STRUCTURE AND FUNCTION STUDIES OF METHYL-CPG BINDING DOMAIN PROTEINS AND THEIR COMPLEXES

Mary Jeannette Sperlazza

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the School of Arts and Sciences.

Chapel Hill
2017

Approved by:

David C Williams Jr

Dorothy Erie

Eric Brustad

Brian Kuhlman

Marcey Waters

ABSTRACT

Mary Jeannette Sperlazza: Structure and Function Studies of Methyl-CpG Binding Domain Proteins and Their Complexes
(Under the direction of David C Williams Jr)

DNA methylation is an epigenetic mechanism of transcriptional silencing of increasing interest for treating human disease. Methyl-CpG binding domain (MBD) proteins recognize 5-methylcytosines (mC) primarily in symmetrical CpG (mCG) dinucleotides. Seven proteins comprise the MBD family, MBD1-6 and MeCP2. MeCP2 is primarily expressed in the brain and plays a critical role in neuron maturation, as mutations disrupting its function account for up to 80% of Rett Syndrome cases. MBD2/3 associates with the nucleosome remodeling and deacetylase complex (NuRD) and modulates gene expression through alteration of the chromatin architecture surrounding the mC mark. In these studies, we examine the behavior of the MBDs of MeCP2 and MBD2, in addition to further characterizing the protein-protein interactions between subunits of NuRD.

Recent work suggests the primary effects of MeCP2 on gene expression in the developing mammalian brain are mediated by binding asymmetrically methylated and hydroxymethylated CpA (h/mCA) dinucleotides. This work establishes that the MeCP2 MBD binds mCA with high affinity in a strand specific and orientation dependent manner. This preference is specific to MeCP2, as the MBD2 MBD does not show high affinity or methyl-specific binding to mCA. Introduction of the Rett Syndrome-associated

mutations T158M, R106W and P101S destabilized the MeCP2 MBD and lessened recognition of mCG and mCA equally. Finally, hydroxymethylation of a high affinity mCA site did not dramatically change binding properties, however hemi-hydroxylation of the same cytosine in mCG significantly decreased affinity. We suggest a model for MeCP2 recognition of mCA and for hydroxymethylation as an epigenetic switch to redistribute MeCP2 among mCG and mCA loci.

Blocking recruitment of NuRD by MBD2 restores expression of developmentally silenced fetal hemoglobin and aberrantly silenced tumor suppressor genes. Additionally, knockdown of the NuRD helicase, CHD4, results in cancer cells growth arrest and increased sensitivity to DNA damage. Therefore, targeting MBD2-NuRD presents a promising avenue for treating β-hemoglobinopathies and cancer. Towards understanding the recruitment of the NuRD components to the complex, this study characterizes the GATA-like zinc-finger domains of the NuRD components GATAD2A and MTA2. We propose a model of NuRD in which MTA2 binds DNA and GATAD2A serves to bridge MBD2 and CHD4.

To my grandfather, Joel C. Koch, who instilled my love for knowledge and role modeled the value of life-long learning.

ACKNOWLEDGEMENTS

I owe the completion of this work to many who have provided their expertise and support. First, my committee chair and surrogate advisor, Dorothy Erie, who provided guidance through the many trying times and obstacles encountered on this journey. Her wisdom and support emboldens all who meet her to be better scientists and braver human beings. Second, my advisor, David Williams, who took a chance on a fourth-year student and trusted me to pursue my own interests. Thank you for providing an encouraging training environment and giving me the independence to follow the story even when it led your laboratory into uncharted territory. From you, I have learned so much more than just how to analyze an HSQC, but how to approach everything with an eye that is both critical and optimistic. Next, I must thank our collaborators, Gordon Ginder (Virginia Commonwealth University) and Joel Mackay (University of Sydney, Australia) on the GATAD2A project and Marcey Waters on the peptide inhibitor project. Their contributions took our questions into new directions and ushered intriguing ideas to experimental realities.

The University of North Carolina at Chapel Hill has provided an outstanding training environment. The Molecular and Cellular Biophysics training program laid an excellent foundation for my pursuits in structural biology and continues to inform my perspective on every biological process. The TIBBS office, especially Patrick Brandt, Joshua Hall, Jessica Harrell, Anna O'Connell and Rebekah Layton, who gave valuable friendship, advice and training throughout my graduate career. Thank you for offering

UNC’s early-career scientists the opportunities to explore all of their career interests and the reassurances as we navigate them. Special thanks to Dr. Theresa Maitland, UNC students are so lucky to have you and your dedication to our success.

Members of many labs have provided essential technical and emotional support over the past six years. I had some excellent role models of successful graduate students and post-doctoral scientists, especially Michael, Monica, Krystle, and Jackie, who provided guidance on everything from project ideas and protocol details to prospective career and parenting advice. To Julianne, thanks for always having my back in the lab and pushing my limits in the gym. Kira and Jake, I must thank for initiating my crossword puzzle addiction and introducing me to road cycling. Thanks to Stephanie for teaching me to not break the NMR magnet, process 2D data and the virtues of Midwestern beer. To Eileen, Logan, and Christina, it was a pleasure mentoring each of you and having the privilege of watching you grow into scientists. For keeping my spirits up and always being accommodating of my crazy schedule, I thank my friends, especially, Angela, Kaiulani and Joan, and my family, especially my Mom, Mom S., Dad S., Alyssa, Jeremy, Sabrina, Jason and Bobby.

Finally, I thank my most important collaborator and confidant, my husband Justin. Thank you for your understanding and patience; for teaching me transfections and western blotting; for every night spent discussing experiments, each weekend passed in the lab and all the miles driven commuting from Richmond. My love for science would not have survived this endeavor without you, and there is no one with whom I would rather be chasing my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

%	percent
°C	degrees Celsius
±	plus or minus
>	greater than
6-FAM/FAM	fluorescein
^{1}H	hydrogen
^{13}C	carbon-13 isotope
^{15}N	nitrogen-15 isotope
α1	alpha helix 1
A	adenine
A, Ala	alanine
ADP	adenosine diphosphate
ATP	adenosine triphosphate
ATPase	adenosine triphosphate hydrolase
BAH	bromo-adjacent homology domain
bdnf	brain derived neurotrophic factor gene
BLASTN	Basic Local Alignment Search Tool for Nucleotide
BLASTp	Basic Local Alignment Search Tool for Protein
bp	base pair
BSA	bovine serum albumin
C-terminal, C-term	carboxyl terminal
C	cytosine

$CaCl_2$	calcium chloride
CC	coiled-coil
CG, CpG	cytosine-guanine dinucleotide
CHD3	chromodomain helicase DNA binding protein 3
CHD4	chromodomain helicase DNA binding protein 4
CHDCT2	chromodomain helicase DNA binding protein c-terminal 2 domain
Cl	chloride
cm	centimeter
CPMG	Carr-Purcell-Meiboom-Gill ☒
CR1	conserved region 1 of GATAD2A
CR2	conserved region 2 of GATAD2A
CRISPR	clustered regularly interspaced short palindromic repeats
CV	column volume
dlx5	distal-less homeobox 5 gene
DMSO	dimethyl sulfoxide☒☒☒
DNA	deoxyribonucleic acid
DNAse	deoxyribonucleic acid hydrolase
DNMT	deoxyribonucleic acid methyltransferase
dsDNA	double stranded deoxyribonucleic acid
DTT	D,L-dithiothreitol
E, Glu	glutamic acid
E1	ubiquitin-activating enzyme
E2	ubiquitin-conjugating enzyme

E3	ubiquitin ligase enzyme
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ELM2	Egl-27 and MTA1 homology 2 domain
Em	emission
EMSA	electrophoretic mobility shift assay
Ex	excitation
f	average fluorescence anisotropy signal
Fu	fluorescence unit(s)
fxyd1	FXYD domain containing ion transport regulator 1
g	g-force, gravitational acceleration
G, Gly	glycine
G	guanine
gamt	guanidinoacetate N-methyltransferase gene
GATAD2A	GATA zinc finger domain-containing protein 2A
GATAD2B	GATA zinc finger domain-containing protein 2B
h	hour(s)
H, His	histidine
H. sapiens	*Homo sapiens*
H^+	hydrogen ion
H_2O	water
H1	histone 1 protein
H2A	histone 2A protein

H2B	histone 2B protein
H3	histone 3 protein
H4	histone 4 protein
HCl	hydrochloric acid
HDAC1	histone deacetylase 1
HDAC2	histone deacetylase 2
HDCC	histone deacetylase core complex
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	6 x histidine affinity tag
hmC	hydroxymethyl-cytosine
hmCA	hydroxymethyl-cytosine-adenine dinucelotide
hmCG	hydroxymethyl-cytosine-guanine dinucelotide
hmCH	hydroxymethyl-cytosine-non-guanine dinucelotide
HPLC	high-performance liquid chromatography
HS	hypersensitivity site
IDR	intrinsically disordered region
IPTG	isopropyl-β-D-thiogalactopyranoside
K, Lys	lysine
KCl	potassium chloride
K_D	dissociation constant
KH_2PO_4	potassium phosphate
KCl	potassium chloride
L	liter

L, Leu	leucine
LB	lysogeny broth
LCR	locus control region
M, Met	methionine
MBD	methyl-CpG binding domain
MBD1	methyl-CpG binding domain protein 1
MBD2	methyl-CpG binding domain protein 2
MBD3	methyl-CpG binding domain protein 3
MBD4	methyl-CpG binding domain protein 4
MBD5	methyl-CpG binding domain protein 5
MBD6	methyl-CpG binding domain protein 1
mC	methyl-cytosine
mCA	methyl-cytosine-adenine dinucelotide
mCAC	methyl-cytosine- adenine- cytosine trinucelotide
mCG	methyl-cytosine-guanine dinucelotide
mCH	methyl-cytosine-non-guanine dinucelotide
me	methyl
MeCP2	methyl-CpG binding protein 2
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
mL	milliliter
mM	millimolar

MSA	multiple sequence alignment
MTA1	metastasis associated 1 family member protein 1
MTA2	metastasis associated 1 family member protein 2
N-terminal, N-term	amino terminal
N, Asn	asparagine
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
NCP	nucleosome core particle
Ni	nickel
Ni^{2+}	nickel cation
NIH	National Institutes of Health
nM	nanomolar
nm	nanometer
NMR	nuclear magnetic resonance spectroscopy
NuRD	nucleosome remodeling and deacetylase complex
OD_{600}	optical density at 600 nm wavelength
oligos	oligonucleotides
P, Pro	proline
PAGE	polyacrylamide gel
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDB ID	PDB identification code
PEG	polyethylene glycol

pH	negative log (base 10) of the molar concentration of hydronium ions
PHD	plant homeodomain
F, Phe	phenylalanine
PTM	post-translational modifications
R, Arg	arginine
RBBP4	retinoblastoma-binding protein 4
RBBP7	retinoblastoma-binding protein 7
RCSB	Research Collaboratory for Structural Bioinformatics
RNA	ribonucleic acid
s^{-1}	per second
SDS	sodium dodecyl sulfate
S, Ser	serine
SANT	Swi3, Ada2, N-Cor, and TFIIIB domain
sc	single-chain
T	thymine
T, Thr	threonine
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
UNC	University of North Carolina at Chapel Hill
UNC-CH	University of North Carolina at Chapel Hill
UV	ultraviolet
V, Val	valine
W, Trp	tryptophan

WD40	WD repeat domain
WT	wild type
Y, Tyr	tyrosine
Zn	zinc
ZnF	zinc-finger domain
α	alpha
β	beta
ε	epsilon
δ	delta
Δ	change
λ	wavelength
μg	microgram
μL	microliter
μM	micromolar
μm	micrometer

CHAPTER 1: INTRODUCTION TO DNA METHYLATION AND METHYL-CpG BINDING DOMAIN PROTEINS

DNA Methylation and Epigenetics

The regulation of gene expression is a highly dynamic process, simultaneously choreographed by an array of signals from transcription factor activation to chromatin remodeling. Epigenetic regulation of transcription enables responsive changes in genetic output relative to environmental conditions. In the development of complex organisms, cells must coordinate correctly timed events and signaling processes to guide cell differentiation. During organ formation, fluctuations in the patterns of genes expressed gradually define the profile that enables the specific functions of each tissue. Even into late fetal development and the early stages of life, transcription is modulated over time to transition from the developmental to the adult proteome. This shift is often accomplished by employing DNA methylation, a heritable epigenetic modification of cytosine bases that directs long-term gene silencing.

In humans, DNA is methylated by DNA methyltransferases (DNMTs) at the 5C position of cytosine (mC), primarily in cytosine-guanine (CpG) dinucleotides. CpG dinucleotides are commonly clustered into “CpG islands” within the 5’ regulatory regions of about 60% of transcribed gene promoters. [1,2] In healthy somatic cells, the majority of CpG islands are unmodified [3], with the approximately 4% of all cytosines in the human genome being methylated in regions of non-coding DNA, repetitive elements, gene

imprinting and X-chromosome inactivation [4,5]. In addition to disrupting transcription factor binding to target sequences, cytosine methylation creates docking sites for methyl-binding proteins and their associated co-repressor complexes to prevent transcription activation [4].

Chromatin Structure and Transcription Regulation

Eukaryotic organisms organized genomic DNA into a dynamic higher ordered structure called chromatin. Chromatin is comprised of repeating units of nucleosomes, ~147 base pairs chromosomal DNA wrapped around histone protein octamers of the histone proteins H2A, H2B, H3, and H4 [6–8]. Nucleosomes assemble in an array connected by linker regions of DNA which are bound by the histone protein H1 to form chromatosomes illustrated in **Figure 1.1**[9,10]. The 1.8 meters of DNA in a cell can then be wound into higher order structures and packaged into a single cell nucleus. The linker DNA between nucleosomes varies in length and the spacing of nucleosomes regulates local genome access to the transcription machinery. The presence of nucleosomes impedes transcription factor binding and RNA Polymerase II transcription initiation [11]. Therefore, genomic areas of active transcription tend to be less densely populated with nucleosomes, known as euchromatin, while in silenced regions nucleosomes spacing is more compact, known as heterochromatin [3].

The structure of chromatic regions is influenced by histone protein post-translational modifications and nucleosome repositioning activity of chromatin remodeling enzymes. The histone octamer core of nucleosomes possesses about 80 sites for several different covalent post-translational modifications (PTMs), such as methylation, acetylation, phosphorylation, and ubiquitination, illustrated in **Figure 1.2** [12].

PTMs constitute the “histone code”, a highly dynamic pattern of epigenetic marks that can be modified by enzymes depositing (“writers”) or removing (“erasers”) PTMs. The presence or absence of a PTM can be recognized by “reader” proteins that in turn signal for the activation or repression of transcription [13]. For example, histone acetylation is heavily associated with transcriptional activation and, conversely, deacetylation correlates with the repression of gene expression [14]. Histone “reader” proteins often recruit ATP-dependent chromatin remodeling enzymes, which assemble, slide, or eject nucleosomes to compact or loosen chromatin structure [15]. The co-repressor complexes associated with methyl-DNA binding proteins often contain histone reading domains, a histone deacetylase enzyme, and a chromatin remodeling ATP-ase to create and maintain heterochromatin [16].

Methyl-CpG Binding Domain Protein Family

In humans, DNA methylation signals are read by members of the methyl-CpG binding domain (MBD) family of proteins encompassing MeCP2 and MBD1 through MBD6 [16]. The protein family members contain a homologous MBD domain of ~70 amino acids with the selectivity for methylated CpG dinucleotides (mCG) varying from a high preference (MBD1, MBD2, MeCP2) or minimal preference (MBD3, MBD4) [17] to none (MBD5, MBD6) [18]. The available crystal and solution structures of MBD proteins bound to DNA reveal mCG motifs are similarly recognized across the family (PDBs: 1IG4, 3C2I, 2KY8, 3VXX, 2MOE), employing two conserved arginines and a tyrosine in methylation-specific binding [19–23]. However, the domain architecture outside of the MBD and the associated complexes vary among family members (except MBD2/MBD3), contributing to differences in the preferred sequence context of the mC motif and their

ability to perform distinct functions (**Table 1.1**) [24]. MBD1, MBD2, and MeCP2 each contain an additional transcriptional repression domain that is thought to mediate interaction with co-repressor complexes, which include histone deacetylases and chromatin remodeling enzymes, to translate the methylation signal into a change in local genome structure, illustrated in **Figure 1.3** [16]. Additionally, MeCP2 is found at levels rivaling histone H1 and has been suggested to compact higher-order chromatin structure by competing for binding sites of the linker histone in chromatosomes [25].

MBD Proteins and Human Disease

Chromatin remodeling is a dynamic and reversible means of controlling cellular genetic output, therefore the proteins responsible for modulating nucleosome positioning are appealing therapeutic targets [26–29]. Aberrant nucleosome organization can severely disrupt gene expression, DNA repair and cellular differentiation, and is associated with disorders including schizophrenia [30], cardiovascular disease [31], intellectual disability [30,32] and cancer [33–35]. The role of DNA methylation in disease has been most extensively studied in cancer, beginning in the early 1980's. Tumor methylome profiles are globally hypomethylated, possibly to relax the repression of imprinted, viral and repetitive gene elements, while CpG islands within the promoters of tumor suppressor genes often locally hypermethylated, when compared to healthy tissue. This abnormal hypermethylation correlates with transcriptional silencing, and thereby effective inactivation, of tumor suppressor genes [36].

Consequently, the first epigenetic chemotherapies were cytosine analogs, 5-azacitidine, and decitabine, modified at the 5th carbon position to prevent methylation when incorporated into genomic DNA and inhibit DNMTs, to restore the expression of

tumor suppressors [37,38]. Clinically these compounds are approved for use in myelodysplastic syndrome and leukemia, however, their effectiveness in is limited by systemic toxicity, including myelosuppression [39]. Treating cancer continues to be the focus of the epigenetic-based drug discovery and the majority on the market (and in development) target the “writers” and “erasers” of DNA methylation and the histone code (**Table 1.2**) [40]. Epigenetic “readers” and their associated chromatin remodeling enzymes, therefore, represent a powerful class of novel drug targets that have yet to be thoroughly explored. Mutations, downregulation, and overexpression of MBD family proteins have been associated with malignancies of the breast, uterus, lung, brain and gastrointestinal system [16]. MeCP2 and MBD2 have garnered specific interest as targets in cancer and beyond for their roles in cell differentiation, growth, and maturation.

MeCP2 Association with Neurological Disorders

Methyl-CpG binding protein 2 (MeCP2) is most abundantly expressed in the brain and is critical for neural development. It is encoded on the long arm of the X-chromosome, and perturbations of protein expression levels or sequence are associated with schizophrenia, Prader-Willi syndrome, Angleman syndrome and, most notably, Rett syndrome [16]. Missense and frameshift mutations throughout MeCP2 occur in up to 80% of Rett syndrome cases, a disorder that affects about 1 in 10,000 females characterized by stagnated, sometimes regressive neurological development at 6-18 months of age [41–43]. The timing of the onset of Rett syndrome symptoms correlates with the period during which MeCP2 expression levels rise and mCG motifs accumulate in brain cells [44–46]. Recent advancements in base-resolution methylome sequencing have

revealed significant levels of non-CpG methylation also accumulate in developing brain cells and additional specificity for mC(A, T, or G) have been described for MeCP2 [47].

Transcription is globally affected by MeCP2 dysfunction, in addition to misregulation of specific genes involved in differentiation, metabolism and neuronal growth, including BDNF, GAMT, FXYD1, and DLX5 [48–50]. Rett syndrome affects cognitive, sensory, emotional, motor, and autonomic functions, including autism-like features in learning, speech and social deficits, as well as seizures, digestive problems, and irregular breathing patterns [41]. Symptoms vary significantly from patient to patient, however, some identified MeCP2 mutations can be predictive of the severity of disability [51,52]. No drugs have been approved to treat the underlying defect in MeCP2 or manipulate downstream effectors and clinical trials testing methods to minimize symptoms have not yet yielded significant results [53].

About 20% of Rett syndrome-associated mutations in MeCP2 occur in the MBD domain [54]. Investigations of missense mutations in the MBD have revealed destabilizing effects on protein structure, reductions in methylated DNA binding selectivity and impaired transcriptional repression activity [54–60]. Neurological symptoms in immature and mature conditional MeCP2 knock-out mice can be reversed upon re-activation of wild-type protein expression [61], therefore strategies to re-stabilize the MBD domain and restore methylated DNA binding function hold promise for treating this population of Rett Syndrome patients.

MBD2 Developmental Silencing of Globin Genes

Hemoglobinopathies are the world's most common diseases caused by a single gene defect. The genes of the human β-globin locus are encoded in the order which

they are expressed during development: 5′- ε - γ - δ- β 3′. Embryonic ε-globin expression by the yolk sac is silenced as γ-globin expression rises in the fetal liver, which is in turn silenced during post-natal development as hematopoiesis shifts to the bone-marrow, activating adult β-globin expression, illustrated in **Figure 1** [62]. The switch from fetal to adult hemoglobin expression is not irreversible, therefore sickle cell anemia and β-thalassemia patients suffering from mutations in the β-globin gene can receive therapeutic benefit from the induction of γ-globin production. Hydroxyurea is currently used to treat sickle cell anemia through increasing γ-globin expression, however, its efficacy varies between individuals and it is not effective for treating β-thalassemia [63]. Therefore, epigenetic mechanisms of globin gene silencing have become of interest as a novel therapeutic approach.

DNA methylation and histone modifications both play roles in developmental globin gene silencing [64–67]; and 5-azacytidine treatment increases γ-globin synthesis in sickle cell anemia and β-thalassemia patients [68,69]. Although an MBD protein has not been demonstrated to bind directly to the globin gene promoters, MBD2 and its' associated Nucleosome Remodeling and Deacetylase complex (NuRD) have been shown to play an indirect role in silencing human embryonic ε- globin and fetal γ-globin gene expression [70–73]. The NuRD co-repressor complex is composed of the histone deacetylase core complex of HDAC1/2, MTA1/2, RBBP4/7 (HDCC), GATAD2A/B and a chromatin remodeling enzyme, CHD3/4 [74–76]. Disrupting interactions of MBD2 with NuRD through the HDCC or GATAD2A/CHD4 is sufficient to release MBD2-directed transcription repression, thereby making these protein-protein interfaces of therapeutic interest [72,77].

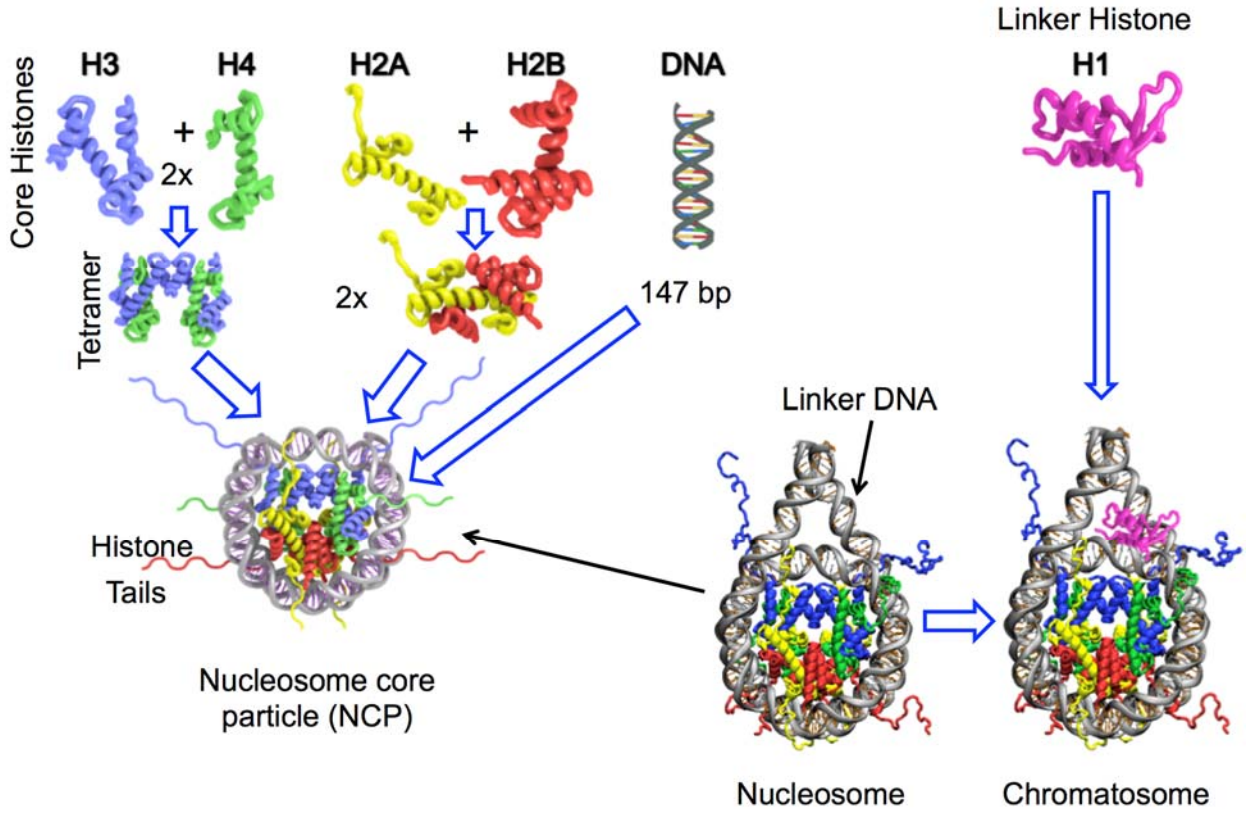

Figure 1.1 Nucleosome Structure.
The core histone proteins (H2A, H2B, H3, H4) form tetramers of heterodimers of H2A/H2B and H3/H4. These tetramers assemble into an octomer which is wrapped 1.7 times with ~147 base pairs of DNA to form the nucleosome core particle (NCP). NCPs form arrays connected by linker DNA which is bound by histone protein H1 to form higher order chromatin structures. Image reprinted with permission [9,10].

