration while the amount of disorder remains constant according to circular dichroism. Percent secondary structure is plotted against % TFE after deconvolution of CD data using CDpro software (see experimental procedures). The % disordered (▲) remained constant at ~30% over the course of TFE titration. The % α-helix (♦) increased from 5% to 34%, and % β strand/turn (■) decreased from 54% to 37% in 0% and 70% TFE respectively.

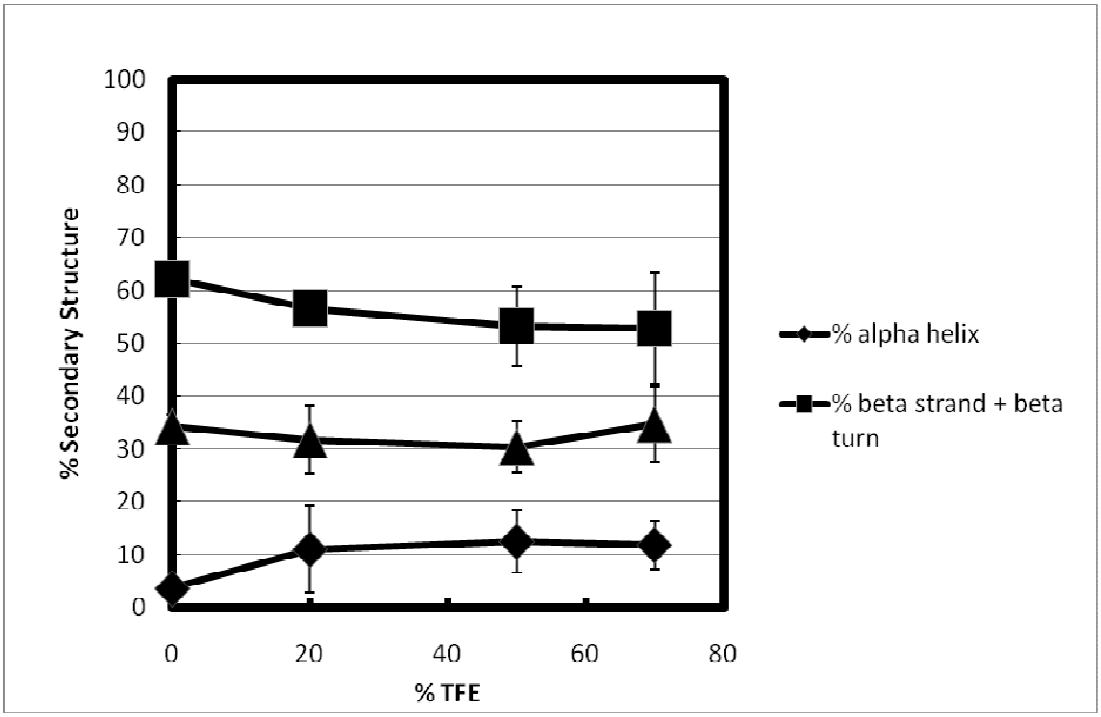

Figure 2.5. The Intervening Domain (ID) of MeCP2 does not undergo detectable change in secondary structure content during TFE titration according to circular dichroism. Percent secondary structure is plotted against % TFE after deconvolution of CD data using CDpro software (see experimental procedures). The % disordered (▲), % α-helix (♦), and % β strand/turn (■) all remained constant at ~32%, ~12% and ~55% respectively.

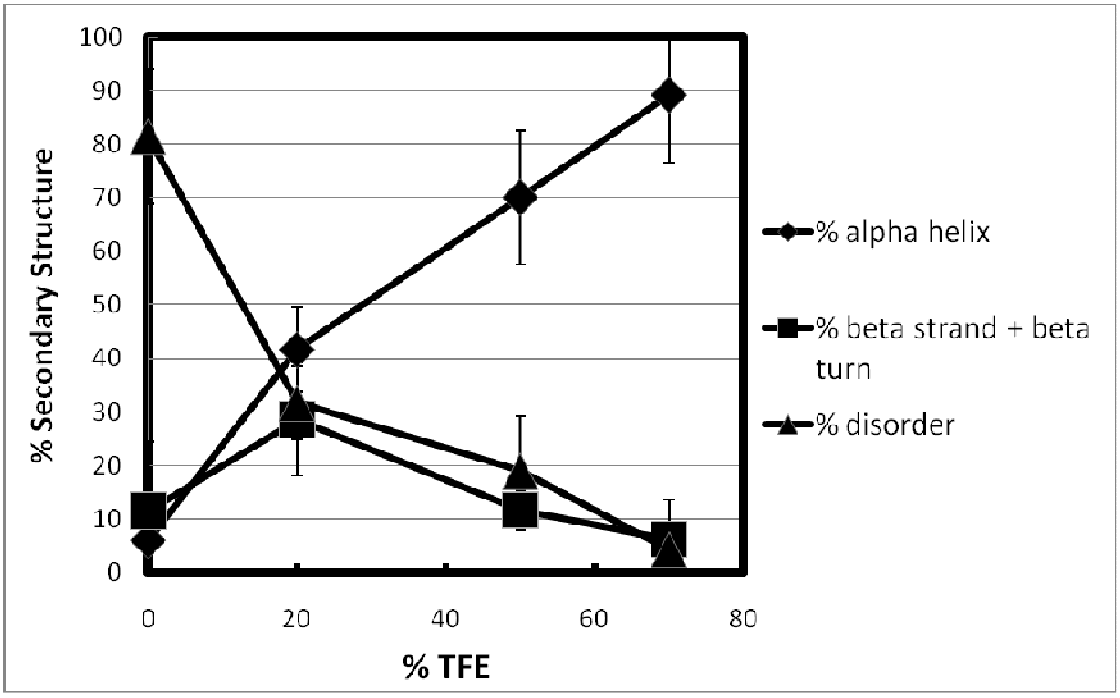

Figure 2.6. The Transcription Repression Domain (TRD) of MeCP2 almost entirely converts to α-helix in 70% TFE. Percent secondary structure is plotted against % TFE after deconvolution of CD data from using CDpro software (see experimental procedures). The % disordered (▲) decreased from 81% to 5% in 0% and 70% TFE respectively. Most of the diminished disorder transformed to alpha helix as % α-helix (♦) increased from 7% to 89% , and % β strand/turn (■) decreased slightly from 12% to 7% in 0% and 70% TFE respectively.

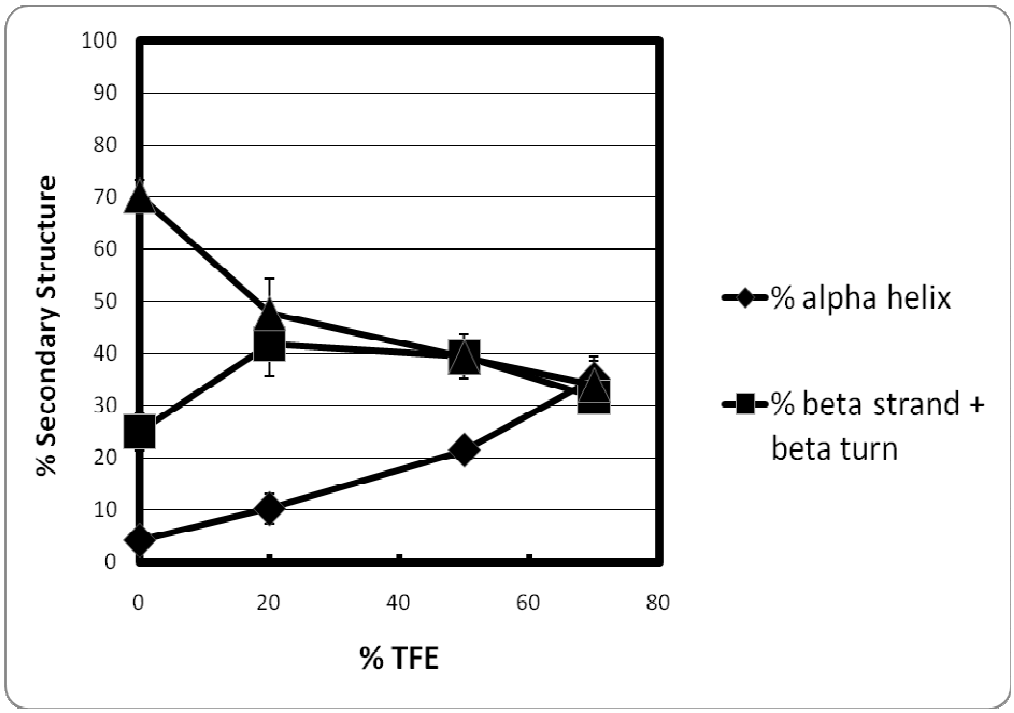

Figure 2.7. The C-terminal Domain (CTD) of MeCP2 went from a mostly disordered state to a structure with equal parts disorder, α-helix, and β-strand/turn with increased TFE concentration. Percent secondary structure is plotted against % TFE after deconvolution of CD data from using CDpro software (see experimental procedures). The % disordered (▲) decreased from 70% to 36% in 0% and 70% TFE respectively. The % α-helix (♦) increased from 5% to 32%, and % β strand/turn (■) increased slightly from 25% to 34% in 0% and 70% TFE respectively.

2.4 DISCUSSION

In the present study we have examined how the hydrogen-bond stabilizing solvent, TFE (99,100,111-113), influences the α-helical content of MeCP2 and its isolated domains. CD experiments in the presence of TFE have proven useful in studying coil-to-helix transitions in other intrinsically disordered proteins (103-106). Our working hypothesis was that TFE would mimic the effects of molecular recognition and induce α-helix formation in those regions of MeCP2 that contain α-MoRFs (63). α-MoRFs are short (i.e., ~10-70 residue) stretches of an intrinsically disordered polypeptide chain that have the ability to gain α-helical structure concomitant with macromolecular interactions (97,98,114,115). MeCP2 is predicted to have one or more α-MoRFs in the NTD, MBD, TRD and CTD but none in the ID (63). Our results indicate that full-length MeCP2 and all of its domains except the ID acquire α-helical structure in a dose-dependent manner in TFE. Thus, there was a direct correlation between α-helix induction by TFE and the presence of predicted α-MoRF(s) in the domain. This suggests that many different regions of the MeCP2 polypeptide chain form α-helices and act as binding interfaces under the right conditions. Toward this end, the NTD interacts with HP1 (78), the MBD with unmethylated and methylated DNA (36,61,85,96,116) and the ATRX protein (74), the TRD with many different co-repressors (16,22,70,75,76,79) and unmethylated DNA (85,96), and the CTD with unmethylated DNA (96), chromatin (96), and RNA-splicing machinery (34,66). Moreover, some proteins interact with multiple MeCP2 domains, e.g., PU.1 binds to the NTD and TRD (80), CDKL5 binds to the TRD and CTD (70), and DNMT1 binds to the MBD, ID and TRD (79).

Based on the CD (present studies) and H/DX (13) results, we envision the following paradigm for the structural basis of MeCP2 multifunctionality. When free in solution MeCP2 has regions of quasi-stable secondary structure located throughout the length of its polypeptide chain. Consequently, the protein sequence rapidly samples multiple secondary structures and the full-length protein is highly conformationally malleable. This intrinsic flexibility, together with strategically located coil-to-helix transitions allows the formation of many different macromolecular complexes. Finally, many nuclear regulatory proteins are components of large macromolecular assemblies that themselves are very malleable (117). Thus, its marked conformational flexibility may be required for the proper assembly and function of higher order MeCP2-containing macromolecular complexes.

CHAPTER 3

Non-specific unmethylated DNA binding activity in the intervening and carboxy-terminal domains of MeCP2 has been narrowed to residues 168-188 and 300-354 respectively.

3.1 INTRODUCTION

MeCP2 has historically been known and, in fact, named for its ability to bind methylated DNA, but recent *in vivo* experiments reveal it to be bound to both actively transcribed and repressed regions of the genome (see Chapter 1). The significance of MeCP2 binding non-methylated DNA is further emphasized by the observation that it binds ubiquitously throughout the entire genome with a stoichiometry of one MeCP2 molecule per every two nucleosomes in neuronal tissue (54). These observations have focused experimental efforts in this chapter on examining which regions of the MeCP2 molecule are responsible for binding non-methylated DNA and chromatin *in vitro*. My interest in MeCP2 lies in its ability to bind DNA and to condense nucleosomal arrays into locally folded secondary chromatin structures and oligomeric suprastructures (24).

MeCP2 consists of six protease-resistant domains organized into an intrinsically disordered, random-coil-like tertiary structure (85). These domains in linear order are the N-terminal domain (NTD), methyl DNA binding domain (MBD), intervening domain (ID), transcriptional repression domain (TRD), C-terminal domain-alpha (CTD-alpha) and C-terminal domain-beta (CTD-beta), depicted in Chapter 1, Figure 1.2. The full length protein is 486 amino acid residues in length, has an unusually high overall average pI (~11.6), and exists as a monomer in physiological conditions.

The observation that MeCP2 does not self associate, the overall low incidence of large hydrophobic amino acid residues, and enrichment of lysine residues leading to a higher than average isoelecric point provide further evidence that MeCP2 is a model intrinsically disordered protein (IDP). These descriptors align with those outlined by Dr. Ulrich Hartl (44).

> “Disordered proteins should also be protected from aggregation because, unlike globular proteins, they contain few hydrophobic amino acids, which tend to stick together — and are instead rich in 'polar' amino acids that are happy swimming in water.”

Like other model IDPs such as p53 and Sic1, MeCP2 also has numerous binding partners. Reviewing the literature reveals in excess of 20 direct binding partners. Though disruption of binding to these partners undoubtedly plays a role in pathological phenotypes of Rett syndrome, the observation that four out of the six protease-resistant domains bind unmethylated DNA focused our investigation into the role of MeCP2 in chromatin architecture (8,36,61,96,116,118-120).

The most clearly understood DNA-binding motif of MeCP2 is the MBD (11,12,121). When expressed as an isolated peptide (amino acids 77-167) MBD, compared to all other MeCP2 peptides, is least susceptible to degradation after purification (personal experience). The crystal structure of the MBD bound to a methylated 20mer double-stranded DNA molecule from the BDNF promoter reveals a three-stranded beta sheet along with a single 11 amino acid α-helix forming a wedge tertiary structure that fits non-specifically in the major groove of DNA. This wedge makes contact by proxy to the methyl-group on a methylated CpG via a coordinated water molecule. The portion of the MBD consisting of recognizable secondary structure is 34 amino acids in length accounting for 7% of the entire MeCP2 molecule. Little structural information is known of the remaining 93% of the protein. Clinical reports of RTT-causing MeCP2 mutations show two widespread types of Rett mutations: missense mutations in the MBD, and nonsense mutations which result in truncation of the carboxyl-terminal portion of the protein. These are however not the only mutations

causing Rett syndrome. Mutations along the entire length of MeCP2 have been reported (see frequency of amino acid mutation at the Rett syndrome database - http://bit.ly/fep0kk).

Experimental observations presented in Chapter 2 revealed that MeCP2 has TFE-inducible alpha-helix distributed along the entire length of the protein excluding the ID. The coincidence of DNA-binding activity and regions of inducible alpha helix by TFE titration in previous studies of the C-termini of histones H1 and H5 (103) led me to ask similar questions of MeCP2 domains. Does the presence of inducible α-helices in MeCP2 domains coincide with unmethylated DNA-binding activity? I went on to use the same domains used in Chapter 2 and several additional fragments to search for unreported DNA-binding domains and narrow the regions of MeCP2 that bind unmethylated DNA. Though MeCP2 exists as 60% disordered in the absence of a binding partner or secondary structure-inducing solvents, it retains the ability to bind an array of partners, primarily unmethylated DNA.

To better understand the molecular basis for MeCP2 interaction with unmethylated DNA, I characterized the DNA- and chromatin-binding properties of eleven MeCP2 domains or domain pairs when expressed and purified as individual polypeptides. Electrophoretic mobility shift assays (EMSAs) were used to assay DNA and chromatin binding and chromatin compaction. Results indicate that that the MBD, ID, TRD, and CTD-alpha all bind to unmethylated DNA. Mutants in the full length protein confirmed binding sites identified in the ID and CTD-alpha ((96) and unpublished data from Dr. Mary Porter-Goff respectively). Quantification of the binding affinity yielded nanomolar dissociation constants for all unmethylated-DNA-binding regions (96). In regions of MeCP2 where new DNA-binding domains were observed, point mutations were made to further narrow the DNA binding region. The major contribution to this work came by in-depth investigation of the intervening domain (ID) and the strong

methylation-independent DNA-binding activity of this domain. Point mutations in this domain indicate that the DNA binding is facilitated by residues that are N-terminal to residue 188 (aa numbered in context of full-length MeCP2).

Interestingly, parallel results agree that the ability of MeCP2 to compact nucleosomal arrays into secondary chromatin structures can be reproduced by the TRD-CTD fragment of the protein. There is consensus in the published results (see Chapter 4) that the MBD acts as an autonomous domain that can recruit MeCP2 to methylated DNA (96). However, extensive engagement of a large internal region of the protein with unmethylated DNA, followed by binding to either other nonspecific DNA or histones by the CTD is necessary for full-length MeCP2 to condense nucleosomal arrays into novel locally folded secondary chromatin structures.

3.2 Materials and methods:

3.2.1 MeCP2 Domain cloning and expression:

Domains were expressed and purified using either the Intein-Mediated Purification with an Affinity Chitin-binding Tag (IMPACT) system or the Histidine-tagging modified pET28a vector system. Both methods are described in detail in Chapter 2 of this dissertation. Constructs used in the present study are listed in table 3.1 with vector and corresponding amino acids indicated.

Table 3.1 MeCP2 constructs used in electrophoretic mobility shift assays are listed. Corresponding amino acid numbers in relation to the full length MeCP2 protein are listed in middle column. The vector plasmid housing each construct is listed in the right column. [1]-Unable to express this construct.

Domain – as defined by the Hansen Lab (85).	**MeCP2 residues**	**Vector Plasmid**
MeCP2 full length wt	1-486	ptyb1– (chitin binding protein)
NTD	1-78	ptyb1 – (chitin binding protein)
MBD	78-168	ptyb1 – (chitin binding protein)
ID (intervening domain)	168-206	ptyb1 – (chitin binding protein)
TRD	198-305	ptyb1 – (chitin binding protein)
CTD (α+β)	300-486	ptyb1 – (chitin binding protein)
R188E full length MeCP2	1-486	ptyb1 – (chitin binding protein)
R188E - ID	168-206	ptyb1 – (chitin binding protein)
GRP189-191AAA - ID	168-206	ptyb1 – (chitin binding protein)
CTDβ (Adams)	354-486	petP – His-tag
[1]CTDα	310-354	petP – His-tag
Acidic domain	400-450	petP – His-tag
TRD-CTD	198-486	petP – His-tag
CTDβ (Ghosh)	335-486	ptyb1 – (chitin binding protein)

An alternate expression system was implemented for four MeCP2 constructs. This modified pet 28a plasmid system includes a PreScission protease site added between the histidine tag and the coding region. The vector, known as petP plasmid, was modified and provided to us by Dr. Wayne Lilyestrom. The modified pET28a vector is about 3,000 base pairs shorter than the pTYB1 vector and is less cumbersome for making point mutations.

Fragments spanning residues 310-354, 355-486, 400-450 were synthesized, codon-optimized, and sub-cloned by geneart (www.geneart.com). Fragments spanning residues 300-486 and 206-486 were sub-cloned by traditional cloning methods described in Chapter 2 of this dissertation.

Point mutants were prepared in the context of the full-length MeCP2 protein and the ID. Site-directed mutagenesis was done as described before (122). Single-stranded DNA primers were ordered from Integrated DNA technology (IDT). For full-length and ID R188E point mutations the following primers were employed, forward 5' – GGC AGA GGC GAG GGA CGC CCC AAA G – 3', and reverse 5' – CCC TTT GGG GCG TCC CTC GCC TCT GC – 3'. To generate the GRP189-191AAA triple alanine replacement mutant in the context of ID the following primers were used, forward 5' – GGC AGA GGC CGG GCA GCC GCC AAA G – 3', reverse 5' – TGC CGC TCC CTT TGG CGG CTG CCC GG – 3'. For the reactions described in the Carrigan protocol, Pfu Turbo® was used instead of PfuUltra® high-fidelity polymerase. Both enzymes are available through Agilent Technology Genomics.

3.2.2 Reconstitution of 208-12 Nucleosomal arrays:

Chromatin templates were reconstituted by combining histone octamers purified from *Gallus gallus* erythrocytes and a DNA template consisting of twelve tandem repeats of a 208-base pair sequence derived from the sea urchin *Lytechinus variegatus* 5 S rRNA. Arrays were reconstituted according to the salt dialysis method of Hansen and Lohr (123). Experiments were repeated using recombinant wild-type *Xenopus laevis* octamers instead of the native *Gallus gallus* octamers. *Xenopus* histones were obtained from Teri McLain of the W. M. Keck Protein expression and purification facility for histones**.**

3.2.3 Electrophoretic Mobility Shift Assay (EMSA):

Nucleoprotein complexes were analyzed by EMSA in native 0.8% agarose gel run in Tris, acetic acid, EDTA (TAE) buffer at 5 V/cm for 4-5 hours. Complexes were formed by mixing nucleosome arrays with indicated molar ratios of various purified MeCP2 constructs and allowed to incubate at room temperature for 30 minutes before loading on the gel. A molar ratio of 1 corresponded to one protein molecule per one repeat of 208 base pairs of DNA or per nucleosome repeat for arrays. This translates to 12 molecules of the respective protein per array or naked 208-12 DNA molecule. In all cases 200 ng of DNA was used if the substrate was naked DNA, while 400 ng of arrays were used. Twice the amount of array was used because nucleosomes hinder the binding of ethidium bromide and impair visualization. Thus, more material is needed to see the nucleoprotein complexes as bands when an array is used as the substrate. Glycerol at 10% by volume concentration was used as a loading buffer to keep the complexes in a native state.

3.3 Results and Discussion:

3.3.1 DNA binding in the C-terminus of MeCP2 lies between residues 300-354:

Specific focus has been placed on the C-terminus of MeCP2 as there is a high frequency of clinically observed C-terminal deletion mutations in the MeCP2 gene in Rett patients (87). After our collaborators initially observed amino acids 335-486 (which they termed the CTDβ) bound to nucleosomal arrays but not to methylated DNA, I wanted to test the CTDα+β domain fusion (simply termed CTD) for this unique chromatin binding ability. I also wanted to compare this binding with the binding of the TRD-CTD domain fusion. The TRD is known to have DNA binding ability as our collaborators had shown this domain fusion to have ability to compact chromatin templates nearly as efficiently as wt MeCP2 (85). I also hypothesized that the short acidic portion of the C-terminal of MeCP2 spanning residues 400-450 may be responsible for histone binding based on its anomalously low isoelectric point (4.86) relative to the overall basic isolelectric point of full-length MeCP2 (9.5).

The 0.8% agarose gel imaged in the Figure 3.1 contains three different DNA-based templates; 208-12 DNA alone in the top left section of the image, subsaturated 208-12 arrays loaded with ~11 recombinant tailless *Xenopus* octamers in the bottom left section, and subsaturated 208-12 arrays loaded with ~11 recombinant Xenopus octamers (with intact tails) in the top and bottom sections on the right third of the image (separated by the DNA ladders). Each of the three DNA-based templates listed in the last sentence were incubated with either 2,4, or 6 molecules of wt MeCP2, CTD, TRD-CTD, and 400-450 and run on the same agarose gel. Experiments were conducted on one large agarose gel rather than split up on smaller gels so that direct comparisons could be made between electrophoretic mobility. This technique ensured that run-time, buffer concentration, and agarose percentage were uniform. Each band in this results section is a trans-UV illuminated DNA-based template stained with ethidium bromide

after binding the listed fragments of MeCP2 for 30 minutes at room temperature in a final volume of either 20 or 40 µl for DNA alone and chromatin templates respectively. The concentration of DNA template in all experiments was held constant at 8 nM. Mixtures were incubated in a sealed Eppendorf tube to maintain equal molar concentration of reagents in all cases.