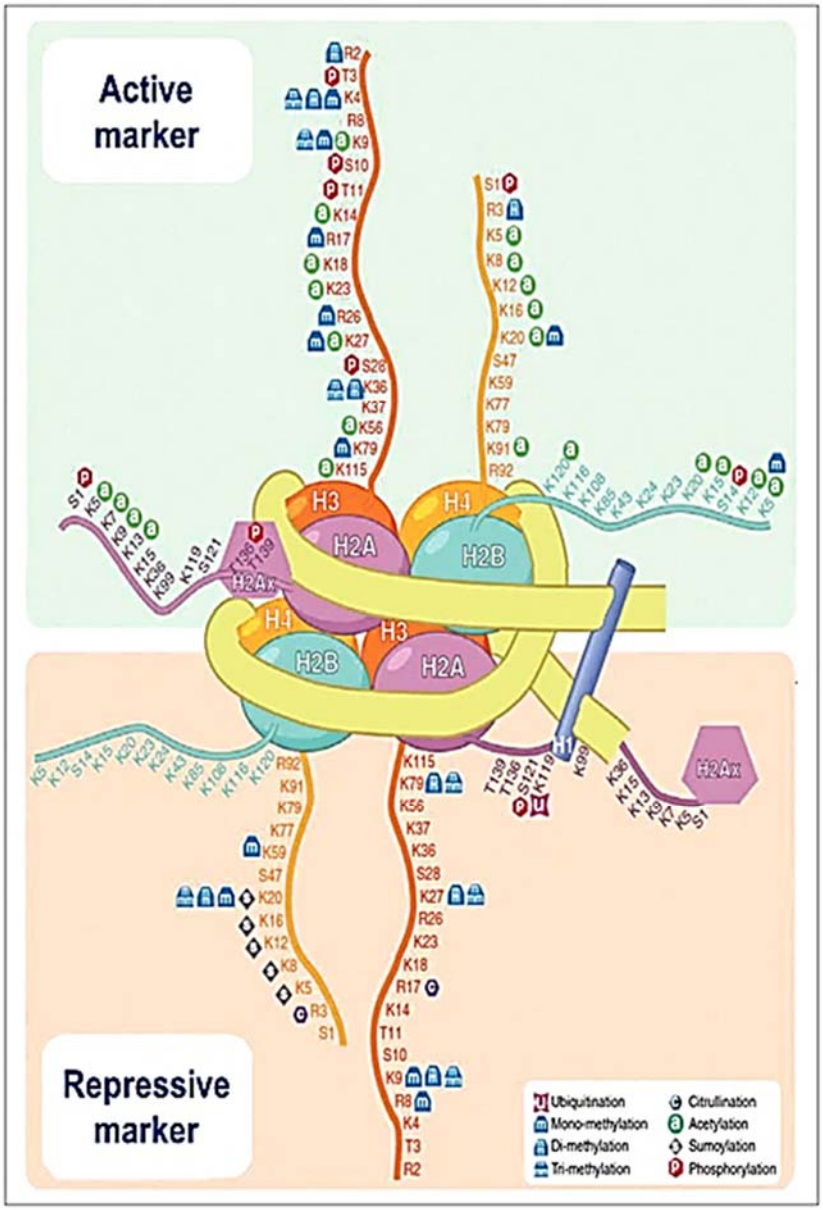

Figure 1.2 Active and Repressive Histone Modifications
N- and C-terminal histone tails extend from the globular domains of the histones H2A, H2B, H3, and H4. The tails are accessible to enzymes for post-translational covalent modification (PTM), including acetylation, methylation, phosphorylation, and ubiquitylation. Specific PTMs of certain amino acids within the tails are associated with actively transcribed or repressed regions of DNA, depicted on the top and bottom halves of the diagram, respectively. Histone tail acetylations are commonly found in active gene promoters, while sumoylation tends to repress transcriptional activity. Image reprinted with permission [78].

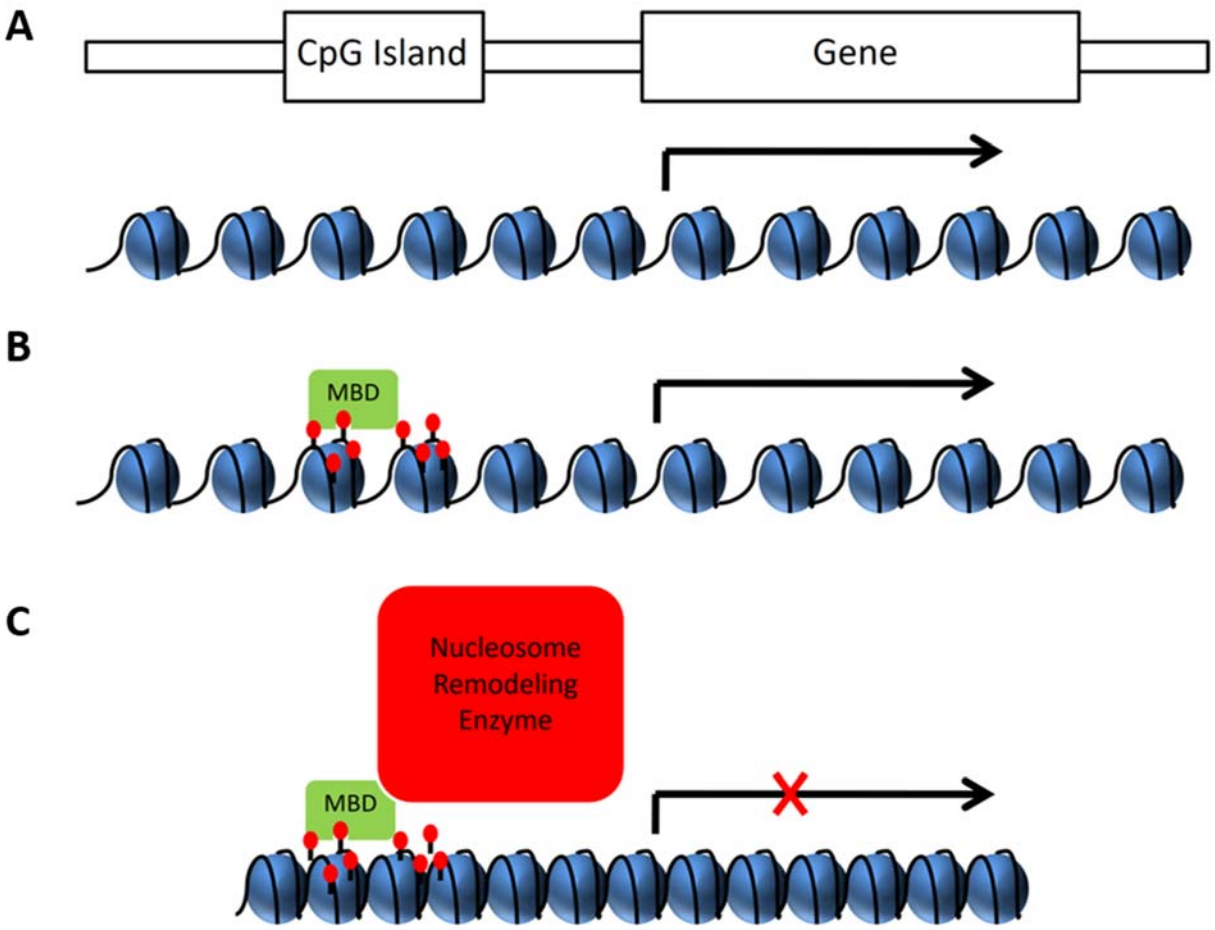

Figure 1.3 Methylation-Dependent Transcription Repression
A) DNA sequences rich in CG motifs (CpG Islands) are often found in promoter and enhancer regions upstream of coding sequences. At actively transcribed genes, nucleosomes (blue spheres) are spread out on the DNA (black line), making euchromatin accessible to transcription factors. B) When CpG islands are methylated (red circles), methyl-CpG binding proteins (MBD) bind to the DNA. C) MBD proteins recruit their associated remodeling enzymes to compact the nucleosomes into heterochromatin and repress transcription initiation.

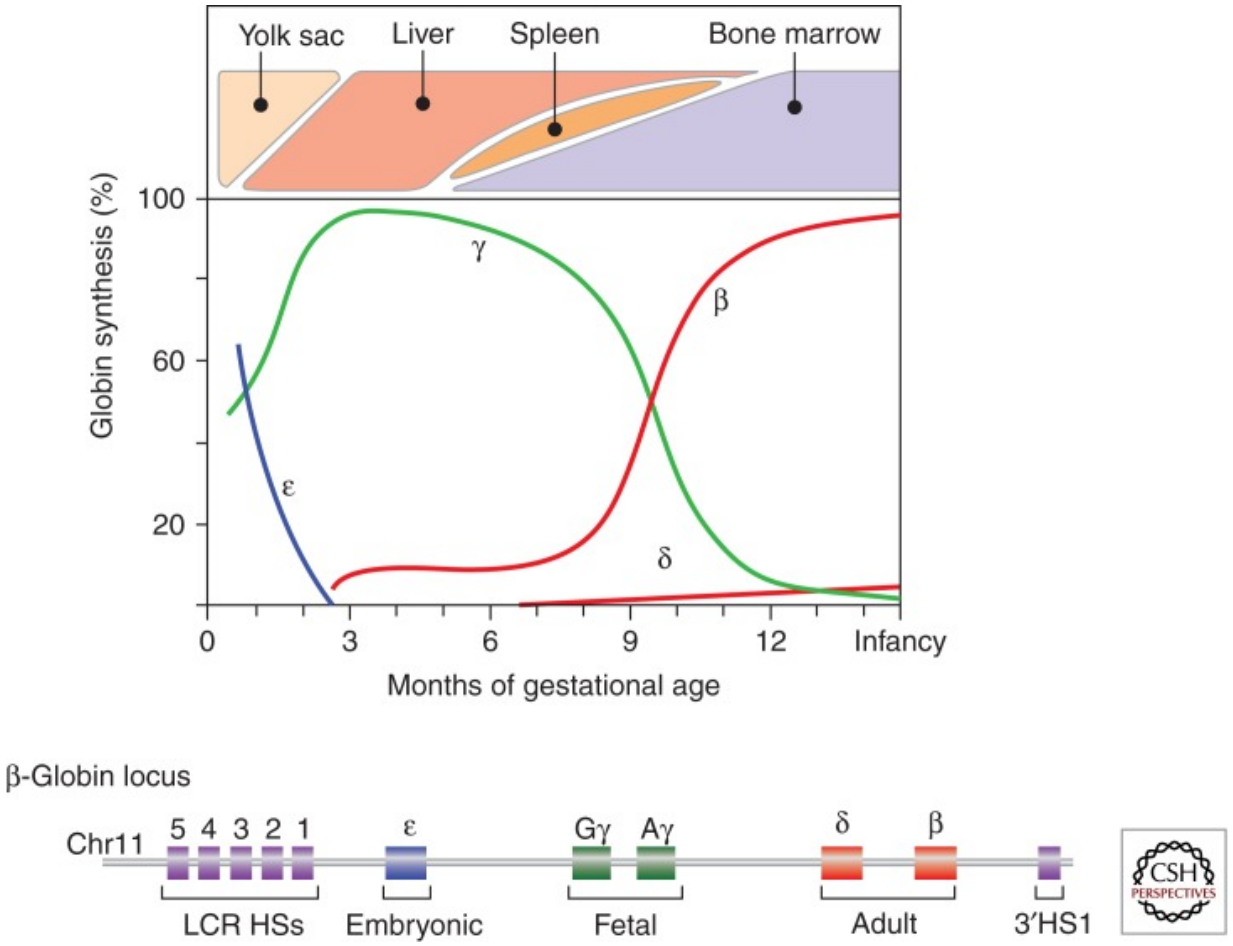

Figure 1.4 The Human Hemoglobin Switch
The globin locus on chromosome 11 is diagrammed in the bottom panel and the timing of the developmental hemoglobin switch is illustrated in the top panel, using the genes corresponding color (embryonic ε-globin in blue, fetal γ-globins in green, and adult α- and β-globins in red). The globin locus upstream enhancer, the locus control region (LCR), and its corresponding DNAse I hypersensitivity sites (HSs) and a downstream HS, known as the 3′HS1, are depicted in purple.
Image reprinted with permission [79].

Table 1.1 MBD Family Proteins Binding Preferences, Complexes, and Functions

MBD protein*	Sequence preference	Associated Complex	Function
MeCP2	$[A/T]_4$ + mCG or mCAC	Sin3A, NCoR	Histone modification and gene silencing
MBD1	TmCGCA or TGmCGCA	SETDB1 and CAF-1	Chromatin Modifying
MBD2	mCGG	NuRD	Nucleosome remodeling and histone deacetylation
MBD3	Non-specific	NuRD	Nucleosome remodeling and histone deacetylation
MBD4	mCG·TG mismatch	Thymine glycosylase domain	Mismatch repair of spontaneously deaminated methyl-cytosines

* MBD5 and MBD6 do not bind methylated DNA or have a known regulatory complex association

Table 1.2 Epigenetic enzyme drug target classes and drug discovery efforts

Epigenetic enzyme classes	Approved drugs	Clinical trials
DNA Methylation		
DNA methyltransferases	X	X
Histone Modification		
Deacetylases class I, II, IV	X	X
Deacetylases class III		
Acetyltransferases		X
Mono-ADP ribosyltransferases		
Poly-ADP ribosyltransferases		X
Biotin ligase		
Deiminases		
Glycosyltransferases/glycosidases		
Methyltransferases		X
Demethylase		X
Kinases/phosphatases		
E1, E2, and E3 enzymes		
microRNA expression		
miRNA-regulating proteins		
Chromatin Remodeling		
ATPase/Helicase		

REFERENCES

1. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3740–5 (2002).

2. Svedružić, Ž.˘M. & Reich, N. O. The mechanism of target base attack in DNA cytosine carbon 5 methylation. *Biochemistry* **43,** 11460–11473 (2004).

3. Herman, J. G. & Baylin, S. B. Gene Silencing in Cancer in Association with Promoter Hypermethylation. *N. Engl. J. Med.* **349,** 2042–2054 (2003).

4. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics {&} Chromatin* **8,** 24 (2015).

5. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* **99,** 451–454 (1999).

6. Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science (80-.).* **184,** 868–871 (1974).

7. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).

8. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423,** 145–50 (2003).

9. Shaytan, A. K., Landsman, D. & Panchenko, A. R. Nucleosome adaptability conferred by sequence and structural variations in histone H2A–H2B dimers. *Curr. Opin. Struct. Biol.* **32,** 48–57 (2015).

10. Shaytan, A. K. *et al.* Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. *J. Mol. Biol.* **428,** 221–37 (2016).

11. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology* **339,** 258–266 (2010).

12. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).

13. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).

14. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).

15. Piatti, P., Zeilner, A. & Lusser, A. ATP-dependent chromatin remodeling factors

and their roles in affecting nucleosome fiber composition. *International Journal of Molecular Sciences* **12,** 6544–6565 (2011).

16. Du, Q., Luu, P.-L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7,** 1051–73 (2015).

17. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–9 (2012).

18. Laget, S. *et al.* The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One* **5,** e11982 (2010).

19. Ho, K. L. *et al.* MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).

20. Ohki, I. *et al.* Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105,** 487–497 (2001).

21. Scarsdale, J. N., Webb, H. D., Ginder, G. D. & Williams, D. C. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* **39,** 6741–52 (2011).

22. Walavalkar, N. M., Cramer, J. M., Buchwald, W. A., Scarsdale, J. N. & Williams Jr., D. C. Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic Acids Res.* **42,** 11218–11232 (2015).

23. Otani, J. *et al.* Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.* **288,** 6351–6362 (2013).

24. Fatemi, M. & Wade, P. a. MBD family proteins: reading the epigenetic code. *J. Cell Sci.* **119,** 3033–3037 (2006).

25. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Shlyakhtenko, L. S. & Woodcock, C. L. MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. *Mol. Cell. Biol.* **30,** 4656–70 (2010).

26. Blancafort, P., Jin, J. & Frye, S. Writing and rewriting the epigenetic code of cancer cells: from engineered proteins to small molecules. *Mol. Pharmacol.* **83,** 563–76 (2013).

27. Lockett, G. A. *et al.* Epigenomics and allergic disease. *Epigenomics* **5,** 685–99 (2013).

28. Qureshi, I. A. & Mehler, M. F. Epigenetics and therapeutic targets mediating neuroprotection. *Brain Res.* **1628,** 265–272 (2015).

29. Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. TL - 20. *Nat. Struct. Mol. Biol.* **20,** 274–281 (2013).

30. McCarthy, S. E. *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19,** 652–8 (2014).

31. Bevilacqua, A., Willis, M. S. & Bultman, S. J. SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease. *Cardiovascular Pathology* **23,** 85–91 (2014).

32. López, A. J. & Wood, M. A. Role of nucleosome remodeling in neurodevelopmental and intellectual disability disorders. *Front. Behav. Neurosci.* **9,** 100 (2015).

33. Malik, K. & Brown, K. W. Epigenetic gene deregulation in cancer. *Br. J. Cancer* **83,** 1583–8 (2000).

34. Kumar, R., Li, D.-Q., Müller, S. & Knapp, S. Epigenomic regulation of oncogenesis by chromatin remodeling. *Oncogene* **35,** 1–14 (2016).

35. Sperlazza, J. *et al.* Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation. *Blood* **126,** 1462–1472 (2015).

36. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301,** 89–92 (1983).

37. Glover, A. B. & Leyland-Jones, B. Biochemistry of azacitidine: a review. *Cancer Treat Rep* **71,** 959–964 (1987).

38. Ferguson, A. T. *et al.* Role of estrogen receptor gene demethylation and DNA methyltransferase.DNA adduct formation in 5-aza-2'deoxycytidine-induced cytotoxicity in human breast cancer cells. *J. Biol. Chem.* **272,** 32260–6 (1997).

39. Götze, K., Müller-Thomas, C. & Peschel, C. The role of azacitidine in the management of myelodysplastic syndromes (MDS). *Cancer Manag. Res.* **1,** 119–30 (2009).

40. Valdespino, V. & Valdespino, P. M. Potential of epigenetic therapies in the management of solid tumors. *Cancer Manag. Res.* **7,** 241–251 (2015).

41. Bienvenu, T. *et al.* MECP2 mutations account for most cases of typical forms of Rett syndrome. *Hum. Mol. Genet.* **9,** 1377–84 (2000).

42. Cheadle, J. P. *et al.* Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location.

Hum. Mol. Genet. **9,** 1119–29 (2000).

43. Huppke, P., Laccone, F., Krämer, N., Engel, W. & Hanefeld, F. Rett syndrome: analysis of MECP2 and clinical characterization of 31 patients. *Hum. Mol. Genet.* **9,** 1369–75 (2000).

44. Mullaney, B. C., Johnston, M. V. & Blue, M. E. Developmental expression of methyl-CpG binding protein 2 is dynamically regulated in the rodent brain. *Neuroscience* **123,** 939–949 (2004).

45. Shahbazian, M. D., Antalffy, B., Armstrong, D. L. & Zoghbi, H. Y. Insight into Rett syndrome: MeCP2 levels display tissue-and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.* **11,** 115–124 (2002).

46. Skene, P. J. *et al.* Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Mol. Cell* **37,** 457–468 (2010).

47. Kulis, M. *et al.* Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet* **47,** 746–756 (2015).

48. Bienvenu, T. & Chelly, J. Molecular genetics of Rett syndrome : when DNA methylation. *Nat. Rev. Genet.* **7,** 415–426 (2006).

49. Nuber, U. A. *et al.* Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Hum. Mol. Genet.* **14,** 2247–2256 (2005).

50. Deng, V. *et al.* FXYD1 is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Hum. Mol. Genet.* **16,** 640–650 (2007).

51. Cuddapah, V. A. *et al.* Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J. Med. Genet.* **51,** 152–8 (2014).

52. Neul, J. L. *et al.* Specific mutations in Methyl-CpG-Binding Protein 2 confer different severity in Rett syndrome. *Neurology* **70,** 1313–1321 (2008).

53. Wang, H., Pati, S., Pozzo-Miller, L. & Doering, L. C. Targeted pharmacological treatment of autism spectrum disorders: fragile X and Rett syndromes. *Front. Cell. Neurosci.* **9,** 55 (2015).

54. Kucukkal, T. G., Yang, Y., Uvarov, O., Cao, W. & Alexov, E. Impact of Rett Syndrome Mutations on MeCP2 MBD Stability. *Biochemistry* **54,** 6357–6368 (2015).

55. Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Gierasch, L. M. & Woodcock, C. L. Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J. Biol. Chem.* **283,**

20523–20534 (2008).

56. Ballestar, E., Yusufzai, T. M. & Wolffe, A. P. Effects of Rett Syndrome Mutations of the Methyl-CpG Binding Domain of the Transcriptional Repressor MeCP2 on Selectivity for Association with Methylated DNA. *Biochemistry* **39,** 7100–7106 (2000).

57. Yusufzai, T. M. & Wolffe, A. P. Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res.* **28,** 4172–9 (2000).

58. Goffin, D. *et al.* Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat. Neurosci.* **15,** 274–83 (2012).

59. Kudo, S. *et al.* Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.* **40,** 487–93 (2003).

60. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25,** 558–570 (2016).

61. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science (80-.).* **315,** 1143–1147 (2007).

62. Stamatoyannopoulos, G. Control of globin gene expression during development and erythroid differentiation. *Exp Hematol* **33,** 259–271 (2005).

63. Quarmyne, M.-O. *et al.* Hydroxyurea effectiveness in children and adolescents with sickle cell anemia: A large retrospective, population-based cohort. *Am. J. Hematol.* **0,** 1–5 (2016).

64. Singal, R., Ferris, R., Little, J. A., Wang, S. Z. & Ginder, G. D. Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13724–9 (1997).

65. Pikaart, M. J., Recillas-Targa, F. & Felsenfeld, G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev.* **12,** 2852–2862 (1998).

66. Forsberg, E. C. *et al.* Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 14494–14499 (2000).

67. Mabaera, R. *et al.* Developmental- and differentiation-specific patterns of human gamma- and beta-globin promoter DNA methylation. *Blood* **110,** 1343–52 (2007).

68. Ley, T. J. *et al.* 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta-thalassemia. *N. Engl. J. Med.* **307,** 1469–1475 (1982).

69. Charache, S. *et al.* Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the gamma-delta-beta-globin gene complex. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4842–4846 (1983).

70. Rupon, J. W., Wang, S. Z., Gaensler, K., Lloyd, J. & Ginder, G. D. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6617–6622 (2006).

71. Rupon, J. W., Wang, S. Z., Gnanapragasam, M., Labropoulos, S. & Ginder, G. D. MBD2 contributes to developmental silencing of the human ε-globin gene. *Blood Cells, Mol. Dis.* **46,** 212–219 (2011).

72. Gnanapragasam, M. N. *et al.* p66α–MBD2 coiledcoil interaction and recruitment of Mi2 are critical for globin gene silencing by the MBD2–NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7487–7492 (2011).

73. Amaya, M. *et al.* Mi2 -mediated silencing of the fetal -globin gene in adult erythroid cells. *Blood* **121,** 3493–3501 (2013).

74. Kloet, S. L. *et al.* Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.* **282,** 1774–1785 (2015).

75. Tenenhaus, C., Subramaniam, K., Dunn, M. A. & Seydoux, G. PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of Caenorhabditis elegans. *Genes Dev.* **15,** 1031–1040 (2001).

76. Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26,** 843–51 (2006).

77. Desai, M. A. *et al.* An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* **43,** 3100–3113 (2015).

78. Kim, Y. Z. Altered histone modifications in gliomas. *Brain tumor Res. Treat.* **2,** 7–21 (2014).

79. Sankaran, V. G., Xu, J. & Orkin, S. H. Advances in the understanding of haemoglobin switching: Review. *British Journal of Haematology* **149,** 181–194 (2010).

CHAPTER 2: STRUCTURAL BASIS OF MECP2 DISTRIBUTION ON NON-CPG METHYLATED AND HYDROXYMETHYLATED DNA [1]

Introduction

As discussed in Chapter 1, methyl-CpG binding protein 2 (MeCP2) functions as a methylation-dependent transcriptional regulator, tuning the expression of methylated target genes throughout neuron maturation. Although MeCP2 was initially identified by its affinity for 5-methylcytosines in CpG (mCG) dinucleotides, advancements in methylation sequencing methods have enabled the examination of the developing neuronal methylome with greater resolution of non-CG methylated sites (mCH, where H = A, C, or T) and locations of hydroxymethylation. The results of these studies suggest the MeCP2 activity most relevant to development may, in fact, be at non-CG methylated sites. Healthy human and murine neurons accumulate mCH and hydroxymethylcytosine (hmC) marks while maintaining relatively stable levels of mCG from fetal to adult development [1]. MeCP2 binding correlates with mCH density and has higher occupancy of genes with elevated mCH levels. These loci appear to have functional roles in regulating neuronal gene expression as mouse models of MeCP2 disorders demonstrate elevated mCH levels in misregulated genes [2]. The most biologically significant MeCP2 mCH binding events appear to be to mCA and hmCA. By EMSA,

[1] This chapter adapted from a manuscript accepted in the Journal of Molecular Biology.