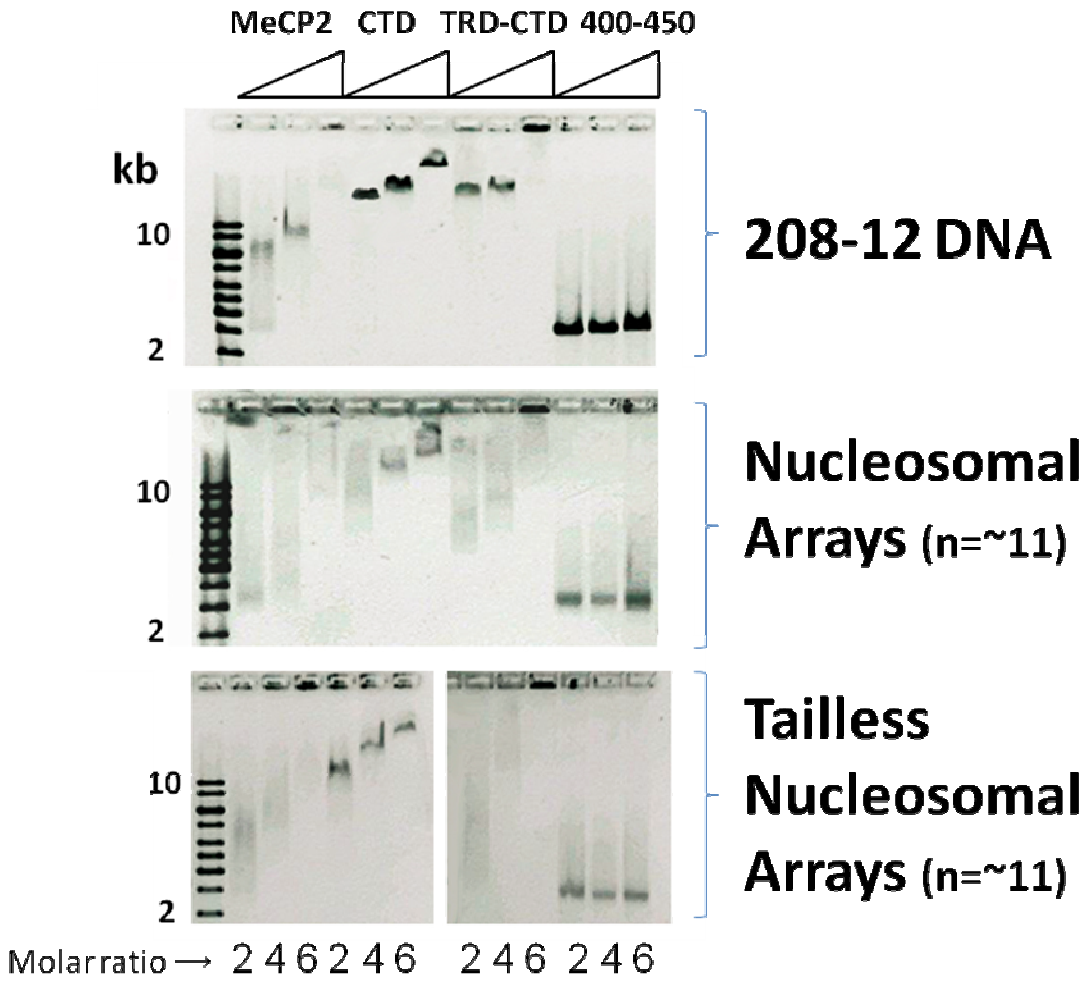

Figure 3.1 CTD shifts DNA alone, nucleosomal arrays, and tailless arrays while the acidic region spanning residues 400-450 does not. The paneled pictures above were captured from the same 0.8% agarose gel. They show the relative electrophoretic mobility of 208-12, DNA alone, nucleosomal arrays with and without N-terminal tails. Each MeCP2 construct was incubated with 200 ng of array or DNA alone at 2,4, and 6 molecules MeCP2 fragment per 208 b.p repeat at room temp for 30 minutes.

3.3.2 There is no observable histone-tail dependence for wt MeCP2, CTD, TRD-CTD or 400-450 MeCP2 peptides by electrophoretic mobility shift assay.

According to the results in Figure 3.1 the absence of the histone N-terminal tails does not change the shifting capacity of wt MeCP2, CTD, TRD-CTD, or the 400-450 peptide. The relative mobility of each nucleoprotein complex matched in tailed vs. tailless arrays. This is a preliminary result indicating the role of histone tails in compaction is non-essential for nucleosomal array compaction, though it does not rule out a direct interaction between MeCP2 and the histone amino-terminal tails.

3.3.3 The TRD-CTD domain fusion can bind and compact both DNA and nucleosomal arrays similarly to wt MeCP2 while the CTD cannot.

The EMSA in Figure 3.1 demonstrates DNA, nucleosomal arrays, and tailless nucleosomal arrays were compacted by both wild-type MeCP2 and the TRD-CTD domain fusion and unable to enter the agarose gel by a molar ratio of 6. This is not the case for the CTD alone. At the highest molar ratio assayed, CTD was able to shift all templates but not to the same degree as wild-type or the TRD-CTD fusion. The observation of the TRD-CTD having similar ability to compact nucleosomal arrays as wild-type MeCP2 is in agreement with the findings of our collaborators in Figures 4.3 and 4.4 of Chapter 4. It is important to note that the domain nomenclature used in that paper differs significantly from the nomenclature used in the rest of this dissertation. The differences are in the precise amino acid residue numbers that constitute the CTDα. Based on previously published trypsin digestion resistance mapping conducted by Dr. Valerie Adams, I define the CTDα as residues 310-354, whereas our collaborators define the CTDα as residues 261-330. Examining “Table 4.1. Domains of Human MeCP2 Used in this Study” in Chapter 4 reveals confusing nomenclature surrounding

the CTDα. The CTDα is listed in the table but no corresponding residue numbers are listed and the “CTD” is there defined as residues 261-330.

I was unable to express what the Hansen lab has defined as the CTDα (310-354) but I was able to express and purify residues 300-486 which is named CTD in this dissertation. The observation that this CTD fragment (300-486) shifts DNA, nucleosomal arrays, and tailless arrays is being reported here for the first time. Making a direct comparison between the shifting capacity of TRD-CTD fusion, wild-type MeCP2 and the CTD is also a novel observation.

In the next set of gel shift experiments I further narrowed the region in the C-terminal portion of MeCP2 that can shift DNA and chromatin templates. Figure 3.2 refines the initial observation in Figure 3.1 by examining smaller and further C-terminal pieces of MeCP2. As Figure 3.1 shows, there is no significant difference between MeCP2 fragment-binding to tailed vs. tailless arrays. Therefore tailless arrays were not used in the experiment depicted in Figure 3.2. I tested the ability of fragments 354-486 and 335-486 in MeCP2 to bind and shift either 208-12 DNA alone (depicted in the top ppanel of Figure 3.2) or 208-12 nucleosomal arrays (depicted in the lower panel of figure 3.2).

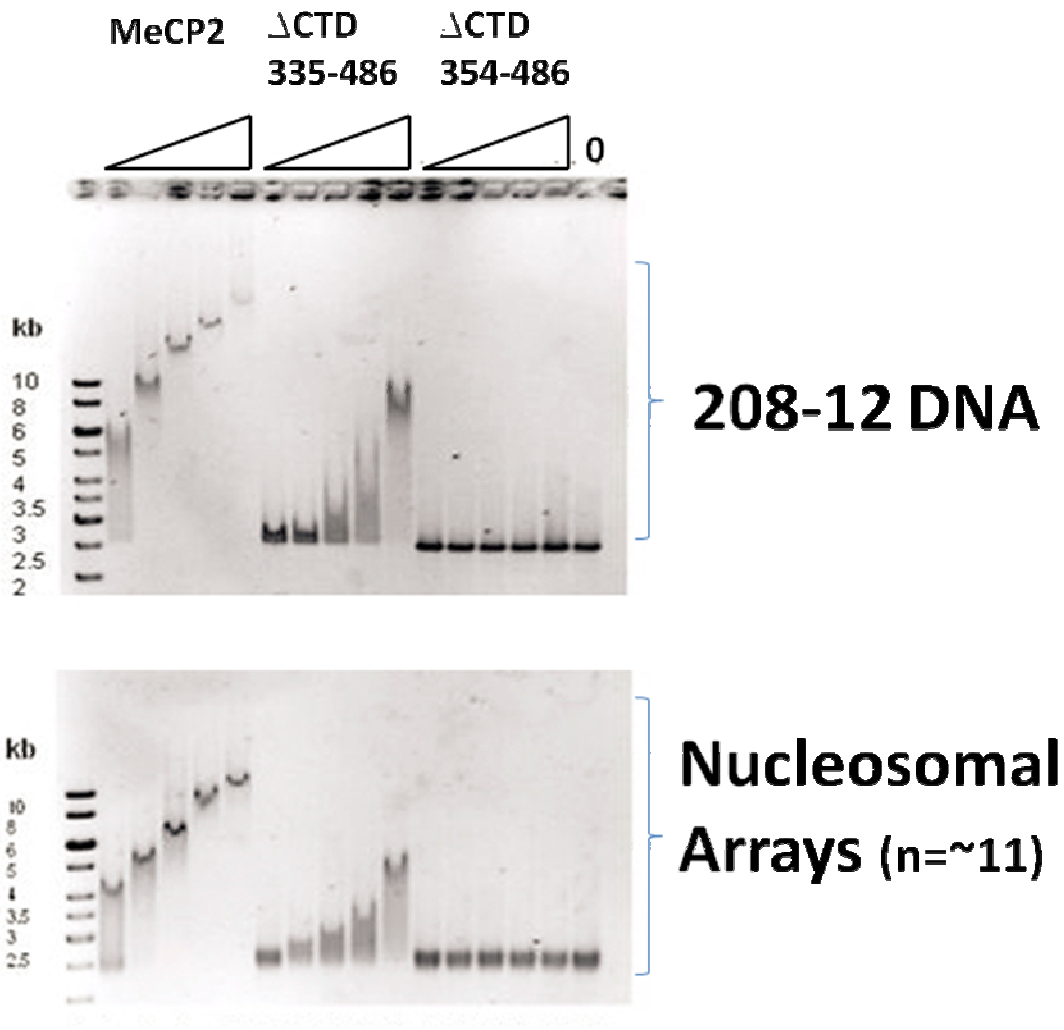

Figure 3.2 MeCP2 fragment 335-486 shifts 208-12 DNA and nucleosomal arrays while fragment 354-486 does not. Each MeCP2 construct was incubated with 200 ng of array or DNA alone at 2,4, and 6, 8 and 12 molecules MeCP2 fragment per 208 bp repeat at room temp for 30 minutes.

3.3.4 MeCP2 residues 400-450 cannot bind either DNA or nucleosomal arrays. MeCP2 residues 400-450 were cloned and expressed as an isolated fragment because of the anomalously low isoelectric point of that fragment compared to the high average isoelectric point of the protein overall (4.86 compared to 9.95 respectively). It was hypothesized that this acidic bit of MeCP2 may be the piece responsible for histone binding. However, this was not the case *in vitro* as 400-450 expressed alone could not bind DNA or chromatin templates. This is clearly demonstrated in both Figures 3.1 and 3.2 in which the 400-450 fragment has no effect on the electrophoretic mobility of DNA, nucleosomal arrays, or tailless arrays at any of the assayed molar ratios. There were no observable shifts by the 400-450 fragment.

3.3.5 Carboxy-terminal domain mutants 300-486 and 335-486 of MeCP2 bind unmethylated DNA and nucleosomal arrays, while residues 354-486 do not. A pTYB1 plasmid containing a construct coding for MeCP2 amino acid residues 335-486 was obtained from the lab of Chris Woodcock. This fragment, which they named the CTDβ, has 20 more amino acids on the N-terminal side than does the CTDβ (corresponding to residues 354-486) as named by our lab in 2007. In the study mentioned earlier (see Appendix I), the CTDβ as defined by Woodcock and Ghosh was used to establish a unique motif in the carboxy terminal of MeCP2 capable of shifting nucleosomal arrays but not DNA alone. The data I collected and presented in Figure 3.2 does not agree with these published results. I used the CTDβ (according to Ghosh; 335-486) and the CTDβ (according to Adams; 354-486) and compared their ability to bind either 208-12 DNA alone or nucleosomal arrays. I found the fragment 354-486 incapable of shifting DNA alone or nucleosomal arrays, while fragment 335-486 was

able to shift both nucleosomal arrays and DNA alone. This observation indicates the unmethylated-DNA-binding ability in the carboxy terminal of MeCP2 lies between residues 335 and 354. When considered in conjunction with the observed ability of fragment 300-486 to bind both DNA and arrays in Figure 3.1 there is the possibility of additional DNA binding activity between residues 300-335. Current investigation by Dr. Mary Porter-Goff is testing this hypothesis by deleting this region of MeCP2 from the full-length protein.

It is worth noting that the EMSAs I presented in Figure 3.1 and 3.2 were conducted using the 208-12 tandem-repeat 5S DNA alone, arrays reconstituted with this DNA and recombinant wild type *Xenopus laevis* octamers, and arrays reconstituted with tailless *Xenopus* octamers. These reagents differed from those used by our collaborators (Dr. Chris Woodcock) as they used the 601 DNA sequence alone and reconstituted into arrays. There is a difference in primary sequence of the reagents used which may account for the discrepancy between Figure 3.2 of this dissertation and panels a. and d. in Figure 4.3 of Chapter 4. However, this is unlikely as Dr. Uma Muthurajan tested the same C-terminal constructs as I on 601 and 207 mono-nucleosomes and obtained EMSA results that agreed with mine (personal communication). Also, fragment 354-486 did bind to DNA and arrays at extremely high molar ratios but needed to be mixed at an order of magnitude higher concentration (ratios between 40-60 were required) than wild-type to have the same effect. This observation shows the extreme carboxy-terminus of MeCP2 has weak unmethylated-DNA-binding ability.

Parallel studies conducted in the lab of our collaborator Dr. Christopher Woodcock show that CTD-beta (beginning at residue 335 extending to residue 486) can bind to nucleosomal arrays but does not bind to naked DNA. It was concluded from this study that there is a chromatin-specific binding site in this domain. However, when I

attempted to repeat this result I was unable to reproduce this and instead showed the MeCP2 fragment extending from amino acid 335-486 to bind and shift both 208-12 DNA alone and reconstituted chromatin arrays. The results are published, though this discrepancy in the data is not resolved.

3.3.6 DNA binding in the intervening domain (ID) of MeCP2 resides in residues 168-188:

Prior to the observation that the TRD-CTD domain fusion within MeCP2 was the minimum required region to facilitate nucleosomal array compaction (96) our lab had hypothesized that a motif within the intervening domain (ID) of MeCP2 (aa 168-206) might facilitate array compaction. This hypothesis prompted me to perform a scanning BLAST search in the human proteome using a frame of 10 amino acids within the ID of MeCP2. I was looking for motif homologues of this region of MeCP2 that might also have chromatin architectural functionality. Here I found a seven-amino acid stretch of identical homology between MeCP2 and the HMGA family of architectural transcription factors. That sequence is RGRPKGS and corresponds to a.a. 188-194 in the human MeCP2 gene product (see Figure 3.3). Consequently, these seven amino acids exactly match the middle of the three AT-hook motifs in the HMGA1 proteins and only roughly match the core consensus GRP within the other two flanking AT-hook motifs.

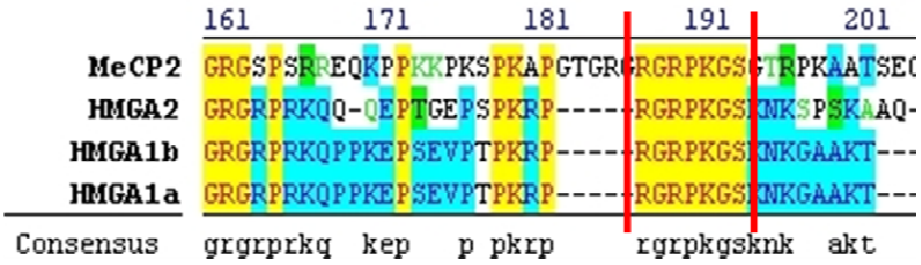

Figure 3.3 A seven amino acid portion of an AT-hook motif from HMGA proteins (seen between the red bars) is identically homologous to amino acid residues 188-194 in MeCP2. This alignment was performed using Jellyfish Sequence Analysis software.

I conducted a series of gel-shift assays to determine if this homologous AT-hook motif was a genuine DNA-binding domain in MeCP2. I began by generating point mutant - R188E, a residue at the N-terminal end of the AT-hook homology, in the context of full-length MeCP2. This point mutation significantly reduced the ability of the full-length protein to bind and shift non-methylated DNA alone. Figure 3.4 shows an image of a 0.8% agarose gel run at 5V/cm for four hours with 200 ng 208-12 DNA in lanes 2-9 with labeled λ BSTE II marker in lane 1 and 10. DNA was mixed and incubated with the indicated molar ratios of wild-type MeCP2 or R188E point mutant respectively, indicated above the gel image.

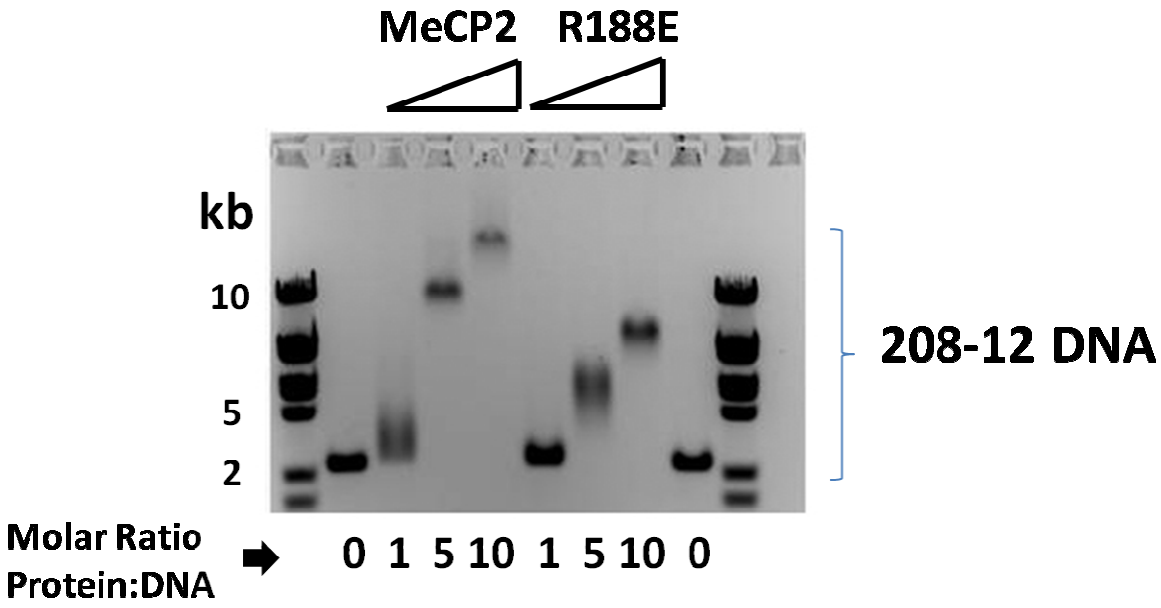

Figure 3.4 MeCP2 R188E has reduced ability to shift nucleosomal arrays in EMSA. In this 0.8% agarose gel Lambda BSTE II marker is shown in lanes 1 and 10, 200 ng 208-12 5S DNA alone in lanes 2 and 9, and the same amount of DNA incubated with human MeCP2 full length wild-type at a molar ratio (moles protein: moles 208bp repeat) of 1, 5 , and 10 in lanes 3-5, and the R188E mutant full length MeCP2 at the same increasing ratios in lanes 6-8.

The result depicted in Figure 3.4 led me to clone the intervening domain peptide alone (residues 168-206) (as described in the materials and methods section of Chapter 2 of this dissertation) and assay its ability to bind DNA. The 38-a.a. peptide was able to shift 208-12 DNA alone as shown in Figure 3.5. I also observed this to be reversible by competing the peptide off the 208-12 DNA by incubating it with 208-1 DNA. ID peptide came off the 208-12 and shifted the smaller DNA template as well (data not shown).

.

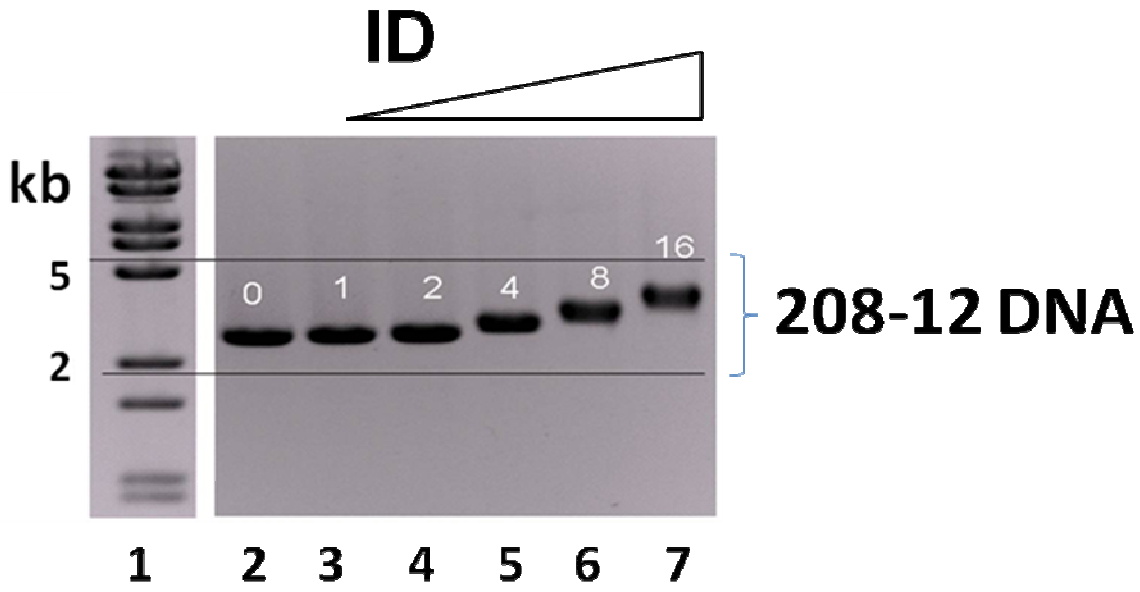

Figure 3.5. ID alone binds and shifts 208-12 DNA. Pictured is a 0.8% agarose gel. Lane 1 is a λ BstEII DNA marker. Lane 2 contains 208-12 DNA alone. Lanes 3-7 show decreased 208-12 DNA mobility when a higher ratio of ID molecules per 208 b.p. repeat is present. The molar ratios of ID molecules per 208 b.p are shown in white.

3.3.7 The Intervening Domain (ID) of MeCP2 has a robust non-specific DNA binding motif between residues 168-188.

There are two regions inside the primary sequence of MeCP2 that have homology to an AT-hook. These correspond to MeCP2 residues 188-194 (within the ID) and residues 267-274 (within the TRD) (Figure 3.3). An AT-hook is a novel peptide motif that binds non-specifically to the minor groove of DNA. There are three such domains in each of the high mobility group A (HMGA) family of non-histone nuclear proteins (124). These novel AT hook motifs all share a consensus core of amino acids with a glycine, arginine and proline (GRP) making up the three amino acids that are most evolutionarily conserved. Of the two 'AT hook'-like motifs in MeCP2 the one from residues 188-194 has the GRP core while the other has a lysine inserted between the R and the P of the motif. I had hypothesized that this homologous domain was responsible for the DNA-binding activity observed in the isolated ID peptide apparent in the gel shift in Figure 3.5. I disproved this hypothesis and instead showed that the DNA-binding activity present in the ID is most likely located between residues 168-188. When I made a triple point mutation substituting three alanines in place of the GRP consensus core in the context of the isolated ID peptide, I observed an increased DNA binding affinity in gel shift assays (Figure 3.5). A single point mutation at residue 188, R188E decreased DNA shifting activity of the peptide. During the course of these experiments I hosted Dr. Raymond Reeves of Washington State University to present a seminar on AT-hook motifs. During the course of his visit the seemingly counter-intuitive results made sense. He explained that nanomolar affinity of the AT hook motifs in the HMGA proteins require that the GRP consensus core be flanked on the amino side by a proline in order to give the motif the steric pucker necessary for the motif to wedge into the minor groove of DNA. The 'AT-hook' –like region of the ID, though it has seven identical amino acids to

the central AT-hook of the HMGA proteins it is missing a proline on the amino side of the GRP core, indicating this is not the region of the ID responsible for the DNA-binding activity of the isolated ID peptide. The observed homology suggests the HMGA family of non-histone nuclear proteins as ancestors to MeCP2. Further studies are needed to narrow the DNA binding activity of the ID between residues 168 and 188. This observed binding activity of the ID was observed in agreement with our collaborators' results and was published jointly. The primary data of our collaborators was used in Figure 4.3 of Chapter 4. Results of point mutations generated and tested in Figures 3.4 and 3.5 are published for the first time in this dissertation.

Chapter 4

Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2)

4.1 Preface:

Chapter 4 is reprinted from a paper I jointly authored that first appeared in Biochemistry online April 20th, 2010 (96). The entire paper can be accessed freely at the following shortened URL - http://bit.ly/gc3Qup. My contributions to this collaborative effort included corroborating results using each of the six MeCP2 fragments listed in electrophoretic mobility shift assays to determine DNA and chromatin binding activity in MeCP2 domains. In depth discussion about agreement and discrepancies among these concurrent results can be found in Chapter 3 of this dissertation.

4.2 Introduction:

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 (125,126)). The finding that mutations in human MeCP2 result in Rett syndrome (RTT), a debilitating neurodevelopmental disorder (23), and that its mis-regulation is common in other patients with Autism Spectrum Disorders (127,128), 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 (128). 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 (129). In humans, MeCP2 deficits have been linked to conditions other than RTT, including cancer (118,130,131)

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 (85). Early work identified a short (~90 residue) ordered region of MeCP2 (between residues 75-164) with the ability to bind methylated DNA (8). 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 (11,12,132). DNA

methylation is typically associated with transcriptional repression, and indeed MeCP2 has been shown to repress methylated genes in artificial systems *in vitro* (3). This transcriptional down-regulation activity was mapped to amino acids 207-310 and the region was accordingly termed the transcriptional repression domain (TRD) (3). The MBD and TRD together comprise ~40% of the 486 residue hMeCP2 sequence (Fig. 4.1a, 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 (133). 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 (134). 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 (22), DNMT1 (79), histone H3K9 methyltransferase (75), PU1 (80), splicing factors (34), BRM (77), RNA (65), and the RNA splicing machinery (64). 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 (85), which characteristically has large number of binding partners and multiple functions (135).

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 (5), and that MeCP2 binding can lead to both repression and up-regulation depending on the gene context (4). Evidence has also been presented that MeCP2 is involved in the maintenance of large-scale chromatin loops, perhaps by physically anchoring loop bases (136). 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 (24,36,137). 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 (11,88). 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 (78). 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 (86), and a study of RTT patients showed that mutations tend to be located in this region (87). Finally, the C- terminal portion of MeCP2 required for chromatin interactions *in vitro* (36) also harbors the Group II WW domain binding motif required for binding to splicing factors (34), 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 (87).

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 *Mo*lecular *R*ecognition *F*eatures (MoRFs) (97,138), 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.