MeCP2 demonstrates a lower affinity for mCC, mCT, hmCG and unmethylated DNA, than sequences containing mCG, mCA and hmCA sites [3,4]. Additionally, ChIP-seq analysis of the murine frontal cortex revealed high mCA and hmCA levels at loci of MeCP2 enrichment, suggesting they serve as MeCP2 binding sites *in vivo* [5].

The report of MeCP2 mCH binding events holds important implications for the structure-function relationship of methylation-dependent DNA binding. Seven proteins comprise the human methyl-CpG binding domain (MBD) family, the first discovered, MeCP2, and MBD1-6 [6]. The protein family members share a homologous MBD domain of about seventy amino acids with varying degrees of specificity for symmetrically methylated mCG dinucleotides [7]. The available crystal and solution structures of MBD proteins bound to DNA reveal mCG motifs are similarly recognized across the family (PDBs: 1IG4, 3C2I, 2KY8, 3VXX) [8–11]. Two conserved arginine residues form bidentate hydrogen bonds with the symmetrical guanine bases and pack their guanidinium groups against the methyl-cytosine on the respective strand. A highly-conserved tyrosine at the protein-DNA interface interacts with the methyl-cytosine methyl group on one strand and is critical to forming a highly-specific interaction with mCG DNA (Figure 2.1A) [9]. The genetic, biochemical and biophysical properties of mCG and hmCG binding by MBD proteins have been extensively studied [7,12–16]. MeCP2, however, is the first MBD protein with demonstrated selectivity for asymmetrically methylated mCH and hmCH [2–5]. The reported evidence even suggests binding methylated and hydroxymethylated CpA (h/mCA) dinucleotides mediates the primary effects of MeCP2 on gene expression in the developing mammalian brain. Therefore, we sought to understand the features that

enable binding these ‘alternative’ motifs, if they are unique to MeCP2 and how they may affect the distribution of MBD proteins across the genome.

In a few sequence-specific contexts, MBD-containing proteins have an enhanced selectivity for mCG sites. MBD2 has been shown to have an increased affinity for mCGG [10], while SELEX experiments with MeCP2 mCG suggested a preference for an adjacent [A/T] run of at least four bases [17]. In the crystal structure (PDB: 3C2I) [9], the MeCP2:mCG complex assembles such that the visible c-terminal residues of the MBD make contacts with the phosphate backbone of the AATT motif in the BDNF promoter sequence. We hypothesized this flanking A/T region serves to orient the MeCP2 MBD on the mCH site in one direction, positioning Arg^{111} and Arg^{133} to interact with a guanine on a particular strand. Previous work has suggested the conserved MBD arginines contribute unequally to DNA binding [18–20]. Therefore, we anticipated the binding of asymmetrically methylated mCA and hmC sites would be a strand-specific recognition event and asked if that could be differentially affected by Rett Syndrome-associated mutations.

Here, we investigated the interactions of MeCP2, and the related protein MBD2, with a defined single mCG or mCA site in a native high-affinity target sequence (Figure 2.1B and C). We employed nuclear magnetic resonance spectroscopy and isothermal calorimetry to obtain robust, quantifiable binding measurements and elucidate the molecular details of the accommodation of mCH. As predicted, both MeCP2 and MBD2 better recognize an adenine substitution of the 3’-strand guanine (mCA-3) than on the 5’-strand (mCA-5). MeCP2, however, uniquely maintains a high affinity for mCA-3 possibly reflecting the dynamic flexibility of the bound state. The strand specificity is

strongly affected by local sequence context and not primarily determined by the direction of the AATT motif, as MeCP2 demonstrated very similar affinities for mCAC on either stand. Additionally, we examined the effects of Rett syndrome-associated mutations in the MeCP2 MBD on the ability to recognize methylation of CG and CA. While the mutations consistently destabilized the folded domain, we found at least three of the more common mutations (T158M, P101S, and A140V) retain recognition of both mCG and mCA sites. Finally, we compared MeCP2 binding to hemi-hmCG and hmCA, and demonstrate the effects of hmC on the distribution among local binding sites, giving insight into a potential mechanism of regulating MeCP2 directed transcriptional control.

Materials and Methods

Cloning, Expression, and Purification of the MeCP2 Methyl Binding Domain

Amino acid residues 77–167 from human MeCP2 was cloned into the pet28a (Novagen) vector, untagged. Sequence confirmed constructs were transformed into *Escherichia coli* BL21 Rosetta (DE3) cells (Novagen). Bacterial cultures were grown in 1 L of lysogeny broth (LB) or isotopically labeled M9 media, in the presence of kanamycin, at 37 °C with shaking, until the culture reached an A_{600} of 0.6 absorbance units. Protein expression was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 2-4 h. Cells were spun at 4,000 x *g* for 30 minutes and pellets were stored at -80 °C. Pellets were resuspended in 30 mL of lysis buffer [20 mM Tris-HCl (pH 8.0), 6 M Guanidine HCl] and homogenized by sonication. Cell extracts were centrifuged at 20,000 × *g* for 15 min. The supernatant was dialyzed against 4 L of refolding buffer [20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM 2-mercaptoethanol] for 16 hr at room temperature. Precipitates were removed by centrifugation at 20,000 × *g* for 15 minutes. The

supernatant was incubated with 35% saturated ammonium sulfate by mixing for 20 minutes at room temperature and precipitates were removed by centrifugation at 10,000 × *g* for 15 minutes. MeCP2 was precipitated from the supernatant with 70% ammonium sulfate, slowly added and mixed for 20 minutes, then centrifuged at 20,000 × *g* for 15 minutes. MeCP2 protein was resuspended in gel filtration buffer [20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1mM dithiothreitol]. The protein solution was syringe filtered and loaded onto a pre-equilibrated HiLoad 26/600 Superdex 200 column (GE Healthcare). Fractions containing protein as assessed by UV absorbance were combined and flowed over a MonoS 10/100 GL column (GEHeathcare), pre-equilibrated with 20 mM Tris-HCl (pH 8.0). Bound protein was eluted with a linear gradient of elution buffer [20 mM Tris-HCl (pH 8.0), 1 M NaCl]. Fractions containing protein as assessed by UV absorbance were analyzed via SDS-PAGE and those containing >95% pure MeCP2 protein were combined and stored at 4 °C.

MeCP2 Rhett Syndrome-associated mutations (T158M, R106W, P101S, or A140V) were introduced to the MBD construct by site-directed mutagenesis with the QuikChange Lightning Kit (Agilent). Mutant proteins were expressed and purified as described above.

MeCP2 Rhett Syndrome-associated mutations (T158M, R106W, P101S, or A140V) were introduced to the MBD construct by site-directed mutagenesis with the QuikChange Lightning Kit (Agilent). Mutant proteins were expressed and purified as described above.

Cloning, Expression, and Purification of the MBD2 Methyl Binding Domain

Amino acid residues 2–72 from chicken MBD2 were previously cloned by this lab as a fusion protein with thioredoxin and a hexahistidine N-terminal tag using a modified pET32a (Novagen) vector [10]. Sequence confirmed constructs were transformed into *Escherichia coli* BL21 Rosetta (DE3) cells (Novagen). Bacterial cultures were grown in 1 L of lysogeny broth (LB) or isotopically labeled M9 media, in the presence of ampicillin, at 37 °C with shaking, until the culture reached an A_{600} of 0.6 absorbance units. Protein expression was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 2-4 h. Cells were spun at 4,000 x g for 30 minutes and pellets were stored at -80 °C. Individual pellets were resuspended in 30mL of Bacterial Protein Extraction Reagent (B-PER) (ThermoFisherScientific) along with protease inhibitor tablets (Roche). Cells were homogenized by sonication and the resulting extract was centrifuged at 20,000 × *g* for 15 min. The soluble fraction was filtered and flown over a 5 mL HisTrap column (GE Healthcare), pre-equilibrated with Buffer A [20 mM Tris HCl (pH 8.0), 300 mM NaCl, 30 mM imidazole]. The column and bound protein were washed with 5 column volumes (CV) of Buffer A, then eluted with Buffer B [20 mM Tris HCl (pH 8.0), 300 mM NaCl, 300 mM imidazole]. Fractions containing protein as assessed by UV absorbance were combined and buffer exchanged into 20 mM Tris HCl (pH 8.0). The protein solution was flowed over a Resource-S column (GE Life Sciences) and eluted with a linear gradient of elution buffer [20 mM Tris-HCl (pH 8.0), 1 M NaCl]. Fractions containing protein as assessed by UV absorbance were combined and applied to a HiLoad 26/600 Superdex 75 column (GE Healthcare). Fractions containing protein as assessed by UV absorbance were analyzed via SDS-PAGE and those containing >95% pure MBD2 protein were combined and stored at 4 °C.

DNA Preparation

The DNA sequence used for binding studies was derived from the BDNF promoter, known to be a native target sequence for MeCP2. Nineteen base complementary oligonucleotides with a central methylated cytosine (Integrated DNA Technologies) or hydroxymethylated cytosine (Midland Inc) were purchased and resuspended in 20 mM Tris HCl (pH 8.0). Complementary strands were annealed by heating to 98 °C for one minute and then cooling the solution to room temperature. Double stranded DNA was isolated by purification over a MonoQ 10/100 GL ion exchange column (GE Healthcare).

NMR Sample Preparation

Isotopically labeled MeCP2 or MBD2 purified protein was combined with 10% excess double stranded oligonucleotide, or the amount calculated based on measured binding affinity to reach the desired fraction bound. The complex was buffer exchanged into 10 mM NaPO4, pH 6.5, 1 mM DTT, 10% $^{2}H_2O$ and concentrated to 0.2-1 mM. Samples were stored at 4 °C until time of analysis.

NMR Structure Assignment and Binding Experiments

Spectra from standard experiments for resonance assignments, distance, and torsional angle restraints were collected at 25°C with Bruker Avance III NMR spectrometers equipped with a cryogenic probe and operated at a ^{1}H frequency of 500, 700 or 850 MHz. The data were processed with the NMR-Pipe program [21] and analyzed with the CcpNMR Analysis program [22]. For both the free and DNA-bound MeCP2 proteins, backbone resonance was assigned based on 3D HNCO, HNCA, CBCA(CO)NH, and HNCACB spectra using standard techniques. Side-chain ^{1}H and ^{13}C

resonances were assigned based on 3D HBHA(CO)NH, HCCH-TOCSY, H(CCO)NH and C(CO)NH spectra [23]. The resonance assignment data were deposited to Biological Magnetic Resonance Bank with the accession number 26978 for the DNA-bound protein.

For chemical shift perturbation (CSP) measurements, assigned peaks in the mCG-bound protein HSQC spectrum were transferred to mCH-bound HSQCs and confirmed by 15NOESY. CSPs and peak heights were calculated with the CcpNMR Analysis program and plotted with pro Fit 7.0.9 (Quansoft).

Relaxation dispersion

Relaxation dispersion spectra (REFS) were collected at 25°C on instruments operating at ^{1}H frequencies of 500, 700, and 850 MHz using a constant time CPMG delay of 60 ms [24,25]. A total of 13 pulse frequencies (ν_{CPMG}) ranging from 67 to 1800 Hz, with a repeat of two frequencies for error analysis, were collected in an interleaved fashion with heating compensation. Data was analyzed using the relax-nmr software [26,27] and fit with either the CR72 [28] or TSMFK01 [29] 2-site models. Model selection was based on the AIC technique [30] as implemented in relax-nmr [31].

Isothermal Calorimetry Binding Affinity

Binding affinities were determined by isothermal calorimetry analysis on a MicroCal Auto-iTC200 (Malvern). Proteins and DNA were purified as described above and individually buffer exchanged into ITC buffer [10mM potassium phosphate (pH 6.0), 100mM KCl, and 1mM EDTA]. Protein, DNA, and additional ITC buffer were each filtered and degassed under vacuum immediately prior to analysis. In a 37 ºC sample

cell, 400 µL of DNA (10-30 µM) was titrated with 20 injections of 2 µL of protein (100-300 µM) with intervals of 120 s between successive injections. Binding isotherms were generated by plotting the corrected heats of binding against the molar ratio of the protein to DNA. Data plots were fit to a one-site binding model and used to calculate dissociation constants (Kd), enthalpies of binding (ΔH), stoichiometry (n), and entropy of binding (ΔS) with the manufacturer's software Origin 7.0.

Dynamic Light Scattering Protein Aggregation Studies

MeCP2 A140V mutant MBD protein was purified and concentrated in buffer [10mM potassium phosphate (pH 6.0), 1 mM EDTA] with increasing concentrations of KCl (100, 300, and 600 mM). The hydrodynamic radii of A140V protein in each condition was measured by a DynoPro DLS system (Wyatt Technology Corporation). All samples and buffers were filtered through 0.2 µM filters (Millipore) and centrifuged at 14,000 × g for 2 min before measurement. Three replicates were performed for each sample. The hydrodynamic radii and molecular weights of samples were estimated using the assumption of globular protein shape.

Circular Dichroism Secondary Structure Analysis

Purified MeCP2 MBD and MBD2 MBD protein was buffer exchanged into 100 mM sodium phosphate pH 7.5 and concentrated to 0.1 mg/ml. CD spectra were collected from 190 to 260 nm (0.2 nm interval, 24 nm/min, 0.1 cm path length, 25 °C) on a Chirascan Plus CD spectrometer (Applied Photophysics). CD spectra were normalized to provide mean residue molar ellipticity in degrees cm^{-2} $dmol^{-1}$ $residue^{-1}$. Helical content for each peptide was calculated from the ratio of the observed molar

ellipticity at 222 nm to the expected for 100% helix as given by $40{,}000 \times \left((n-4) \div n\right)$, where n is the number of amino acid residues [32,33].

Protein Structure Modeling

Text coordinate files were obtained from the RSCB Protein Data Bank (www.rcsb.org) [34] and modified to alter DNA bases in the MeCP2 binding pocket. Images were produced with PyMOL [35].

Results

Thermodynamic analysis of MeCP2 and MBD2 MBD-mCH interactions

We asked if the binding of asymmetrically methylated mCA and hmC sites by MeCP2 would be a strand-specific recognition event and if it is unique to MeCP2 among the MBD family of proteins. We compared the binding affinities of oligonucleotides based on a native MeCP2 binding site in the BDNF promoter with either a symmetrically methylated mCG, or an asymmetrically methylated mCA, substituting an adenine for the guanine on either the 5' or 3' strand. Models are shown of the expected interactions for each interface based on the x-ray crystal structure of the MeCP2:mCG complex (PDB: 3C2I) (Figure 2.1A-C) [9]. Binding thermodynamics and dissociation constants were measured by isothermal calorimetry (ITC). MeCP2 binds the BDNF promoter sequences with a stoichiometry of 1:1, regardless of the methylated cytosine position. MeCP2 binding affinity has an apparent K_d value of 50 nM for mCG and 81 nM for mCA-3, and has a significantly lower affinity for mCA-5 with a K_d value of 409 nM (Figure 2.2A-C and Table 2.1). The differences in binding affinity of MeCP2 for the mCA-5 and mCA-3 oligomers reflect an unfavorable change in entropy upon binding ($-T\Delta S$ = 2.8 and 7.7 kcal/mol for mCA-5 and m-CA3, respectively) that is or is not fully

compensated by a more favorable change in enthalpy (ΔH = −11.9 and -17.8 kcal/mol for mCA-5 and m-CA3, respectively).

However, when the base following the mCA is replaced with a cytosine, high affinity binding is maintained for mCAC-3 with a K_d value of 36 nM and the affinity increases for mCAC-5 a K_d value of 79 nM (Figure 2.3 and Table 2.1). As with mCA-3, a large unfavorable change in entropy for mCAC-3 and mCAC-5 (−TΔS = 5.7 and 8.0 kcal/mol, respectively), is compensated by a larger favorable change in enthalpy (ΔH = −16.3 and -18.1 kcal/mol, respectively). These results show that MeCP2 maintains high affinity binding to mCA-3, mCAC-3, and mCAC-5 through the development of additional enthalpically favorable interactions at the expense of a greater reduction in entropy. In contrast, MeCP2 does not effectively adapt to mCA-5 with a less favorable change in enthalpy as compared to mCpG (ΔH = −11.9 kcal/mol) with a similar change in entropy (−TΔS = 2.8 kcal/mol).

For comparison, we assessed mCA recognition by MBD2 (Figure 2.4), which shows very high selectivity for mCG [13,36] and represents the most evolutionarily ancient MBD protein [37,38]. We found MBD2 has a slightly lower affinity than MeCP2 for the mCG BDNF sequence, with an apparent a K_d value of 157 nM (Figure 2.2D). Unlike MeCP2, however, the binding affinity decreased substantially with the substitution of adenine on the 3'-strand. The ITC measurements presented an apparent K_d value of at least 1.5 μM (Figure 2.2E), however, the isotherm of binding did not reach a plateau, and therefore was unable to be adequately fit with a one-site model. The failure to accomplish binding site saturation, regardless of increased DNA concentration, indicates a low-affinity and non-specific interaction.

Chemical shift perturbations map the transition from the non-specific to methylation-specific binding mode

The NMR chemical shift parameter is highly sensitive to changes in the local environment and, thereby, subtle changes in the three-dimensional structure of proteins. The residue-specific changes in the amide backbone of MeCP2 bound to the BDNF promoter sequences were monitored by $^{1}H/^{15}N$-Heteronuclear Single-Quantum Correlation (HSQC) NMR spectroscopy and compared using standard chemical shift perturbations (CSPs). In NMR studies of the MBD family proteins, we identified linear chemical shift changes that correlate with methylated DNA binding and present in every ortholog and paralog examined to date [10,38–40]. In particular, two of these resonances (Gly^{27} and Ala^{30} of MBD2) show large CSPs between the non-specific and methylation-specific binding modes, such that we have used these resonances as reporters of mCG binding [39]. We assigned the MeCP2 chemical shifts (BMRB: 26978) and likewise found the signature linear CSP of Gly^{114} and Ala^{117} upon binding methylated DNA (Figure 2.5). To confirm that the chemical shift changes reflect the bound state, we collected HSQCs for MeCP2:CpG at 200:220, 400:400, and 400:800 μM which show that MeCP2 is fully bound under the conditions of our analysis (Figure 2.6). Therefore, these shifts provide a unique opportunity to directly observe the domain's non-specific and methylation-specific binding modes.

We compared the positions MeCP2 Gly^{114} and Ala^{117} or MBD2 Gly^{27} and Ala^{30} peaks when fully bound to mCA-5 and mCA-3. The chemical shifts for both residues were almost equivalent in the MeCP2:mCG and MeCP2:mCA-3 spectra, but in the MeCP2:mCA-5 complex Gly^{114} shifted from 102.8 ppm downfield in ^{15}N to 103.5 ppm and Ala^{117} shifted downfield in ^{1}H from 6.8 ppm to 6.9 ppm. Based on the measured

binding affinity of MeCP2 for mCA-3 and mC-5, MeCp2 is saturated (99.6% and 98.3% bound, respectively) under the conditions of the experiment, such that the CSPs reflect the bound state. Similar to the CSPs observed in MeCP2, binding to mCA-5 shifted MBD2 resonances for Gly^{27} downfield from 102.5 ppm in ^{15}N to 106.2 ppm and Ala^{30} from 6.85 to 7.15 ppm in ^{1}H. Surprisingly, Gly^{27} and Ala^{30} also had large CSPs with an adenine substituted on the 3'-strand, shifting Gly^{27} downfield in ^{15}N to 104.5 ppm and Ala^{30} to 7.00 ppm in ^{1}H (Figure 2.5). Compared to mCG bound chemical shifts of MeCP2, mCA-5 and mCA-3 binding generated CSPs throughout the MBD, plotted in Figure 2.7A. The largest perturbations clustered in three regions of the protein sequence, seen mapped to the crystal structure in Figure 2.7B (PDB: 3C2I) [9].

Domain structures accommodating of mCA are dynamic in the mCG bound conformation

Relaxation dispersion experiments enable the quantitation of protein motions, detecting conformational exchange of dynamic residues. Motions in the MeCP2:mCG complex were measured with Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion pulse sequences. Data was analyzed by relax-nmr software (see representative relaxation curves in Figure 2.8A) and fit with either the CR72 or TSMFK01 2-site model [26,27,41]. Residues exchanged on different timescales and could not be uniformly fit to a single model. Therefore, the residues with faster exchange rates were fit to the CR72 model [28] and those with very slow exchange rates, within the range of millisecond to second time scale, were fit to the TSMFK01 model [29]. The extracted fit parameters for MeCP2 are listed in Table 2.3 and 2.4, for the CR72 and TSMFK01 models, respectively. A large fraction of the well-resolved peaks displayed dispersion and residues with observable exchange are mapped to the MeCP2:mCG complex crystal

structure in Figure 2.8B (PDB: **3C2I**) [9]. Regions displaying dynamic motion were the flexible loops at the n-terminus (R85, G92, M94, E102, T105) and the c-terminus (N153, D157, T160, G161, R162); the major-groove binding loop and DNA interface (R111, K112, G114, R115, R133, S134, V136, E137); and residues in the hydrophobic core (V122, L124, I125, K130, A131, F132). Residues with slower exchange are highlighted as yellow spheres in Figure 2.8B and clustered in the C-terminal loop, α-helix, and β-loop.

The extracted fit parameters for MBD2 are listed in Table 2.5 and 2.6, for the CR72 and TSMFK01 models, respectively. A smaller fraction of the well-resolved peaks displayed dispersion and residues with observable exchange are mapped to the MBD2:mCG complex solution structure in Figure 2.8C (PDB: 2KY8) [10]. The majority of residues displaying dynamic motion were found to be surface exposed (E20, E21) or in flexible loops (D3, K4, L28, A58, D60, L61); and in the hydrophobic core (S33, D121, V122, Y37, F45, S47, K48, Q50, Y54. Only two were found at the DNA binding interface (K44 and R46). Residues with slower exchange are highlighted as yellow spheres in Figure 2.4C and clustered in the C-terminal loop, α-helix, and β-loop.

Rett Syndrome mutations do not affect MBD preference for mCG and mCA

Several of the Rett Syndrome-associated mutations (T158M, R016W, and P101S) were examined for recognition of mCG and mCA-3 by HSQC. All three mutants had less defined and dispersed spectra than wild-type, to varying degrees. The spectrum for the R106W mutant lacked peaks for reporter residues Gly^{114} and Ala^{117}, which indicates significant line-broadening and a lack of stable interaction with mCG or mCA. The spectra for T158M and P101S, however, did show similar chemical shifts for

reporter residues, indicating these mutants retain the ability to selectively and preferentially bind mCG and mCA (Figure 2.9A-C).

Rett Syndrome mutation A140V forms MBD protein aggregates

The A140V mutant MBD was examined for DNA binding activity and recognition of mC motifs by HSQC. In the absence of DNA, A140V MBD protein had a significantly less ordered spectrum than wild-type off of DNA, however, in the presence of 10% excess BDNF mCG DNA, the peaks became as defined and well dispersed as wild-type under the same conditions, seen in Figure 2.10. The spectra also displayed similar chemical shifts for reporter residues Gly^{114} and Ala^{117} to wild-type, indicating this mutant retains the ability to selectively and preferentially bind mCG and mCA (Figure 2.11). Despite demonstrating specific binding in the presence of DNA under no salt conditions, ITC experiments in the presence of 100 mM KCl yielded very noisy data with extremely low heat (Figure 2.12). We suspected the protein may be aggregating in the absence of DNA, therefore we ran several salt titration experiments. First, we added 100 mM KCl to the A140V:mCG NMR sample and visually observed the immediate precipitation of protein. Next, we ran dynamic light scattering experiments on MeCP2 A140V MBD protein diluted in buffer with 100 uM and added 100, 300 or 600 μM KCl. The radius measurements indicated a positive trend towards increasing particle size and polydispersity with increasing salt concentration (Figure 2.12 and Table 2.7).

Hydroxymethylation reduces MeCP2 affinity for CG but not CA and shifts the distribution of MBD populated sites

To compare the effects of hydroxymethylation on CG and CA DNA binding, we synthesized additional oligonucleotides, one hemi-hydroxymethylated at the CG position and one with a hydroxymethylated CA pair, both on the 3’ strand (Figure 2.13A). We

examined binding thermodynamics and dissociation constants of MeCP2 by ITC and measured an apparent K_d value of 667 nM for hemi-hmCG and 167 nM for hmCA-3 (Figure 2.13B and Table 2.2). When inspected by HSQC, methylation-specific reader residues' resonances shifted in the MeCP2:hemi-hmCG complex to 103.1 ppm in ^{15}N for Gly^{114} and 6.8 ppm in ^{1}H for Ala^{117}. By the same method, in the MeCP2:hmCA-3 complex the resonances of Gly^{114} and Ala^{117} are observed at 102.8 ppm in ^{15}N and 6.8 ppm in ^{1}H, respectively (Figure 2.13C). The CSPs throughout the domain clustered in three regions of the protein sequence (Figure 2.14A), seen mapped to the crystal structure in Figure 2.14B (PDB: 3C2I) [9].

By NMR spectroscopy one can directly measure the relative population of distinct states and use differences in chemical shifts to detect the distribution between binding sites. To test whether the differences in relative binding affinities of MeCP2 for mCH and hmCH correlate with preferential localization, we examined MeCP2 incubated with a 1:1 mixture of mCG:mCA (Figure 2.15A) and hemi-hmCG:hmCA (Figure 2.15B). Resonances for both populations of complexes were consistent with a slow exchange between mCG and mCA DNA molecules. Hence, the peak height of six residues demonstrating non-overlapping chemical shifts was measured to derive the relative populations of the mCA and mCG bound states. In the presence of equal molar quantities, the fraction bound to mCG was 68.6 ± 4.1 % and to mCA was 31.4 ± 4 % consistent with preferential binding to mCG. When incubated with the hydroxymethylated DNAs, 33.4 ± 3.1 % bound to hemi-hmCG and 66.6 ± 3.1 % bound to hmCA-3, demonstrating MeCP2 now preferentially localizes to the hmCA site (Figure 2.16A).

Discussion

MeCP2 uniquely recognizes mCA sites in an orientation-dependent and methylation-specific binding mode

Our results suggest that the MeCP2 methyl-binding domain orients to interact specifically with the 5'-strand guanine through Arg^{111}, anchored by Asp^{121}, and with the 3'-strand guanine via Arg^{133}. Tyr^{123}, in a hydration shell around the methyl group, recognizes the 5'-strand mC. Substituting adenine for the 5'-strand guanine disrupts hydrogen bonding with Arg^{111}, leaving intact the 3'-G: Arg^{133} and 5'-mC: Tyr^{123} interactions. In contrast, the change of the 3'-strand guanine to adenine results in the loss of hydrogen bonds between 3'-G: Arg^{133} while maintaining 5'-G: Arg^{111} interaction and substituting the 5'-mC with thymine to preserve the methyl-recognition of Tyr^{123}.

We originally hypothesized that the orienting of the MeCP2 MBD on DNA is driven by the proximal [A/T] region identified in SELEX experiments [17]. However, a recent preprint presents evidence that the primary MeCP2 bound mCA sequences in vivo occur in the mCAC context [42], therefore we asked if the neighboring base could alter the strand-specific preference. We found that MeCP2 binds mCAC on either strand with high affinity irrespective of the relative position of the BDNF promoter sequence's AATT motif, however, a slight preference was observed for the 3'-mC orientation. This suggests that MeCP2 binding to mCA is strand-specific and orientation-dependent, driven by a combination of the proximity to an A/T rich region and a preference for sites followed by a C or T to G.

As expected of a major-groove DNA interaction, the ITC experiments demonstrate mCG, mCA-5, and mCA-3 binding is an exothermic and enthalpically

driven process [43,44]. The ITC measurements also reveal the methylated cytosine environment does not affect MeCP2:DNA binding stoichiometry but does affect binding affinity. We anticipated an mCA-3 preference for MeCP2 binding and found a less than two-fold change in binding affinity with an adenine substitution of the 3'-strand guanine, compared to a ten-fold decrease with a substitution on the 5'-strand (Table 2.1). This observation is consistent with computational quantum mechanical modelling that suggested Arg^{133} is more flexible than Arg^{111}, which is strongly fixed by hydrogen bonding with the Asp^{121} side chain, and that the Arg^{133} guanidinium group rotates towards Arg^{111} in the presence of a 5'-strand methylated base to create a 'closed/locked' state (Figure 2.1) [18]. Our relaxation dispersion measurements of the MeCP2 MBD bound to mCG DNA provide additional support for this model, measuring an exchange rate approximately 860x faster for Arg^{133} than Arg^{111} (Table 2.3 and 2.4).

Remarkably, binding to mCA-3, mCAC-5 or mCAC-3 is more exothermic than mCG to compensate for a substantial decrease in entropy of 2-5 kcal/mol (Table 2.1). This suggests that MeCP2 forms additional enthalpically favorable hydrogen bonds and/or ionic interactions to adapt to binding alternative substrates with high affinity, however, the mCH-bound state is more restricted with fewer degrees of freedom. This leads us to speculate that symmetrical mCG binding sites allow more flexibility in orientation on the DNA and perhaps switching between the recognized methyl group on either strand.

In the methylation-specific binding mode, the MBD's major-groove binding β-loop is stabilized by the hydrogen bonding between the 5'-strand guanine and Arg^{111}, at the base of β-strand 1. Therefore, the resonances of loop residues are reporters of

methylation-specific and non-specific binding by analysis of chemical shifts [39]. The linearity of the chemical shifts of $Gly^{114}$and Ala^{117} of MeCP2 between the different DNAs is characteristic of fast exchange between the major conformations (Figure 2.5). We varied the ratio of bound MeCP2 and observed no measurable change in peak position (Figure 2.6), therefore confirming that the shift of reporter residues do not reflect exchange between bound and free but instead reflect exchange between methylation specific and non-specific binding modes (as we have described previously for MBD2 and MBD3) [39]. When bound to mCA-5, the substitution of the 5'-strand guanine for adenine and Arg^{111} fails to stabilize the loop, thus MeCP2's Gly^{114} and Ala^{117} chemical shifts occur closer to their conformation on unmethylated DNA. When compared across the domain, residues with CSPs above 0.05 ppm between mCG and mCA bound MeCP2 clustered around the DNA interface and the major-groove binding loop, as well as the more distant helix 1 (Figure 2.7). In these regions and overall, the measured CSPs are fewer and smaller in the presence of mCA-3 than mCA-5, consistent with the mCA-3 bound state more closely resembling the mCG-bound MeCP2 MBD.

In contrast, the affinity of the MBD2 MBD for BDNF DNA is severely decreased by substitution of adenine, even on the 3'-strand (Figure 2.2 D and E). The methylation selectivity for mCA is also diminished, although less so for mCA-3, as indicated by the chemical shifts of Gly^{27} and Ala^{30} away from the specifically mCG-bound conformation (Figure 2.5) and towards the non-specifically CG-associated conformation [39]. Taken together, the HSQC measurements demonstrate methylation-specific binding is orientation dependent for both MeCP2 and MBD2. They also imply mCA-3 selectivity is

unique to MeCP2 among the MBD family proteins, and indicate that changes in helix 1 could contribute to accommodating mCA.

DNA binding proteins can have greater promiscuity in target sequences through conformational flexibility within the protein interaction surface [45]. Studies of transcription factor binding of DNA target motifs have shown flexibility can be essential to function [46] and simulations illustrate that target sequence recognition is a dynamic process [47,48]. Relaxation dispersion measurements can detect dynamic processes, such as side chain reorientation, loop motion, secondary structure changes and hinged domain movements. The dispersion observed in the MeCP2:mCG complex indicates residues at the DNA interface undergo microsecond to millisecond motions (Table 2.3 and 2.4). This exchange likely represents the rapid breaking and reforming of salt bridges and hydrogen bonds between coordinating amino acids, as well as direct amino acid-base interactions. The loops of the N- and C-terminal unstructured regions demonstrated the fastest exchange rates which likely reflects a dynamic transient interaction between the amino-acid residues and the DNA backbone, leading to large chemical shifts, noted with orange in Figure 2.8 [49]. The majority of residues with distinguishable chemical shifts along helix 1 and the β-loop display slower exchange, noted in yellow in Figure 2.8, on the millisecond to second time scale, consistent with larger coordinated motions. Taken with the chemical shift analysis, the relaxation dispersion measurements further indicate flexibility and dynamics in helix 1. In contrast, helix 1 of MBD2 is shorter and held closer to the core β-sheet by a shorter loop, possibly restricting its structural plasticity and ability to form alternative interactions with DNA (Figure 2.4).

Therefore, we postulate MeCP2 retains high affinity for non-CG methylated DNA due to flexibility in its MBD when bound to DNA along with an increase in the length of helix 1. As we have described previously for MBD4 [40], MeCP2 has a small insertion of 4 amino acids that extends helix 1 by one additional turn as compared to MBD2. This addition contributes to a larger hydrophobic core possibly stabilizing the isolated domain. Consistent with this hypothesis, the circular dichroism spectra demonstrate that when free in solution MeCP2 MBD adopts a better-defined structure than the MBD2 MBD, which more resembles a random coil (Figure 2.17). Likewise, the HSQC spectrum of MeCP2 MBD in the absence of DNA is well-dispersed indicative of a folded domain, while that of MBD2 MBD is significantly broadened until bound to DNA, indicating it undergoes a disorder-to-order transition upon binding (Figure 2.18). We suggest that increased stability of the isolated MeCP2 MBD allows the protein to more readily adapt to alternative binding sequences by allowing the residues involved in DNA binding to sample a larger conformational space, without an increase in entropy sufficient to overcoming the free energy barrier to unfolding [45,48]. The relaxation dispersion observed in the hydrophobic core of MeCP2 is consistent with the ability of the protein structure to breathe, which enables recognition of more than one binding sequence and allows for accommodation of mCA motifs.

The MBD2 MBD is mostly disordered off of DNA and may behave like an intrinsically disordered protein in targeting CG motifs by rapidly sampling partially folded states for finding the most favored conformation [50]. This conformation is often stabilized upon binding more from forming intermolecular interactions, than from the energetics of folding [51], therefore MBD2 may be less tolerant of adjustments to the binding

interactions available in an mCA dinucleotide. We conclude the larger entropic cost of ligand-induced folding limits the range of sequences that provide sufficient interactions to stabilize the methylation-specific binding conformation of MBD2.

Rett Syndrome associated mutations in the MeCP2 MBD to not eliminate methylation recognition function

Interestingly, mutations in the MeCP2 that disrupt the domain architecture do not preferentially affect selectivity for mCG or mCA sequences. The R016W mutation inserts a bulky tryptophan residue into the domain's hydrophobic core, destabilizing the MBD as evident from the decrease in discretely defined peaks in an HSQC spectrum. The presence of neither mCG or mCA-3 DNA ordered the R106W MBD sufficiently to observe peaks for the Gly^{114} and Ala^{117} reporter residues (Figure 2.12C). The less disruptive mutations of T158M, P101S, and A140V increase the domain motions, as evident from the broadening of residue peaks, but they retain the ability to bind mC in the methylation-specific binding mode in the preferred orientation. The chemical shifts indicate the loss of hydrogen bonds between Arg^{133} and the 3'-strand guanine of the mCA-3 sequence does not inhibit mC recognition even when the MBD is moderately destabilized (Figure 2.9 and Figure 2.11). We were unable to sufficiently concentrate the A140V mutant protein for ITC analysis due to aggregation (Figure 2.12); however, our results support the recent suggestion that mutations in the MBD may not cause Rett by affecting selectivity for methyl-cytosines. Instead, our data are consistent with the dysfunction associated with changes in MeCP2 levels in both MeCP2-null and -overexpressing mice [2,52–54]. Furthermore, Rett Syndrome-associated mutations lead to much lower MeCP2 cellular concentrations as compared to wild-type [55]. Perhaps aggregates of destabilized MeCP2 more readily signal for degradation and the

remaining soluble protein cannot fully occupy high affinity sites, creating a dilution effect that is further amplified by the lower overall affinity of the mutant protein for both mCpA and mCpG sites. Our findings further support the therapeutic potential of re-stabilizing mutant MeCP2 to restore native protein function, but this model of MeCP2 dysfunction needs to be rigorously tested via extensive investigation.

Hydroxymethylation reduces MeCP2 affinity for CG while maintaining selectivity for CA

By HSQC, there were clear chemical shift differences between the hemi-hmCG and hmCA-3 bound states, however, the reporter residues of the methylation bound state were relatively close to each other and mCG (Figure 2.13C). The ITC derived K_d is a highly sensitive measure of global binding affinity, whereas the chemical shift changes of Gly^{114} and Ala^{117} reflect distribution between CG specific and non-specific binding modes. Our data indicates MeCP2 recognizes hmC in a similar manner as mC, despite not retaining a high affinity for hemi-hmCG. From examining the 3D models based on the MeCP2:mCG crystal structure (Figure 2.13A) (PDB: **3C2I**) [9], we expected both hemi-hmCG and hmCA-3 would retain the 5'-G:Arg^{111} interaction; supply a methyl-group for recognition by Tyr^{123}; and, by adding a hydroxyl group to the mC that packs against Arg^{133}, potentially disrupt the hydrophobic interaction. The preservation of the 5'-G:Arg^{111} interaction continues to stabilize the β-loop in which Gly^{114} and Ala^{117} reside. As such, their corresponding methylation specific reporter peaks in the hemi-hmCG and hmCA HSCQs do not shift far from mCG (Figure 2.13C). Furthermore, the hemi-hmCG bound conformation remains highly similar to mCG with the fewest and least extreme changes in chemical shift (Figure 2.14).

In contrast, hydroxymethylation of the 5'-mC significantly changes the binding affinity of MeCP2. Our ITC measurements demonstrate only a two-fold difference in the apparent K_d between mCA-3 and hmCA-3, however, hydroxymethylation of the same cytosine position in the hemi-hmCG oligo resulted in a greater than ten-fold decrease from the mCG binding affinity (Figure 2.13B and Table 2.2). This observation implies that the anchoring of Arg^{133} through hydrogen bonding with the mCG 3'-strand guanine impedes the accommodation of a hydroxyl group on the 5'-strand mC. When adenine replaces the 5'-strand guanine, Arg^{133} is 'unlocked' and released from the hydroxyl-imposed steric tension.

We asked if the changes in binding affinity with hydroxymethylation would reflect a change in the distribution of MeCP2 among mC sites. The K_d measurements would suggest MeCP2 equitably occupies mCG and mCA-3 sites. Peak heights in an HSQC provide information about the relative populations of states. In the presence of a 1:1 mixture of mCG:mCA-3 DNA, peaks with distinct chemical shifts between mCG and mCA-3 bound conformations (Figure 2.15A) indicated nearly 70% of the MeCP2 present was mCG bound (Figure 2.16A). Alternatively, upon addition to a 1:1 mixture of hemi-hmCG:hmCA-3 DNA (Figure 2.15B), the MeCP2 populations shifted to over 65% hmCA-3 bound (Figure 2.16A), consistent with the discrepancy in binding affinity. To date, hmCA has rarely been observed in base-level resolution mapping of the hydroxymethylome of the human neurons [5], however, hmCG is known to accumulate in the brain throughout development, especially in the gene bodies of highly expressed genes [1,14,56]. Therefore, our findings support a model in which hydroxymethylation could serve as a molecular switch to effectively redistribute MeCP2 from mCG to

h/mCA-3 sites, modeled in Figure 2.16B, and warrant further investigation into the role of MeCP2 function at mCA-3 and hmCA-3 loci.

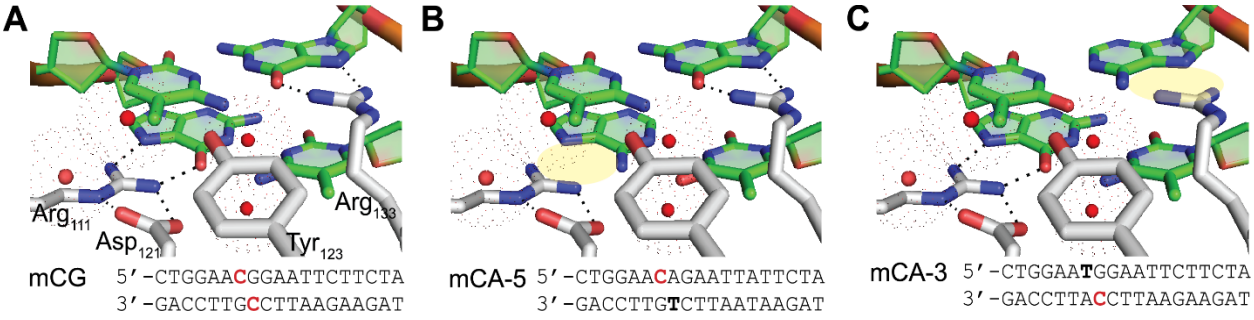

Figure 2.1 MeCP2 MBD binding site with mCH
Models comparing the (a) mCG complex, and highlighting in yellow the hydrogen bonds disrupted by substituting an AT base pair for one CG on the (b) 5' or (c) 3' strand. The sequence of the oligonucleotide used in the study is below each model, methylated bases are in red and substitutions are bolded.

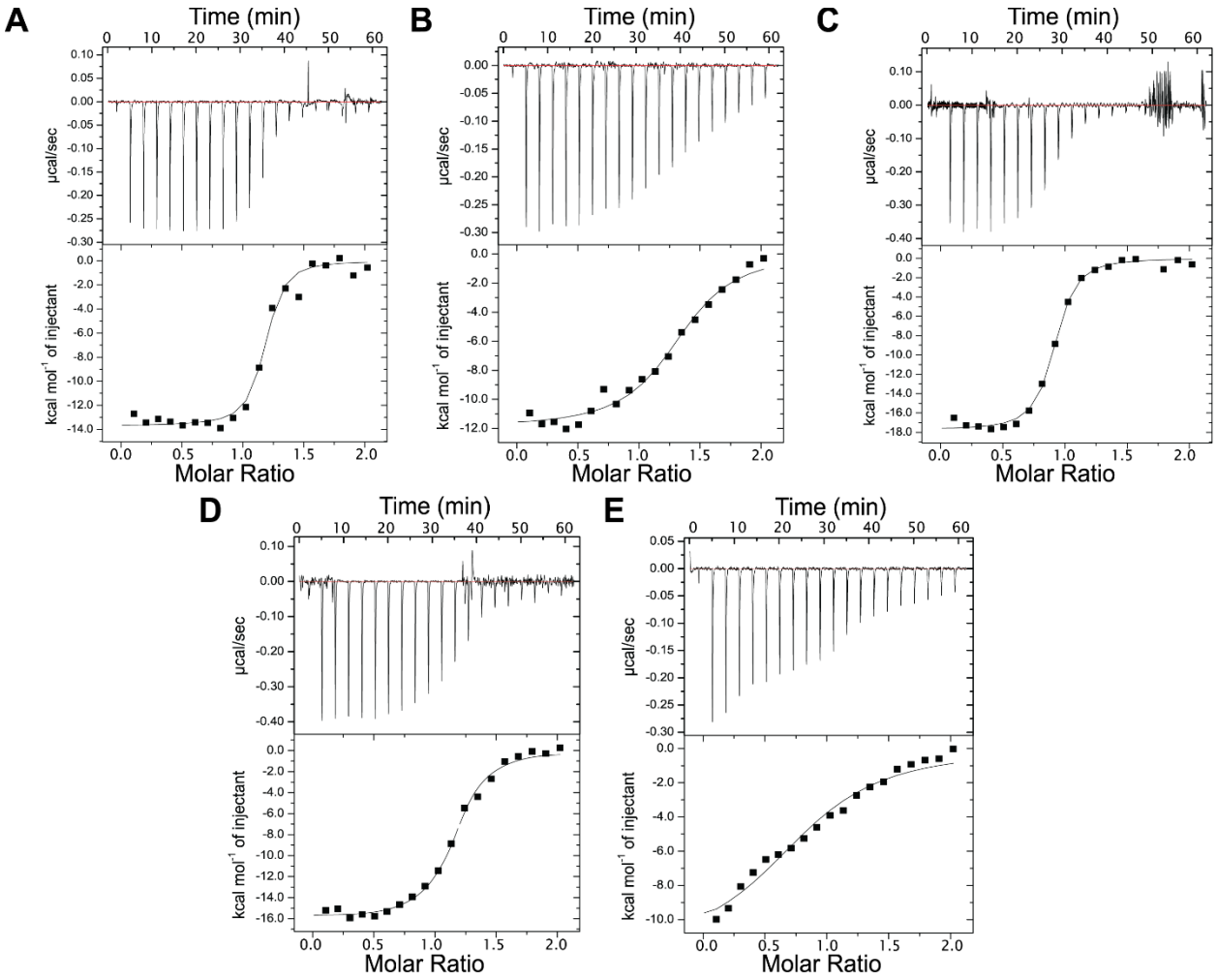

Figure 2.2 MeCP2 and MBD2 binding affinity for mCH
ITC titrations of the MeCP2 MBD with (a) mCG, (b) mCA-5 and (c) mCA-3 oligonucleotides, and of the MBD2 MBD with (d) mCG and (e) mCA-3. The lines represent the fit of the data (Table 2.1).

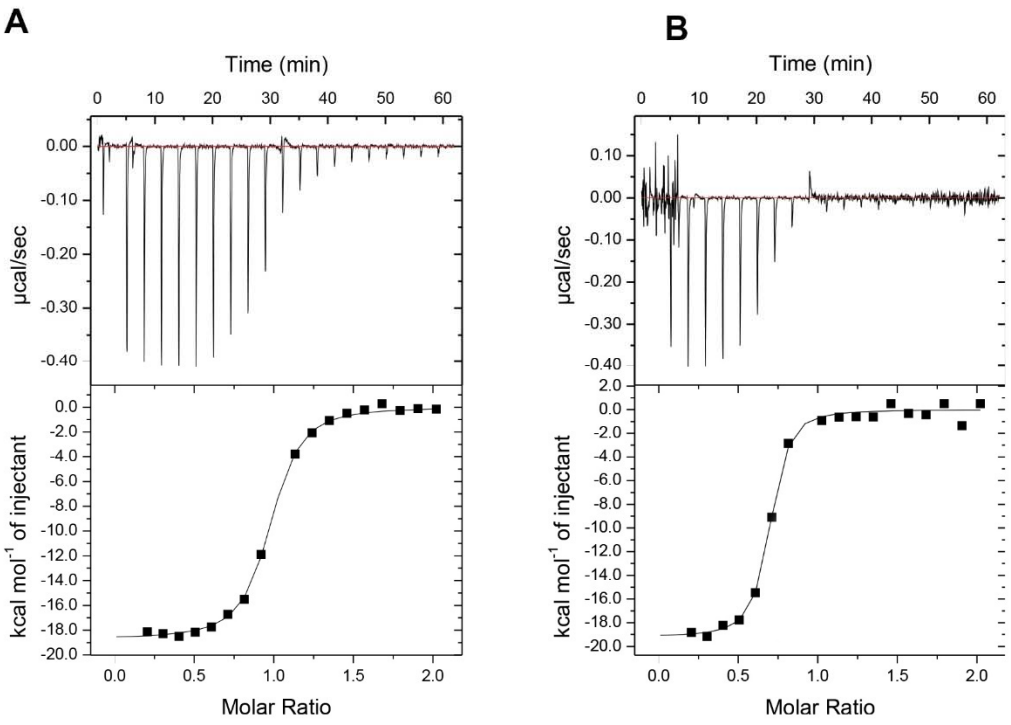

Figure 2.3 MeCP2 binding affinity for mCAC
ITC titrations of the MeCP2 MBD with (a) mCAC-5 and (b) mCAC-3 oligonucleotides. The lines represent the fit of the data (Table 2.1).

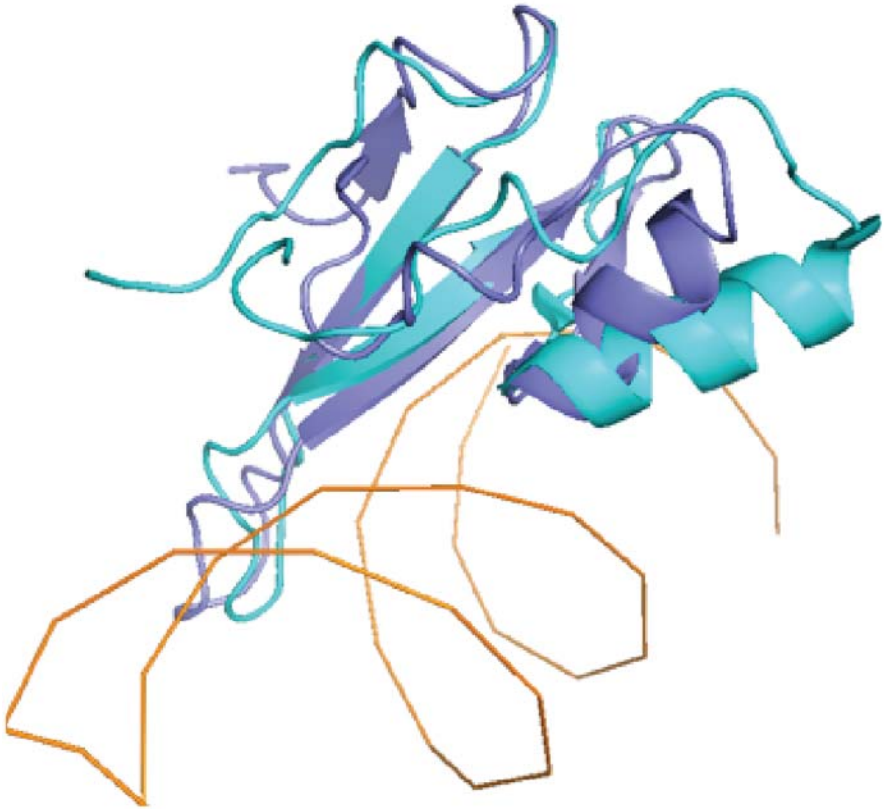

Figure 2.4 Structural alignment of MeCP2 and MBD2
Overlay of the crystal structure of the MeCP2 MBD:mCG DNA complex (cyan) and NMR ensemble average structure of the MBD2 MBD:mCG DNA complex (purple).