4.3 Materials and methods

4.3.1 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 (GACCGTGAATTC) 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′ CCGGTTTAAACCGGGGATCTCGATCC 3', NTD Reverse 5′ GTTAGAGAATTCGTCACGGATGATGGAGCGCCGCTG 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 (residues165-210), TRD (residues 207-310), CTD-α (residues 261-330) and CTD-β (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'CAATGACATATGGAAGCTTCTGCCTCCCCCAAACAGC 3',
Reverse primer 5' GTTAGAGAATTCGCTCCCTCTCCCAGTTACCGTGAAG 3'
ID Forward primer 5' CAATGACATATGCCCTCCCGGCGAGAGCAGAAACC 3',
Reverse primer 5' GTTAGAGAATTCCACCTGCACACCCTCTGACGTGGC 3'
TRD Forward primer 5' CAATGACATATGGTGCAGGTGAAAAGGGTCCTGGAG 3',
Reverse primer 5' GTTAGAGAATTCCTCCCGGGTCTTGCGCTTCTTGATG 3'
CTD-α Forward primer 5'CAATGACATATGCCTCAGGCCATTCCCAAGAAACGGG3',
Reverse primer 5' GTTAGAGAATTCCTCACCGAGGGTGGACACCAGCAG 3'
CTD-β Forward primer 5'CAATGACATATGGGACTGAAGACCTGTAAGAGCCCTGG 3',
Reverse primer 5' GTTAGAGAATTCGCTAACTCTCTCGGTCACGGGCGTC 3'

The domain combinations NTD-MBD (residues 1-164), MBD-ID (residues75-210) 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-CTDα-CTDβ: TRD forward primer, CTD-β reverse primer). The triple domain NTD-MBD-ID was synthesized using the primers 5' - CCCGGTTTAAACCGGGGATCTCGATCCCGC - 3' forward and 5' - TTTCAG AATTCCTGCACACCC TCTGACGTGGCCGC - 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. AEAAHRWCCPGCCKTF (GTTAGA*GAATTC* GCTGCTCATCGTTGGTGTTGTCCTGGTTGTTGTAAAACTTTT *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 XhoI 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 (139).

4.3.2 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 (36,88). For NTD and CTDβ, 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).

4.3.3 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 (88). 601-12 DNA was purified, methylated and reconstituted into saturated/ undersaturated nucleosomal arrays as described (36).

4.3.4 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-CTDα-CTDβ 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℃ 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-α, CTD-β, TRD-CTDα-CTDβ) 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.

4.3.5 Electron Microscopy (EM)

Sample fixation, grid preparation, and darkfield EM imaging were as described (88).

4.3.6 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 (140). Estimates of secondary structure were calculated using CONTINLL and reference set 7 on Dichroweb (141,142), which, in addition to a set of structured proteins (140,143), 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 (88) that estimates of secondary structure in MBD and full-length MeCP2 derived using CONTILL and LIN-COMB is almost identical (144). 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+CTDα, NTD+ID, ID+TRD, ID+CTDα, NTD+TRD) CD data were acquired separately for the two domains and also for their mixture, keeping the concentrations constant (145). 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 [((A)+(B))-(A+B)]/((A)+(B))].

4.3.7 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 (88). 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 4.4 are averages of three independent sets of data.

4.3.8 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 0.5nm 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 295nm. At this wavelength there is no inner filter effect due to acrylamide (146). In agreement with this, at 295nm and at 324 nm the maximum concentration of acrylamide used in the assay (310μM) 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 2.5μM protein in a final volume of 600μl. 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:

$F0/F=1+ K_{sv}*Q$ (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 *Ksv* is the effective quenching constant. Fit of this equation to the raw data (F0/F) yielded the K_{sv} and a y intercept of 1.

4.3.9 Fluorescence labeling of tetraCys-MBD and anisotropy

MBD-tetraCys stock solution was incubated with 10mM DTT for 6hr at 4℃ 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. ~50µM MBD-tetraCys was then incubated overnight at room temperature with 2 molar equivalents of FlAsH EDT2 (LumioGreenTM labeling reagent, Invitrogen) and dialyzed extensively against the same buffer in the dark at 4℃. A two-fold molar excess of F lAsH label for purified tetraCys fusions has been shown to be sufficient for efficient labeling (147,148). FlAsH-EDT alone has negligible fluorescence whereas the tetraCys peptide-bound FlAsH undergoes a boost in quantum yield in excess of 10^4 fold (149), 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 (λex) 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 10µM,

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 (146). 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 (150). 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.

4.3.10 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/GGAATTCATAATG - 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 λex used was 480 nm and the emission was collected over a window of 4nm (518-521nm). A constant amount of labeled DNA (100pM) was mixed with various amounts of MBD, MBD-ID, NTD-MBD, and TRD-CTDα-CTDβ covering a range of 100pM to 60nM and NTD, ID, TRD, CTD-α and CTD-β 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-β to the other domains (e.g. ID) anisotropy values were normalized using the equation $r_{norm}= r_n/r_0$.

4.3.11 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 (151) 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.

4.3.12 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 (152). To this end, the fractional difference in composition between MeCP2 (or a set of disordered proteins from the DisProt database, (153) 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 (153).

4.3.13 Disorder, α-MoRF prediction, and modeling

Disorder predictions for MeCP2 were made using PONDR® VLXT (152). 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 α-MoRF predictor which detects short (≤20 residue) stretches within long regions of disorder with the potential for helical structure acquisition upon binding (97,114). 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) was used to visualize the MoRF residues within the structure of a MBD-methylated DNA complex (PDB 3c2i) (12).

4.4 Results

4.4.1 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 4.1 lists the salient features of the MeCP2 polypeptides we prepared, and Figure 4.1a (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 (85) denoted CTD-α and CTD-β.

Table 4.1. 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-α			
CTD-β		335-486	156
NTD-MBD		1-164	164
MBD-ID		75-210	136
NTD-MBD-ID		1-208	208
TRD-CTD		207-486	280
$MeCP2^{1-294}$		1-294	294

4.4.2 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) (152,154)(Fig. 4.1a, 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 (11,12). Other short segments of predicted order occur throughout the protein and are found in all the domains except the short ID (Fig. 4.1a). This alternating pattern of disorder and order is also predicted by the FoldIndex algorithm (85). Unstructured proteins tend to have an amino acid composition that favors structure-disrupting residues (153,155,156) and MeCP2 represents an extreme case of this skewing, exceeding that of the unstructured proteins in the DISPROT (153) data base (Fig. 4.1b).

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. 4.2a). Of the six domains, the MBD was the only one showing a characteristic positive band in its CD spectrum at ~197 nm, indicative of significant ordered secondary structure (Fig. 4.2a). 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 (88,141,142) deconvolution (see Materials and Methods for details). Deconvolution produced highly reproducible estimates of secondary structure for each domain (Table 4.2).

In agreement with previous findings (11,85,88), the MBD is ~60% structured, with ~45% β-sheet/turn, and ~15% α-helix. The proportion of predicted unstructured sequence for the NTD, ID, and CTD ranged from ~62% to ~78% (Table 4.2), consistent with the net 60% unstructured sequence in intact MeCP2 (85,88). The amount of disorder in each of the domains determined by CD (Table 4.2) is close to that predicted by the PONDR (Fig. 4.1a) and FoldIndex (85) algorithms.

Table 4.2. Secondary structure content of MeCP2 domains based on CONTINLL deconvolution of CD data

Domain	% ordered secondary structure (standard error)				% unstructured (standard error)
	α-helix	β-strand	β-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-α	7 (0.5)	12 (1.0)	12 (0)	31	69 (0.5)
CTD-β	9 (0.5)	16 (0)	15 (0)	40	60 (0.5)

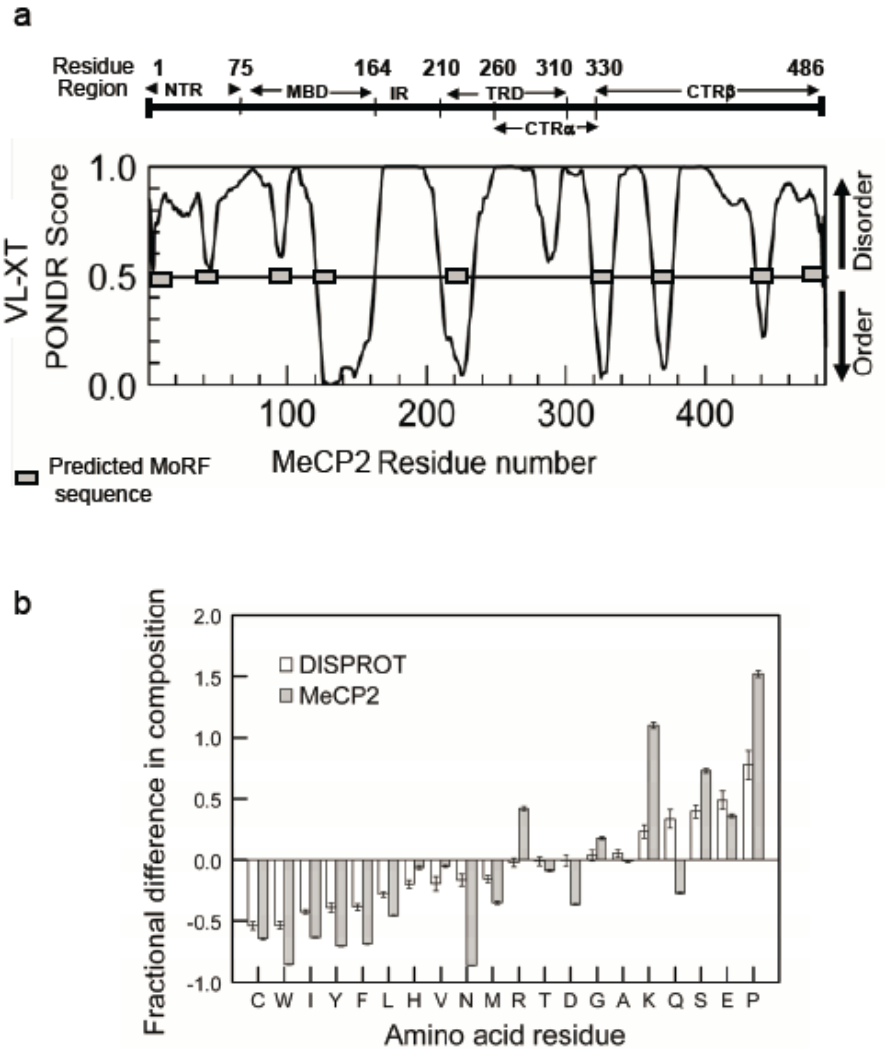

Figure 4.1. Organization of MeCP2 and relation to disorder predictions
(a) Upper panel – map of MeCP2 showing the six major domains identified by partial proteolysis (*11*). The graph shows the order-disorder score of MeCP2 predicted by PONDR VLXT, a neural network predictor of native disorder (*39*). 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) show 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.

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

Native electrophoretic mobility shift assays (EMSAs) provide a qualitative estimate of the DNA and chromatin interaction properties of MeCP2 (36,157), 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) (36). 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 (158), 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 (119). Also, we have found that closely spaced methylated CpGs favor cooperative DNA binding of MeCP2 (RPG and CLW, in preparation).

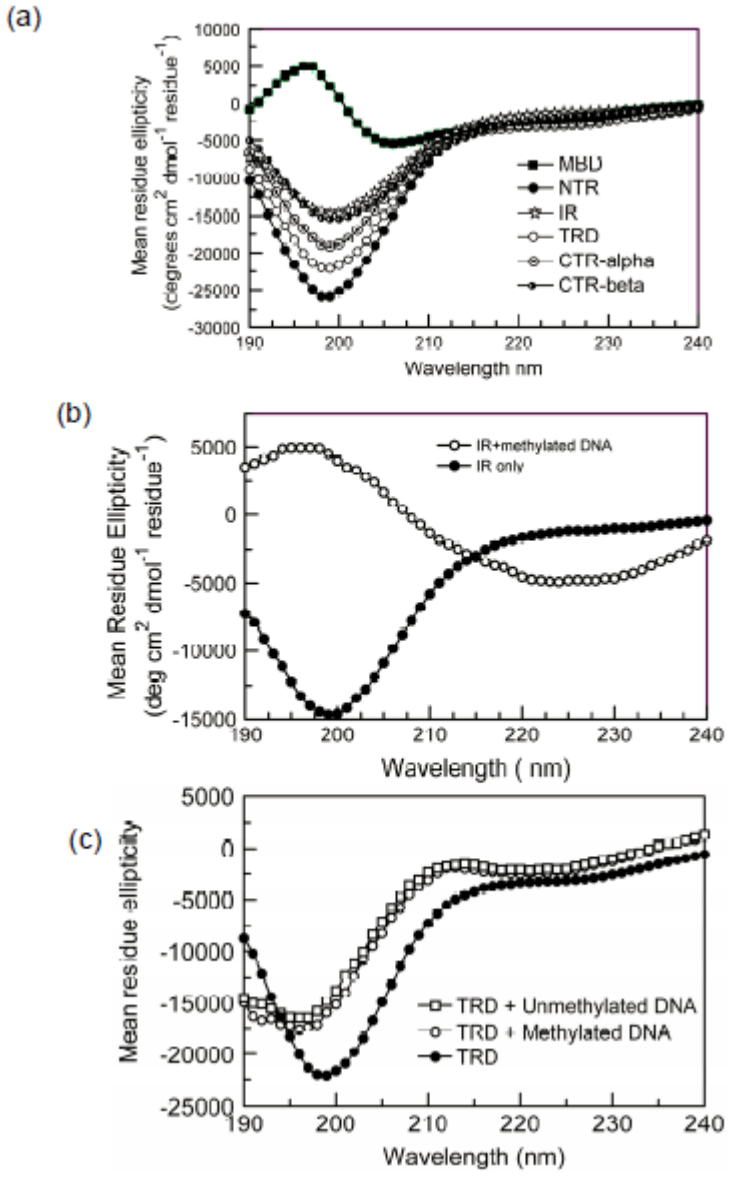

Figure 4. 2. Circular dichroism spectra of MeCP2 domains reveal marked differences in secondary structure content.
CD spectra are representative of two - four separate acquisitions. (a) Compared to the 195nm peak indicative of β-sheet structure within the MBD (black squares), all the other domains show a negative band in the 195nm-198nm region indicative of disorder. NTD (black circles), TRD (white circles) and CTDα (black rhombi) have lower structure content than ID (stars) and CTDβ (half filled circles) (see Table 4.1 for quantitation).(b) Addition of DNA (methylated as well as unmethylated) to the ID induces changes typical of the formation of α-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.

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. 4.2b far right). Among the individual domains, the NTD and CTD-β stand out as inducing only very minor mobility shifts in DNA, whereas the ID, TRD, and CTD-α fragment all induce marked shifts (Fig. 4.3a). The shifts induced by these three domains are methylation-independent (Figure 4.3c), accounting for the substantial methylation-independent binding observed with intact MeCP2 (Fig. 4.3b) (24,85,137). 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 (85). In general, the shifts with DNA and NAs are qualitatively similar. However, the CTD-β is a clear exception, inducing a moderate but reproducible shift with NAs but not with naked DNA (Fig. 4.3a, 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 (36), and suggests that the CTD-β contains a unique histone binding region(s).

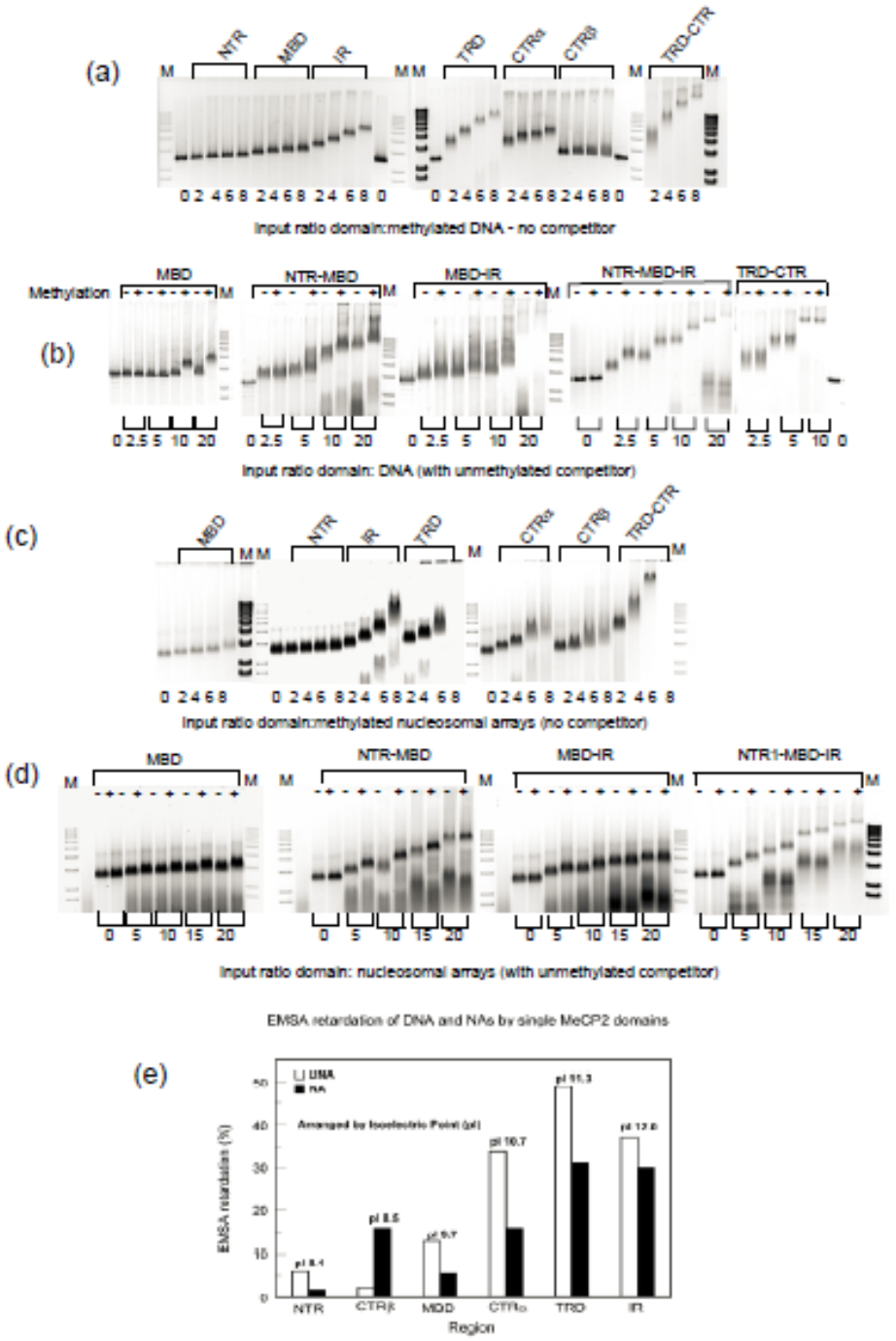

Figure 4.3. MeCP2 domains induce electrophoretic mobility shifts upon addition to DNA or chromatin.

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, CTDα, CTDβ, TRD-CTDα-CTDβ) 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 CTDα 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-β 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.

4.4.4 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. 4.3a, far right) and NAs (Fig. 4.3d, far right). This is consistent with DNA binding by TRD and CTD-α, and chromatin binding by the CTD-β (Fig. 4.3a, 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 (36,88). 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. 4.3b) and chromatin (Fig. 4.3d 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. 4.3a, 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. 4.3b). 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. 4.3 b, e). It should be stressed that methylation-enhanced shifts occur only with MBD-containing polypeptides. Examples of this are shown in Fig. 4.3c where, for ID and the TRD-CTD fusion (TRD-CTDα-CTDβ), the methylation state of the DNA has no effect on the induced shift.

4.4.5 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 (24,36). 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 (159). 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 (151). 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 (±2) S (Fig. 4.4). Full-length MeCP2 at a 1:1 input ratio induced a dramatic increase to ~63S (Fig. 4.4), similar to our earlier observation in the 5S 208-12 NA system (24). 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. 4.4. The MBD alone increased the sedimentation coefficient by only ~2(±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. 4.3d), caused an increase in sedimentation coefficient of ~4.5 (±0.5) S, whereas MBD-ID resulted in a sedimentation coefficient increase of ~6 (±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 (±1) S was seen with the NTD-MBD-ID construct (Fig. 4.4), and the most striking result was obtained with TRD-CTD. Binding of the latter caused an increase of ~26 (±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. 4.4, 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).

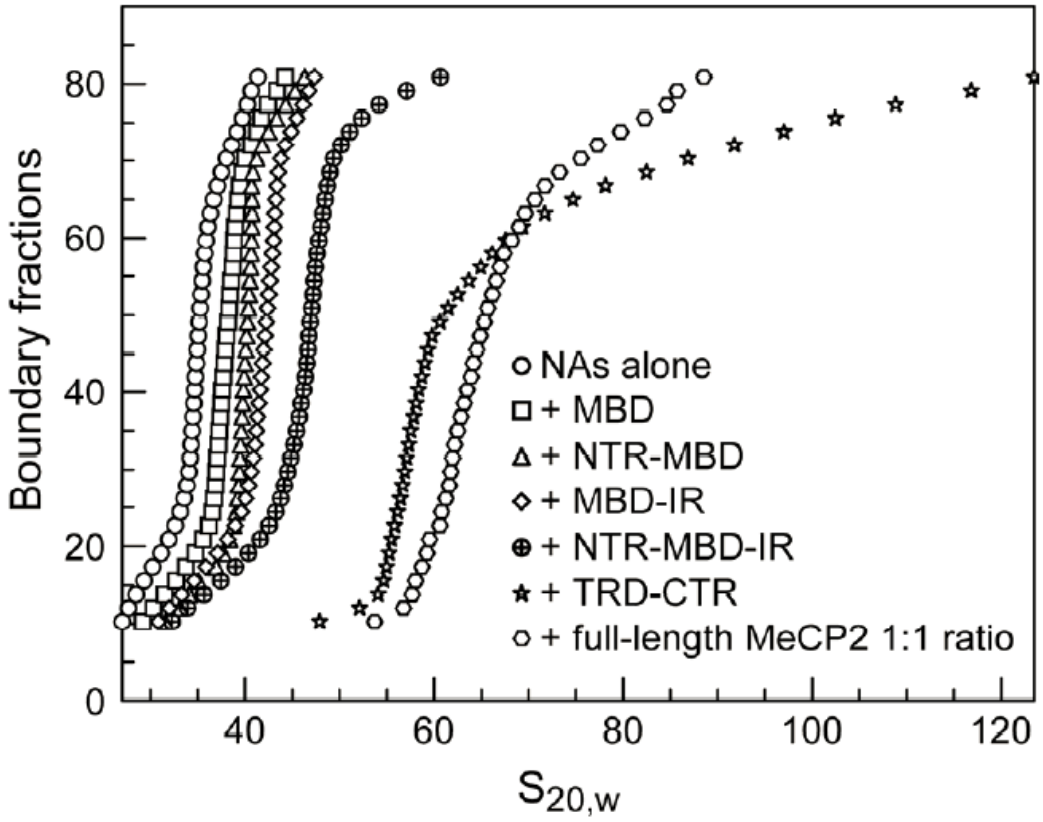

Figure 4.4. Sedimentation velocity reveals differences in the ability of MeCP2 domains to compact nucleosomal arrays

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.

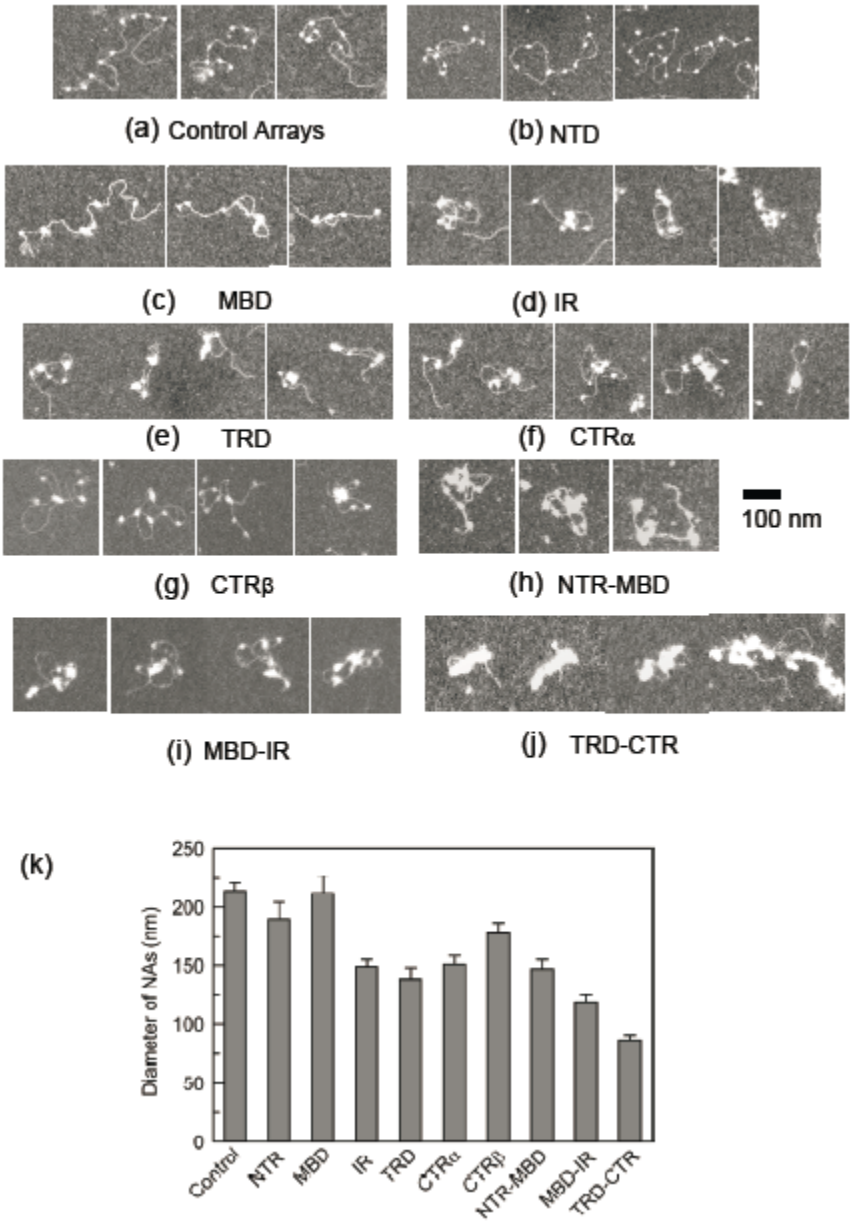

Figure 4.5. Direct EM observation reveals differences in conformational changes induced in undersaturated nucleosomal arrays by MeCP2 domains

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.

4.4.6 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 (36). 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. 4.5a). Little change in conformation was seen with either MBD or NTD alone (Fig. 4.5b, c), consistent with the EMSA data on saturated NAs (Fig. 4.3). All the other individual domains induced partial clustering of nucleosomes within arrays (Fig. 4.5d-g), leading to significant ($p<0.001$) reductions in array diameter (diameter of the smallest circle that fully encloses the array; Fig. 4.5k). The three contiguous constructs examined (MBD-ID, NTD -MBD, and TRD-CTD) also showed nucleosome clustering and array compaction (Fig. 4.5h-j), with decreased array diameters (Fig. 4.5k) that parallel the increased sedimentation coefficients seen in Fig. 4.4. 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. 4.5j, far right panel), consistent with our observation of a large fraction of heterogeneous rapidly sedimenting material (Fig. 4.4). The full-length

protein induces extensive array oligomerization under these conditions, (*32*) precluding accurate measurements of array compaction.

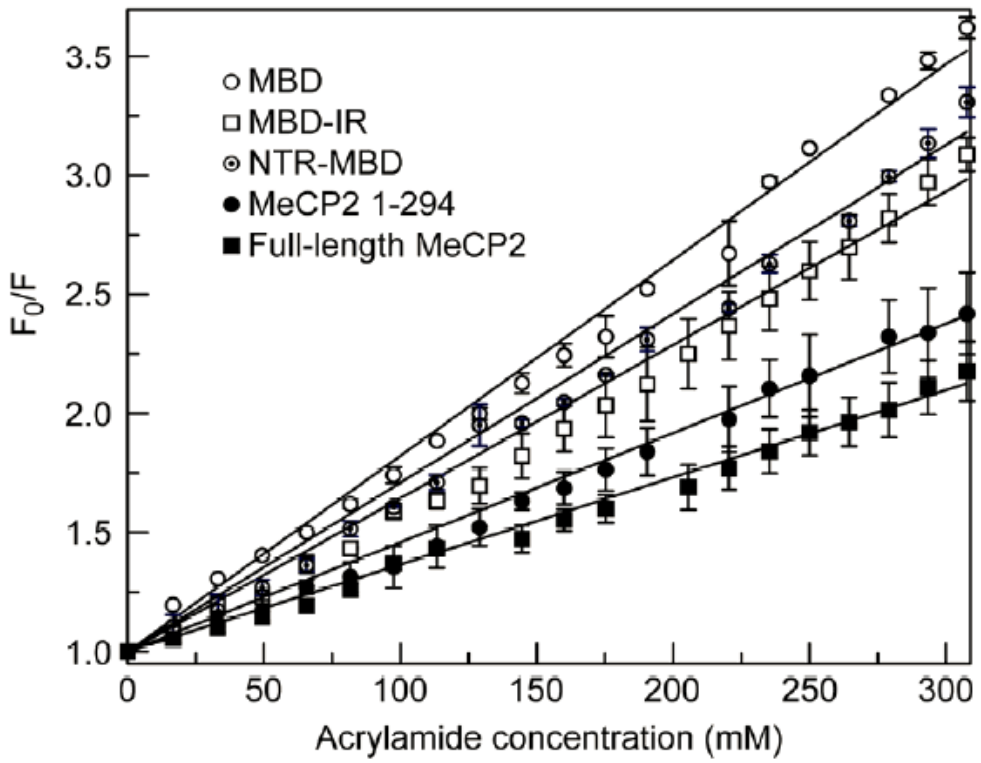

Figure 4.6. Interactions between the MBD and flanking domains revealed by tryptophan accessibility

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.

4.4.7 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 (146) of the single tryptophan at position 104, and reported that domains flanking the MBD provided solvent protection to W104 (88). 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. 4.6) show that acrylamide most effectively quenches the fluorescence signal of tryptophan in the MBD alone (quenching constant K_{sv} 8.4 M^{-1}), and least effectively in full-length MeCP2 (K_{sv} 3.9 M^{-1}). Intermediate levels of quenching (K_{sv} 7.3 M^{-1} and 6.8 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 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 (160).

4.4.8 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 (161). 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 (149,162) at a level that produced a robust emission (Fig. 4.7A 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 μM and ~2.5 μM respectively (Fig. 4.7a). In contrast, no interactions were observed between MBD and NTD. A weak effect was seen with the CTD-β, 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 (163). 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. 4.7b, 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-β pairs (Fig. 4.7c). 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).

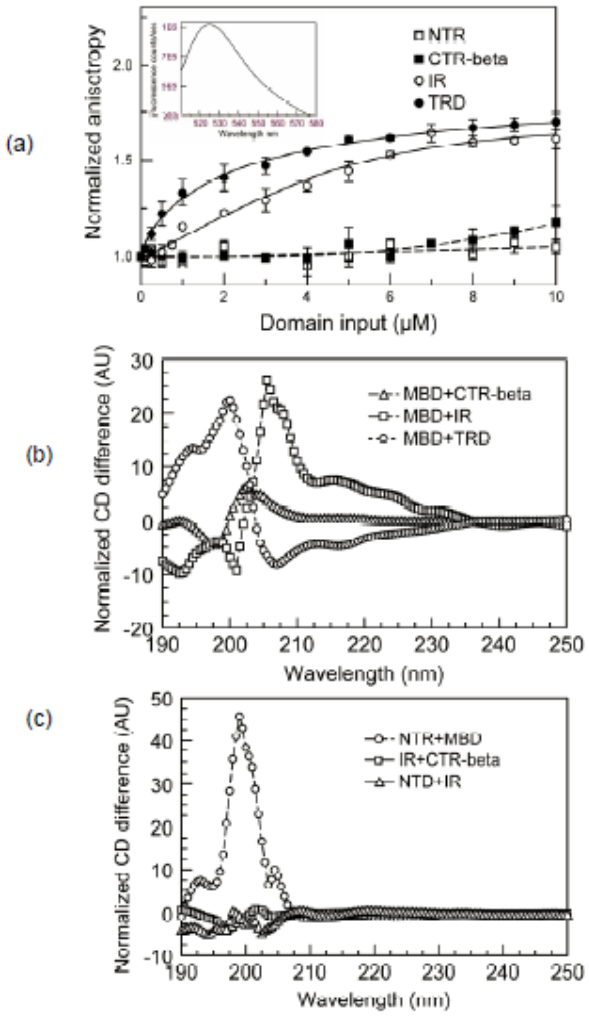

Figure 4.7. In trans interactions between MeCP2 domains revealed by fluorescence anisotropy and CD

(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. CTDβ (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+CTDβ pair shows only a minor change. (c) The NTD+MBD pair (circles) show a distinct difference at ~198nm while NTD+ID (triangles) and ID+CTDβ (squares) show no differences.

4.4.9 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 (72,119). A similar stretch of BDNF DNA was used in the MeCP2 complex for which the x-ray structure has been determined (12). 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. 4.2 b, 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 4.3)). For ID, the acquired structure is approximately equally divided between α-helix and β-strand, while for the TRD; the increase is in the β-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 α-helix (88).

Table 4.3. Changes in secondary structure of MeCP2 domains upon DNA binding.

Domain	% ordered secondary structure	
	Without DNA	With DNA
Full-length MeCP2	35[1]	42[1]
NTD	22	23
MBD	60[1]	66[1]
ID	38	59
TRD	22	30
CTD-α	31	30

[1] From (*34*)

Table 4.4 Thermal stability of MeCP2

MeCP2 polypeptide	T_m alone (℃)	T_m (apparent) with unmethylated DNA	ΔT_m (unmet)	T_m (apparent) with methylated DNA	ΔT_m (met)
MBD^{75-164}	44.9 (±0.1)	46.6 (±0.6)	1.7	54.3 (±0.3)	9.4
$NTD\text{-}MBD^{1-164}$	40.4 (±0.6)	46.3 (±0.1)	5.9	54.5 (±0.3)	14
$MBD\text{-}ID^{75-210}$	46.7 (±0.2)	55.7(±0.3)	9	63.8 (±0.2)	17.1
$NTD\text{-}MBD\text{-}ID^{1-208}$	43.1 (±0.1)	58.6(±0.4)	15.5	67.1(±0.2)	24
$MeCP2^{1-294}$	45.3 (±0.1)	56.3 (±0.3)	11	64.6 (±0.2)	19.3
Full-length $MeCP2^{1-486}$	44.5 (±0.2)	55.7 (±0.5)	11.2	63.1 (±0.4)	18.6

ΔT_m – increase in melting temperature upon DNA binding

T_m values in the presence of DNA are denoted ‘apparent’ since reversibility cannot be tested.

4.4.10 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. 4.8 a-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. 4.8a, 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. 4.8a, 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. 4.8c), 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-α, which bound with affinities of 20 nM and 96nM respectively (Fig 8b), CTD-β induced small increases in fluorescence anisotropy and only at very high protein concentrations (~600nM) (Fig. 4.8c). 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-β, CTD-α, ID, TRD, and MBD.

Since DNA binding to MeCP2 also confers thermal stability to the protein (88), 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 t he 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$^{1-294}$ which includes most of the TRD. Unexpectedly, the T_m of NTD-MBD domain fusion was lower than MBD (Table 4.4), 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 4.4). 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$^{1-294}$ and full-length MeCP2$^{1-486}$ indicate that the C-terminal domain of the molecule does not contribute to the overall thermal stability of the protein when bound to DNA.

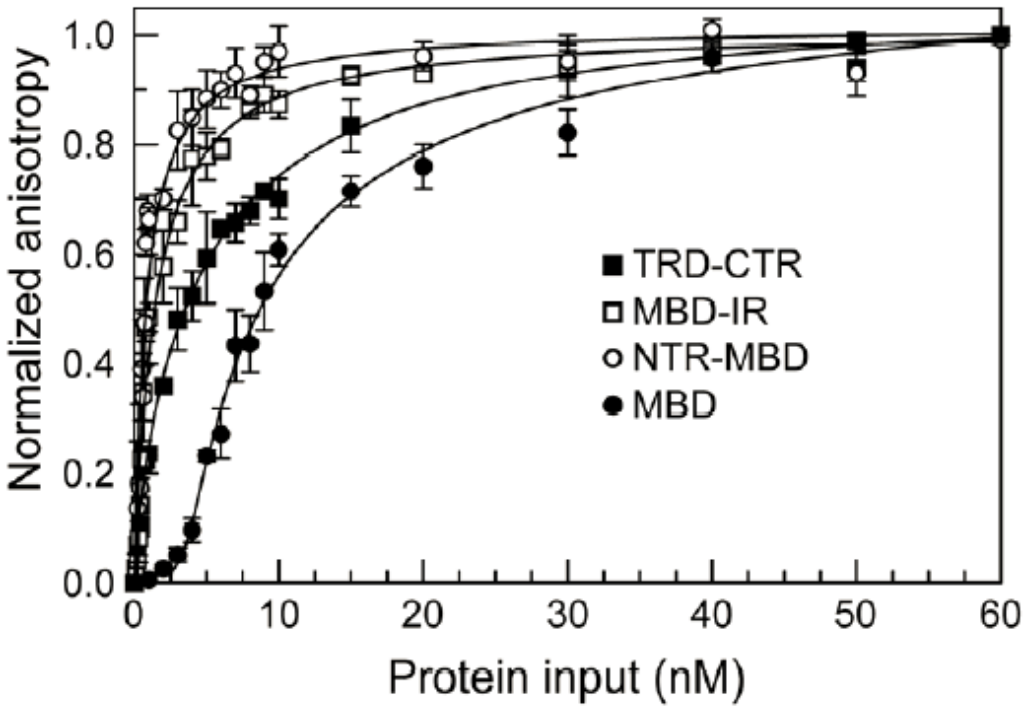

Figure 4.8. Quantitation of DNA binding affinity of MeCP2 fragments
(a) Normalized fluorescence anisotropy, $r_{norm}= (r_n-r_0) / (r_{max}-r_0)$, of a 5'-fluorescein labeled 22bp fragment of BDNF promoter DNA with a single methylated CpG was measured in the presence of increasing concentrations of MeCP2 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-α (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-β (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.

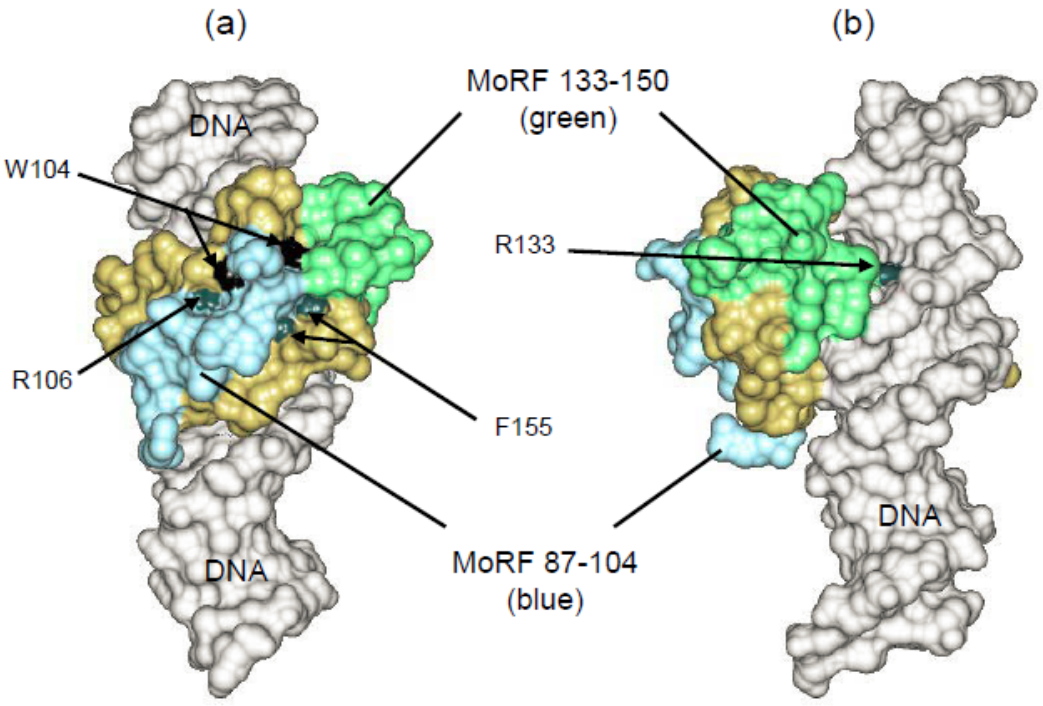

Figure 4.9. Structure of MBD bound to DNA suggests that MoRFs flank interaction surfaces.

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.

4.5 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.

4.5.1 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 (164-167). In this respect, it is significant that MeCP2 gains secondary structure and undergoes striking thermal stabilization upon binding to DNA (Table 4.4)*.* α-MoRF predictors (97,114) (see materials and methods) predict nine α-MoRFs in MeCP2. This is an unusually large number for a protein of this size (165), 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 α-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 (α-thalassemia/mental retardation X-linked) syndrome (74). In the x-ray structure of the MBD-DNA complex (12) the two discontinuous MoRFs, 87-104 and 133-150 (Fig 1), form a continuous surface (Fig 4.9). Both α-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. 4.9).

4.5.2 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 (5,24,36,162). 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 (104,168,169). MBD is solely responsible for methylation specific binding (Fig. 4.3b). The autonomous non-specific DNA binding capabilities of the ID, TRD and CTD-α-β 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-β, over DNA with which there was virtually no shift (Fig. 4.3d), 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 (24,61,170).

4.5.3 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. 4.3b, e and 4.8a, 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 (171). 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-α. It is also the most potent in terms of inducing mobility shifts of DNA and NAs (Fig. 4.3a, 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* and is likely to contribute to the folding behavior and tertiary structure of MeCP2. 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 as structure is likely promote simultaneous interaction with multiple partners conferring considerable functional flexibility to MeCP2.

4.5.4 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-α and CTD-β.

While both the TRD (207-310) and CTD-α (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-α. In support of the former, deletion mutagenesis of the isolated TRD suggests its DNA binding region lies between residues 245-270 (data not shown). The exclusive binding of CTD-α to NAs (Fig. 4.3c), strongly suggest that this domain specifically interacts with nucleosomal histones. 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.

4.5.5 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 (78), 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 (172). 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 (16), N-CoR and Ski (81), H3 methyl transferase (173), p20, a putative Xenopus protease (71), as well as the DNA methyltransferase Dnmt1 (22). A

fluorescence recovery after photobleaching study (86) 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. 4.5) (36) is consistent with the growing *in vivo* evidence for a role of MeCP2 in stabilizing large chromatin loops (174-176). 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 (34). 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 (133) 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.

Chapter 5

DNA Binding Restricts the Intrinsic Conformational Flexibility of MeCP2 Exclusively in its Methyl DNA Binding Domain

5.1 Preface

Chapter 5 is reprinted from a paper I coauthored that first appeared in the Journal of Biological Chemistry online April 4th, 2011. The entire paper can be accessed freely at the following shortened URL - http://bit.ly/ig4A7s. My contributions to this collaborative effort were to express and purify all protein, and protein fragments analyzed. Additionally, where nucleoprotein complexes of MeCP2 and 198-1 DNA (methylated and unmethylated) were used, I methylated the DNA and reconstituted the complexes at a stoichiometry I had worked out through trial and error that provided complexes with ample DNA:MeCP2 ratio to ensure free protein was not present to distort the Hydrogen Deuterium exchange and subsequent mass-spectroscopic analyses. Please see experimental procedures; *Protein expression and purification,* and *Formation of MeCP2--DNA complexes* subsections for a precise description of my contributions to this work.

5.2 Introduction

Methyl CpG binding protein 2 (MeCP2) is named for its ability to selectively bind to methylated CpG dinucleotides (2), acting through its methyl DNA binding domain (MBD)(8). MeCP2 is a 53 kDa nuclear protein that is present in high amounts in neuronal tissues, where its stoichiometry approaches one MeCP2 per nucleosome throughout the genome (54). In addition to binding unmethylated and methylated DNA (36,63,116,120) chromatin (24,36,61,63) and RNA (65), MeCP2 interacts with many different nuclear proteins, including Sin3a (16), SUV39H1 (75), HP1 (78), DNMT1(79), Ski and N-COR (22), PU.1 (80), BRM (77), and ATRX (74). While originally hypothesized to be a specific repressor of methylated genes (16), MeCP2 is now recognized to be multifunctional, with roles in transcriptional activation *and* repression (4), RNA processing (64), and chromatin organization (5,24,54,177). Consistent with these results, MeCP2 is localized to both promoter and intergenic regions in the nuclei of neuronal cells (5). Of note, mutations located throughout the entire length of the MeCP2 amino acid sequence are associated with Rett Syndrome (RTT), a severe X-linked neurodevelopmental disorder that afflicts about one in 10,000 girls (23,82). MeCP2 dysfunction is also involved in autism spectrum disorders (127,128) and certain cancers (130,131). Hence, it is important to understand the structure/function relationships that apply to MeCP2 in both the health and disease states.

Intrinsically disordered proteins lack well-folded traditional tertiary structure over some or all of their polypeptide sequence (39,89,91,92,135). Several biophysical techniques that measure averaged solution behavior have recently documented that MeCP2 is an intrinsically disordered protein. Steady-state circular dichroism (CD) measurements revealed that MeCP2 in solution is ~60% unstructured, 35% β-strand/turn, and 5% α-helix (85,88). Sedimentation equilibrium studies showed that free MeCP2 is a

monomer over a wide concentration range and sedimentation velocity analysis yielded a sedimentation coefficient of 2.2S (85). Taken together, the analytical ultracentrifugation results demonstrated that MeCP2 has a frictional coefficient that would be expected for a random coil-like molecule. Thus, one can infer from these averaged data that considerable portions of the MeCP2 polypeptide chain lack secondary structure. However, while the atomic structure of the isolated MBD fragment has been solved alone (11) and in complex with methylated DNA (12), because of its highly disordered nature there is no fine resolution structural information available for *full length* MeCP2, either when free in solution or when bound to DNA.

Mass-spectrometry-based hydrogen/deuterium exchange (H/DX) has emerged as a powerful technique for studying the structure of monomeric intrinsically disordered proteins and their higher-order complexes at high resolution. For instance, the dramatic reduction of H/DX rates along the polypeptide backbone of amyloid forming proteins has been characterized at amino acid resolution following their assembly into amyloid (178). For full length MeCP2, H/DX has the promise to substantially extend the averaged biophysical measurements by providing fine resolution mapping data that locates regions of secondary structure throughout the protein sequence..

In the present studies, we apply H/DX to the analysis of full length MeCP2 when free in solution and when bound to unmethylated and methylated DNA. Results demonstrate that the MBD is the only domain within free MeCP2 that shows even modest protection from H/DX, and even the H/DX of the MBD is fast compared to a typical globular protein. This indicates that full length MeCP2 rapidly samples many different conformational states and tertiary structures when free in solution. We further show that binding of MeCP2 to unmethylated DNA substantially decreases the global conformational

flexibility of the MBD, while the rapid H/DX exchange elsewhere in the protein was unaffected. Thus, even with the increased H/DX protection in the MBD, full length MeCP2 remains a very intrinsically disordered protein when bound to DNA. Relative to binding to unmethylated DNA, binding to methylated DNA only slightly increases the conformational rigidity in a local region within the N-terminal portion of the MBD. Finally, we to examine the effects of several common RTT mutants on the properties of the isolated MBD, and find widely varied effects. Taken together, the H/DX experiments have yielded high resolution structural dynamics data characterizing the extreme intrinsic conformational flexibility of full length MeCP2, and also provided accompanying information about how the structure and stability of the MeCP2 MBD is affected by DNA binding and certain specific RTT mutations. For the latter case, the implications for treatment of RTT are discussed.

5.3 EXPERIMENTAL PROCEDURES

5.3.1 Protein expression and purification :

Full length human MeCP2 isoform e2 and the isolated R106W, F155S, and T158M mutant MBD polypeptides were purified using a modification of the protocol described previously (85). Recombinant proteins were expressed using the IMPACT system (New England Biolabs). Constructs were subcloned into the ptyb1 plasmid vector. MBD mutant constructs were kindly supplied by Dr. C.L. Woodcock. All purified proteins contained a vestigial sequence, EFLEGSSC, on their C-terminal ends as a result of cloning methods previously described (85). *Escherichia coli* BL21RP+ was used as host expression bacteria grown in lysogeny broth at 37℃ to an optical density of 0.5 absorbance units, induced with 0.4 mM IPTG and cooled to 18℃ for 2-3 hours prior to harvest. Expression hosts were pelleted in an Avanti J-26 XPI preparative centrifuge

(Beckman Coulter) in a JLA 8.100 rotor at 5,000 g for 15 minutes. Pellets were resuspended in wash buffer (25 mM TRIS pH 7.5/100 mM NaCl) and repelleted under the same conditions. Clean pellets were resuspended in column buffer (25 mM Tris–HCl pH 8.0, 500 mM NaCl) supplemented with 0.1% Triton X-100, 0.2 mM PMSF, and Protease Inhibitor Cocktail Set II (Calbiochem), followed by two rounds of sonication, 90s each, using a Branson Sonifier 450 with a large tip at 50% duty cycle and a power output of 7. Lysate was transferred to Oakridge tubes and spun at 21,000 *g* for 25 min in the preparative centrifuge in a JA-17 rotor (Beckman). The supernatant was bound overnight to chitin beads (New England Biolabs) previously equilibrated in column buffer. Chitin beads were washed with five column volumes column buffer, decanted and washed with an equal volume column buffer supplemented with NaCl to 900 mM NaCl final concentration to wash off errant bacterial DNA left from sonication. Chitin beads were washed with an additional 5 column volumes 500 mM NaCl Column buffer. Chitin binding protein-MeCP2 chimeras were cleaved on the column (40)(179)(179)(179)[180][180][179](*179*)(*179*)(*177*)(*125*)(125). Column buffer supplemented with 50 mM DTT was passed through the column such that 1 cm buffer remained between the top of the column bed and the meniscus in a 10 cm Kontes FlexColumn (Fischer) and left for 48-72 hours for complete cleavage. Protein was eluted from the chitin column with column buffer, diluted from 500 mM to 300 mM NaCl and loaded onto a HiTrap Heparin HP column (GE healthcare). Proteins were eluted from the heparin column via step gradient from 300 mM NaCl to 1 M NaCl buffer using 100 mM NaCl steps in 25 mM Tris, pH 7.5, 10% glycerol background buffer. Peak fractions were pooled and dialyzed into 10 mM Tris pH 7.5.

5.3.2 Formation of MeCP2-DNA complexes:

Unmethylated and methylated 198 bp DNA fragments derived from the sea urchin 5S rRNA gene were prepared as described (85). This DNA has 12 methylatable CpGs. Purified MeCP2 (155 µl; 0.80 mg/ml) was added to either an unmethylated or methylated 198 bp DNA fragment (430 µl; 0.2 mg/ml) in a total of 585 µl of 10 mM Tris (pH 7.5), 10 mM NaCl buffer. MeCP2 minimally binds 11 bp of DNA (61). Under these conditions, unmethylated and methylated DNA binding sites were present in molar excess over MeCP2, and DNA concentrations always were above the K_d (120) such that H/DX was being measured under saturated binding conditions and dual population (i.e. bound and unbound to DNA) affects were avoided.