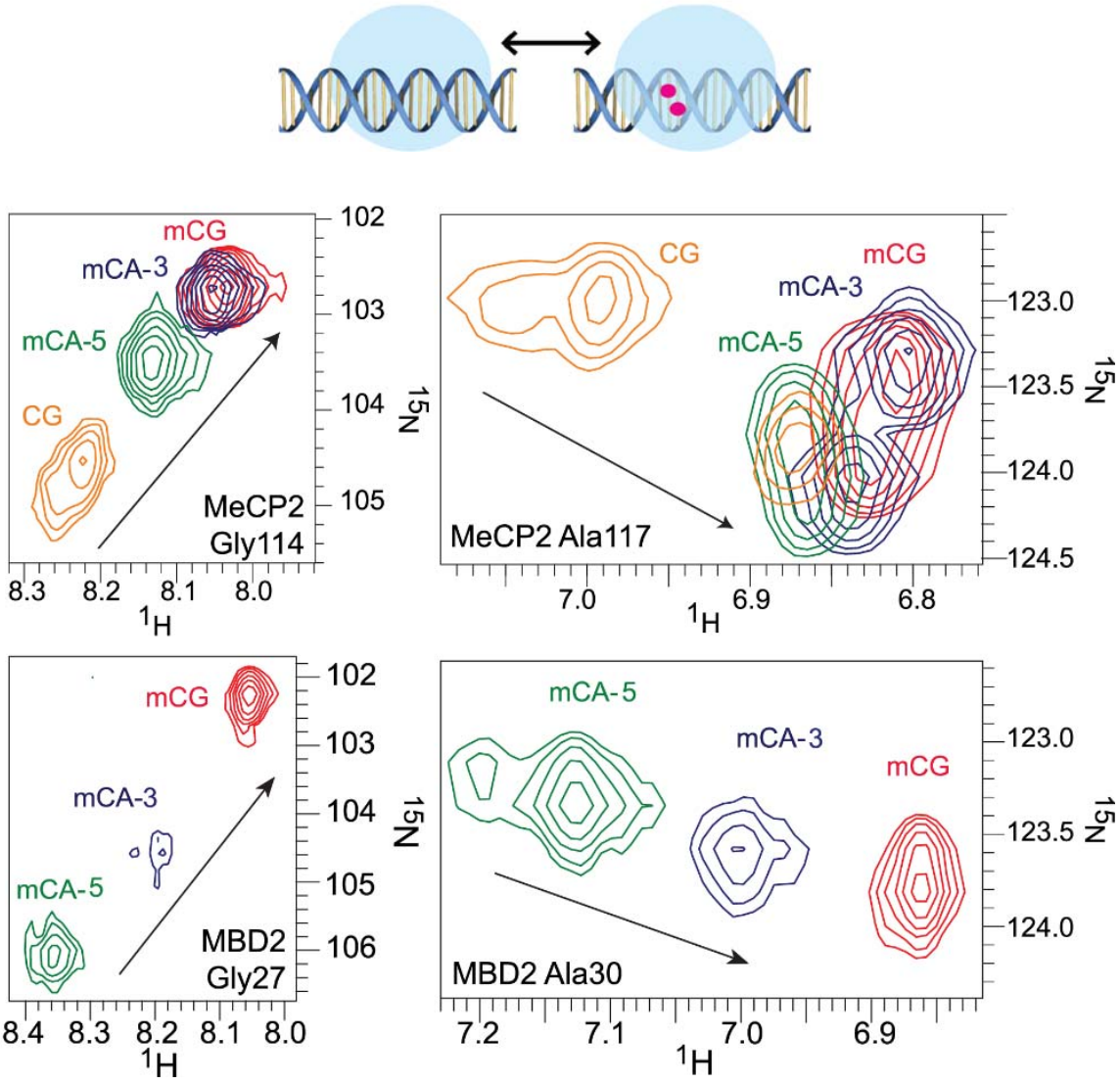

Figure 2.5 Binding modes of MeCP2 and MBD2 to mCH
Top: A schematic illustrating the exchange of an MBD between the non-specific (left) and methylation-specific (right) binding modes. mC bases are represented as red circles.
Bottom: An overlay of ^{15}N HSQC spectra for key reporter residues of wild-type MeCP2 and MBD2 bound to CG, mCG, mCA-5, and mCA-3.

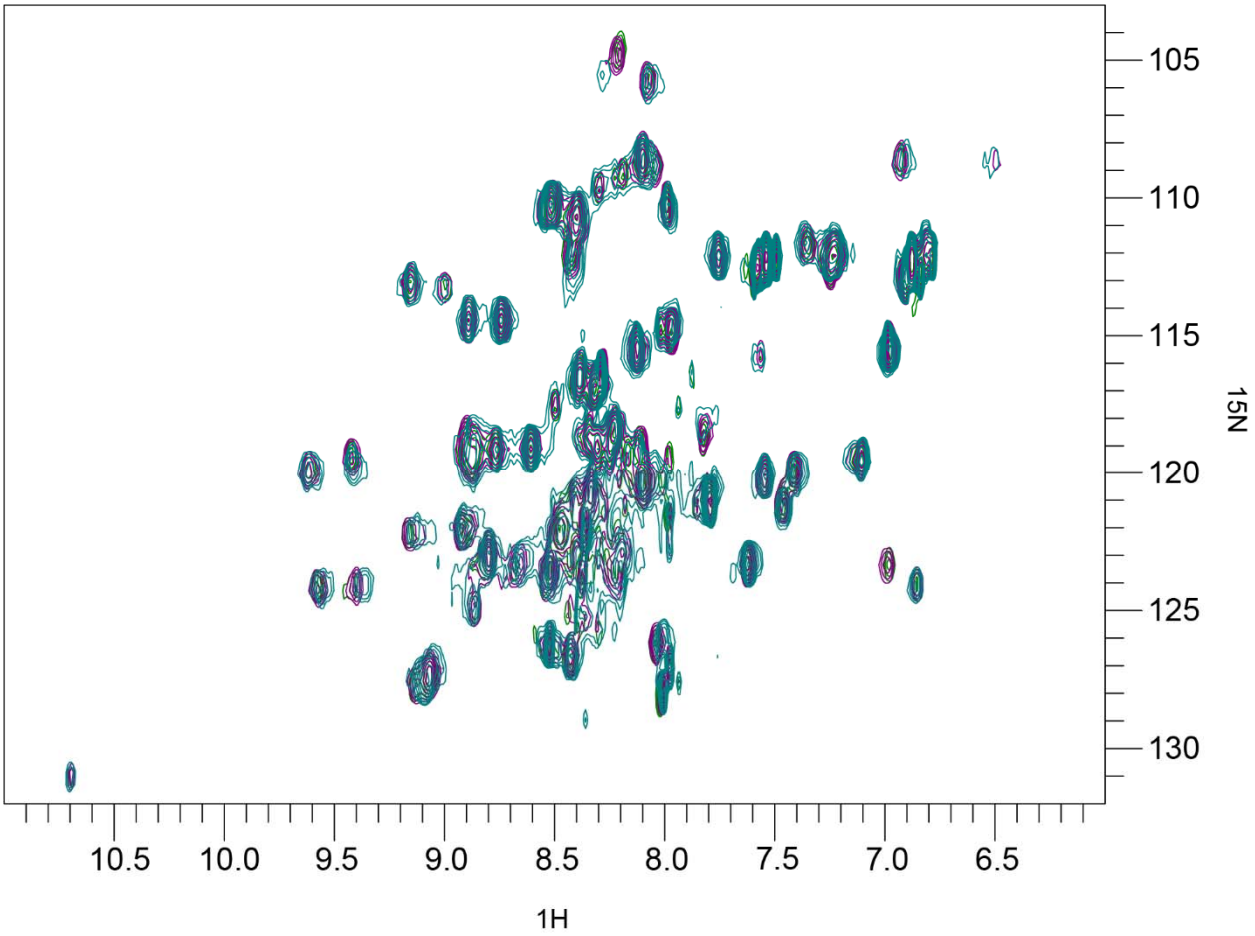

Figure 2.6 MeCP2 chemical shifts upon binding unmethylated DNA at varying concentration ratios

^{15}N HSQC spectra of the methyl-binding domain of MeCP2 on unmethylated BDNF DNA at ratios of 200:220 (teal), 400:400 (green) and 400:800 (purple) micromolar.

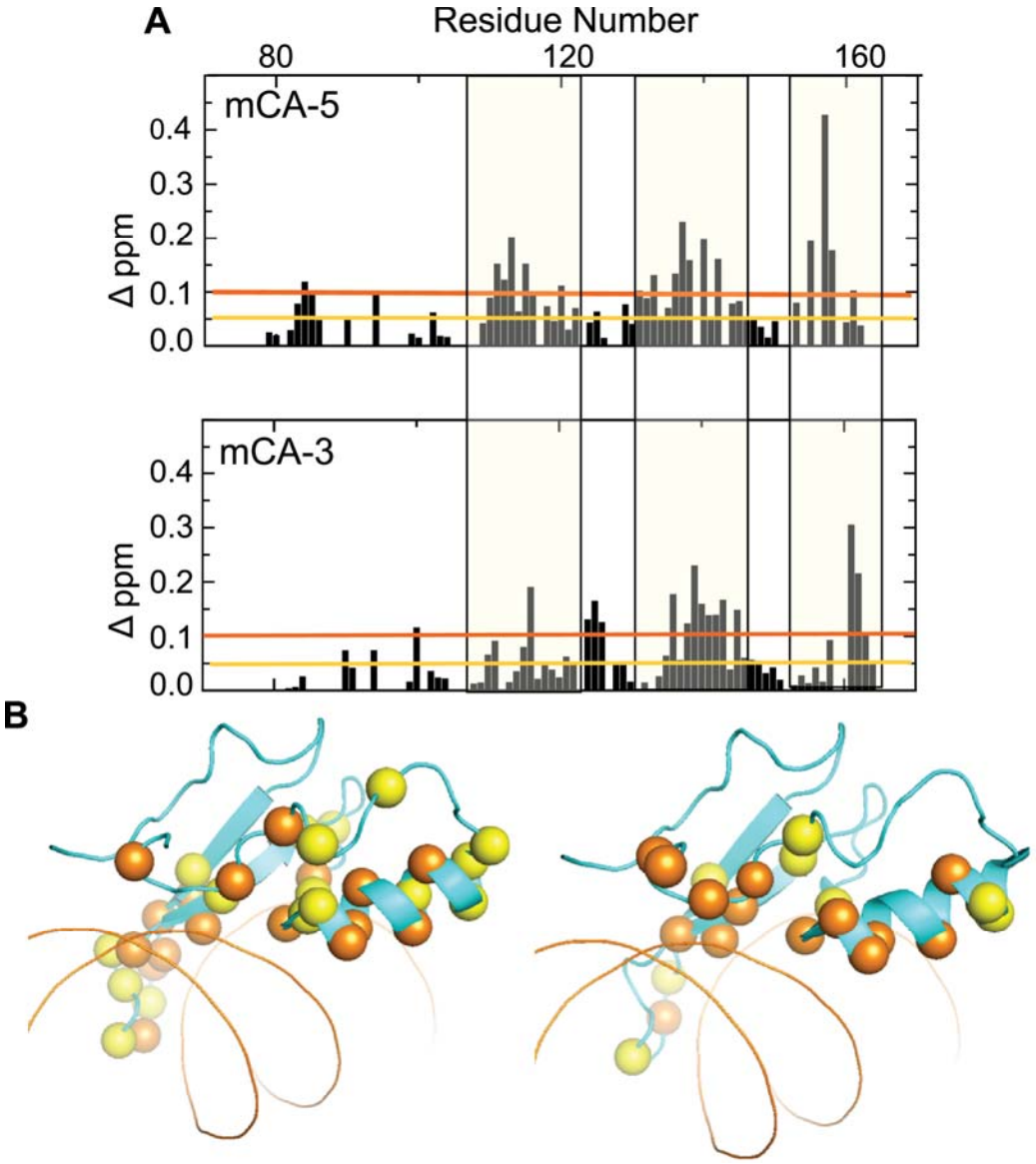

Figure 2.7 Structural changes induced by binding of MeCP2 to mCH
(a) Bar graphs showing the CSPs induced by MeCP2 binding to mCA-5 (top) and mCA-3 (bottom), compared to mCG bound chemical shifts. Sequence regions with the greatest CSPs are indicated. (b) Mapping of the CSPs induced by mCA-5 (left) and mCA-3 (right) binding to MeCP2 on the crystal structure of mCG-MeCP2 (PDB 3C2I). Residues with CSPs larger than 0.5 (yellow) and 1.0 (orange) p.p.m. are represented as sticks.

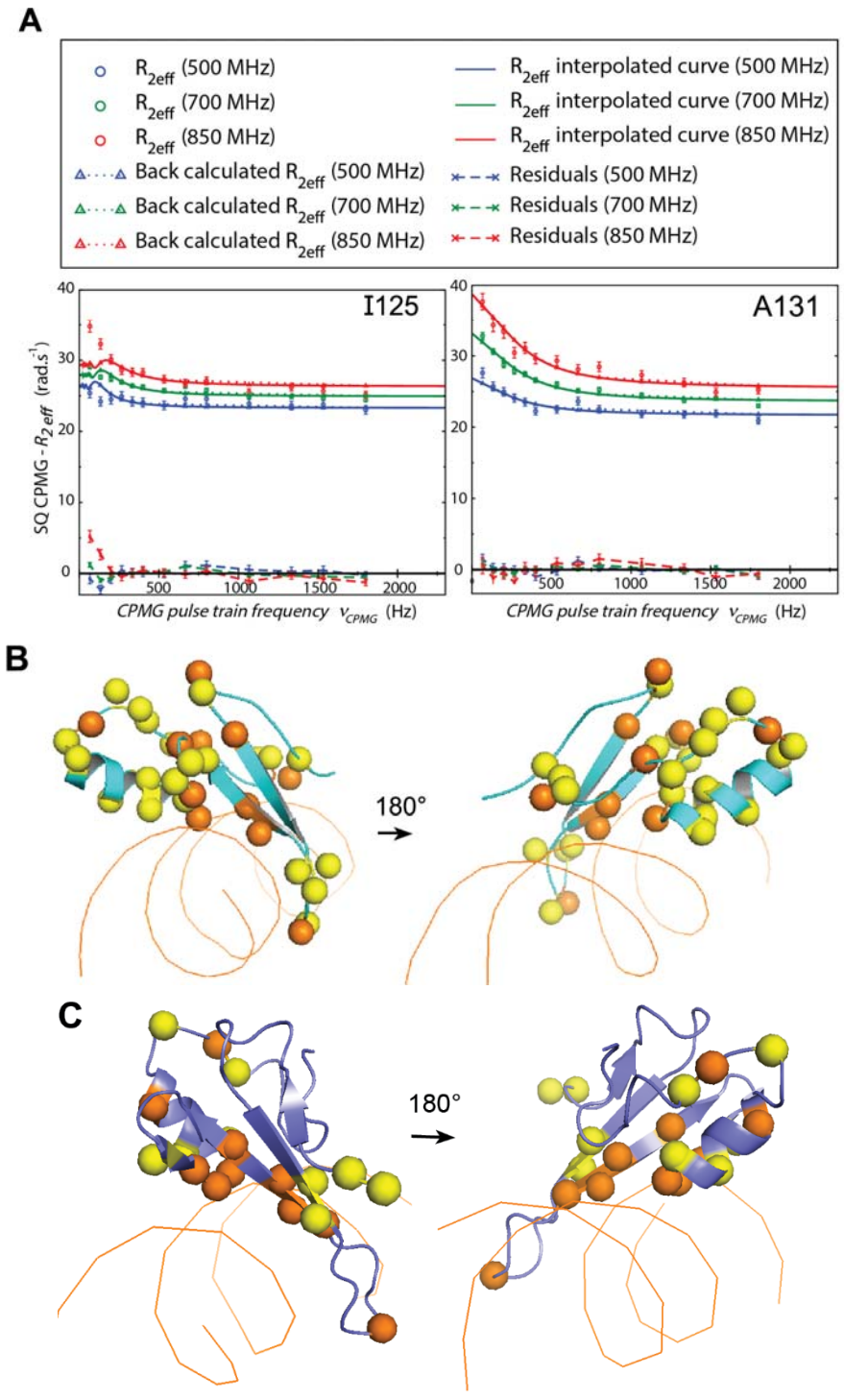

Figure 2.8 Internal dynamics of MeCP2 and MBD2 bound to mCG
(a) Representative relaxation dispersion curves of residues in the MeCP2:mCG complex, demonstrating slow exchange fit with the TSMFK01 model (left) or fast exchange fit with the CR72 model (right). Residues of (b) MeCP2 and (c) MBD2

displaying slow (yellow) and fast (orange) exchange are shown as spheres in the structural maps.

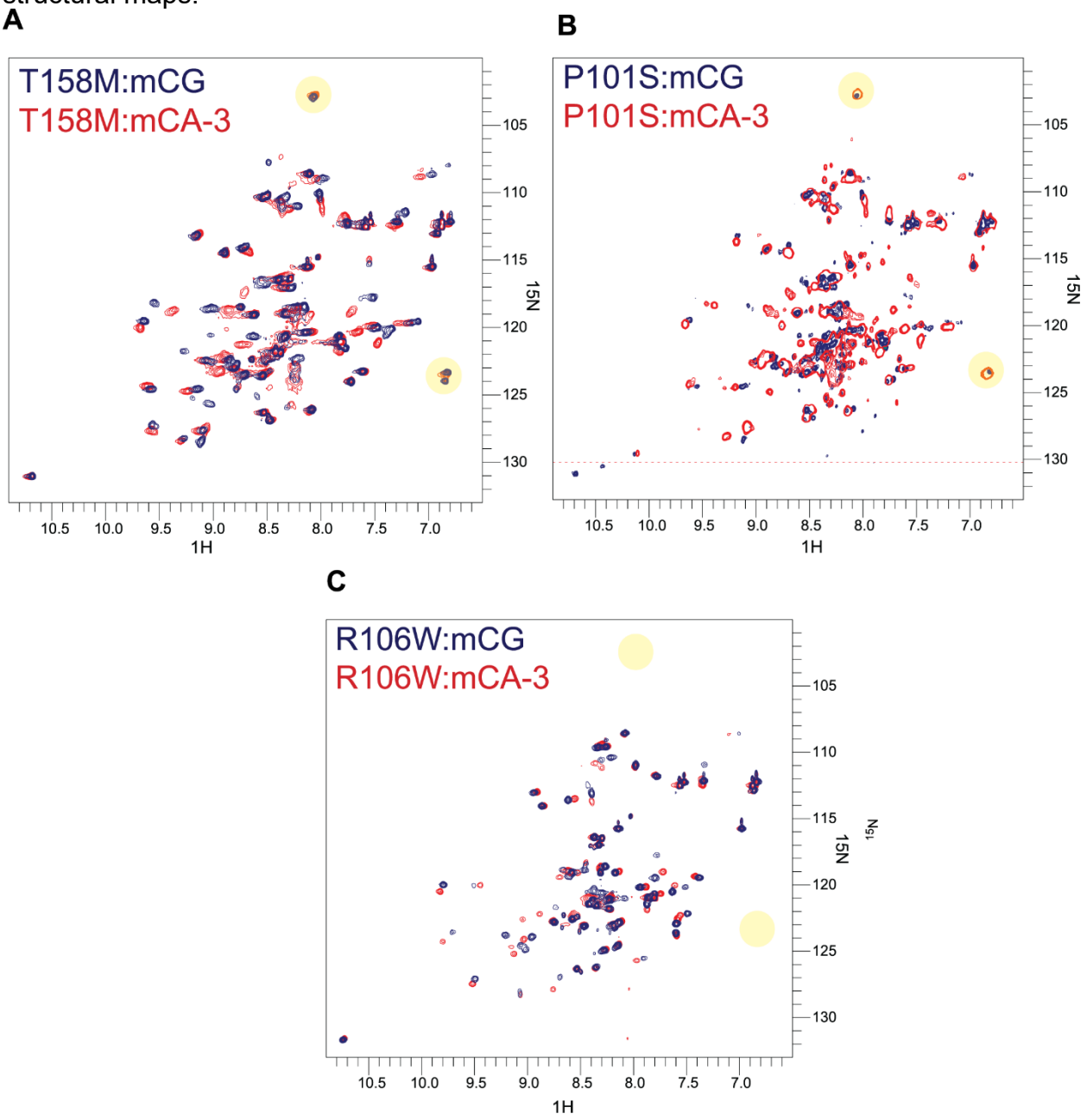

Figure 2.9 mCH recognition by Rett Syndrome-associated mutation in the MBD of MeCP2

^{15}N HSQC spectra of the MeCP2 MBD with Rhett Syndrome-associated mutations (a) T158M, (b) P101S, (c) R106W, and (d) A140V on mCG (navy) and mCA-3 (red) DNA. Location of reporter residues for methylation-specific binding are highlighted in yellow.

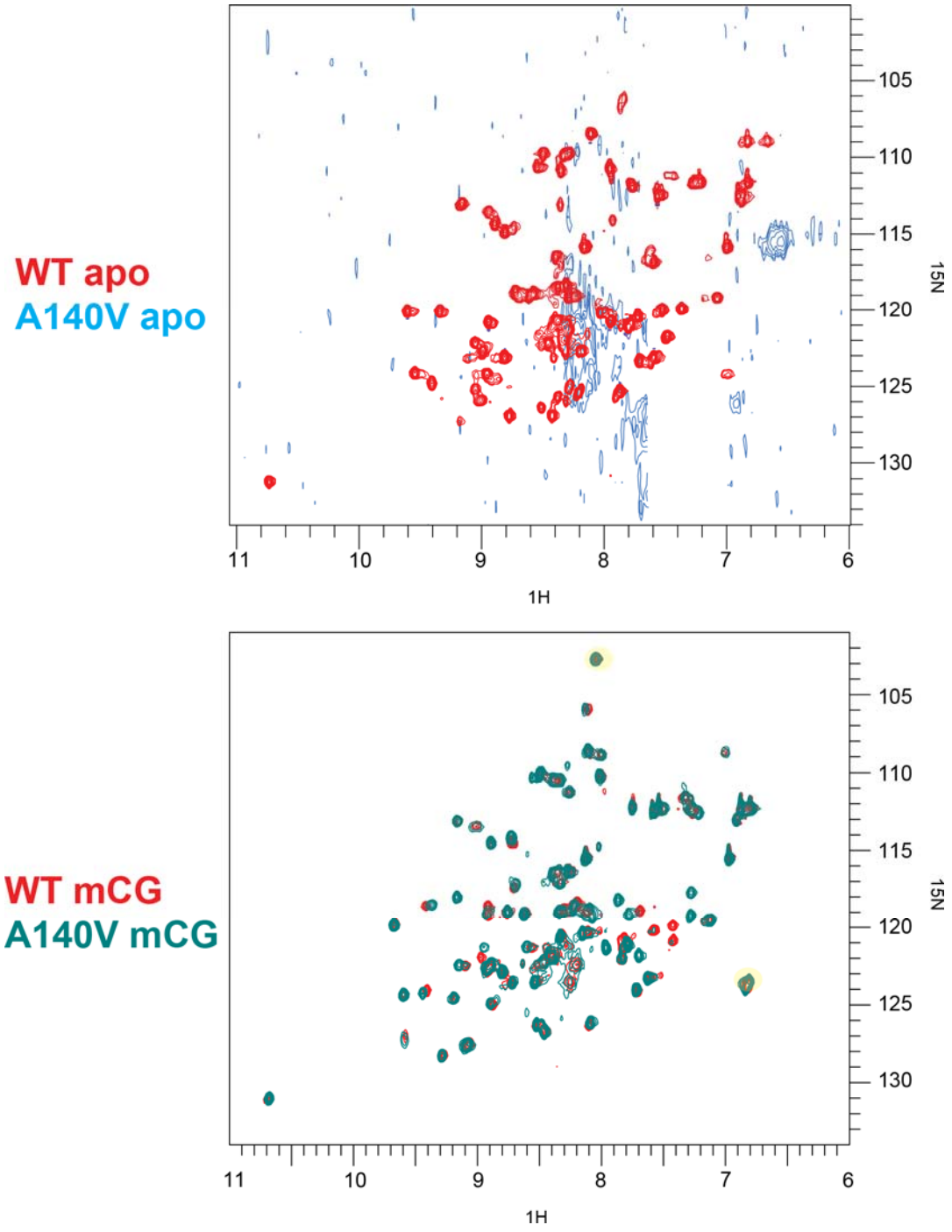

Figure 2.10 DNA binding by Rett Syndrome-associated mutation A140V in the MBD of MeCP2

^{15}N HSQC spectra comparing free protein (top) and mCG bound (bottom) wild-type MeCP2 MBD (red) with Rhett Syndrome-associated mutation A140V (blue or teal). Location of reporter residues for methylation-specific binding are highlighted in yellow.