5.3.3 H/DX reactions

A total of 10 µl of each sample (4-10 µg full length MeCP2 [alone or in a complex with the indicated DNA fragment] or 0.4-3 µg MBD fragments [the wild type version or indicated RTT-associated mutation]) was mixed with 30 µl of D_2O containing 10 mM Tris (pD 7.2), 10 mM NaCl and incubated at the indicated temperature. At each indicated timepoint, the H/DX samples were added to vials containing 60 µl of a quench solution (0.8 M guanidinium-HCl, 0.8% formic acid, 10% glycerol) at 0^{o}C and immediately frozen in liquid N_2. The samples were stored at -80^{o}C until analysis by MS.

5.3.4 Protein fragmentation and MS:

H/DX samples were individually melted at 0^{o}C, then injected (100 µl) and pumped through tandem immobilized protease (pepsin and fungal protease XIII; both from Sigma) columns (50 µl/min, 1 X 20 mm [16 µl] columns of each protease coupled to Poros 20 AL support [Applied Biosystems]). Protease generated fragments were collected onto a C18 HPLC trap column (2.5 X 0.5 mm). Peptides were eluted into and through an analytical

C18 HPLC column (0.3 X 75 mm) by an acetonitrile gradient (12-55% B; 6 µl/min; solvent A, 0.1 % formic acid; solvent B, 0.1% formic acid, 99.8% acetonitrile) and the effluent was directed to the mass spectrometer (LTQ Orbitrap XL, ThermoFisher Scientific). The SEQUEST software program (ThermoFisher Scientific) was used to identify the likely sequence of the parent peptides using nondeuterated samples via tandem MS.

5.3.5 H/DX analysis

DXMS software (Sierra Analytics) was used for searching the MS1 data from H/DX samples and calculating the centroid of the isotopic envelope of each peptide using a general scheme that is described elsewhere (180,181). Peptides that score highly in the DXMS program were checked for matching of calculated versus known mass, charge state, and the retention time of the peptide on the C18 column, and peptides that satisfied these criteria were selected for further analysis. The level of H/DX occurring at each timepoint is expressed as either the number of deuterons or the percentage of exchange within each peptide. In each case, corrections for loss of deuterium label by individual peptides during H/DX-MS analysis were made through measurement of loss of deuterium from reference samples that had been deuterated under denaturing conditions as described elsewhere (182,183). For generating deuterium exchange profiles (181,184), maps of rate-classes along the polypeptide sequence was assembled using the H/DX data, employing a strategy in which the (generally smaller) peptides containing one or two rate classes were first placed in primary sequence register, followed by placement of peptides with two, and then three rate classes, in a manner that required that placement of the rate classes of the amides in each peptide conform to the preceding placements.

5.4 RESULTS

We first used H/DX to analyze the polypeptide backbone dynamics of full length MeCP2 when free in solution (Fig. 5.1). Because MeCP2 binds strongly to unmethylated and methylated DNA (63), we also determined the dynamics of H/DX when MeCP2 was bound to an unmethylated 198 bp DNA fragment derived from the sea urchin 5S rRNA gene (85)(Figs. 1-3 and supplemental Figs. 5.1-3), and to the methylated version of the same DNA fragment (Figs. 5 and 6 and supplemental Figs. 5.4-6). In Figures 5.2-6 the H/DX experiments characterize full length MeCP2, focusing on the MBD region of the protein.

5.4.1 H/DX demonstrates the extreme conformational plasticity of full length MeCP2 when free and bound to DNA:

The H/DX approach consisted of incubation of the samples in heavy water (D_2O) at 4℃ to exchange deuterium with the amide protons alon g the polypeptide backbone of MeCP2. At time points spanning 10^1 s to 10^4 s, the exchange reactions were quenched, MeCP2 fragmented by proteolysis, and deuterium incorporation measured by mass spectrometry. Protection from H/DX in native proteins or protein/DNA complexes is expected in those regions that are folded into stable secondary structures. This is due to the fact that measurable amide protons are hydrogen bonded and must transiently lose secondary structure, locally or globally, in order for exchange to occur (185).

In our experiments, peptides covering 87% (free MeCP2) or 73% (MeCP2 bound to unmethylated DNA) of the entire length of the MeCP2 sequence were initially identified by MS/MS and then successfully monitored over the entire time course of exchange.

Areas where coverage was not achieved tend to be unstructured (for example, compare residues ~160-200 and ~250-280 in the -/+ DNA maps in Fig. 5.1). In the absence of DNA, H/DX along nearly the entire length of MeCP2 was rapid, with 64% of all peptides measuring >90% of complete exchange by the first time point (10 s). The sole region that exhibited measurable protection from H/DX was found in the peptides spanning amino acids ~90-160, closely overlapping the MBD of MeCP2 (which encompasses residues 78-162). These results indicate that when free in solution full-length MeCP2 has only one region of even marginally stable secondary structure, corresponding to the MBD. Moreover, the H/DX of even the MBD was faster than would be expected for a typical globular protein in which most of the amide backbone was engaged in hydrogen bonding. Thus, the entire polypeptide chain of free full length MeCP2 is intrinsically disordered.

When bound to unmethylated DNA under saturating conditions, the MBD-derived peptides were much slower to exchange, while there was no change observed in the very rapid H/DX at any other location in the MeCP2 protein sequence (Fig. 5.1 and supplemental Fig. 5.1). Thus, although the stability of the MBD was greatly increased when MeCP2 was bound to unmethylated DNA, the remainder of the polypeptide chain stayed extremely disordered. The MBD only accounts for 17% of the 486 residue full length protein. Thus, the data in Figure 5.1 demonstrate that MeCP2 remains very intrinsically disordered even when bound to DNA. Given that the MBD was the sole region to show evidence of secondary structure formation, we focused our next experiments on high resolution H/DX characterization of the dynamics of the MBD in full length MeCP2 when free and bound to DNA.

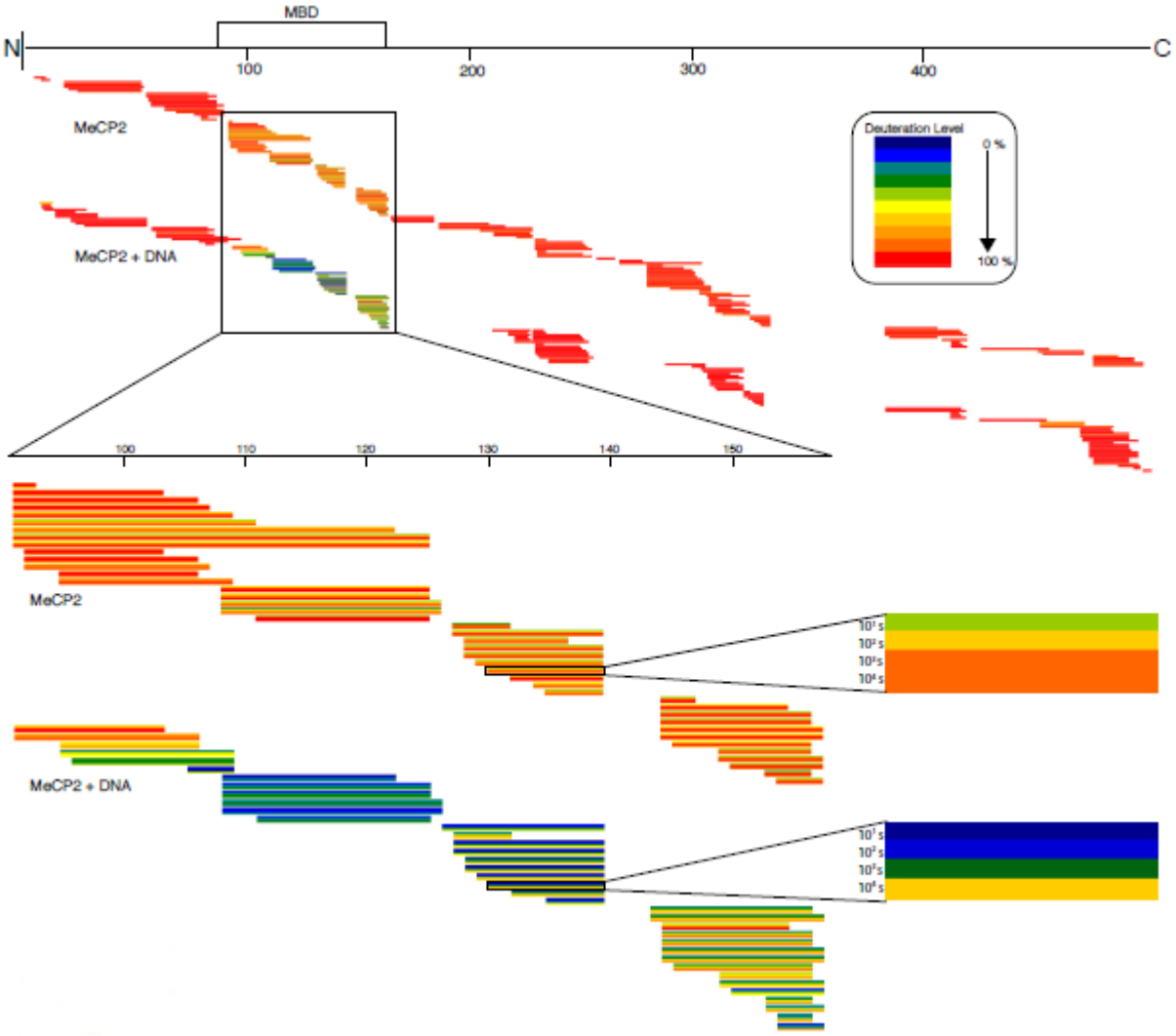

FIGURE 5.1. Protection from H/DX before and after DNA binding is detectable only within the MBD of MeCP2. Each horizontal bar represents an individual peptide derived from MeCP2 when free in solution (MeCP2), or bound to unmethylated DNA (MeCP2 + DNA). They are color coded for the percent deuteration at 4℃ at each timepoint (10^1, 10^2, 10^3, and 10^4 s), as represented by an individual stripe within each bar. The MBD is enlarged on the lower left, with an individual peptide enlarged further and labeled to illustrate the inclusion of each time point for every peptide in the data set.

5.4.2 Rapid sampling of partially unfolded states occurs within the MBD when MeCP2 is free in solution:

Since the MBD region was particularly well represented with many partially overlapping peptides, we could employ a strategy with three rate classes (0-10 s, 10-100 s, >100 s) to finely map H/DX exchange. Similar exchange profile strategies that utilize partially overlapping peptide information have been especially informative when combined with crystal structures (184,186). In the absence of DNA, only small regions (i.e. stretches of 1-4 amino acids.; blue positions in Fig. 5.2A and B) exchange amide protons for deuterons slower than 100 s. Thus, the stability at many locations within the MBD is the same or only marginally greater than the remainder of the MeCP2 polypeptide chain when the protein is free in solution. Further, when the H/DX profile of the MBD of free MeCP2 is mapped (Fig. 5.2C) onto the crystal structure of the MBD (12) it is clear that even in the most rigid portion of each β-strand and the α-helix in this region must rapidly sample partially unfolded and partially folded conformations in order to allow for the relatively rapid H/DX (i.e. nearly complete exchange by 1000 s at every location) that is observed in the MBD of the free full length protein (Fig. 5.1).

5.4.3 Rapid sampling of partially unfolded states within the MBD is restricted when MeCP2 binds to unmethylated DNA:

MeCP2, when bound to unmethylated DNA under saturating conditions, has many more residues within the MBD that are substantially protected from H/DX (Fig. 5.2D). On the level of individual peptides, while exchange is complete by 10^4 s within the MBD of the free protein, when MeCP2 is bound to DNA, several MBD residues remain protected from H/DX at the same 10^4 s time point (Fig. 5.3 and supplemental Figs. 5.2 and 5.3).

When we examined the H/DX protection imposed upon the MeCP2 MBD due to unmethylated DNA binding in closer detail (Fig. 5.2E), we noted that the H/DX profile closely matched the known secondary structural elements (11,12) of the isolated MBD (Fig. 5.2F, blue residues labeling the most protected regions of MeCP2 when bound to DNA). In the crystal structure of the MBD bound to methylated DNA (12), the majority of the surface of the isolated MBD is exposed to solvent, not buried with the surface of the bound DNA. The amide protons of R111, D121, and R133 (labeled red in Fig. 5.2, panels G and H) are the only positions predicted from the crystal structure (12) to become protected by DNA (or water) contacts, which cannot explain the extent of protection of H/DX we observed upon unmethylated DNA binding by MeCP2 (compare protection in Fig. 5.2, panels B and E). Our scheme for generating H/DX profiles of the MBD allowed for resolution down to a small number of amino acids (see EXPERIMENTAL PROCEDURES). As described above, there is a strong correlation between the H/DX of the MBD of MeCP2 when bound to unmethylated DNA and the secondary structure of the MeCP2-methylated DNA complex (Fig. 5.2E and 5.2F). Our H/DX profiles (Figs. 2B and 2E) were generated without bias toward any structural information, but if we include predicted protection (Figs. 2G and 2H) in positioning the measured H/DX on each peptide, the final profile (Fig. 5.2I) only differs from the unbiased profile (Fig. 5.2E) at seven positions. This indicates that the slowed backbone exchange upon unmethylated DNA binding corresponds to global stabilization of the same folded secondary structural elements observed in the structural studies of the MBD.

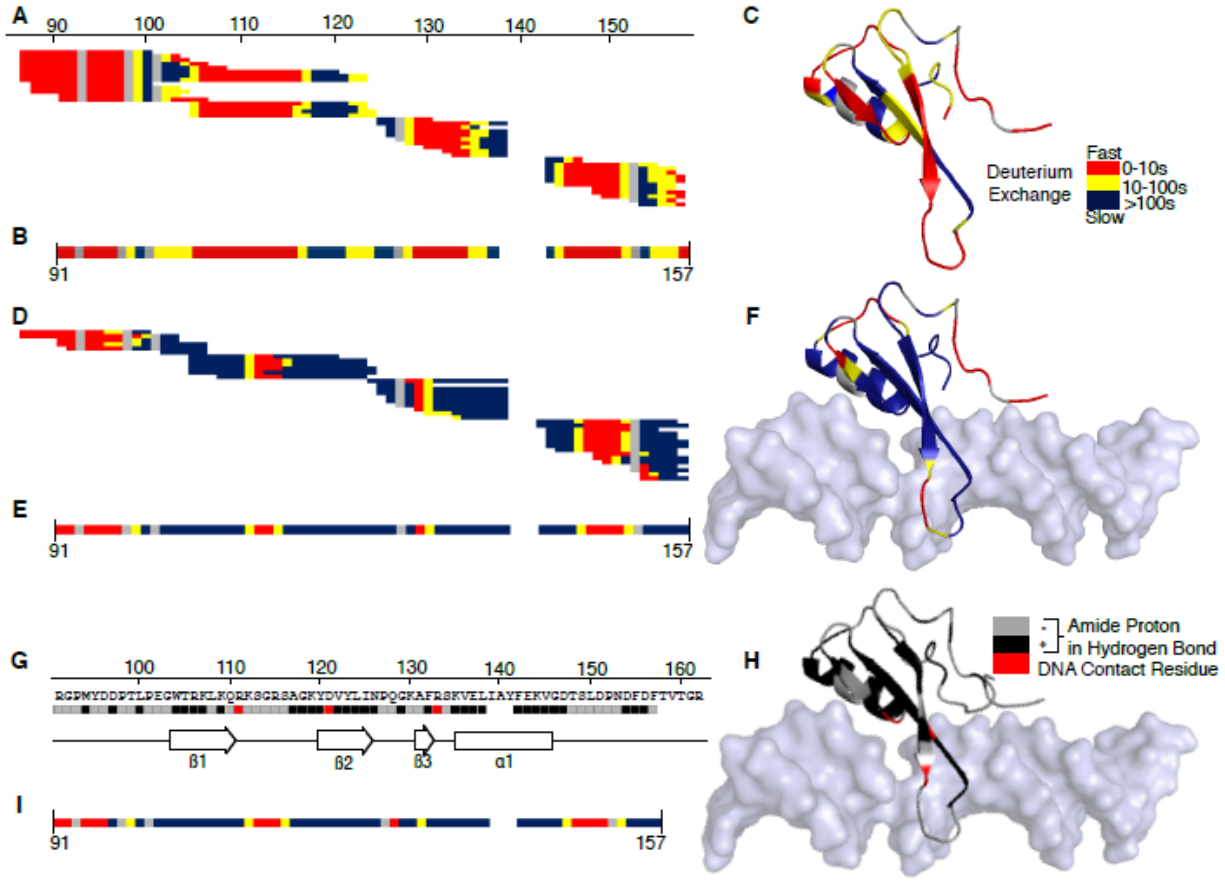

FIGURE 5.2. **Stabilization of the MBD upon unmethylated DNA binding results from restricted sampling of intrinsic unfolding rates.** *A*, Deuterium exchange profile maps for each peptide of the MBD of free MeCP2 in solution are color coded as described in EXPERIMENTAL PROCEDURES. *B*, The consensus exchange rate at each position (spanning residues 91 [the N-terminal residue of the MBD clear in the crystal structure (37) to residue 157 [the C-terminal residue where our high-density coverage of the MBD ends]) from panel *A* is shown in a single linear representation. *C*, The consensus map from panel *B* overlaid onto the crystal structure of the MBD (the DNA is not shown; PDB ID 3C2I). *D*, The exchanged deuterons for each MBD peptide bound to unmethylated DNA, mapped as in panel *A*. *E*, The consensus map of panel *D*. *F*, The consensus map from panel *D* overlaid onto the crystal structure of the MBD. *G*, The MBD sequence with the location of residues where the amide proton is engaged in a hydrogen bond in the crystal structure (black), as well as the position of the reported residues (37) with direct DNA contacts. *H*, Mapping of the residue labeling from panel *G* onto the MBD crystal structure. *I*, A structure-biased consensus map using the same number of slow (blue), medium (yellow), and fast (red) exchanging amino acid position as in panel *E*. Note how closely the slow (blue) residues in the final consensus map match amide proton protection predicted by the stable structure obtained by others by crystallography (37). This indicates that the increase in protection upon binding to unmethylated DNA result from stabilizing the overall fold of the MBD rather than increasing protection locally at its relatively small DNA binding surface.

To summarize, our data indicate that the H/DX protection of the MBD upon binding of MeCP2 to unmethylated DNA occurs mainly via domain-wide restriction of rapid sampling of its secondary structural elements, not by inaccessibility to solvent due to steric hindrance imposed by DNA. That is, the increase in H/DX protection upon binding unmethylated DNA is due to stabilization of the overall fold of the MBD rather than increased local protection at its relatively small DNA binding surface. In addition, there are fundamental changes in the flexibility of the MBD structure that occur upon DNA binding that occur even when the DNA is unmethylated.

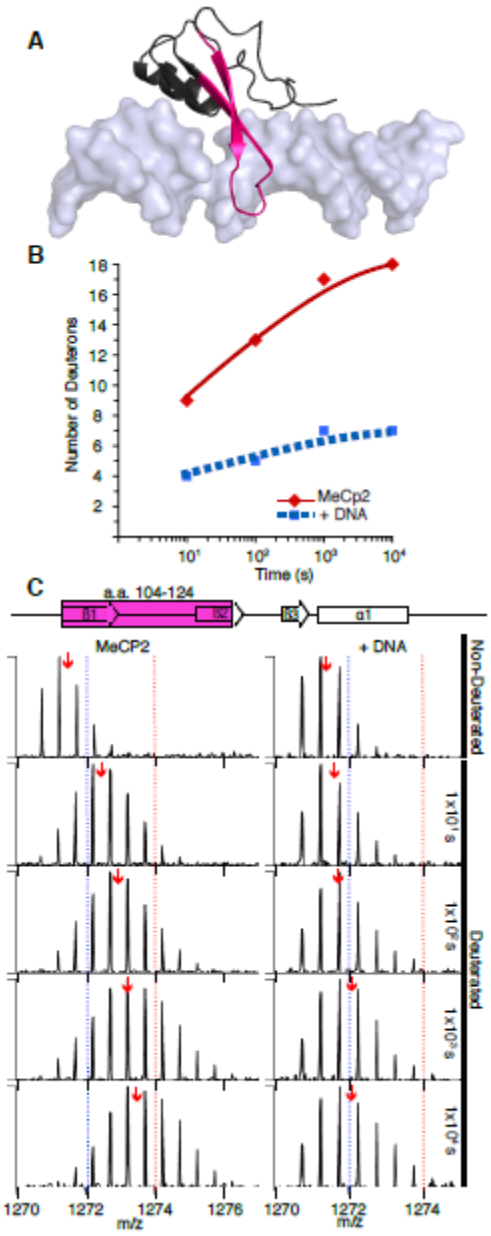

FIGURE 5.3. The MBD is protected from H/DX upon binding unmethylated DNA. *A*, The location on the crystal structure (PDB ID 3C2I) of a representative MBD peptide (amino acids 104-124). *B*, Comparison of H/DX for this peptide from free MeCP2 and MeCP2 bound to DNA. *C*, Side-by-side analysis of raw data for the peptide from panel *B* from free MeCP2 (left) or MeCP2 bound to unmethylated DNA (right). Dotted blue and red lines are guideposts to highlight the differences in m/z shifts reflecting the magnitude of H/DX. Red arrows mark the location of the centroid value of the peptide in each case.

5.4.4 Binding to methylated DNA locally restricts conformational flexibility within the MBD:

Given that MeCP2 only has a modest preference for methylated DNA over unmethylated DNA (120), we were curious of the extent to which binding to methylated DNA would correspond to increased H/DX protection within the MBD (or elsewhere in the protein). Our initial experiments (Figs. 1-3) were performed at 4^{o}C to slow the chemical exchange rate and maximize our ability to see any protection at all for free MeCP2. Since many residues were completely protected from exchange at 10^4 s when MeCP2 was bound to unmethylated DNA at 4^{o}C (Figs. 1 and 3 and supplemental Figs. 2 and 3), we predicted that increasing the temperature would be necessary to perform a reasonable time course that spanned the exchange rates of most of the protected portions of the MBD in the case of DNA (either unmethylated or methylated)-bound MeCP2. To determine if this was the case, we first analyzed MeCP2 at a single time point at three varied temperatures (4^{o}C, 23^{o}C, and 37^{o}C) and measured H/DX on peptides within the MBD (Fig. 5.4). This experiment indicated that increased temperature (most likely corresponding to predictable higher chemical rates of H/DX (185)), increased the amount of H/DX within the MBD. Thus, we chose to perform an expanded time course (1×10^1, 1×10^2, 1×10^3, 1×10^4, 2.5×10^4, and 1×10^5 s time points) of H/DX at 37^{o}C. With peptide coverage spanning much of the full length of MeCP2 in all three cases (MeCP2 alone, +unmethylated DNA, and +methylated DNA; supplemental Fig. 5.4), the only observed region with any measurable protection from H/DX once again was within the MBD (Fig. 5.5A). In the absence of DNA, all MBD peptides were >90% deuterated at 10 s, with substantial slowing (most greater than 10^3-fold) of exchange for all MBD peptides when MeCP2 was bound to either unmethylated or methylated DNA (Fig. 5.5A).

In the MeCP2-methylated DNA crystal complex (12), structured waters connect the methyl groups on the DNA with specific MBD residues (R111, D121, Y123, and R133) via hydrogen bonding. To test how the conformational flexibility of the MBD is affected by methylated DNA-specific contacts, we carefully examined representative peptides for which we had a comprehensive MS data set for all samples and time points. Peptides spanning the N-terminal ~half of the MBD, i.e., amino acids 90-106 (Fig. 5.5B and supplemental Fig. 5.5) and amino acids 104-123 (Fig. 5.5C and 5.6A-D), showed clear reduction in H/DX rates when MeCP2 was bound to methylated versus unmethylated DNA, while peptides spanning the C-terminal ~half of the MBD, i.e., amino acids. 124-138 (Fig. 5.5D and supplemental Fig. 5.6) and amino acids 142-157 (Figs. 5E and 6A,E-G), showed little or no alteration. The N-terminal peptides showed additional protection exceeding that which would be expected if the residues only were involved in hydrogen bonding with the structured waters. For instance, none of the residues in the amino acids 90-106 peptide, and only three (R111, D121, and Y123) in the amino acids 104-123 peptide, are directly bonded to the structured waters implicated in methylated DNA specificity, yet several additional amide protons exhibit slowed exchange that was dependent upon binding to methylated DNA (Fig. 5.6B and supplemental Fig. 5.5). On the other hand, the third residue (R133) that participates in a hydrogen bond with one of the structured waters appeared to have almost no influence on H/DX exchange within the MBD (see a representative peptide corresponding to amino acids 124-138; supplemental Fig. 5.6). These results indicate that: 1) relative to binding to unmethylated DNA, binding to methylated DNA affects H/DX locally (only the N-terminal ~half of the MBD) and not globally throughout the entire domain, 2) the additional protection of the N-terminal MBD peptides is consistent with additional local rigidity, not solely backbone protection resulting from the residues directly hydrogen bonded to the structured waters

that are coordinated with the methyl groups on the DNA, and 3) the local added protection from H/DX upon binding to methylated DNA versus unmethylated DNA is not due to differences in binding kinetics with the MBD since the C-terminal peptides (spanning amino acids ~124-157) exchange with indistinguishable slow kinetics when MeCP2 is bound to either DNA fragment compared to the rapid H/DX when MeCP2 is free in solution.