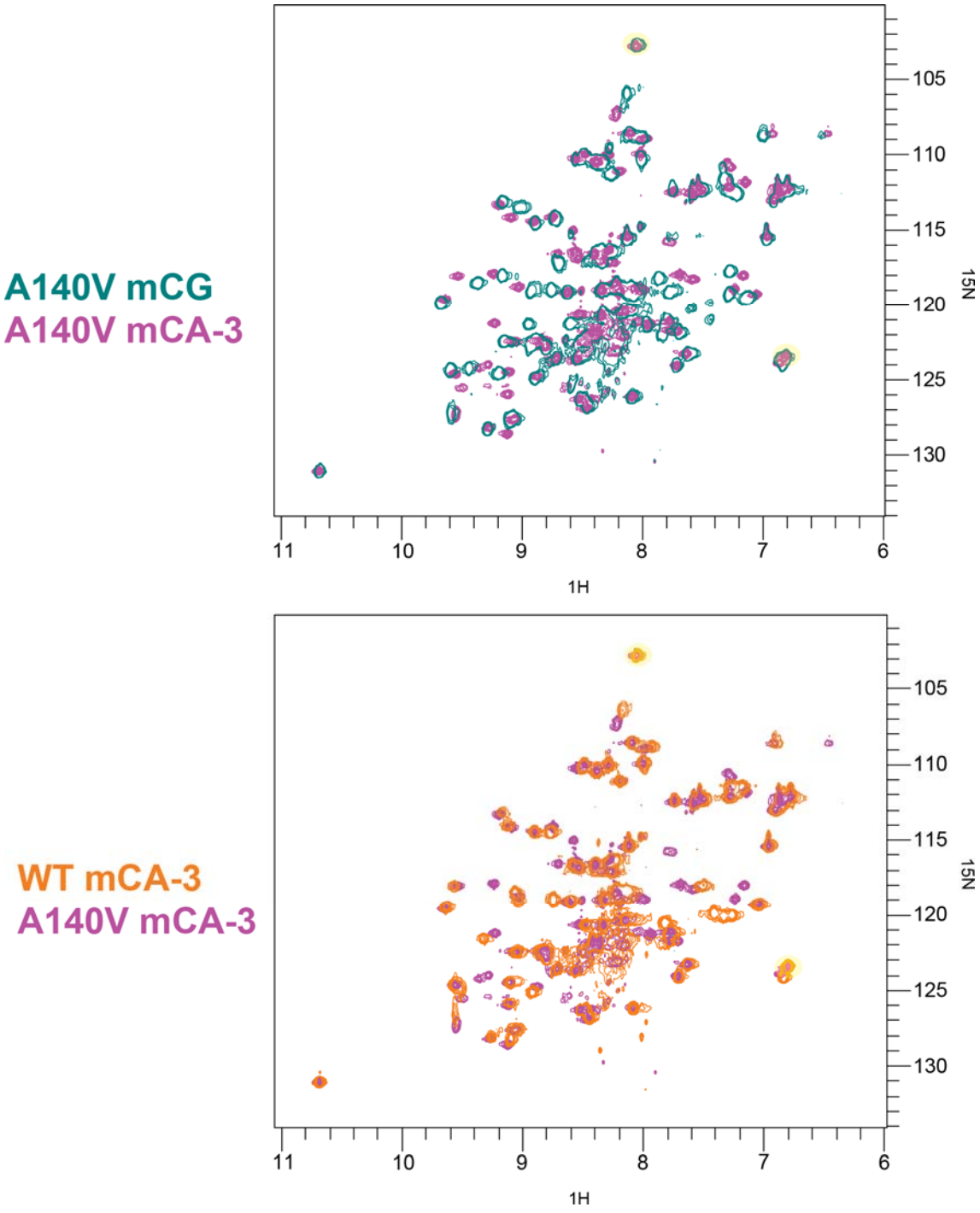

Figure 2.11 mCH recognition by A140V mutation in the MBD of MeCP2
^{15}N HSQC spectra comparing MeCP2 A140V (top) binding mCG (teal) with mCA-3 (magenta) and (bottom) wild-type MeCP2 MBD (orange) with A140V (magenta) on mCA-3. Location of reporter residues for methylation-specific binding are highlighted in yellow.

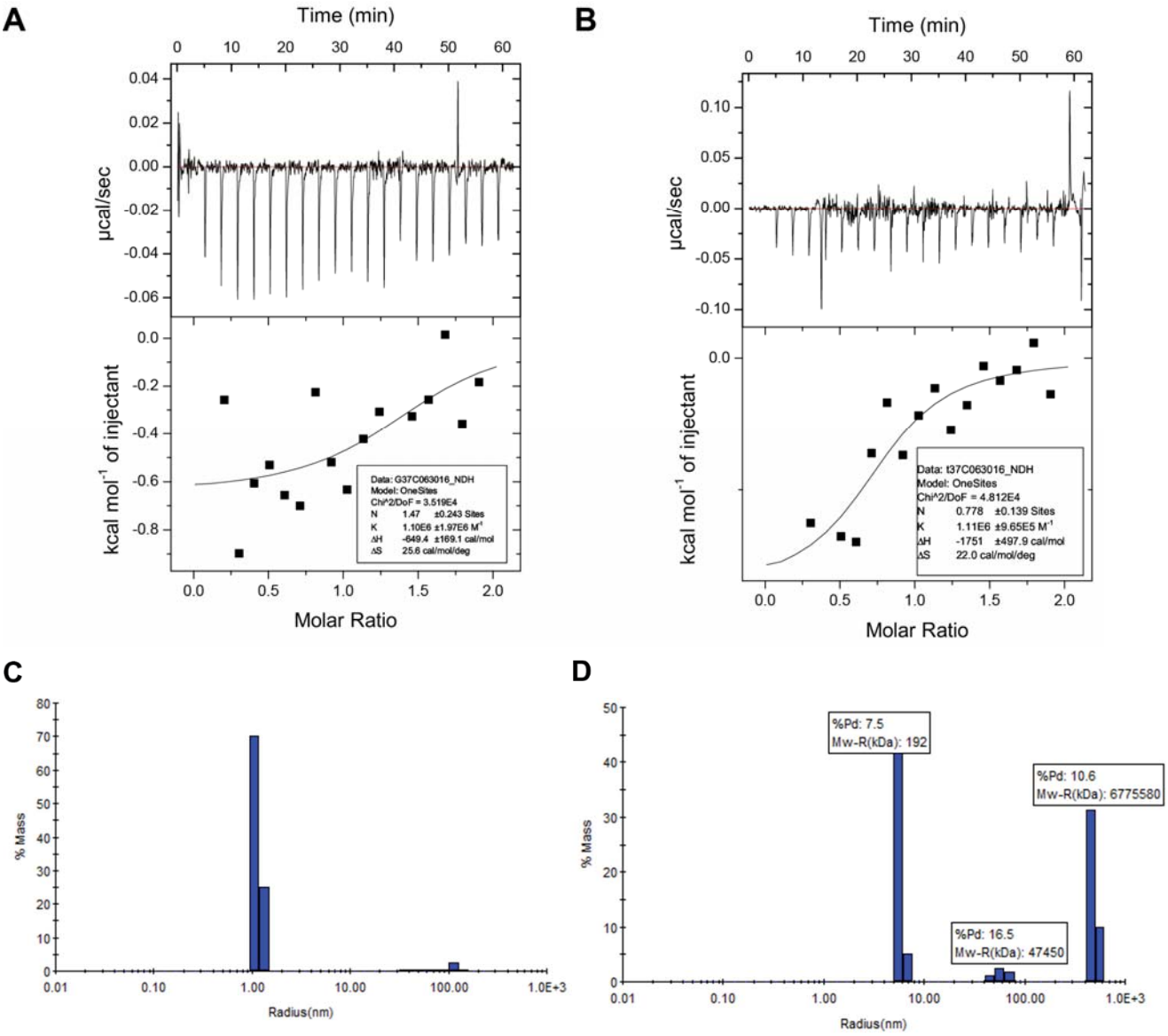

Figure 2.12 Aggregation study of A140V Rett Syndrome-associated mutation in the MBD of MeCP2

ITC titrations of the MeCP2 MBD with (a) mCAC-5 and (b) mCAC-3 oligonucleotides. The lines represent the best estimated fit of one-site binding model. Representative DLS histograms of the hydrodynamic radii of MeCP2 A140V MBD protein in the presence of (c) 100 mM and (d) 300 mM KCl.

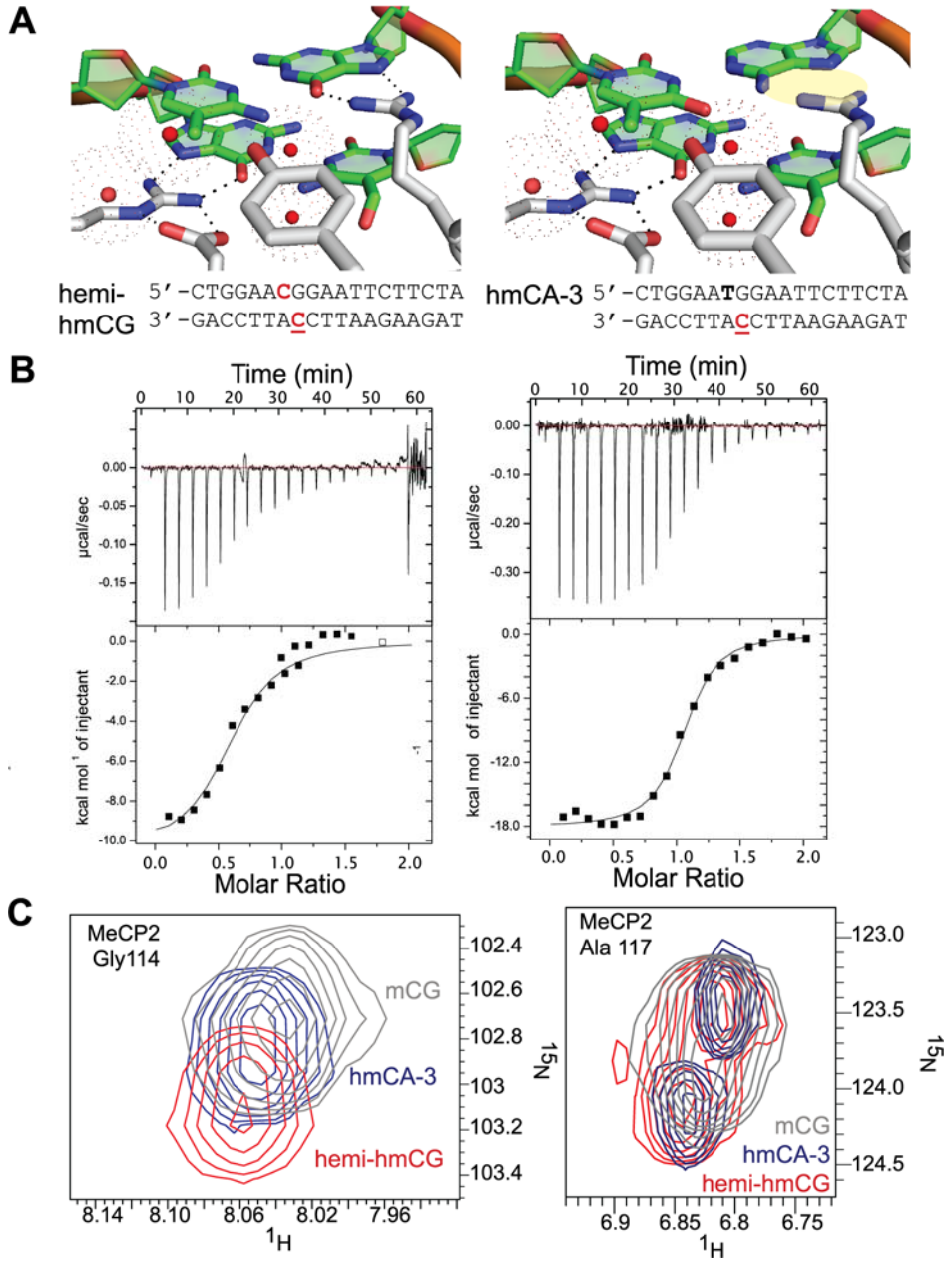

Figure 2.13 MeCP2 MBD binding to hydroxymethylated CH
(a) Models of the MeCP2 MBD binding site comparing the hemi-hydroxymethylated CG (left) and hydroxymethylated CA-3 (right) complexes, highlighting in yellow the predicted hydrogen bonds disrupted compared to the mCG bound crystal structure. Sequence of the oligonucleotide used in the studies are below each model, including methylated cytosines (red) and hydroxymethylated bases (underlined). (b) ITC titrations of MeCP2 MBD with hemi-hmCG (left) and hmCA-3 (right). The lines represent the fit to a one-site model (Table 2.2). (c) An overlay of ^{15}N HSQC spectra for key reporter residues Gly114(left) and Ala117(right) of wild-type MeCP2 bound to mCG (gray), hemi-hmCG (red) and hmCA-3 (navy).

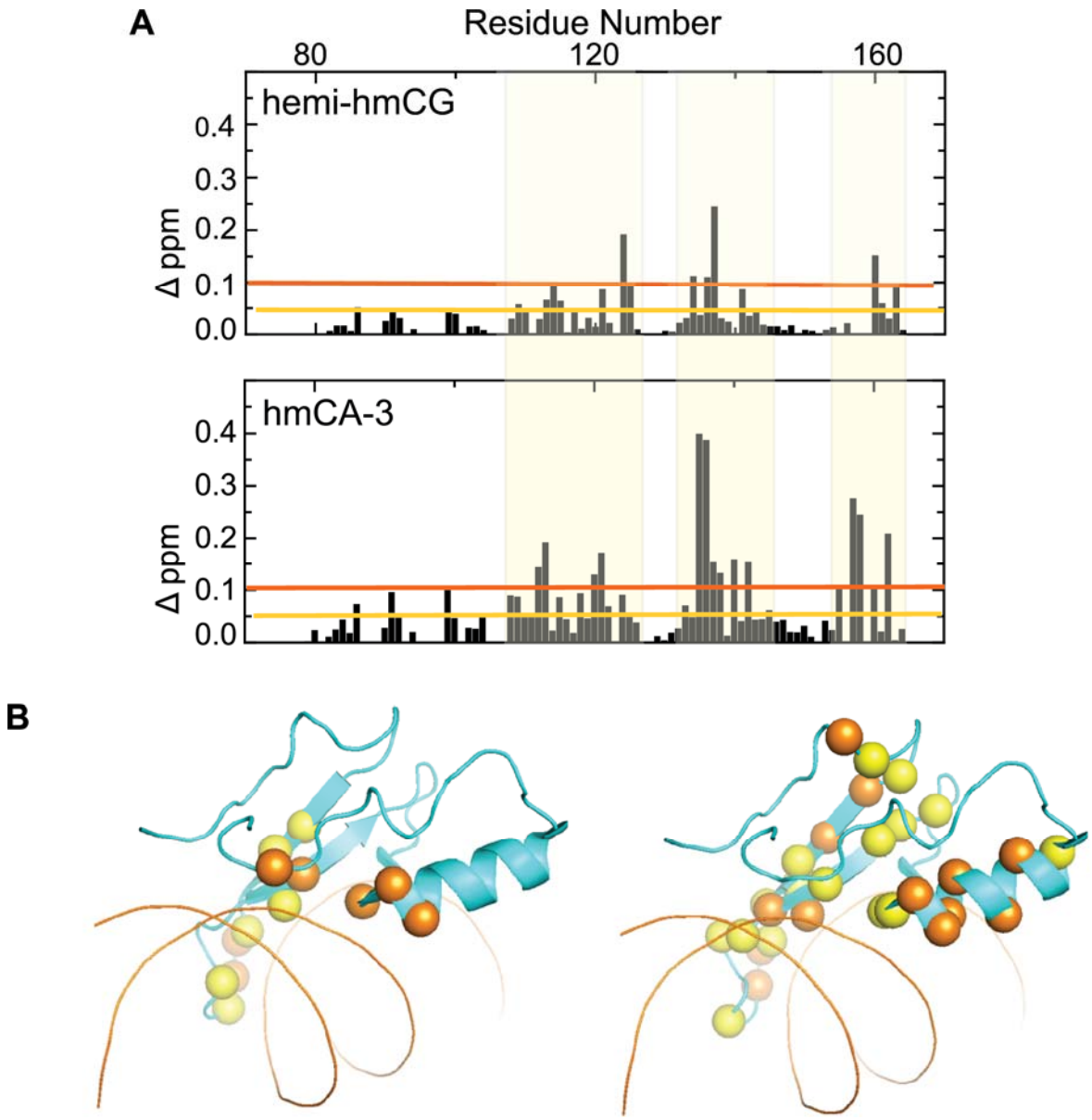

Figure 2.14 Structural changes induced by binding of MeCP2 to hmCH
(a) Bar graphs showing the CSPs induced by MeCP2 binding to hmCA-3 (top) and hemi-hmCG (bottom), compared to mCG bound chemical shifts. Sequence regions with the greatest CSPs are indicated. (b) Mapping of the CSPs induced by hemi-hmCG (left) and hmCA-3 (right) binding to MeCP2 on the crystal structure of mCG-MeCP2 (PDB 3C2I). Residues with CSPs larger than 0.5 (yellow) and 1.0 (orange) p.p.m. are represented as sticks.

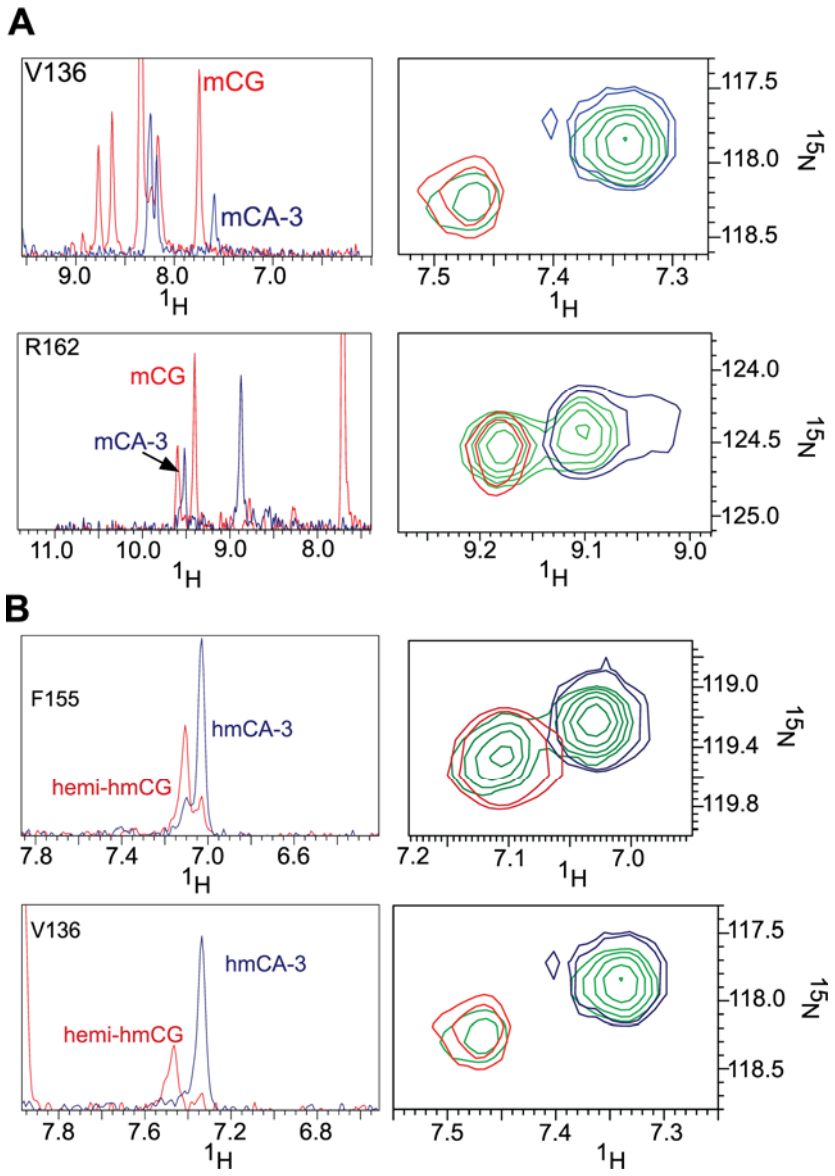

Figure 2.15 HSQC difference peaks in MeCP2 MBD on methylated and hydroxymethylated CH

^{15}N HSQC spectra of key reporter residues of MeCP2 bound to (a) mCG (red), mCA-3 (blue) and a 1:1 mixture (green) with 1D slices on left and 2D overlays on right. Below, in (b) are reporter residues of MeCP2 bound to hemi-hmCG (red), hmCA-3 (blue) and a 1:1 mixture (green) with 1D slices on left and 2D overlays on right.

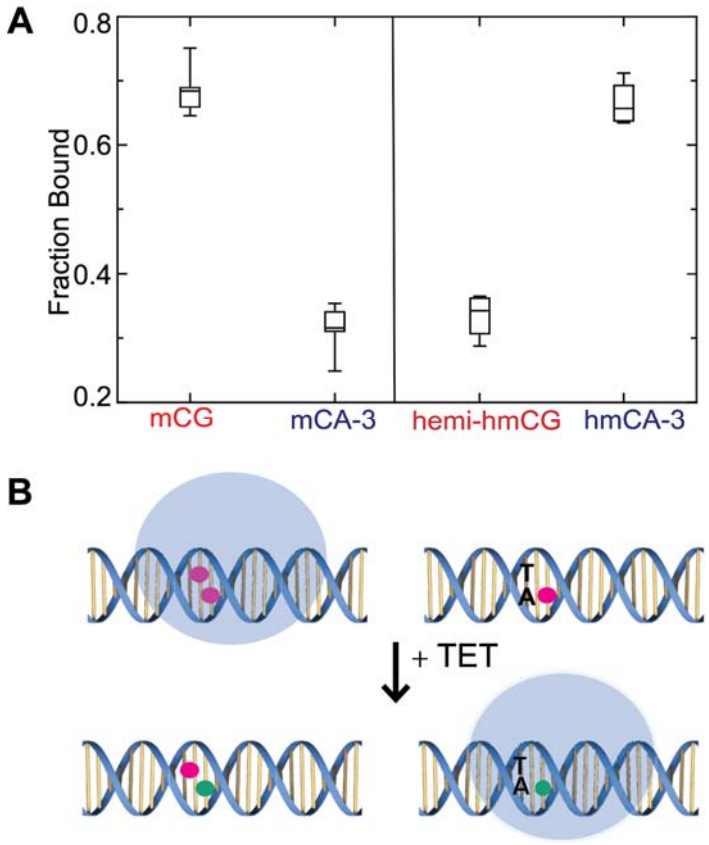

Figure 2.16 MeCP2 MBD distribution on methylated and hydroxymethylated CH
(a) Box plot of the fraction of MeCP2 bound to mCG:mCA-3 and hmCG:hmCA-3 (n=6). (b) Model of distribution shift from mCG motifs to mCA-3 motifs (red circles) upon hydroxymethylation (green circles) of both sites.

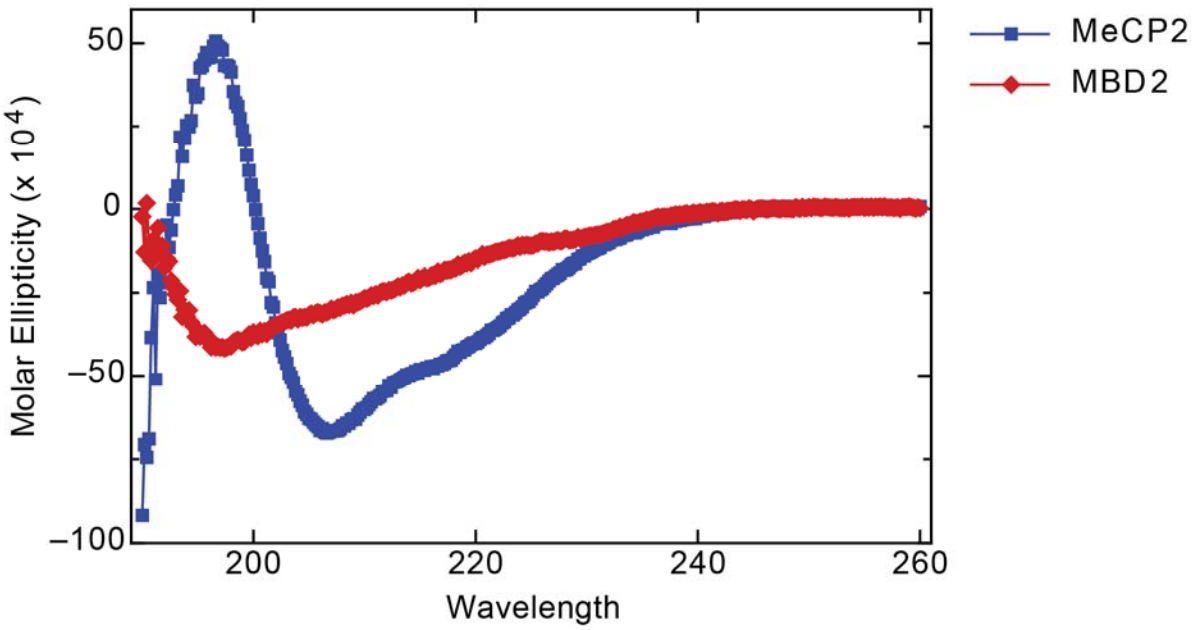

Figure 2.17 Secondary Structure of MeCP2 and MBD2 MBDs off DNA
Far-UV circular dichroism spectra of the isolated methyl-binding domain of MeCP2 (blue) and MBD2 (red) with the characteristic shape and magnitude of an alpha-helical protein and random coil structure, respectively.