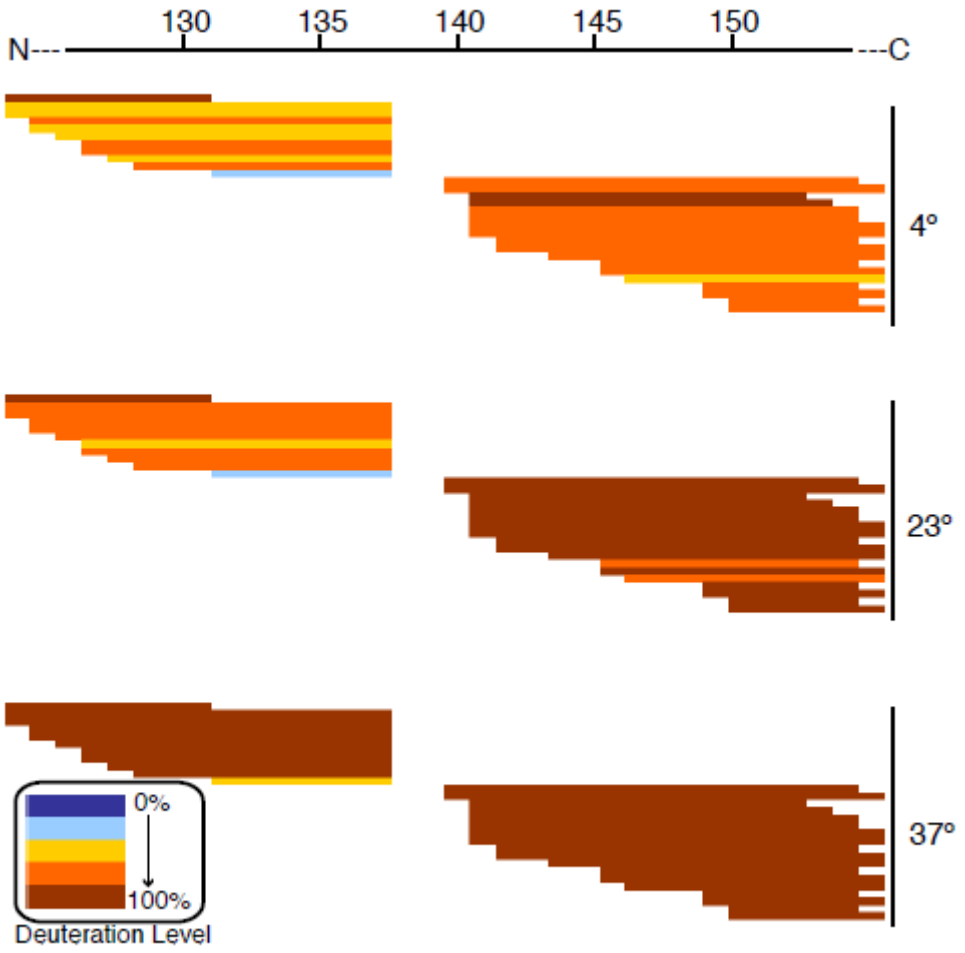

FIGURE 5.4. Temperature dependency of H/DX within the MeCP2 MBD. Peptides from a representative region within the MBD of free MeCP2 (i.e. unbound to DNA) are shown as single horizontal bars. Color-coding corresponds to the level of H/DX after 10^4 s incubation at each indicated temperature.

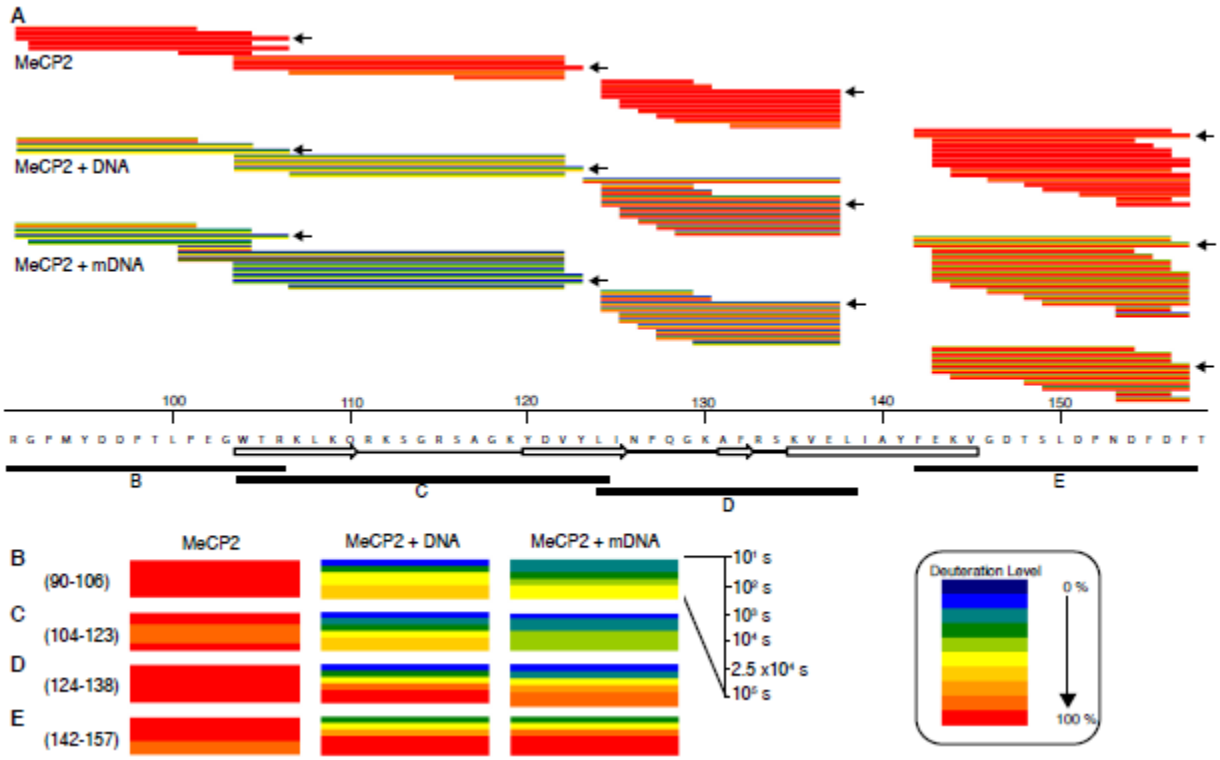

FIGURE 5.5. Binding to methylated DNA further increases protection from H/DX within the N-terminal but not C-terminal portion of the MBD. *A*, Peptides from the MBD are shown for free MeCP2 (top), MeCP2 bound to unmethylated DNA (middle), and MeCP2 bound to methylated DNA (bottom). Data are presented as in Fig. 5.1. *B-D*, Representative peptides that span the MBD with the indicated amino acid positions are enlarged and compared. The residues boxed in red indicate positions that directly contact DNA and/or form bonds with water molecules that are coordinated with the methyl groups on methylated DNA.

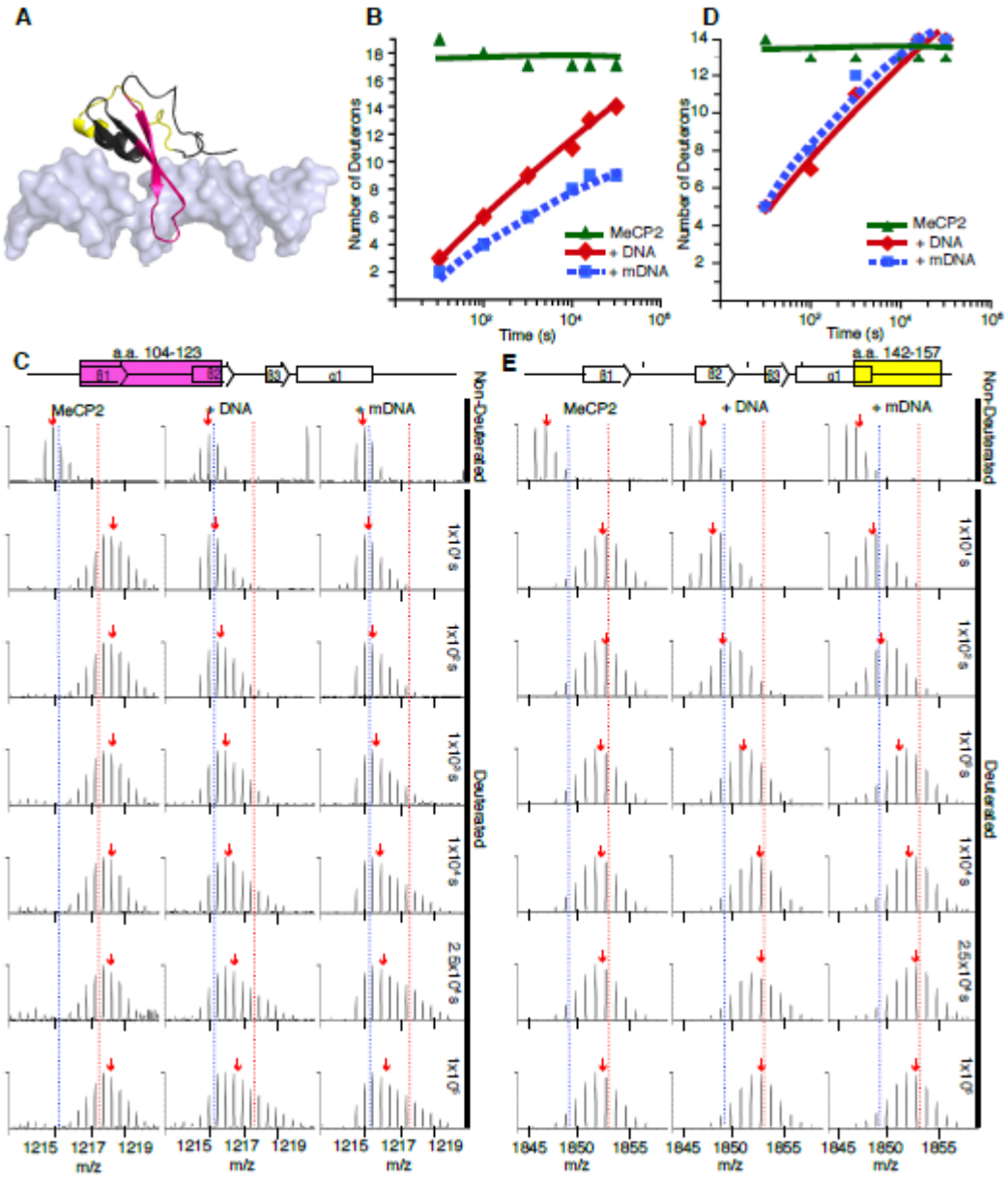

FIGURE 5.6. **Representative MBD peptides showing protection from H/DX when MeCP2 binds to unmethylated DNA and methylated DNA.** *A*, The location on the crystal structure of the MBD bound to methylated DNA (PDB ID 3C2I) of representative MBD peptides (amino acids 104-123 [pink] and 142-157 [yellow]). *B* and *D*, Comparison of H/DX for peptides amino acids 104-123 (panel *B*) and amino acids 142-157 (panel *D*) from free MeCP2 and MeCP2 bound to unmethylated DNA and methylated DNA. *C* and *E*, Side-by-side analysis of raw data for the indicated peptides from free MeCP2 (left), MeCP2 bound to unmethylated DNA (middle), or MeCP2 bound to methylated DNA (right). Labeling is as in Figure 5.3C.

5.4.5 RTT Mutations in the MBD Have Varied Effects on Local and Domain-wide Flexibility:

Mutations within the MBD are common in RTT patients, with many of the most frequent missense mutations (such as R106W and T158M) predicted to disrupt stabilizing interactions within the domain (12). For instance, the side-chain of R106 forms hydrogen bonds with the polypeptide backbone at positions T158 and Val159 (12). The extent to which these hydrogen bonds affect dynamic protein behavior is not easily discernable from static structures, but we reasoned that H/DX would provide a fine resolution dynamic measurement to directly assess the local and domain-wide effects of RTT mutations in the MBD (Fig. 5.7). For this analysis, H/DX experiments were performed at X℃ with isolated MBD peptides bearing specific RTT mutations.

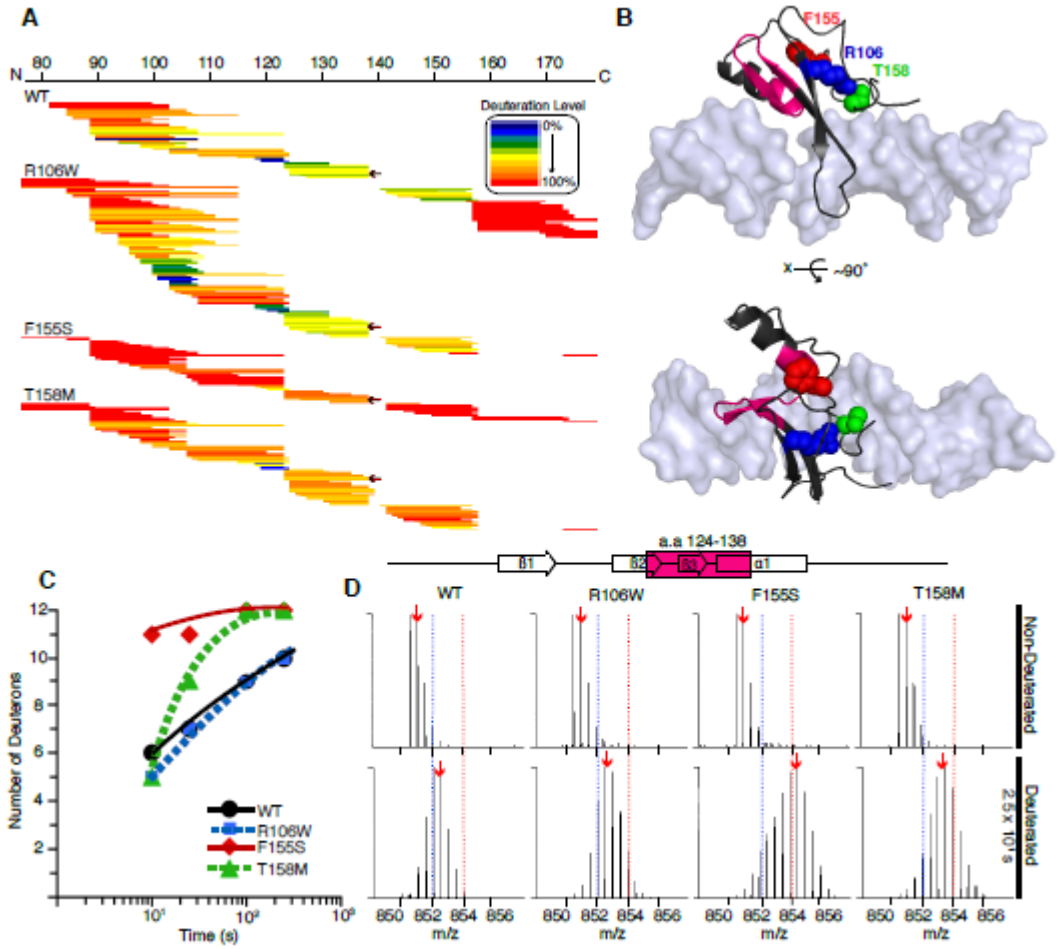

FIGURE 5.7. Effects of Rett Syndrome-associated missense mutations on the conformational flexibility of the isolated MBD. *A*, The level of H/DX at 2.5 × 10^1 s are color-coded as indicated on horizontal bars representing peptides derived from the wild type (WT) MBD, and the MBDs harboring the indicated point mutations. Data from all time points monitored (1 × 10^1, 2.5 × 10^1, 1 × 10^2, and 2.5 × 10^2 s) are shown in Supplemental Figs. 7-10. The black arrows indicate the representative peptide (amino acids 124-138, compared panels *C* and *D*, below). *B*, The location on the MBD crystal structure (PDB ID 3C2I) of the peptide (amino acids 124-138) highlighted in panels *C* and *D*. The positions of the amino acid side-chains in the RTT MBD mutants that were investigated are shown as colored spheres. *C*, Comparison of H/DX for a peptide (amino acids 124-138) from the MBD. *D*, Side-by-side analysis of raw data for the WT MBD and the MBDs harboring the indicated mutations. Labeling is as in Fig. 5.3C.

Our results indicate that the R106W MBD mutant had generally similar H/DX behavior as the wild type (WT) MBD (Fig. 5.7A, C, and D), with increased H/DX protection in the β1 peptide and only minor acceleration of exchange occurring in peptides spanning amino acids ~140-155 (Fig. 5.7A). By contrast, the F155S mutation, which removes hydrophobicity from within the MBD protein fold (see the position of F155 in Fig. 5.7B) and has been shown previously to greatly destabilize the MBD by fluorescence anisotropy measurements and melting experiments (88), forces the entire MBD to sample unfolded conformations at rates that are several orders of magnitude higher than for the protected residues in the wild type MBD. Consequently, for the F155S mutant, even at the earliest time point (10 s) there is complete exchange throughout the domain (Fig. 5.7C and supplemental Fig. 5.9). The side-chain of T158 is predicted from the co-crystal structure to be important for the integrity of the so-called Asx-ST motif (located at the C-terminal end of the MBD), in part by making hydrogen bonds with the polypeptide backbone at positions G161 and R162 (12). T158M is the site of the most common point mutation in RTT, accounting for ~9% of all cases (Rettbase; http://mecp2.chw.edu.au/). Despite residing in the MBD C-terminus, the T158M mutation surprisingly only affects conformational dynamics spanning portions well into the N-terminal folded portion of the MBD, including amino acids ~125-155 (refer to Fig. 5.7D and the representative peptide [amino acids 124-138] that fits into the major grove of DNA [Fig. 5.7B]). The H/DX behavior of the RTT MBD mutants indicates the variable extent to which these common mutations affect the local and global conformational dynamics of the MBD of MeCP2.

5.5 DISCUSSION

5.5.1 Full length MeCP2 structure:

The structure of full length, wild type MeCP2 is not consistent with a conventional protein comprised of stable, folded domains. Based on averaged biophysical measurements the conformation of full length MeCP2 is very disordered, e.g., deconvolution of CD data yields 60-65% unstructured content and the protein sediments as if it were a random coil (85). Prediction algorithms (PONDR-VXT, FoldIndex) also yield estimates of ~65% disorder, with short regions of predicted order interspersed throughout an otherwise intrinsically disordered polypeptide chain (85,96). MeCP2 is composed of six biochemically defined domains: the N-terminal domain (NTD), the MBD, the intervening domain (ID), the transcriptional repression domain (TRD), and the C-terminal domain alpha and beta (CTDα and β) (Fig. 5.8;(85,96)). CD analysis of the isolated MeCP2 domains indicates that each has between 60-80% unstructured content with the exception of the MBD, and even the MBD is ~40% unstructured (47; Fig. 5.8). While the steady-state CD experiments have documented the extensively unstructured nature of full length MeCP2, they also have raised an important question: Is the 35-40% averaged secondary structure content calculated to be present in MeCP2 and its domains in the form of stable three-dimensional structures, or does it result from an ensemble of folded structures that each rapidly samples unfolded, random-coil state(s) during the course of the steady-state measurements? Our H/DX experiments indicate the latter possibility is correct. We obtained nearly total peptide coverage of the MeCP2 polypeptide chain in our experiments, and all domains of MeCP2 except the MBD underwent essentially complete H/DX at the earliest measureable time point (Fig. 5.1). While H/DX was slower in the MBD, in the absence of DNA binding it still was faster than the H/DX of a typical globular protein. Accordingly, we envision that the full length MeCP2 polypeptide chain rapidly samples many different secondary structures, and

equilibrates between multiple tertiary structures, even when bound to DNA. In this regard, various domain-domain interactions in *cis* and *trans* have been detected previously (88,96), consistent with the conformational malleability of full length MeCP2 observed by H/DX.

What are the functional advantages of the extreme conformational flexibility and fast conformational sampling exhibited by MeCP2? This is a key question, as it is relevant to the subject of intrinsically disordered proteins in general. Given that intrinsically disordered proteins often acquire secondary structure in conjunction with macromolecular interactions (171), we propose that enhanced conformational flexibility and rapid structural sampling: 1) provide the biochemical basis of MeCP2 multifunctionality by allowing a 53 kDa monomeric protein to bind to so many different macromolecular partners (16,22,74,75,77,78,80,173) facilitate the formation of higher order macromolecular complexes involved in MeCP2 function that themselves are conformationally malleable (see (117)).

5.5.2 MeCP2-DNA interactions

Very few proteins have been investigated where H/DX has been used to probe the conformational rigidity imposed by binding to DNA. One case is the Lac repressor, which is capable of binding to non-specific DNA sequences, but binds specifically to its operator sequence with $>10^6$-fold higher affinity (187). When bound to DNA nonspecifically, the conformational flexibility of the folded portion of the Lac repressor DNA binding domain is unaffected as measured by H/DX (188). Only upon forming the base specific hydrogen bonds with its operator sequence (189), is there imposed a substantial (10^5-fold) increase in conformational rigidity throughout the Lac repressor DNA binding domain (188). Another example is the rigidity imposed upon core histones

during nucleosome assembly. Sub-nucleosomal histone complexes, upon assembly into nucleosomes, undergo >10^3-fold protection from H/DX, spanning the globular domains of each of the core histones (180). For both Lac repressor and histone H/DX protection upon DNA binding, the effects appear to be global throughout domains (i.e. not local at the points directly contacting DNA). Moreover, the H/DX data for these proteins cannot be explained generally by solvent inaccessibility in the bound state, but rather by inflexibility gained in secondary structural elements upon either specific (Lac repressor) or non-specific (histone) DNA binding.

In the case of MeCP2, we found that global, domain-wide inflexibility is conferred to the MBD upon non-specific binding to a 198 bp unmethylated DNA fragment. Moreover, binding to the methylated version of the same DNA (containing 12 CpG sites) subsequently leads to additional local protection in the MBD. These results are consistent with previous steady state CD measurements showing increased secondary structure due to DNA binding (88). The global inflexibility gained upon unmethylated DNA binding is substantial—at least 1000-fold at many locations throughout the MBD (Figs 1 and 3 and supplemental Fig. 5.2). Importantly, a clear feature upon comparing the effects on MBD polypeptide backbone dynamics resulting from binding unmethylated or methylated DNA is not the differences in H/DX, but the striking similarities. That is, our H/DX experiments strongly suggest that the MBD folds and recognizes double-stranded DNA in a manner that is very similar when it is bound to either unmethylated or methylated DNA. There is somewhat more protection in the N-terminal half of the MBD when bound to methylated DNA, but no measurable differences in the C-terminal half (Figs. 5 and 6 and supplementary Figs. 5 and 6). Thus, we can conclude that in addition to the changes imparted by unmethylated DNA binding, the methyl groups coordinate waters with the MBD in a manner that increases the affinity of the DNA interaction

several-fold and increases backbone rigidity in the N-terminal half of the domain, which contains three of the four residues that interact with the structured waters. The NMR solution structures (11) and the crystal structure (12) provide clear views of a favored folded state of the molecule, but in order to understand the dynamics of unfolding/refolding of the MBD, a measurement that is capable of characterizing dynamic protein behavior, such as the H/DX approach described here, needed to be employed. In this regard, CD analysis has showed that MeCP2 and many of its isolated domains gain secondary structure content when bound to unmethylated and methylated DNA (88,96). These averaged measurements are likely assaying the same global and local decreases in MBD flexibility observed by H/DX.

Previous studies have demonstrated that the isolated ID, TRD, and CTDα domains of MeCP2 each are capable of binding unmethylated DNA in native gel mobility shift assays (96). However, as discussed above, the H/DX experiments find no evidence of protection outside the MBD when MeCP2 is bound to unmethylated DNA. Thus, the ID, TRD, and CTDα may continue to sample multiple conformational states even as they interact non-specifically with DNA. It is also possible that the incomplete regions of the H/DX profiles seen when MeCP2 is bound to unmethylated DNA (Fig. 5.1) result from peptides that remained bound to DNA even after denaturation and digestion. MeCP2 is an abundant chromatin associated protein *in vivo* (54,86). In terms of chromatin we speculate that the MBD binds to linker DNA (36,61,120), while other DNA binding site(s) are used to interact with nucleosomal DNA and help MeCP2 physically envelop the nucleosome (190). This in turn promotes compaction of MeCP2 bound chromatin fibers (24,61). In sum, our H/DX experiments allowed us to access site-specific backbone dynamic information that complements the earlier structural studies and significantly

extends our understanding of the physical and chemical mechanisms of DNA binding of this highly abundant and functionally important component of neuronal chromatin.

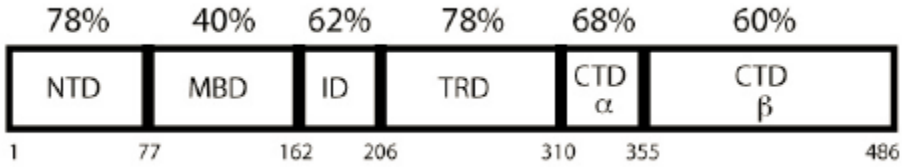

FIGURE 5.8. Diagram of MeCP2 domain organization with the percentage of each domain that is unstructured, as measured by CD , indicated for each domain.

5.5.3 MBD RTT mutations:

The H/DX approach has yielded novel information about the flexibility of several common RTT MBD mutants that cannot be discerned from the static co-crystal structure or averaged solution measurements. All three mutations (R106W, F155S, and T158M) that we chose to study replace side-chains that do not directly contact DNA (or waters coordinated by the methyl groups on methylated DNA), but rather were predicted from static models (11,12) to be important for the structural integrity of the MBD. For example, on the basis of direct hydrogen bonding seen in the crystal structure, the T158 residue was proposed to stabilize the Asx-ST motif in the extreme C-terminal region of the MBD along with residue R106 in the N-terminal portion of the MBD, yet with the T158M mutant the H/DX clearly showed increased flexibility of peptides that span the secondary structural elements (β1 and β2 strands, and α2 helix) that lie between the MBD termini (Fig. 5.7). The R106W mutant is predicted to be disruptive of the Asx-ST motif (12) and potentially disruptive to the integrity of the entire MBD. However, for this mutant the H/DX was virtually identical to the wild type MBD, clearly not generating instability that leads to a higher frequency of sampling an unfolded state that is transmitted throughout the domain. Ultimately, the H/DX data suggest that interactions between the N-terminus and C-terminus of the MBD that are likely to be disrupted by either the T158M or R106W mutations may affect DNA binding, and indeed, this has recently been demonstrated (88). The F155S mutant does not affect any proposed hydrogen bonds with DNA or water-mediated DNA contacts, yet this mutation yields the most rapid conformational sampling of any of the three mutations tested probably by its disruption of hydrophobic contacts with the many neighboring side-chains lining the interior of the MBD fold. As might be expected, this mutant potently disrupts DNA binding (88).

Our H/DX studies of three frequent RTT mutants indicate that MBD mutations that each cause RTT in a physiological context can do so in a manner that either disrupts or preserves the flexibility of the folded state of the MBD. Final structures are useful in predicting the local and global effects of mutation of amino acid sequence, and our H/DX studies of the MBD mutant proteins illustrates the further utility of site-resolved dynamic protein measurements to determine the extent to which point mutations affect backbone dynamics. Our studies of the F155S and T158M mutations also raise the tantalizing prospect that a subset of MBD-localized RTT mutations may benefit by so-called small molecule chaperones to return the MBD to the relative stability of the wild type protein, and we propose that RTT should be added to the growing list of diseases where such pharmacological stabilization of mutant disease causing gene products is pursued as a viable therapeutic approach.

Chapter 6

Experimental observations of an intrinsically disordered protein - MeCP2; critical discussion of observed DNA binding activity and secondary structure acquisition in MeCP2 and constituent fragments

6.1 Introduction

The work outlined in this dissertation represents one part of a contemporaneous collaborative effort to experimentally observe the MeCP2 gene product behaving as an intrinsically disordered protein (IDP). An IDP is a protein or protein region that does not continuously hold a well-defined secondary or tertiary structure. Diverse methods were used, including hydrogen-deuterium exchange (H/DX) (appendix 1), high-speed atomic force microscopy (hsAFM) (collaborative manuscript in preparation) and circular dichroism (CD). Each method of investigation was performed in solution state rather than in crystal state. This allowed us to directly observe the flexibility of MeCP2 and to corroborate it accordingly to disorder prediction algorithms (89,90,191). FoldIndex© was the primary prediction algorithm used. The method I have used in this collaborative effort to measure extent and type of disorder in MeCP2 is CD. I will first discuss the definition of an intrinsically disordered protein, then I will hypothesize about the role of intrinsic disorder in MeCP2 function and finally, from CD analysis and non-methylated DNA-binding assays, I will make conclusions and propose models for how MeCP2 binds at least 20 different protein partners and how MeCP2 may use a folding tertiary structure to orchestrate chromatin architecture.

6.1.1 How is intrinsic disorder defined?

An intrinsically disordered protein (IDP) is defined as a whole protein or protein region that does not naturally form and maintain regular secondary or

tertiary structure when isolated in solution at physiological conditions, yet is functional. IDPs can form classical α-helix, *β*-strand, polyproline II helix, or irregular secondary structure conformations while encountering binding partners. This is not always the case, as IDPs may also remain disordered while binding or not binding other proteins. It is predicted by multiple disorder prediction algorithms, including FoldIndex©, that nearly 40% of all human proteins harbor at least one intrinsically disordered region of 30 amino acids or more and that 25% of all human proteins are entirely disordered (94). These predictions contradict a long-standing central tenet of molecular biology which states that the function of a protein depends on it maintaining a thermodynamically stable three-dimensional structure in which intermolecular non-covalent interactions confer a narrow range of conformers. Studies using high-resolution nuclear magnetic resonance and small-angle x-ray scattering techniques have revealed that some functional regions were disordered during enzymatic activity (192,193). Inherent in the emerging paradigm is that protein segments can function when transiently or "durably disordered" (44). A durably disordered protein or protein fragment is defined as a peptide that maintains disorder constitutively, even during activity, including engagement with a binding partner.

MeCP2 shares distinct characteristics with predicted IDPs. Namely, MeCP2 has a low incidence of large hydrophobic amino acid residues (2.4%), and enrichment of lysine residues (13.4% lysine). Shared physical characteristics along with 60% predicted disorder according to Foldindex© reveal MeCP2 is an archetypical IDP (85).

Disorder is observed in two major classes of chromatin architectural proteins (ChAPs): ATP-dependent complexes and AT- independent factors (ATFs). These are reviewed extensively on page 154 of a book on IDPs by Dr. Peter Tompa (92). MeCP2 represents an ATP-independent ChAP regulating accessibility of transcriptional machinery to the genome. The structural state of chromatin can be remodeled through bending and distortion by ATFs such as MeCP2. Indeed higher order chromatin organization is enabled by the flexible regions which can simultaneously bind DNA and proteins, facilitating inter-loci bridging between strands of chromatin. This phenomenon was observed recently using a chromatin immunoprecipitation–combined loop assay in which MeCP2 bridged chromatin loci 11 kb apart (136). Interestingly, the HMGA family of chromatin architectural proteins discussed in Chapter 3 not only have sequence homology to MeCP2 but also share a similar degree of overall disorder and functionality of DNA bending and chromatin-strand bridging. The results presented in the preceding chapters provide a more complete understanding of how specific sub-domains of MeCP2 bind non-methylated DNA and whether they have evolved to undergo complete transition from disorder to order or maintain a certain amount of durable disorder no matter how favorable the solvent conditions are for inducing secondary structure.

The MeCP2 monomer previously has been observed to be 60 percent disordered by circular dichroism (CD) analysis and this observation is in agreement with the disorder percentage predicted by FoldIndex© (85). Although

the methyl-CpG-binding domain (MBD) and transcription repression domain (TRD) were also studied by CD as isolated domains in this previous report, the manner in which intrinsically disordered regions of MeCP2 either acquire secondary structure or remain disordered was not studied. The work here tries to paint a more complete picture of either induced secondary structure or durably maintained disorder along with a more complete map of non-methylated non-specific DNA-binding activity among sub-domains of MeCP2.

6.1.2 Method used to observe intrinsic disorder in MeCP2 and constituent fragments.

The primary method used was circular dichroism (CD) spectroscopy to observe full-length MeCP2 and constituent fragments as they underwent disorder-order transition upon titration with trifluoroethanol (TFE), an organic solvent known to stabilize α-helices in protein and polypeptide solutes (194). Using this method, I sought to observe MeCP2 and constituent fragments as individual objects of study in a highly controlled and purified *in vitro* setting. Attempts were made to use a 60mer double-stranded DNA molecule in order to observe structural changes upon binding; however, the DNA addition resulted in large contributions to the CD signal that could not be reliable subtracted as a blank. I chose to perform titration of MeCP2 with TFE, as TFE has been called the “magic” solvent for its unique ability to overcome this signal to noise hurdle in the far UV range. Another option might have been to use a non-polar solvent, but this was not an option due to the insolubility of small peptides in polar solvents.

Peptides as short as 38 amino acid residues were used in the present work. TFE solvates peptides well and has the unique ability to stabilize intra-H-bonded secondary structures like α-helices or β-turns (195). I hypothesized the regions of MeCP2 with the most numerous binding partners would acquire the most secondary structure upon TFE titration.

The other aspect of investigation in this dissertation has been observation of MeCP2 and constituent fragments as they bind to non-methylated DNA and chromatin templates in electrophoretic mobility shift assays.

6.2 Overall discussion and conclusions

The data presented in Chapter 2 of this dissertation allow conclusions to be made about the disorder-to-order transition of full-length MeCP2, the six protease resistant domains and one domain fusion (the TRD-CTD). The historical reproducibility of the technique was established by comparing the acquisition of secondary structure by purified native *Gallus gallus* histone - H5 to recombinant human wild-type MeCP2 protein. A shift in predicted alpha-helical percentage of full-length H5 was first reported in 1988 (103). In the absence of TFE histone, H5 had 17% predicted alpha helix, and in the presence of 50% TFE had 50% predicted alpha helix. The results presented here match that initially reported result within experimental error and add titration points of 20 and 70% TFE.

Rather than repeating exactly the previously reported method (103), I chose to use an additional percentage of TFE in the titration. The use of four

titration points (0, 20, 50, and 70%TFE) provided a wider view of the disorder-to-order transitions than the use of three percentages in the previous method (0, 50, and 65%). Using a broader concentration range also allowed observation of pieces of MeCP2 that reached a plateau in percent acquired secondary structure and those that acquired more structure at each increasing percentage TFE. This allowed me to differentiate between those regions that durably maintain disorder and those that do not.

Interestingly, the transcriptional repression domain (TRD) alone had the most dramatic shift from disorder to alpha helix in the presence of TFE. TRD changed from less than 10% alpha-helical in the absence of TFE to nearly 89% alpha helix in 70% TFE. I initially hypothesized the observation of dramatic induction of alpha helix in the TRD indicated the TRD had more protein binding partners than the rest of the fragments. However, upon review of the literature this is not the case. I have listed in Table 6.1 all proteins and nucleic acids reported to bind to MeCP2, and noted the amino acid residues within MeCP2 to which the interactions were narrowed, as well as the method or methods used in the respective studies. Examination of this Table reveals that two out of twenty one MeCP2-protein interactions are exclusively assigned to the TRD domain of MeCP2. In the four other cases where TRD is mentioned as the region of interaction the result also includes other domains expressed as domain fusions with TRD. Despite the observed difference in observed secondary structure induction in TRD being 45% higher than any other domain, this does not seem to correlate to the number of binding partners. In fact, according to the current set of

reported proteins that interact with MeCP2, the incidence of binding region seems to be equally distributed along the entire length of MeCP2. This observation brings added understanding to Rett syndrome pathology, explaining why mutations along the entire length of MeCP2 lead to a diseased state. Indeed, if each domain has reported protein binding partners, mutations in each domain can affect the role of MeCP2 in maintaining dynamic chromatin architecture.

6.2.1 Alternative hypothesis and future directions.

If the TRD does not have more binding partners than the other domains, does 45% more α-helix acquired by TRD in TFE titration compared to all other domains have physiological relevance? Further experiments need to be done to determine this. Considering the result and the reported number of binding partners I hypothesize that the role of the TRD is to act as a collapsible hinge upon DNA- and/or factor-binding bringing the two halves of MeCP2 in closer proximity. This speculative model entails MeCP2 folding in half at the center of the TRD. There is a known bipartite nuclear localization sequence (NLS) within the TRD between residues 255 to 271 (RKAEADPQAIPKKRGRK) (196). The two proline residues located in the middle of this NLS may give the TRD a pivot point for two long inducible alpha helices (55 amino acid residues N-terminal and 45 C-terminal to the NLS) to emanate and fold similar to a pair of open to closed scissors upon trans-cis isomerization of the prolines at residues 261 and 265. This folding event may be triggered by binding nuclear transport receptors. MeCP2 folding in half may allow it to pass through the nuclear pore

complex. If this folding point exists it may play a key role in the chromatin condensation activity of MeCP2. The significance of this region is highlighted by the observation that two of the most common Rett mutations occur on either side of this hypothetical “folding” center, the R255X and the R270X which are the 3rd and 4th most common Rett mutations with an incidence of 265 and 245 cases reported respectively world-wide. I recommend atomic force microscopy be conducted on full-length MeCP2 and TRD in the presence of increasing amounts of TFE, or DNA to allow direct visualization of this proposed folding helix model. These experiments would verify or reject the proposed model of prolines 261 and 265 acting as a point of symmetrical folding for MeCP2 nuclear transport and chromatin condensation activity.

Table 6.1 MeCP2 domain interactants.

Interactant of MeCP2	MeCP2 region of interaction	Method used in observation
RNA (64,65)	Not defined	Native 5% PAGE and autoradiography (gel shifts) used.
DNA (8,36,54,116,118-120)	MBD, ID, TRD, CTD	Electrophoretic mobility Shift Assay (EMSA), Analytical Ultracentrifugation (AUC), Fluorescence anisotropy, Capillary electrophoresis mobility shift assay, Chromatin Immunoprecipitation (ChIP)
formin-binding protein (FBP) and HYPC 11(34,66)	CTD - amino acid residues 325-486 "WW binding region"	Co-Immunoprecipitation and western blot
YB-1 (64) Co-IP and western blots used.	Full length MeCP2 and a MeCP2-308X truncation mutant used. 308X altered but did not ablate binding to YB-1.	Co-Immunoprecipitation and western blot
Sin3a and HDACs (16)	TRD, residues 207−310	Glutathione-*S*-transferase (GST) pulldown experiments and western blot
SUV39H1 (81)	Not defined, other than MBD alone does not bind	Co-Immunoprecipitation and western blot
HP1 (78)	NTD (aa 1-55)	Co-Immunoprecipitation and western blot
DNMT1 (79)	MBD, ID, and TRD	Glutathione-*S*-transferase (GST) pulldown experiments with rat MeCP2 and rat DMNT1
Ski and N-COR (22)	ID-TRD loosely defined as residues 162-309	Glutathione-*S*-transferase (GST) pulldown experiments with human MeCP2 and mouse Ski and N-COR
PU.1 (80)	NTD (1-92) or TRD (207−309)	Glutathione-*S*-transferase (GST) pulldown experiments with mouse MeCP2.
Brahma (Brm), catalytic component of SWI/SNF (197)	Not defined	Co-Immunoprecipitation and western blot
ATRX (a SWI2/SNF2 DNA helicase) (74)	MBD, residues 78 -169	Glutathione-*S*-transferase (GST) pulldown experiments with mouse MeCP2
Histone H3K9 methyltransferase (75)	MeCP2 carries a methyl transferase activity methylating H3 lysine9 in the MBD and extreme c-terminus of the TRD 279-486 also, methylates H3K9	Glutathione-*S*-transferase (GST) pulldown experiments
TFIIB (76)	TRD, residues 204-310	Glutathione-*S*-transferase (GST) pulldown experiments
HIPK2 - homeodomain-interacting protein kinase 2 (68)	Binding region not defined but, MeCP2 is specifically phosphorylated at Ser 80 by HIPK2	Glutathione-*S*-transferase (GST) pulldown experiments
HMGB1 (73)	CTD, residues 380–386	Glutathione-*S*-transferase (GST) pulldown experiments
CDKL5 (70)	TRD-CTD, residues 202–486	Glutathione-*S*-transferase (GST) pulldown experiments
HDAC1 (72)	Not defined	Glutathione-*S*-transferase (GST) pulldown experiments
CBF1 (69)	Not defined	Glutathione-*S*-transferase (GST) pulldown experiments
lamin B LBR (67)	ID, residues 162 to 202	Glutathione-*S*-transferase (GST) pulldown experiments
Xenopus p20 (71)	ID, residues 162 to 202	Glutathione-*S*-transferase (GST) pulldown experiments

6.2.2 Induced α-helix in MeCP2 sub domains correlates precisely with predicted molecular recognition features (MoRFs).

A molecular recognition feature (MoRF) is a relatively short piece of a protein (an average length of 20 amino acid residues) that acquires secondary structure upon recognition of a binding partner. The binding partner could be another protein or a small molecule. These interacting partners form a well-defined complex where either one or both partners acquire secondary structure from a previously disordered state (97). There are nine predicted MoRFs along the length of MeCP2: two in the NTD, two in the MBD, one in the TRD and four in the CTD (63). The ID is the only MeCP2 domain without a predicted α-MoRF. There is a direct correlation between observed acquisition of α-helix upon TFE titration of MeCP2 domains and incidence of predicted MoRFs. Each domain, except the ID, acquires at least 30% α-helix in 70% TFE. The ID which has no predicted αMoRF maintains a constant percentage of all secondary structures throughout TFE titration. This suggests that the region of the ID has evolved to durably maintain disorder as it stays disordered even in a solvating environment extremely conducive to secondary structure formation. This correlation may indicate that the presence of an αMoRF automatically predisposes a peptide to undergo a disorder-to-order transition. However this correlation cannot be labeled causative until further tests are done. Observation of a significant

number of peptides with predicted αMoRFs that did not acquire α-helix in CD TFE titration, or the observation of a significant number of peptides without predicted αMoRFs but still able to acquire alpha helical character would nullify this hypothesis.

6.2.3 Isolated Methyl CpG Binding Domain (MBD) loses β-strand/turn while gaining α-helical secondary structure upon TFE titration.

Unlike the other four MeCP2 domains investigated, MBD was unique in losing β-strand/turn character while acquiring α-helical character. Disorder remained constant at the level of 30% in the presence of 0 to 70% TFE. The MBD is known to house the most rigidly structured motif in the entire MeCP2 molecule, as evidenced by the NMR and crystal structures of this three-stranded β strand/turn, one α-helix wedge, and the hydrogen deuterium exchange data presented in appendix 1. This motif is 34 amino acids long while the entire MBD fragment contains 90 amino acids. The flanking 50 amino acids are predicted to maintain some order though they do not show up in NMR or the crystal structure. According to the CD analysis, in 0% TFE the MBD is found to have over 50% β-strand/turn. The conversion from β-strand/turn to α-helix may be achieved due to these flanking regions, separating and settling MBD into a more stable thermodynamic equilibrium. The intra-helix hydrogen bonds of an α-helix are more stable than inter-strand hydrogen bonds inherent to β-sheet structure within a less polar solvent. The maintenance of 30% disorder in the MBD throughout TFE titration indicates that a certain amount of the MBD has durable disorder.

The MBD is unique in that the α-helix and β-strand/turn are themselves transient and not derived from preexisting disorder but from each other.

6.2.4The carboxy terminal domain (CTD) of MeCP2 converts disorder to both α-helix and β-strand/turn.

The CTD, unlike the other domains, represents a truly transiently disordered peptide. In this case having a wider range of titration points is valuable, allowing subtleties in the disorder-to-order transition to be observed. In the absence of TFE, the CTD is 70% disordered, with 25% β-strand, and 5% α-helix. The initial drop in disorder at 20%TFE was not converted into alpha helix, but shifted β-strand/turn percentage to 42% while bumping the α-helix up to 10%. As more TFE was added the β-strand/turn came down to 39% to match disorder, while there was a more dramatic conversion to α-helix at this step to 22%. In the presence of 70% TFE α-helix barely takes over with just over 33%, whereas β-strand/turn and disorder are depleted equally, allowing α-helix to form. Though there is only a 5% overall increase in β-strand/turn percentage from 0-70% TFE, seeing the middle titration points of 20 and 50% sheds light on the nature of the conversion and the lowest energy states of the peptides. This observation in conjunction with the hypothesis that there are molecular recognition features that have β-strand/turn, complex, and irregular structure suggest that a given region of disorder may be able to interact with multiple binding partners and upon binding different partners acquire different secondary structures itself (97).

6.2.5 MeCP2 sub-domains have different combinations of durable and transient disorder.

Reflecting on the above observations, it is apparent that MeCP2 domains each have a unique mixture of durable and transient disorder. The TRD seems to have no durable disorder as it is able to almost entirely convert to α-helix in 70% TFE. The MBD has extremely durable disorder, as the propensity for high percentage TFE to stabilize alpha helix was forced to draw on existing β-strand/turn to form more α-helix while the durably disordered portion of MBD held disorder at 30%. Unlike the MBD, the ID seems to have dedicated order as well as durable disorder, because TFE titration has no effect on the percentage of order of this peptide. ID holds its secondary structure and lack thereof robustly over the course of the TFE titration. The CTD and NTD are mixtures of transient and durable disorder. Both fragments start at very high percent disorder (~70%) and converge on 33% disorder, 33% β-strand/turn, and 33% α-helix.

The existence of various amounts of durable or transient disorder in each domain studied in this work did not correlate to the ability of the respective domains to bind DNA or nucleosomal arrays. This suggests each domain capable of binding DNA does so either by acquiring secondary structure upon binding DNA or maintaining disorder throughout the course of the interaction.

6.2.6 MeCP2 and constituent fragments bind to unmethylated DNA and chromatin templates.

In Chapter 3 of this dissertation it is reported that full-length MeCP2 , and its truncated versions containing residues 206-486, 300-486, and 335- 486 all decrease mobility of double-stranded DNA alone and nucleosomal arrays progressively when incubated together at increasing molar ratios by electrophoretic mobility-shift assay. Residues 354-486 and 400-450 were not observed to have any effect on electrophoretic mobility of any DNA-based template assayed. No difference in mobility retardation was observed between MeCP2 fragments incubated with nucleosomal arrays reconstituted with tailless or wild-type histones. It is worth noting that wild-type MeCP2 and residues 206-486 nucleo-protein complexes were so large and interconnected at a molar ratio of 6 moles MeCP2 fragment to 1 mole 208 base pair repeat that they did not enter the gel matrix. This observation could indicate compact complex formation and is observed when these constructs are incubated with DNA alone, wild-type nucleosomal arrays, or tailless arrays. In my hands histone N-terminal tails were not required for array compaction by wild type MeCP2 or residues 206-486 (TRD-CTD). This fact does not rule out the possibility of a direct interaction of MeCP2 and histone N-terminal tails. The observation suggests this potential histone NTD - MeCP2 interaction is not necessary for array compaction.

The minimum MeCP2 domain combination required for nucleosomal array compaction is the TRD-CTD. The further C-terminal truncation fragment 300-486 was unable to form complexes large enough not to enter the 0.8% agarose gel

matrix and therefore did not compact nucleosomal arrays to the same degree. This result is consistent with electrophoretic mobility-shift assay and analytical ultracentrifugation observations made by our collaborators (see Figure 4.3. panels a. and d. and Figure 4.4 in Chapter 4) (96).

If MeCP2 folds in half by the collapsible hinge mechanism I hypothesized in section 6.2.1 and this mechanism is required for nucleosomal array compaction, then the existence of at least three independent DNA-binding domains between residues 206-486 should exist. At least one DNA binding motif has been observed in the TRD while another was observed in the carboxy-terminal domain. Further deletion mutations need to be assayed to determine the precise location of independent DNA-binding domains in the TRD CTD fragment not previously reported.

Though residues 300-486 were not able to compact nucleosomal arrays this fragment did exhibit robust ability to shift naked DNA, nucleosomal arrays and nucleosomal arrays lacking N-terminal tails. This led to the question: where does DNA-binding activity stop in the C-terminal portion of MeCP2? Two further C-terminal truncation mutants, residues 335-486 and residues 354-486 were assayed (see Figure 3.2). It was observed that 335-486 was able to shift DNA, and nucleosomal arrays, while residues 354-486 were unable to shift either DNA template. From these results at least one DNA-binding motif between residues 300-354 can be inferred. This result leaves open the possibility of two independently acting DNA-binding motifs between residues 300-354. Observation of residues 335-486 binding both DNA and nucleosomal arrays

differed slightly from the result of our collaborators. They observed residues 335-486 as able to shift nucleosomal arrays but not DNA alone.

The results I present here in Figure 3.2 clearly show MeCP2 fragment 335-486 shifting both nucleosomal arrays and naked DNA. Though preference of this fragment for arrays over DNA alone is published, I was unable to reproduce this. I conclude that MeCP2 does not require histone binding for chromatin compaction activity, but rather depends on at least four non-specific, unmethylated-DNA binding motifs in the MBD, ID, TRD and CTD in the majority of binding sites among neuronal cell nuclei.

Though the intervening domain (ID) was previously shown to be essential for MeCP2 localization in chromatin via fluorescence recovery after photobleaching (FRAP) method in an *in vivo* study (86), direct observation of the ID binding DNA as an isolated domain was not reported until our collaborative publication (see Chapter 4). Though my results (see Figure 3.3-3.5) did not appear in this publication and the fragment I cloned and expressed was slightly smaller (matching 38 MeCP2 residues - 168-206) than that used by our collaborators (matching 46 MeCP2 residues -165-210), our results were in agreement. Indeed, we both concluded from EMSAs conducted with the ID incubated with DNA that the ID robustly shifted DNA. My work with this region of MeCP2 between the MBD and TRD continued to probe for a more precise location of DNA binding activity. I found that the MeCP2 residues with exact homology to the central AT-hook motif of HMGA proteins are not the primary residues engaged in DNA binding. This conclusion is based on the observation

that mutating the AT-hook consensus core sequence with three alanines actually increased DNA shifting activity of the ID peptide (see center three lanes in Figure 3.5). This observation is consistent with the observation that the R188E point mutation severely inhibited DNA binding both in full length MeCP2 Figure 3.3) and the ID peptide (Figure 3.5) suggests DNA binding activity of the ID is housed between residues 168-188. Further truncations of the ID would test this hypothesis.

This result is interesting from an evolutionary perspective. It can be inferred from homology that MeCP2 and the HMGA family of non-ATP dependent ChAPs share a common ancestral gene. This gene product probably had an AT-hook motif and at a certain point after the divergence of MeCP2 and HMGA proteins, the AT-hook retained DNA-binding ability in the HMGA proteins while the MeCP2 gene product maintained DNA-binding activity but shifted that activity just N-terminal to the original binding region. Keeping this functional drift idea in mind while comparing the CD results of the ID, I hypothesize that the ID binds DNA while maintaining durable disorder but has evolved into this state of maintained disorder by a series of mutations from a protein ancestor that once had the requisite palindrome for a “true AT-hook” motif. A “true” AT-hook has two prolines that bookend the GRP consensus core, giving the novel motif the steric pucker required to bind the minor groove of AT rich DNA regions (198). I hypothesize further that over the course of evolution MeCP2 lost this AT-rich tracking ability, trading this functionality for more ubiquitous genomic distribution as was observed by recent *in vivo* studies of neuronal cells (54).