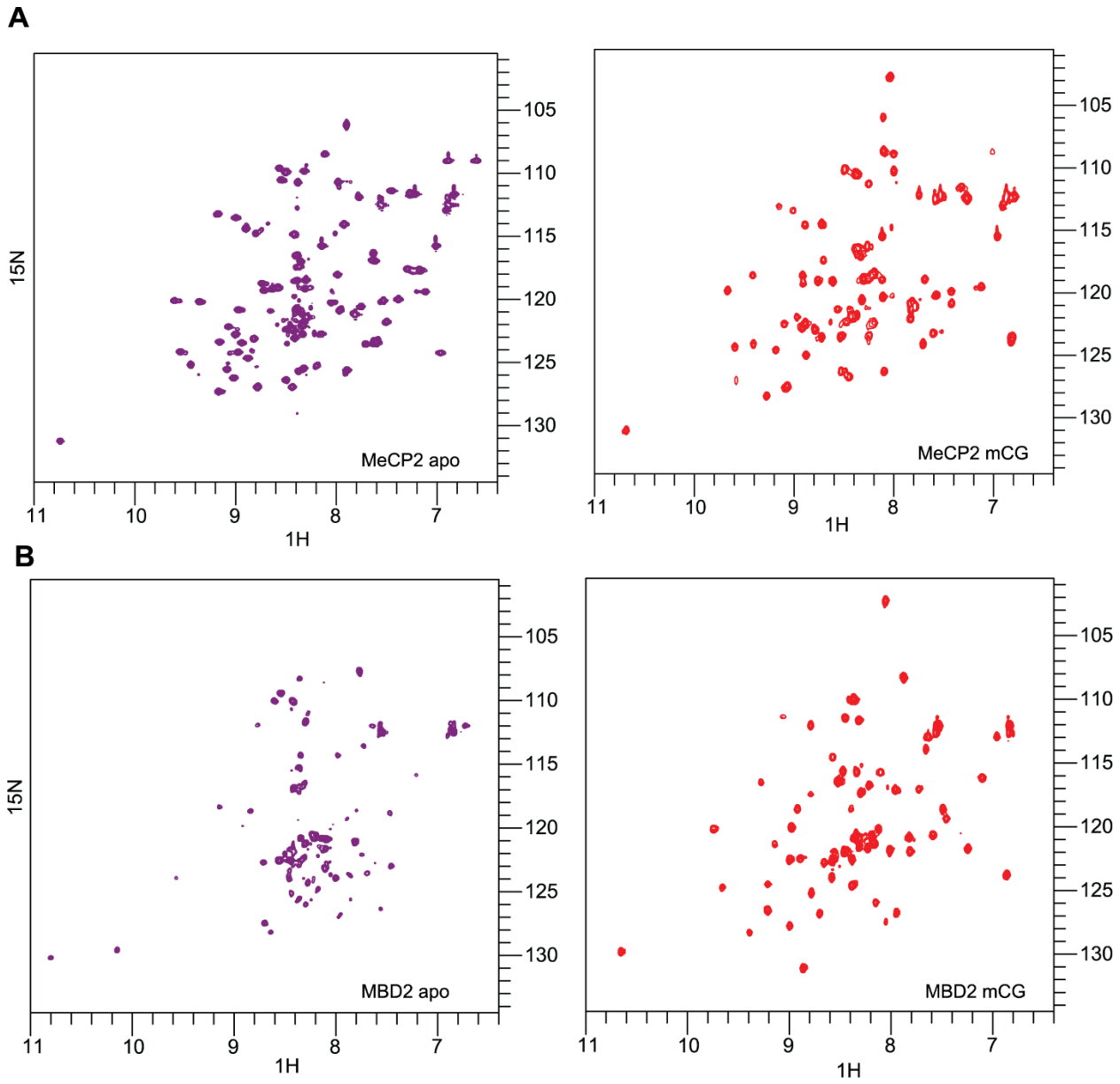

Figure 2.18 MeCP2 and MBD2 structural changes induced by binding DNA
^{15}N HSQC spectra of the methyl-binding domain of (a) MeCP2 off DNA (left, purple) or on mCG DNA (right, red) and (b) MBD2 off DNA (left, purple) or on mCG DNA (right, red).

Table 2.1 Binding affinity of methyl-binding domains to methylated BDNF DNA

Complex	K_D(nM)	n	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)
MeCP2:mCG	50 ± 17	1.1	-13.7 ± 0.3	3.3	-10.4 ± 0.2
MeCP2:mCA-5	409 ± 87	1.3	-11.9 ± 0.3	2.8	-9.1 ± 0.1
MeCP2:mCA-3	81 ± 13	0.9	-17.8 ± 0.2	7.7	-10.1 ± 0.1
MBD2:mCG	157 ± 23	1.1	-15.9 ± 0.2	6.2	-9.7 ± 0.1
MBD2:mCA-3	>1900	0.9	-11.8 ± 0.9	3.7	-8.1 ± 0.2
MeCP2:mCAC-5	79 ± 17	0.9	-18.1 ± 0.2	8.0	-10.1 ± 0.1
MeCP2:mCAC-3	36 ± 10	0.8	-16.3 ± 0.3	5.7	-10.6 ± 0.2

Table 2.2 Binding affinity of MeCP2 to hydroxymethylated BDNF DNA

Substrate	K_D(nM)	n	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)
hemi-hmCG	667 ± 129	0.7	-10.5 ± 0.4	1.7	-8.8 ± 0.1
hmCA-3	161 ± 26	1.0	-18.1 ± 0.3	8.4	-9.6 ± 0.1

Table 2.3 ^{15}N NMR relaxation data for the backbone N-H groups of the MeCP2:mCG complex fit to the CR72 model

Residue	Rate (sec^{-1})	Fraction Population A	ω (ppm)
R 85	637.32 ± 137.11	0.87 ± 0.07	0.86 ± 0.33
E 102	1347.57 ± 66.03	0.99 ± 0.00	2.82 ± 0.11
T 105	3240.66 ± 807.13	1.00 ± 0.00	9.04 ± 0.94
G 114	4673.68 ± 970.73	1.00 ± 0.00	13.32 ± 1.13
D 121	1217.47 ± 356.34	0.97 ± 0.04	2.97 ± 0.71
V 122	878.57 ± 303.22	0.98 ± 0.04	1.45 ± 0.55
N 126	382.60 ± 243.82	0.99 ± 0.07	2.28 ± 0.27
A 131	1794.87 ± 187.65	0.96 ± 0.05	1.63 ± 0.52
R 133	2968.64 ± 778.36	0.99 ± 0.03	2.67 ± 1.38
D 147	4227.73 ± 692.19	0.99 ± 0.04	2.03 ± 1.28
F 155	2380.65 ± 511.80	1.00 ± 0.01	3.59 ± 0.91
T 160	1091.09 ± 337.70	0.84 ± 0.09	1.25 ± 0.84

Table 2.4 ^{15}N NMR relaxation data for the backbone N-H groups of the MeCP2:mCG complex fit to the TSMFK01 model for very slow exchange

Residue	Rate (sec^{-1})	ω (ppm)
A 79	1.04 ± 0.10	3.08 ± 0.26
R 84	4.77 ± 0.53	2.18 ± 0.20
M 94	3.05 ± 0.32	3.39 ± 0.38
G 103	1.96 ± 0.07	8.41 ± 0.28
R 111	3.46 ± 0.40	6.17 ± 0.72
K 112	2.79 ± 0.31	5.77 ± 0.71
R 115	3.15 ± 0.20	5.76 ± 0.31
G 118	2.24 ± 0.34	5.91 ± 0.90
I 125	3.02 ± 0.18	5.58 ± 0.34
G 129	0.82 ± 0.06	11.04 ± 0.68
K 130	2.62 ± 0.06	5.33 ± 0.13
F 132	2.91 ± 0.20	5.51 ± 0.37
V 136	3.54 ± 0.22	9.70 ± 0.58
E 137	1.91 ± 0.30	8.44 ± 1.52
L 138	3.97 ± 0.23	5.92 ± 0.33
A 140	1.76 ± 0.15	6.27 ± 0.45
V 145	4.79 ± 0.13	5.27 ± 0.13
G 146	3.27 ± 0.08	2.76 ± 0.07
S 149	0.88 ± 0.05	9.58 ± 0.44
L 150	1.24 ± 0.06	2.90 ± 0.12
D 151	0.87 ± 0.04	6.19 ± 0.26
N 153	2.78 ± 0.09	3.15 ± 0.90
G 161	3.69 ± 0.45	2.52 ± 0.27
R 162	6.43 ± 0.67	9.51 ± 1.04

Table 2.5 ^{15}N NMR relaxation data for the backbone N-H groups of the MBD2:mCG complex fit to the CR72 model

Residue	Rate (sec^{-1})	Fraction Population A	ω (ppm)
L 28	1246.26 ± 500.75	1.00 ± 0.04	3.52 ± 0.75
S 33	2200.35 ± 436.01	0.98 ± 0.03	2.03 ± 0.91
D 34	1703.77 ± 311.43	0.97 ± 0.04	2.52 ± 0.72
V 35	1268.42 ± 275.68	0.99 ± 0.03	1.96 ± 0.62
Y 37	3075.67 ± 559.82	0.99 ± 0.03	2.67 ± 1.18
F 45	1267.34 ± 495.17	0.98 ± 0.05	3.87 ± 0.79
R 46	1511.04 ± 746.02	0.99 ± 0.08	4.03 ± 1.26
S 47	2031.31 ± 652.61	0.99 ± 0.05	3.16 ± 1.17
Y 54	1531.25 ± 377.27	0.99 ± 0.01	3.31 ± 0.71
D 60	1277.35 ± 472.82	1.00 ± 0.02	3.67 ± 0.68

Table 2.6 ^{15}N NMR relaxation data for the backbone N-H groups of the MBD2:mCG complex fit to the TSMFK01 model for very slow exchange

Residue	Rate (sec^{-1})	ω (ppm)
D 3	1.33 ± 0.04	3.79 ± 0.09
K 4	1.15 ± 0.02	3.56 ± 0.04
E 21	7.22 ± 0.21	3.89 ± 0.09
V 22	11.34 ± 0.56	6.64 ± 0.29
K 44	3.41 ± 0.16	6.38 ± 0.29
K 48	2.18 ± 0.23	6.40 ± 0.63
Q 50	2.90 ± 0.27	10.25 ± 0.93
A 58	1.77 ± 0.06	3.81 ± 0.12
L 61	2.88 ± 0.19	6.05 ± 0.39

Table 2.7 Dynamic light scattering of MeCP2 A140V mutant MBD

[KCl] (mM)	Radius (nm)	Polydispersity (%)
100	62.9 ± 2.3	23 ± 2.1
300	129.4 ± 27.8	Multimodal
600	536 ± 139	Multimodal

REFERENCES

1. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341,** 629–643 (2013).

2. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci.* **112,** 201505909 (2015).

3. Kinde, B. *et al.* Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 6800–6 (2015).

4. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* **17,** 215–222 (2014).

5. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522,** 89–93 (2015).

6. Gigek, C. O., Chen, E. S. & Smith, M. A. C. Methyl-CpG-Binding Protein (MBD) Family: Epigenomic Read-Outs Functions and Roles in Tumorigenesis and Psychiatric Diseases. *J. Cell. Biochem.* **117,** 29–38 (2016).

7. Fraga, M. F. *et al.* The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* **31,** 1765–74 (2003).

8. Ohki, I. *et al.* Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105,** 487–497 (2001).

9. Ho, K. L. *et al.* MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* **29,** 525–531 (2008).

10. Scarsdale, J. N., Webb, H. D., Ginder, G. D. & Williams, D. C. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* **39,** 6741–52 (2011).

11. Otani, J. *et al.* Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.* **288,** 6351–6362 (2013).

12. Du, Q., Luu, P.-L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7,** 1051–73 (2015).

13. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–9 (2012).

14. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to

5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–30 (2012).

15. Heimer, B. W., Tam, B. E. & Sikes, H. D. Characterization and directed evolution of a methyl-binding domain protein for high-sensitivity DNA methylation analysis. *Protein Eng. Des. Sel.* **28,** 543–551 (2015).

16. Esadze, A. *et al.* Changes in conformational dynamics of basic side chains upon protein-DNA association. *Nucleic Acids Res.* **44,** 6961–70 (2016).

17. Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19,** 667–678 (2005).

18. Mezei, P. D. & Csonka, G. I. Features of the interactions between the methyl-CpG motif and the arginine residues on the surface of MBD proteins. *Struct. Chem.* **27,** 1317–1326 (2016).

19. Kudo, S. *et al.* Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.* **40,** 487–93 (2003).

20. Free, A. *et al.* DNA recognition by the methyl-CpG binding domain of MeCP2. *J. Biol. Chem.* **276,** 3353–60 (2001).

21. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–93 (1995).

22. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* **59,** 687–696 (2005).

23. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. & Skelton, N. J. *Protein NMR Spectroscopy*. (Academic Press, 1996).

24. Loria, J. P., Rance, M. & Palmer, A. G. A Relaxation-Compensated Carr - Purcell - Meiboom - Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* **121,** 2331–2332 (1999).

25. Long, D., Liu, M. & Yang, D. Accurately probing slow motions on millisecond timescales with a robust NMR relaxation experiment. *J. Am. Chem. Soc.* **130,** 2432–2433 (2008).

26. d'Auvergne, E. J. & Gooley, P. R. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* **40,** 107–119 (2008).

27. d'Auvergne, E. J. & Gooley, P. R. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR* **40,** 121–133 (2008).

28. Carver, J. P. & Richards, R. E. A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J. Magn. Reson.* **6,** 89–105 (1972).

29. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123,** 11341–11352 (2001).

30. Akaike, H. Information theory and an extension of the maximum likelihood principle. in *International Symposium on Information Theory* 267–81 (Springer New York, 1973). doi:10.1016/j.econlet.2011.12.027

31. d'Auvergne, E. J. & Gooley, P. R. The use of model selection in the model-free analysis of protein dynamics. *Journal of Biomolecular NMR* **25,** 25–39 (2003).

32. Walavalkar, N. M., Gordon, N. & Williams, D. C. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66α). *J. Biol. Chem.* **288,** 3419–27 (2013).

33. Kumita, J. R., Smart, O. S. & Woolley, G. A. Photo-control of helix content in a short peptide.

34. Westbrook, J., Feng, Z., Chen, L., Yang, H. & Berman, H. M. The Protein Data Bank and structural genomics. *Nucleic Acids Res.* **31,** 489–91 (2003).

35. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

36. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18,** 6538–47 (1998).

37. Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* **19,** 269–277 (2003).

38. Cramer, J. M. *et al.* Methylation specific targeting of a chromatin remodeling complex from sponges to humans. *Sci. Rep.* **7,** 40674 (2017).

39. Cramer, J. M. *et al.* Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. *J. Biol. Chem.* **289,** 1294–1302 (2014).

40. Walavalkar, N. M., Cramer, J. M., Buchwald, W. A., Scarsdale, J. N. & Williams Jr., D. C. Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. *Nucleic Acids Res.* **42,** 11218–11232 (2015).

41. Morin, S. *et al.* Relax: The analysis of biomolecular kinetics and thermodynamics using NMR relaxation dispersion data. *Bioinformatics* **30,** 2219–2220 (2014).

42. Lagger, S. *et al.* Domains of methylated CAC and CG target MeCP2 to tune transcription in the brain. *bioRxiv* (2016). doi:https://doi.org/10.1101/087577

43. Spolar, R. S. & Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263,** 777–84 (1994).

44. Ferrari, M. E. & Lohman, T. M. Apparent Heat Capacity Change Accompanying a Nonspecific Protein-DNA Interaction. Escherichia coli SSB Tetramer Binding to Oligodeoxyadenylates. *Biochemistry* **33,** 12896–12910 (1994).

45. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5,** 789–796 (2009).

46. Pufall, M. A. *et al.* Variable Control of Ets-1 DNA Binding by Multiple Phosphates in an Unstructured Region. *Source Sci. New Ser.* **309,** 142–145 (2005).

47. Etheve, L., Martin, J. & Lavery, R. Protein-DNA interfaces: a molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic Acids Res.* **44,** 9990–10002 (2016).

48. Ganji, M., Docter, M., Le Grice, S. F. J. & Abbondanzieri, E. A. DNA binding proteins explore multiple local configurations during docking via rapid rebinding. *Nucleic Acids Res.* **44,** 8376–8384 (2016).

49. Chia, J. Y. *et al.* A/T Run Geometry of B-form DNA Is Independent of Bound Methyl-CpG Binding Domain, Cytosine Methylation and Flanking Sequence. *Sci. Rep.* **6,** 31210 (2016).

50. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19,** 31–38 (2009).

51. Mészáros, B., Tompa, P., Simon, I. & Dosztányi, Z. Molecular Principles of the Interactions of Disordered Proteins. *J. Mol. Biol.* **372,** 549–61 (2007).

52. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27,** 322–6 (2001).

53. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27,** 327–331 (2001).

54. Collins, A. L. *et al.* Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13,** 2679–89 (2004).

55. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25,** 558–

570 (2016).

56. Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* **14,** 1607–1616 (2011).

CHAPTER 3: INTRODUCTION TO THE NURD COMPLEX ARCHITECTURE AND FUNCTION

Nucleosome Remodeling Deacetylase Complex

The nucleosome remodeling and deacetylase complex is a regulator of chromatin structure and gene expression conserved in animals evolutionarily through sponges [1]. As mentioned in Chapter 1, the megadalton complex is comprised of a combination of subunits: MBD2/3, HDAC1/2, MTA1/2, RBBP4/7, GATAD2A/B and CHD3/4. NuRD has generally been considered a transcriptional corepressor complex, as MBD2 targets the histone deacetylase (HDAC1/2) and nucleosome repositioning ATPase (CHD3/4) to induce gene silencing at methylated CpG islands [2]. However, the presence of MBD3 excludes MBD2 from the complex and displays specificity for CG sequences but not enhanced affinity for mC [3–6]. In fact, MBD3-NuRD is found at transcriptionally active enhancers and promoters, suggesting that NuRD plays a diverse set of roles in maintaining and restructuring chromatin organization [7,8].

Proteomic studies provide insight into the stoichiometric ratios and special proximity of NuRD components [9] and structural studies have characterized a few interactions between subunit domains and chromatin [10,11]. Crystallographic and NMR structures illustrate the interfaces of MBD2/3 and methylated DNA [12], the MBD2 coiled-coil domain and GATAD2A conserved region 1 (CR1) [13], the WD40 domain of RBBP4/7 with the tail of H4 [14] and two C-terminal α-helical regions of MTA1 [10,11], the ELM2-SANT domains of MTA1 wrapping around HDAC1 [15], and an N-terminal peptide of H3

methylated at Lys[9] bound to PHD2 finger of CHD4 [16]. Enzymatic assays demonstrate direct deacetylase activity of HDAC1/2 and histone proteins containing acetylated lysine residues [17], as well as the helicase activity of CHD3/4 in the presence of nucleosomes and ATP [18]. Biochemical experiments demonstrate additional interactions between HDAC1/2 and RBBP4/7 [10], the PHD1 finger of CHD4 and unmodified H3 tails [16], the MBD2 intrinsically disordered region (IDR) and the histone deacetylase core complex (HDCC) of HDAC1/2, MTA1/2, and RBBP4/7 [19], and possibly GATAD2A and histones [20] and/or CHD4 [9,13]. The confirmed interactions are mapped in **Figure 3.1**, however, the full picture of complex formation is not complete as the function of conserved domains in MTA1/2, GATAD2A/b and CHD3/4 have yet to be determined.

Methyl-CpG Binding Domain Protein 2 MBD2

MBD2 and MBD3 are the evolutionarily most ancient members of the methyl-CpG binding domain (MBD) protein family. Originating as a single gene product with selectivity for methylated cytosines (mC) in invertebrates, the sequence likely duplicated and subsequently diverged to separate functions in vertebrates [21]. Although the mammalian MBD2/3 proteins share ~70% sequence identity [22], MBD2 retains mC selectivity [23] and changes in critical residues at the DNA interface of MBD3 have reduced binding preference for methylated DNA [6,24]. Both proteins associate with the NuRD complex in a mutually exclusive manner, but only MBD2-NuRD has been implicated in transcription repression [4].

The human MBD2 gene encodes two isoforms with a core 30 kDa protein containing an N-terminal MBD domain, a central intrinsically disordered region, and a C-terminal coiled-coil motif. Knockdown of MBD2 inhibits the growth of human cancer cells

lines [25,26] and xenograft tumors [27] and MBD2-definicient mice are resistant to tumorigenesis [28,29]. MBD2 also contributes to the developmental silencing of embryonic and fetal globin genes [30], re-expression of which are promising avenues for treating hemoglobinopathies. Knockout of several NuRD complex proteins is embryonic lethal (HDAC1,GATAD2A/B, CHD4) [31–33], however, MBD2 knockout mice are viable and fertile, thereby making it a more desirable therapeutic target.

Metastasis associated 1 family member protein 2 MTA2

The founding member of the MTA protein family was identified as a differentially expressed gene in a metastatic rat mammary adenocarcinoma cell line [34]. In human cancers, MTA1 and MTA2 are the most commonly upregulated genes and correlate with tumor invasiveness, chemotherapeutic resistance, poor clinical prognosis and low survival rates [35]. The MTA2 amino acid sequence is 68% homologous to MTA1 and shares the typical architecture with Bromo-Adjacent Homology (BAH), egl-27 and MTA1 homology (ELM2), SWI, ADA2, N-CoR, TFIIIB-B (SANT) and GATA-like zinc-finger (ZnF) domains. The ELM2 and SANT domains of MTA1 have been crystallized in complex with HDAC1 [15] and helices in the less structured c-terminal region of MTA1 have been crystallized in complex with RBBP4/7 [10,11]. The MTA family zinc-finger is atypical with little homology to other GATA domains and the MTA1 ZnF has been shown to immunoprecipitate RBBP4/7 and implicated in facilitating FOG-2 mediated transcriptional repression [36]. The MTA2 ZnF has not been characterized and, in Chapter 4, we present findings of binding assays with various substrates to further elucidate function.

GATA zinc finger domain containing protein 2A GATAD2A

GATAD2A and GATAD2B share 41% protein sequence identity with 2 regions highly conserved in animals and between the two proteins, conserved region 1 (CR1) with 83% identity and conserved region 2 (CR2) with 74% identity. Little work has been done to define GATAD2A's biological function beyond establishing knocking-out the gene in mice is embryonic lethal and confirming its presence in the NuRD complex [31,37]. Our previous studies have extensively characterized the interaction of CR1 with a coiled-coil domain in MBD2/3 and demonstrated its necessity for the recruitment of GATAD2A and CHD4 to the NuRD complex. Recent work has further indicated that the presence of GATAD2A in NuRD is critical to the complex recruitment to sites of DNA damage [38]. A limited examination of CR2 led to the identification of histone tail binding activity [20], however, our previously published work on the MBD2/GATAD2A interaction [13] led to a hypothesis that this region is also involved in the interaction with CHD4, which is investigated in Chapter 4.

Chromodomain helicase DNA binding protein 4 CHD4

The CHD family of proteins belong to the SNF2/RAD54 helicase family of genes and participate in the epigenetic regulation of transcription [39] and facilitation of DNA repair [40–42]. The CHD3 and CHD4 functions within the NuRD complex to modify nucleosome positioning and control target gene expression. Mutations in CHD3/4 have been associated with endometrial tumors [43,44] and developing autoimmunity against them causes dermatomyositis [39], while knockdown and knockout of CHD4 in cell lines and xenograft models of osteosarcoma and acute myeloid leukemia sensitize tumors to genotoxic agents and ionizing radiation [45,46].

CHD3 and CHD4 are 69% identical and share a common domain structure including an n-terminal region, tandem PHD fingers, tandem chromo domains, an ATPase-helicase and c-terminal CHD domain (CTDCHD) or domains of unknown function (DUF) [47]. The n-terminal region of CHD4 contains aHMG-box-like-domain which binds poly(ADP-ribose) in a poly(ADP-ribose) polymerase-dependent manner for recruitment to sites of DNA double-stranded breaks [48]. Of the CHD4 tandem PHD fingers, PHD1 binds unmodified H3 tails and PHD2 binds H3 methylated at Lys^9 and are required to recruit NuRD to target genes for transcriptional silencing [49]. The chromo domains of the homologous CHD4 protein in *Drosophila melanogaster* are not essential for NuRD complex formation but have demonstrated the ability to bind DNA and stimulate ATPase activity of the helicase domain [50]. The helicase alone is sufficient to bind nucleosomes and hydrolyze ATP, and the ability to slide mononucleosomes is enhanced in the presence of the PHD fingers and chromo domains [51]. Truncations that interrupt the CHDCTD domain were shown to disrupt repression or reporter gene expression without changing CHD4 promoter occupation [52], however, the function of this domain has not been further explored. In Chapter 4, we discuss work that investigates the role of the CHDCTD domain in association with the NuRD complex.

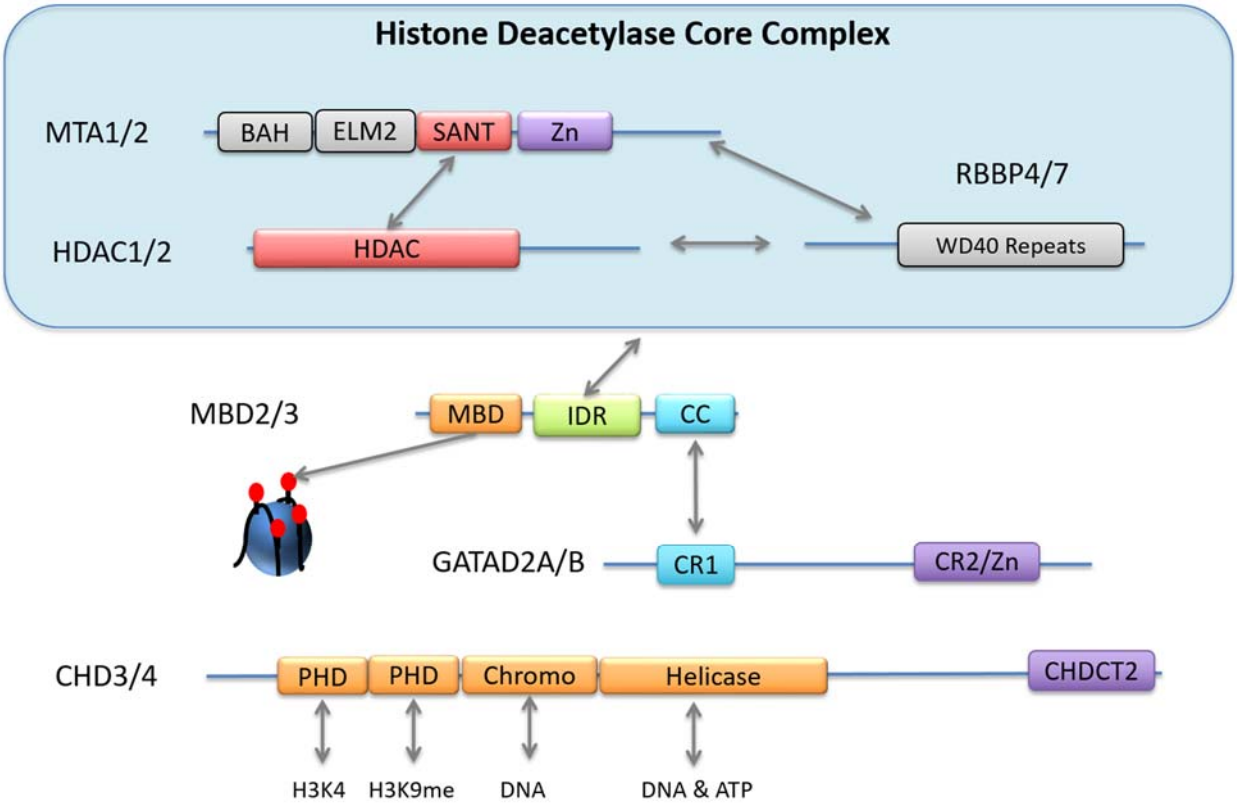

Figure 3.1 NuRD Complex Subunit Domain Interaction Map
The histone deacetylase core subcomplex (HDCC) of NuRD (highlighted in blue) forms through interfaces formed between the enzymatic domain of HDAC1/2 and the ELM2-SANT domains (shown in red), c-terminal regions of MTA1/2 and RBBP4/7, as well as the interaction of HDAC1/2 and RBBP4/7. The HDCC associates with the IDR (green) of MBD2/3 which localizes the complex to methylated (red dots) or unmethylated CpG islands. GATAD2A/B is recruited to the complex through a coiled-coil interaction between the α-helix that comprises its CR1 domain and a c-terminal α-helix of MBD2 (shown in cyan). CHD3/4 binds H3 n-terminal tails either unmodified or methylated at Lys^9 via tandem PHD fingers and binds to DNA through the tandem chromodomains and helicase domain (shown in orange). Domains of unknown function in MTA2, GATAD2A, and CHD4 are shown in purple.