6.3 Summary

The systematic dissection of the MeCP2 gene product from amino terminal through carboxy terminal presented in this dissertation represents a significant expansion of our knowledge of the structural diversity of MeCP2. As a direct result of these studies two DNA-binding domains previously unreported in the literature have been identified and corroborated by collaborators (96). One DNA-binding domain was discovered in the intervening domain between the MBD and the TRD and another between residues 300 and 354 in the C-terminal portion of the protein. The unique physical properties of each protease-resistant domain spanning the entirety of MeCP2 reported allow speculations to be made on how MeCP2 uses transient and durable disorder to bind a variety of partners, including proteins, methylated and unmethylated double-stranded DNA, chromatin, and RNA. Experimental observation of MeCP2 and constituent fragments by CD directly agrees with data obtained by hydrogen/deuterium exchange results presented in appendix 1. The primary conclusion from that work is that MeCP2 maintains robust disorder in the absence of a binding partner and has only a slight increase in observable rigidity when bound to methylated and unmethylated DNA (assigned to the MBD). Comparing results from the CD analysis presented in Chapter 2 of this dissertation and the results of Chapter 4 clearly demonstrates MeCP2 as an archetypical intrinsically disordered protein with functional portions that probably maintain a disordered state while engaged with binding partners. Further, data from the TFE titrations suggest the CTD of MeCP2 is capable of acquiring both α-helix and β-strand secondary structures in

different solvating contexts. Extrapolating this observation to the crowded jungle-like state of the nucleus implies that it is possible that the same disordered region in the CTD of MeCP2 may be able to acquire different secondary structures when binding different partners, depending on the electrostatic context of the interaction.

REFERENCES

1. Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992) *Nucleic Acids Res* **20**, 5085-5092
2. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) *Cell* **69**, 905-914
3. Nan, X., Campoy, F. J., and Bird, A. (1997) *Cell* **88**, 471-481
4. Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., and Zoghbi, H. Y. (2008) *Science* **320**, 1224-1229
5. Yasui, D. H., Peddada, S., Bieda, M. C., Vallero, R. O., Hogart, A., Nagarajan, R. P., Thatcher, K. N., Farnham, P. J., and Lasalle, J. M. (2007) *Proc Natl Acad Sci U S A* **104**, 19416-19421
6. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989) *Cell* **58**, 499-507
7. Meehan, R., Antequera, F., Lewis, J., MacLeod, D., McKay, S., Kleiner, E., and Bird, A. P. (1990) *Philos Trans R Soc Lond B Biol Sci* **326**, 199-205
8. Nan, X., Meehan, R. R., and Bird, A. (1993) *Nucleic Acids Res* **21**, 4886-4892
9. Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A., and Esteller, M. (2003) *Nucleic Acids Res* **31**, 1765-1774
10. Nikitina, T., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., Grigoryev, S. A., and Woodcock, C. L. (2007) *J Biol Chem*
11. Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999) *J Mol Biol* **291**, 1055-1065
12. Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) *Mol Cell* **29**, 525-531
13. Hansen, J. C., Wexler, B. B., Rogers, D. J., Hite, K. C., Panchenko, T., Ajith, S., and Black, B. E. (2011) *J Biol Chem* **286**, 18938-18948
14. Zlatanova, J. (2005) *Biochem Cell Biol* **83**, 251-262
15. Wade, P. A., Jones, P. L., Vermaak, D., Veenstra, G. J., Imhof, A., Sera, T., Tse, C., Ge, H., Shi, Y. B., Hansen, J. C., and Wolffe, A. P. (1998) *Cold Spring Harb Symp Quant Biol* **63**, 435-445
16. Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) *Nature* **393**, 386-389
17. Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998) *Nat Genet* **19**, 187-191
18. Yu, F., Thiesen, J., and Stratling, W. H. (2000) *Nucleic Acids Res* **28**, 2201-2206
19. Urdinguio, R. G., Pino, I., Ropero, S., Fraga, M. F., and Esteller, M. (2007) *Epigenetics* **2**, 11-14
20. Wischnewski, F., Friese, O., Pantel, K., and Schwarzenbach, H. (2007) *Mol Cancer Res* **5**, 749-759
21. Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y. E. (2003) *Science* **302**, 890-893
22. Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001) *J Biol Chem* **276**, 34115-34121

23. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999) *Nat Genet* **23**, 185-188
24. Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A., and Hansen, J. C. (2003) *J Biol Chem* **278**, 32181-32188
25. Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., Berget, S., and Zoghbi, H. Y. (2005) *Proc Natl Acad Sci U S A* **102**, 17551-17558
26. Jeffery, L., and Nakielny, S. (2004) *J Biol Chem* **279**, 49479-49487
27. Pelka, G. J., Watson, C. M., Radziewic, T., Hayward, M., Lahooti, H., Christodoulou, J., and Tam, P. P. (2006) *Brain* **129**, 887-898
28. Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J., Jones, J. R., Scherer, S. W., Schanen, N. C., Friez, M. J., Vincent, J. B., and Minassian, B. A. (2004) *Nat Genet* **36**, 339-341
29. Dragich, J. M., Kim, Y. H., Arnold, A. P., and Schanen, N. C. (2007) *J Comp Neurol* **501**, 526-542
30. Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007) *Science* **315**, 1143-1147
31. Kline, D. D., Ogier, M., Kunze, D. L., and Katz, D. M. (2010) *Journal of Neuroscience* **30**, 5303
32. Adams, V. H., McBryant, S. J., Wade, P. A., Woodcock, C. L., and Hansen, J. C. (2007) *J Biol Chem* **282**, 15057-15064
33. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M., and Tanokura, M. (2004) *J Biol Chem* **279**, 31833-31841
34. Buschdorf, J. P., and Stratling, W. H. (2004) *J Mol Med* **82**, 135-143
35. Dintilhac, A., and Bernues, J. (2002) *J Biol Chem* **277**, 7021-7028
36. Nikitina, T., Shi, X., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., and Woodcock, C. L. (2007) *Mol Cell Biol* **27**, 864-877
37. Chandler, S. P., Guschin, D., Landsberger, N., and Wolffe, A. P. (1999) *Biochemistry* **38**, 7008-7018
38. Klose, R. J., and Bird, A. P. (2004) *J Biol Chem* **279**, 46490-46496
39. Hansen, J. C., Lu, X., Ross, E. D., and Woody, R. W. (2006) *J Biol Chem* **281**, 1853-1856
40. Chen, J. W., Romero, P., Uversky, V. N., and Dunker, A. K. (2006) *J Proteome Res* **5**, 888-898
41. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) *J Mol Recognit* **18**, 343-384
42. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) *Febs J* **272**, 5129-5148
43. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) *J Mol Graph Model* **19**, 26-59
44. Chouard, T. (2011) *Nature* **471**, 151-153
45. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) *Biochemistry* **45**, 6873-6888

46. Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) *PLoS Comput Biol* **2**, e100
47. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) *Genome Inform Ser Workshop Genome Inform* **10**, 30-40
48. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) *Bioinformatics* **21**, 3435-3438
49. Jung, B. P., Jugloff, D. G., Zhang, G., Logan, R., Brown, S., and Eubanks, J. H. (2003) *J Neurobiol* **55**, 86-96
50. Tudor, M., Akbarian, S., Chen, R. Z., and Jaenisch, R. (2002) *Proc Natl Acad Sci U S A* **99**, 15536-15541
51. Traynor, J., Agarwal, P., Lazzeroni, L., and Francke, U. (2002) *BMC Med Genet* **3**, 12
52. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature* **389**, 251-260
53. Ishibashi, T., Thambirajah, A. A., and Ausio, J. (2008) *FEBS Lett* **582**, 1157-1162
54. Skene, P., Illingworth, R., Webb, S., Kerr, A., James, K., Turner, D., Andrews, R., and Bird, A. (2010) *Molecular cell* **37**, 457-468
55. Abuhatzira, L., Makedonski, K., Kaufman, Y., Razin, A., and Shemer, R. (2007) *Epigenetics* **2**, 214-222
56. Moore, H., Leonard, H., Fyfe, S., De Klerk, N., and Leonard, N. (2005) *Ann Hum Biol* **32**, 228-236
57. Glaze, D. G. (2004) *Ment Retard Dev Disabil Res Rev* **10**, 154-158
58. Archer, H., Evans, J., Leonard, H., Colvin, L., Ravine, D., Christodoulou, J., Williamson, S., Charman, T., Bailey, M. E., Sampson, J., de Klerk, N., and Clarke, A. (2007) *J Med Genet* **44**, 148-152
59. Jian, L., Archer, H. L., Ravine, D., Kerr, A., de Klerk, N., Christodoulou, J., Bailey, M. E., Laurvick, C., and Leonard, H. (2005) *Eur J Hum Genet* **13**, 1235-1238
60. Neul, J. L., Fang, P., Barrish, J., Lane, J., Caeg, E. B., Smith, E. O., Zoghbi, H., Percy, A., and Glaze, D. G. (2008) *Neurology*
61. Nikitina, T., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., Grigoryev, S. A., and Woodcock, C. L. (2007) *J Biol Chem* **282**, 28237-28245
62. Yang, C. (2009) Biochemical, biophysical and structural study of the nucleosome-MECP2 complex.
63. Hansen, J. C., Ghosh, R. P., and Woodcock, C. L. (2010) *IUBMB Life* **62**, 732-738
64. Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., and Berget, S. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17551
65. Jeffery, L., and Nakielny, S. (2004) *Journal of Biological Chemistry* **279**, 49479
66. Bedford, M. T., Chan, D. C., and Leder, P. (1997) *the EMBO Journal* **16**, 2376-2383
67. Guarda, A., Bolognese, F., Bonapace, I. M., and Badaracco, G. (2009) *Experimental cell research* **315**, 1895-1903

68. Bracaglia, G., Conca, B., Bergo, A., Rusconi, L., Zhou, Z., Greenberg, M. E., Landsberger, N., Soddu, S., and Kilstrup-Nielsen, C. (2009) *EMBO reports* **10**, 1327-1333
69. Mann, J., Oakley, F., Akiboye, F., Elsharkawy, A., Thorne, A., and Mann, D. (2006) *Cell Death & Differentiation* **14**, 275-285
70. Mari, F., Azimonti, S., Bertani, I., Bolognese, F., Colombo, E., Caselli, R., Scala, E., Longo, I., Grosso, S., and Pescucci, C. (2005) *Human molecular genetics* **14**, 1935
71. Carro, S., Bergo, A., Mengoni, M., Bachi, A., Badaracco, G., Kilstrup-Nielsen, C., and Landsberger, N. (2004) *Journal of Biological Chemistry* **279**, 25623
72. Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y. E. (2003) *Science* **302**, 890
73. Dintilhac, A., and Bernués, J. (2002) *Journal of Biological Chemistry* **277**, 7021
74. Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M. J., Shu, X., Kriaucionis, S., and Bird, A. (2007) *Proc Natl Acad Sci U S A* **104**, 2709-2714
75. Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003) *J Biol Chem* **278**, 4035-4040
76. Kaludov, N. K., and Wolffe, A. P. (2000) *Nucleic acids research* **28**, 1921
77. Harikrishnan, K., Chow, M. Z., Baker, E. K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J. M., Jones, P. L., and Sif, S. (2005) *Nature genetics* **37**, 254-264
78. Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M. C. (2007) *Nucleic Acids Res* **35**, 5402-5408
79. Kimura, H., and Shiota, K. (2003) *J Biol Chem* **278**, 4806-4812
80. Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., and Oikawa, T. (2003) *Oncogene* **22**, 8688-8698
81. Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002) *Science* **298**, 1747-1752
82. Bienvenu, T., and Chelly, J. (2006) *Nature Reviews Genetics* **7**, 415-426
83. Huppke, P., Laccone, F., Krämer, N., Engel, W., and Hanefeld, F. (2000) *Human molecular genetics* **9**, 1369
84. Cheadle, J. P., Gill, H., Fleming, N., Maynard, J., Kerr, A., Leonard, H., Krawczak, M., Cooper, D. N., Lynch, S., and Thomas, N. (2000) *Human molecular genetics* **9**, 1119
85. Adams, V., McBryant, S., Wade, P., Woodcock, C., and Hansen, J. (2007) *Journal of Biological Chemistry* **282**, 15057
86. Kumar, A., Kamboj, S., Malone, B. M., Kudo, S., Twiss, J. L., Czymmek, K. J., LaSalle, J. M., and Schanen, N. C. (2008) *J Cell Sci* **121**, 1128-1137
87. Laccone, F., Huppke, P., Hanefeld, F., and Meins, M. (2001) *Hum Mutat* **17**, 183-190
88. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Gierasch, L. M., and Woodcock, C. L. (2008) *Journal of Biological Chemistry* **283**, 20523
89. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6573-6582
90. Dunker, A. K., and Obradovic, Z. (2001) *Nature biotechnology* **19**, 805-806

91. Dyson, H. J., and Wright, P. E. (2005) *Nature Reviews Molecular Cell Biology* **6**, 197-208
92. Tompa, P. (2009) *Structure and function of intrinsically disordered proteins*, Chapman & Hall/CRC
93. Tompa, P., and Fuxreiter, M. (2008) *Trends in biochemical sciences* **33**, 2-8
94. Uversky, V. N. (2010) *Structure* **18**, 1069-1071
95. Wright, P. E., and Dyson, H. J. (1999) *Journal of molecular biology* **293**, 321-331
96. Ghosh, R. P., Nikitina, T., Horowitz-Scherer, R. A., Gierasch, L. M., Uversky, V. N., Hite, K., Hansen, J. C., and Woodcock, C. L. (2010) *Biochemistry* **49**, 4395-4410
97. Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) *J Mol Biol* **362**, 1043-1059
98. Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005) *Biochemistry* **44**, 12454-12470
99. Tanford, C. (1962) *Journal of the American Chemical Society* **84**, 4240-4247
100. Walters, L., and Kaiser, E. T. (1985) *Journal of the American Chemical Society* **107**, 6422-6424
101. Buck, M. (1998) *Quarterly reviews of biophysics* **31**, 297-355
102. Goodman, M., and Listowsky, I. (1962) *Journal of the American Chemical Society* **84**, 3770-3771
103. Clark, D., Hill, C., Martin, S., and Thomas, J. (1988) *The EMBO Journal* **7**, 69
104. Roque, A., Iloro, I., Ponte, I., Arrondo, J. L. R., and Suau, P. (2005) *Journal of Biological Chemistry* **280**, 32141
105. Bourhis, J. M., Canard, B., and Longhi, S. (2006) *Virology* **344**, 94-110
106. Pattaramanon, N., Sangha, N., and Gafni, A. (2007) *Biochemistry* **46**, 3405-3415
107. Garcia-Ramirez, M., Leuba, S. H., and Ausio, J. (1990) *Protein Expr Purif* **1**, 40-44
108. Clark, D. J., and Thomas, J. O. (1986) *J Mol Biol* **187**, 569-580
109. Smith, R. D. (2009) Characterization of the DNA binding region of the transcription repression domain in MeCP2.
110. Sreerama, N., and Woody, R. W. (2000) *Analytical biochemistry* **287**, 252-260
111. Roccatano, D., Colombo, G., Fioroni, M., and Mark, A. E. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12179
112. Watt, K., Jess, T. J., Kelly, S. M., Price, N. C., and McEwan, I. J. (2005) *Biochemistry* **44**, 734-743
113. Morin, B., Bourhis, J. M., Belle, V., Woudstra, M., Carrière, F., Guigliarelli, B., Fournel, A., and Longhi, S. (2006) *The Journal of Physical Chemistry B* **110**, 20596-20608
114. Cheng, Y., Oldfield, C. J., Meng, J., Romero, P., Uversky, V. N., and Dunker, A. K. (2007) *Biochemistry* **46**, 13468-13477
115. Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) *J Proteome Res* **6**, 2351-2366
116. Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A., and Esteller, M. (2003) *Nucleic acids research* **31**, 1765
117. Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C., and Asturias, F. J. (2008) *Nat Chem Biol* **4**, 728-737

118. Ballestar, E., Paz, M. F., Valle, L., Wei, S., Fraga, M. F., Espada, J., Cigudosa, J. C., Huang, T. H. M., and Esteller, M. (2003) *the EMBO Journal* **22**, 6335-6345
119. Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003) *Science* **302**, 885
120. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Shlyakhtenko, L. S., and Woodcock, C. L. (2010) *Mol Cell Biol* **30**, 4656-4670
121. Ohki, I., Shimotake, N., Fujita, N., Jee, J. G., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) *Cell* **105**, 487-497
122. Carrigan, P. E., Ballar, P., and Tuzmen, S. (2011) *Methods in molecular biology (Clifton, NJ)* **700**, 107
123. Hansen, J. C., and Lohr, D. (1993) *J Biol Chem* **268**, 5840-5848
124. Reeves, R. (2001) *Gene* **277**, 63-81
125. Dhasarathy, A., and Wade, P. A. (2008) *Mutat Res* **647**, 39-43
126. Klose, R. J., and Bird, A. P. (2006) *Trends Biochem Sci* **31**, 89-97
127. LaSalle, J. M., Hogart, A., and Thatcher, K. N. (2005) *Int Rev Neurobiol* **71**, 131-165
128. Moretti, P., and Zoghbi, H. Y. (2006) *Curr Opin Genet Dev* **16**, 276-281
129. Kishi, N., and Macklis, J. D. (2004) *Mol Cell Neurosci* **27**, 306-321
130. Ballestar, E., and Esteller, M. (2005) *Biochem Cell Biol* **83**, 374-384
131. Bernard, D., Gil, J., Dumont, P., Rizzo, S., Monte, D., Quatannens, B., Hudson, D., Visakorpi, T., Fuks, F., and de Launoit, Y. (2006) *Oncogene* **25**, 1358-1366
132. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) *Cell* **105**, 487-497
133. Shahbazian, M. D., Young, J. I., Yuva-Paylor, L. A., Spencer, C. M., Antalffy, B. A., Noebels, J. L., Armstrong, D. L., Paylor, R., and Zoghbi, H. Y. (2002) *Neuron* **35**, 243-254
134. Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., and Mandel, G. (2002) *Science* **298**, 1747
135. Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) *Current opinion in structural biology* **18**, 756-764
136. Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2004) *Nature genetics* **37**, 31-40
137. Chadwick, L. H., and Wade, P. A. (2007) *Curr Opin Genet Dev* **17**, 121-125
138. Cheng, Y., Oldfield, C. J., Meng, J., Romero, P., Uversky, V. N., and Dunker, A. K. (2007) *Biochemistry* **46**, 13468-13477
139. Martin, B. R., Giepmans, B. N. G., Adams, S. R., and Tsien, R. Y. (2005) *Nature biotechnology* **23**, 1308-1314
140. Provencher, S. W., and Gloeckner, J. (1981) *Biochemistry* **20**, 33-37
141. Lobley, A., Whitmore, L., and Wallace, B. (2002) *Bioinformatics* **18**, 211-212
142. Whitmore, L., and Wallace, B. (2004) *Nucleic acids research* **32**, W668
143. Pancoska, P., Bitto, E., Janota, V., Urbanova, M., Gupta, V. P., and Keiderling, T. A. (1995) *Protein science: a publication of the Protein Society* **4**, 1384
144. Perczel, A., Park, K., and Fasman, G. (1992) *Analytical biochemistry* **203**, 83-93
145. Greenfield, N. J. (2004) *METHODS IN MOLECULAR BIOLOGY-CLIFTON THEN TOTOWA*- **261**, 55-78

146. Lakowicz, J. R., and Masters, B. R. (2008) *Journal of Biomedical Optics* **13**, 029901
147. Krishnan, B., and Gierasch, L. M. (2008) *Chemistry & biology* **15**, 1104-1115
148. Luedtke, N. W., Dexter, R. J., Fried, D. B., and Schepartz, A. (2007) *Nature chemical biology* **3**, 779-784
149. Griffin, B. A., Adams, S. R., and Tsien, R. Y. (1998) *Science* **281**, 269
150. Findlay, J. W. A., and Dillard, R. F. (2007) *The AAPS journal* **9**, 260-267
151. Van Holde, K., and Weischet, W. O. (1978) *Biopolymers* **17**, 1387-1403
152. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) *PROTEINS-NEW YORK-* **42**, 38-48
153. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., and Uversky, V. N. (2006) *Nucleic acids research* **35**, D786
154. Iakoucheva, L. M., and Dunker, A. K. (2003) *Structure* **11**, 1316-1317
155. Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., Cortese, M. S., Sickmeier, M., LeGall, T., Obradovic, Z., and Dunker, A. K. (2006) *Proc Natl Acad Sci U S A* **103**, 8390-8395
156. Vacic, V., Uversky, V. N., Dunker, A. K., and Lonardi, S. (2007) *BMC bioinformatics* **8**, 211
157. Georgel, P. T., Fletcher, T. M., Hager, G. L., and Hansen, J. C. (2003) *Genes Dev* **17**, 1617-1629
158. Lowary, P., and Widom, J. (1998) *Journal of molecular biology* **276**, 19-42
159. Hansen, J. C., Lebowitz, J., and Demeler, B. (1994) *Biochemistry* **33**, 13155-13163
160. Heitmann, B., Maurer, T., Weitzel, J. M., Strätling, W. H., Kalbitzer, H. R., and Brunner, E. (2003) *European Journal of Biochemistry* **270**, 3263-3270
161. Berggard, T., Linse, S., and James, P. (2007) *Proteomics* **7**, 2833-2842
162. Adams, S. R., Campbell, R. E., Gross, L. A., Martin, B. R., Walkup, G. K., Yao, Y., Llopis, J., and Tsien, R. Y. (2002) *Journal of the American Chemical Society* **124**, 6063-6076
163. Shuman, C. F., Jiji, R., Akerfeldt, K. S., and Linse, S. (2006) *Journal of molecular biology* **358**, 870-881
164. Cortese, M. S., Uversky, V. N., and Keith Dunker, A. (2008) *Progress in biophysics and molecular biology* **98**, 85-106
165. Mészáros, B., Tompa, P., Simon, I., and Dosztányi, Z. (2007) *Journal of molecular biology* **372**, 549-561
166. Oldfield, C., Meng, J., Yang, J., Yang, M. Q., Uversky, V., and Dunker, A. K. (2008) *BMC genomics* **9**, S1
167. Tompa, P. (2005) *FEBS letters* **579**, 3346-3354
168. Patel, L., Abate, C., and Curran, T. (1990)
169. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990)
170. Chandler, S. P., Guschin, D., Landsberger, N., and Wolffe, A. P. (1999) *Biochemistry* **38**, 7008-7018
171. Wright, P. E., and Dyson, H. J. (2009) *Current opinion in structural biology* **19**, 31-38

172. Marchi, M., Guarda, A., Bergo, A., Landsberger, N., Kilstrup-Nielsen, C., Ratto, G. M., and Costa, M. (2007) *Epigenetics: official journal of the DNA Methylation Society* **2**, 187
173. Kimura, H., and Shiota, K. (2003) *Journal of Biological Chemistry* **278**, 4806
174. Eivazova, E. R., Gavrilov, A., Pirozhkova, I., Petrov, A., Iarovaia, O. V., Razin, S. V., Lipinski, M., and Vassetzky, Y. S. (2009) *Journal of molecular biology* **386**, 929-937
175. Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2005) *Nat Genet* **37**, 31-40
176. Weitzel, J. M., Buhrmester, H., and Stratling, W. H. (1997) *Molecular and cellular biology* **17**, 5656
177. Hite, K., Adams, V., and Hansen, J. (2009) *Biochemistry and Cell Biology* **87**, 219-227
178. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15477
179. Chong, S., Montello, G., Zhang, A., Cantor, E., Liao, W., Xu, M., and Benner, J. (1998) *Nucleic acids research* **26**, 5109
180. Black, B. E., Brock, M. A., Bédard, S., Woods, V. L., and Cleveland, D. W. (2007) *Proceedings of the National Academy of Sciences* **104**, 5008
181. Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., and Cleveland, D. W. (2004) *Nature* **430**, 578-582
182. Coales, S. J. (2010) *Rapid Communications in Mass Spectrometry* **24**, 3585-3592
183. Hamuro, Y., Wong, L., Shaffer, J., Kim, J. S., Stranz, D. D., Jennings, P. A., Woods Jr, V. L., and Adams, J. A. (2002) *Journal of molecular biology* **323**, 871-881
184. Pantazatos, D., Kim, J. S., Klock, H. E., Stevens, R. C., Wilson, I. A., Lesley, S. A., and Woods, V. L. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 751
185. Englander, S. W. (2006) *Journal of the American Society for Mass Spectrometry* **17**, 1481-1489
186. Sekulic, N., Bassett, E. A., Rogers, D. J., and Black, B. E. (2010) *Nature*
187. Lin, S., and Riggs, A. D. (1975) *Cell* **4**, 107-111
188. Kalodimos, C. G., Biris, N., Bonvin, A. M. J. J., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) *Science* **305**, 386
189. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271**, 1247
190. Yang, C., van der Woerd, M. J., Muthurajan, U. M., Hansen, J. C., and Luger, K. (2011) *Nucleic acids research*
191. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) *Bioinformatics* **21**, 3435
192. Ayed, A., Mulder, F. A. A., Yi, G. S., Lu, Y., Kay, L. E., and Arrowsmith, C. H. (2001) *Nature Structural & Molecular Biology* **8**, 756-760
193. Grossmann, J. G., Sharff, A. J., O'Hare, P., and Luisi, B. (2001) *Biochemistry* **40**, 6267-6274

194. Berova, N., and Nakanishi, K. (2000) *Circular dichroism: principles and applications*, Vch Verlagsgesellschaft Mbh
195. Fasman, G. D. (1996) *Circular dichroism and the conformational analysis of biomolecules*, Springer Us
196. Nan, X., Tate, P., Li, E., and Bird, A. (1996) *Molecular and cellular biology* **16**, 414
197. Harikrishnan, K. N., Chow, M. Z., Baker, E. K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J. M., Jones, P. L., Sif, S., and El-Osta, A. (2005) *Nat Genet* **37**, 254-264
198. Reeves, R. (2000) *Environmental Health Perspectives*, 803-809