REFERENCES

1. Cramer, J. M. *et al.* Methylation specific targeting of a chromatin remodeling complex from sponges to humans. *Sci. Rep.* **7,** 40674 (2017).

2. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell* **153,** 480–492 (2013).

3. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–9 (2012).

4. Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26,** 843–51 (2006).

5. Cramer, J. M. *et al.* Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. *J. Biol. Chem.* **289,** 1294–1302 (2014).

6. Fraga, M. F. *et al.* The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* **31,** 1765–74 (2003).

7. Shimbo, T. *et al.* MBD3 Localizes at Promoters, Gene Bodies and Enhancers of Active Genes. *PLoS Genet.* **9,** e1004028 (2013).

8. Gunther, K. *et al.* Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res.* **41,** 3010–3021 (2013).

9. Kloet, S. L. *et al.* Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.* **282,** 1774–1785 (2015).

10. Millard, C. J. *et al.* The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *Elife* **5,** 1–21 (2016).

11. Alqarni, S. S. M. *et al.* Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. *J. Biol. Chem.* **289,** 21844–55 (2014).

12. Scarsdale, J. N., Webb, H. D., Ginder, G. D. & Williams, D. C. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* **39,** 6741–52 (2011).

13. Gnanapragasam, M. N. *et al.* p66α–MBD2 coiledcoil interaction and recruitment of Mi2 are critical for globin gene silencing by the MBD2–NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7487–7492 (2011).

14. Murzina, N. V *et al.* Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* **16,** 1077–85 (2008).

15. Millard, C. J. *et al.* Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol. Cell* **51,** 57–67 (2013).

16. Mansfield, R. E. *et al.* Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J. Biol. Chem.* **286,** 11779–11791 (2011).

17. Taunton, J., Hassig, C. A. & Schreiber, S. L. A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. *Science (80-.).* **272,** (1996).

18. Mahajan, M. C., Narlikar, G. J., Boyapaty, G., Kingston, R. E. & Weissman, S. M. Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the beta-globin locus control region. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15012–7 (2005).

19. Desai, M. A. *et al.* An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* **43,** 3100–3113 (2015).

20. Brackertz, M., Gong, Z., Leers, J. & Renkawitz, R. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res.* **34,** 397–406 (2006).

21. Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* **19,** 269–277 (2003).

22. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18,** 6538–47 (1998).

23. Matsumoto, M. & Toraya, T. cDNA cloning, expression, and characterization of methyl-CpG-binding domain type 2/3 proteins from starfish and sea urchin. *Gene* **420,** 125–34 (2008).

24. Saito, M. & Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* **277,** 35434–9 (2002).

25. Mian, O. Y. *et al.* Methyl-Binding Domain Protein 2-Dependent Proliferation and Survival of Breast Cancer Cells. *Mol. Cancer Res.* **9,** 1152–1162 (2011).

26. Cheishvili, D. *et al.* Synergistic effects of combined DNA methyltransferase inhibition and MBD2 depletion on breast cancer cells; MBD2 depletion blocks 5-aza-2'-deoxycytidine-triggered invasiveness. *Carcinogenesis* **35,** 2436–2446 (2014).

27. Slack, A. *et al.* AntisenseMBD2 gene therapy inhibits tumorigenesis. *J. Gene Med.* **4,** 381–389 (2002).

28. Sansom, O. J. *et al.* Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nat. Genet.* **34,** 145–147 (2003).

29. Campbell, P. M., Bovenzi, V. & Szyf, M. Methylated DNA-binding protein 2 antisense inhibitors suppress tumourigenesis of human cancer cell lines in vitro and in vivo. *Carcinogenesis* **25,** 499–507 (2003).

30. Rupon, J. W., Wang, S. Z., Gnanapragasam, M., Labropoulos, S. & Ginder, G. D. MBD2 contributes to developmental silencing of the human ε-globin gene. *Blood Cells, Mol. Dis.* **46,** 212–219 (2011).

31. Marino, S. & Nusse, R. Mutants in the Mouse NuRD/Mi2 Component P66α Are Embryonic Lethal. *PLoS One* **2,** e519 (2007).

32. O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S. & Hendrich, B. Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. *Development* **142,** 2586–97 (2015).

33. Lagger, G. *et al.* Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J.* **21,** 2672–2681 (2002).

34. Toh, Y., Pencil$, S. D. & Nicolsono, G. L. A Novel Candidate Metastasis-associated Gene, mtal, Differentially Expressed in Highly Metastatic Mammary Adenocarcinoma Cell Lines. *J. Biol. Chem.* **269,** 2295–22963 (1994).

35. Covington, K. R. & Fuqua, S. A. W. Role of MTA2 in human cancer. *Cancer Metastasis Rev.* **33,** 921–928 (2014).

36. Roche, A. E. *et al.* The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex. *J. Mol. Cell. Cardiol.* **44,** 352–60 (2008).

37. Brackertz, M., Boeke, J., Zhang, R. & Renkawitz, R. Two Highly Related p66 Proteins Comprise a New Family of Potent Transcriptional Repressors Interacting with MBD2 and MBD3. *J. Biol. Chem.* **277,** 40958–40966 (2002).

38. Spruijt, C. G. *et al.* ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. *Cell Rep.* **17,** 783–798 (2016).

39. Seelig, H. P. *et al.* The major dermatomyositis-specific mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis Rheum.* **38,** 1389–1399 (1995).

40. Rajagopalan, S., Nepa, J. & Venkatachalam, S. Chromodomain helicase DNA-binding protein 2 affects the repair of X-ray and UV-Induced DNA damage. *Environ. Mol. Mutagen.* **53,** 44–50 (2012).

41. Stanley, F. K. T., Moore, S. & Goodarzi, A. A. CHD chromatin remodelling enzymes and the DNA damage response. *Mutat. Res. Mol. Mech. Mutagen.* **750,** 31–44 (2013).

42. Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc. Natl. Acad. Sci.* **107,** 18475–18480 (2010).

43. Le Gallo, M. *et al.* Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat. Genet.* **44,** 1310–5 (2012).

44. Zhao, S. *et al.* Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl. Acad. Sci.* **110,** 2916–2921 (2013).

45. Larsen, D. H. *et al.* The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. *J. Cell Biol.* **190,** 731–740 (2010).

46. Pan, M.-R. *et al.* Chromodomain Helicase DNA-binding Protein 4 (CHD4) Regulates Homologous Recombination DNA Repair, and Its Deficiency Sensitizes Cells to Poly(ADP-ribose) Polymerase (PARP) Inhibitor Treatment. *J. Biol. Chem.* **287,** 6764–6772 (2012).

47. Marfella, C. G. A. & Imbalzano, A. N. The Chd family of chromatin remodelers. *Mutat. Res. Mol. Mech. Mutagen.* **618,** 30–40 (2007).

48. Silva, A. P. G. *et al.* The N-terminal Region of Chromodomain Helicase DNA-binding Protein 4 (CHD4) Is Essential for Activity and Contains a High Mobility Group (HMG) Box-like-domain That Can Bind Poly(ADP-ribose). *J. Biol. Chem.* **291,** 924–38 (2016).

49. Musselman, C. A. *et al.* Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc. Natl. Acad. Sci.* **109,** 787–792 (2012).

50. Bouazoune, K. *et al.* The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J.* **21,** 2430–2440 (2002).

51. Watson, A. A. *et al.* The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J. Mol. Biol.* **422,** 3–17 (2012).

52. Ramirez, J., Dege, C., Kutateladze, T. G. & Hagman, J. MBD2 and Multiple

Domains of CHD4 Are Required for Transcriptional Repression by Mi-2/NuRD Complexes. *Mol. Cell. Biol.* **32,** 5078–5088 (2012).

CHAPTER 4: CHARACTERIZATION OF NURD COMPLEX GATA-LIKE ZINC-FINGER DOMAINS

Characterization of GATAD2A-CHD4 Interaction Interface

Studies of the interaction between the GATAD2A conserved region 1 (CR1) domain and the MBD2 coiled-coil revealed that a CR1 peptide was sufficient to immunoprecipitate NuRD components with the exclusion of CHD4 [1]. Additionally, reporter assays of CHD4 have demonstrated that the c-terminal CHDCTD domain is required for transcriptional silencing activity yet not required for gene promoter occupation [2]. These observations led us to hypothesize that the conserved region 2 (CR2) of GATAD2A might interact with CHD4, possibly through an interface with the CHDCTD domain. We report here findings demonstrate that CR2, specifically the GATA-like zinc-finger, is sufficient to immunoprecipitate CHD4 to the exclusion of other NuRD components.

Materials and Methods

Cloning, Expression, and Purification of the GATAD2A Zinc-Finger Domain

GATAD2A CR2short (339-464 aa) and zinc-finger short (411-464 aa) constructs were cloned into the bacterial expression vector pGEX4T-1 (GE Healthcare Biosciences), which was engineered for the expression of recombinant proteins with an N-terminal tandem fusion tag of glutathione S-transferase (GST). The resulting GST-GATAD2A fusions were overexpressed in the Rosetta II (DE3) (Invitrogen) *E. coli* strain and grown in Luria Bertani (LB) medium containing 50 µg/ml of ampicillin to an optical

density of 0.5 -0.6 at 600 nm, and then induced with 0.2 mM isopropyl-β-D-thiogalactopyranooside for 2 h at 37 °C. Cells were harvested by centrifugation and lysed by sonication in 30 ml of the B-PER reagent (Thermo Scientific) supplemented with EDTA-free protease inhibitor tablets (Thermo Scientific). The fusion protein was isolated from the soluble fraction by GST affinity chromatography (GSTrap FF, GE Healthcare) and eluted with reduced glutathione buffer (25 mM glutathione, 50 mM Tris pH 8.8, 200 mM NaCl).

Fluorescent Polarization DNA Binding Studies

5'-end 6-FAM labeled DNA oligos were ordered from Integrated DNA Technologies, and annealed and HPLC purified. Oligos and protein were each buffer exchanged into sample buffer [10mM HEPES (pH 7.5), 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT]. The protein and DNA concentrations were measured by UV absorbance and diluted to 5 uM and 70 nM, respectively, in assay buffer [10mM HEPES (pH 7.5), 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 0.005% TWEEN 20]. In a black, flat bottom, half volume 96 well plate (COSTAR), protein was serially diluted 1:2 across each row, except for the protein-free control wells. DNA was added to a final concentration of 10 nM in every well, and samples in each row were mixed with a multi-channel pipette. The protein-DNA complexes were incubated for 10 minutes at room temperature and then crosslinked with glutaraldehyde of varying concentrations and incubation times. Plates were read on a CLARIOstar plate reader (BMG Labtech), by focus adjusting on a protein-free control well and measuring fluorescent anisotropy at 495 nm.

Mammalian Cell Culture

Human embryonic kidney 293T (HEK293T) cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) containing 10% heat-inactivated fetal bovine serum (Hyclone), 2mM L-glutamine and 100 U/ml penicillin and streptomycin. Cells were cultured at 37◦C and 5% CO2.

Co-Immunoprecipitation

GATAD2A constructs were cloned into the pCMV-Tag2b (Stratagene) vector in frame with an N-terminal flag-tag sequence. Plasmids were transfected into HEK293T cells with Turbofect according to the protocol for adherent cells (Thermo Scientific). At 48 h post transfection, HEK293T cells were scraped off of plates in ice-cold 1x PBS, centrifuged at 250 x g for 5 minutes and resuspended in ice cold micrococcal nuclease (MNase) digestion buffer (25 mM HEPES-KOH pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$, 3 mM $CaCl_2$, 10 % glycerol, 0.2 % NP-40 and 1x EDTA-free protease inhibitor cocktail (Roche)). Cell pellets were lysed on ice with a pistol tissue homogenizer followed by MNase digestion using 1500 U/ml of MNase (Worthington Biochemical, Lakewood, NJ, USA) for 2 h on ice. Ethidium bromide was then added to the lysate at 300 µg/ml followed by a spin at 10,000 x g for 15 mins at 4 °C. The supernatant was then pre-cleared by rotating with Protein G agarose beads (Thermo Scientific) for 2 hours at 4 °C. Beads were removed by centrifugation at 5,000 x g for 30 seconds at 4 °C and supernatant were split into two fractions, one each incubated with anti-flag M2 antibody (Sigma F1804) or a mouse IgG (Santa Cruz) control, rotating overnight at 4 °C. Protein complexes were precipitated with Protein G agarose beads (Thermo Scientific) rotating for 3 h at 4 °C. Beads were washed with ice cold high salt buffer (1x PBS, 300 mM NaCl) four times before elution by boiling in laemmli sample buffer.

Western Blot Analysis

Eluted immunoprecipitated proteins were analyzed for different components of the NuRD complex by western blot. Fractions were loaded in a pre-cast 4-20% gradient SDS-PAGE gels (Bio-Rad) and run at 120V for 60-90 minutes. Proteins were transferred to a 0.2 μm PVDF membrane activated in 100 % methanol running at 120V for 2 h in transfer buffer (25 mM Tris pH 8.5, 192 mM glycine, 10% methanol). Membranes were rinsed in 1x PBST and blocked at room temperature in 5% w/v nonfat dry milk for 30 minutes. Blots were probed with primary antibodies against RBBP4 (Novus NBP1-40622), HDAC2 (Millipore 05–814), MTA2 (Santa Cruz sc-28731), MBD2 (Santa Cruz sc-12444), CHD4 (Millipore 06–1306) and HRP-conjugated TrueBlot secondary antibodies (Rockland). After each antibody exposure blots were rinsed with 1x PBST three times for 15 minutes. Blots were incubated in Luminata Crescendo Western HRP substrate (Millipore) and visualized by CCD imaging (ImageQuant LAS 4000, GE Healthcare Biosciences).

Cloning and Co-expression of the GATAD2A Zinc-Finger

The GATAD2A zinc-finger (400-480 aa) was cloned into the pETDuet vector with an n-terminal 6xHis-tag and constructs of the c-terminal domains of CHD4 (1530-1912 aa, 1686-1912 aa) were cloned into the second site of bicistronic pETDuet vector and the bacterial co-expression vector pCDF. Co-expression tests were conducted with Rosetta II (DE3) (Invitrogen) *E. coli* strain grown in Luria Bertani (LB) medium containing 50 μg/ml of ampicillin and/or kanamycin to an optical density of 0.6 at 600 nm. Protein expression was induced with 1 mM IPTG and temperature conditions were varied (2 h at 37 °C, 4 h at 32 °C, or overnight at 23 °C and 16 °C). Cells were lysed in

B-PER reagent, the supernatant was removed and a portion of the membrane fraction was solubilized in 6M Urea. Samples of the soluble and insoluble fractions were boiled in laemmli sample buffer and analyzed by SDS-PAGE.

Sequence Alignments

Amino acid sequences of GATAD2A CR2 and CDH4 CHDCTD were queried NCBI BLASTp to identify homologs in *Amphimedon queenslandic*a and *Hydra vulgaris*. Clustal Omega of the analogous domains was used to generate a multiple sequence alignment (MSA). Sequence conservation was visualized with the Color Align Conservation tool of the Sequence Manipulation Suite [3].

Results

Conserved region 2 (CR2) of GATAD2A contains the GATA-like zinc-finger (GATAD2A ZnF) that gives the protein its name and is evolutionarily conserved in animals through sponges (Figure 4.1). From our published studies of the coiled-coil interaction between MBD2 and GATAD2A conserved region 1 (CR1), as mentioned in Chapter 3, we suspected that the CR2 domain was the critical interface for recruiting CHD4 to the NuRD complex. CR2, however, has also been previously implicated in binding chromatin and histone proteins [4]. To test this prior evidence, we purified a GST-tagged fusion of CR2 (Figure 4.2) that our collaborators in the Strahl lab screened for histone binding activity with an EPITITAN™ modified histone tail peptide array (EpiCypher), which yielded zero hits. We also ran a preliminary fluorescence polarization DNA binding assay, in the presence of the protein-DNA cross-linker glutaraldehyde, and did not detect a non-specific interaction with DNA (Figure 4.3).

To characterize the interaction of CR1 and CR2 with the NuRD complex, we performed immunoprecipitations with an n-terminal or c-terminal segment of GATAD2A and probed for the presence of RBBP4, HDAC2, MTA2, MBD2, and CHD4. We used ethidium bromide and a high salt wash to eliminate bridging interactions through the DNA binding domains of the complex. We confirmed our past finding that CR1 (1- 178 aa) associates with RBBP4, HDAC2, MTA2, MBD2 but not CHD4. We also discovered that the c-terminal segment including CR2 (339-633 aa) associated with CHD4 but not any of the other probed NuRD complex components (Figure 4.4A). Subsequent immunoprecipitations identified that CR2 alone (339-480 aa), specifically the sequence encoding the predicted GATAD2A ZnF (400-480 aa), was sufficient to associate with CHD4 (Figure 4.5).

We hypothesized that the c-terminal domain of CHD4, also known as the CHDCTD, would be the critical region for interacting with GATAD2A. This region is conserved in animals through sponges (Figure 4.6) and recent proteomic studies of the human NuRD complex identified crosslinks between GATAD2A CR2 and amino acids in the CHDCTD region of CHD4 [5]. Our collaborators in the Mackay group determined that GATAD2B interacts with the second half of the CHDCTD domain amino acids 1686-1912, which we refer to as C2 (Figure 4.4B and data not shown).

We were interested in further characterizing the interaction of CHD4 and the GATAD2A ZnF in isolation, therefore, we attempted several approaches for bacterial expression. Initial experiments with cleaving the purified GATAD2A ZnF from the GST-tag resulted in protein precipitation. Additionally, we suspected that the terminal helix of the zinc-finger (464-480 aa) might be involved in the interface, therefore we sought

strategies to express a larger construct. The fusion of a GST-tag to the full GATAD2A ZnF (400-480 aa) and expressing C2 independently both yielded only insoluble protein. We hypothesized that the interaction of the c-terminal region of CHD4 with the GATAD2A ZnF could stabilize both domains and we cloned segments into a bicistronic petDuet vector. By Coomassie staining, we detected bands in the insoluble fraction for CR2 and possibly in the soluble fraction for the full-length CHDCTD domain (Figure 4.7). Efforts to optimize the dual expression system did not achieve improved results, therefore, we moved to a two-plasmid approach. A 6xHis-zinc-finger fusion and isolated C2 sequence were cloned into compatible expression vectors, co-transformed and subjected to double antibiotic selection. Test growths indicated that expression at 23 °C overnight is the optimal condition for obtaining both protein products in the soluble fraction (data not shown).

Discussion and Future Directions

The presented results suggest that CHD4 may be recruited to the NuRD complex through an interaction between the CHDCTD domain and the CR2 domain of GATAD2A. We propose that a function of GATAD2A in the NuRD complex is to serve as the bridge between the two enzymatic components, the HDCC and CHD4 (Figure 4.8A). The zinc-finger domain of GATAD2A may contribute to the localization of CHD4 at methylated CpG-islands through the CR1 domain interaction with MBD2 (Figure 4.8B). In fact, previous studies have shown that early termination of the CHD4 transcript at 1840 aa disrupts foci in the nucleus [2]. Future experiments aim to confirm the direct binding of the GATAD2A ZnF and the C2 region of CHD4 by demonstrating co-purification. If bacterial protein production is insufficient to detect the interaction,

bioluminescence resonance energy transfer (BRET) with NanoLuc- and Halotag- fusion proteins may prove useful to test in cell proximity [6]. Disrupting BRET by removing the characteristic cysteines or the predicted c-terminal helix of the zinc-finger domain would provide further evidence of an interaction and identify the pertinent structures. Clarification of the minimal binding region through experiments with truncated constructs could also reveal a complex more amenable to co-expression and ultimately crystallization of the protein complex to elucidate the atomic-level details of the interaction.

```
Amphimedon IKAKEKQGSADTSLKCPTCDTSVNLPDGK----------------------IEGLTQNLW-FEHKSKEA
HUMAN      ESPASRQAAAKLALRKQLEKTLLEIPPPKPPAPEMNFLPSAANNEFIYLVGLEEVVQNLLETQGRMS--
Hydra      ISAEQKQNAAKTALHRQLEQTLLQIPPPRPPPADWKAIPNVNSMDFMMLVGLDEVVDSILEMDSKPTIK

Amphimedon SIKKKIHSREPILCGKCRVKDSSDAAVVYCCDCGKFLCEYCKNGHKRNPEKADHELIDLGEAARKALKE
HUMAN      --AATVLSREPYMCAQCKTDFTCRWR--EEK-SGAIMCENCMTTNQKKALKVEHTSR-LKAAFVKALQQ
Hydra      KALAELIPYNPRVCNQCNVDFSPCWK--SKDGEDFVLCERCALQNIKRDLKAEHTKR-LKDAFLKALTQ

Amphimedon SVTSCDGVIPPVTEAIANGE
HUMAN      EQE--------IEQRLLQQG
Hydra      EQE--------IEERIKAGE
```

Figure 4.1 Evolutionary conservation of GATAD2A CR2 sequence
Alignment of human GATAD2A amino acids 339-480 with orthologous proteins from early eukaryotes *Amphimedon queenslandica* and *Hydra vulgaris*. Identical residues are highlighted in black and similar residues are shaded in gray.

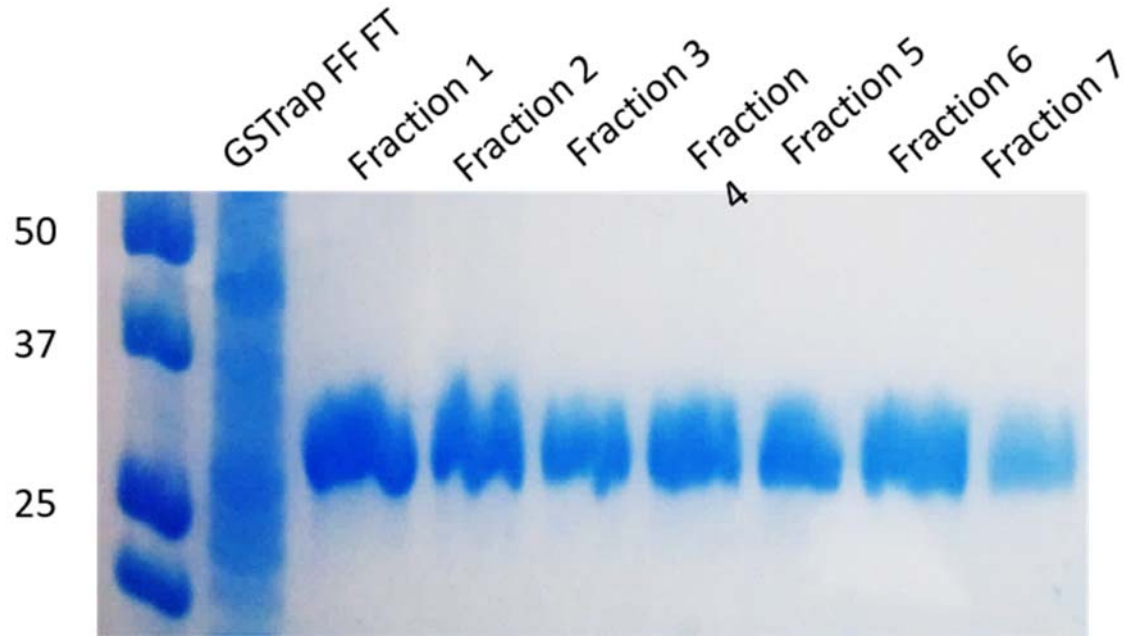

Figure 4.2 GST-fusion GATAD2A CR2 purification
Coomassie-stained SDS-PAGE gel of the flow-through (FT) and elution fractions from a GSTrap FF column extracting the GST-CR2 fusion protein (expected molecular weight 33 kDa).

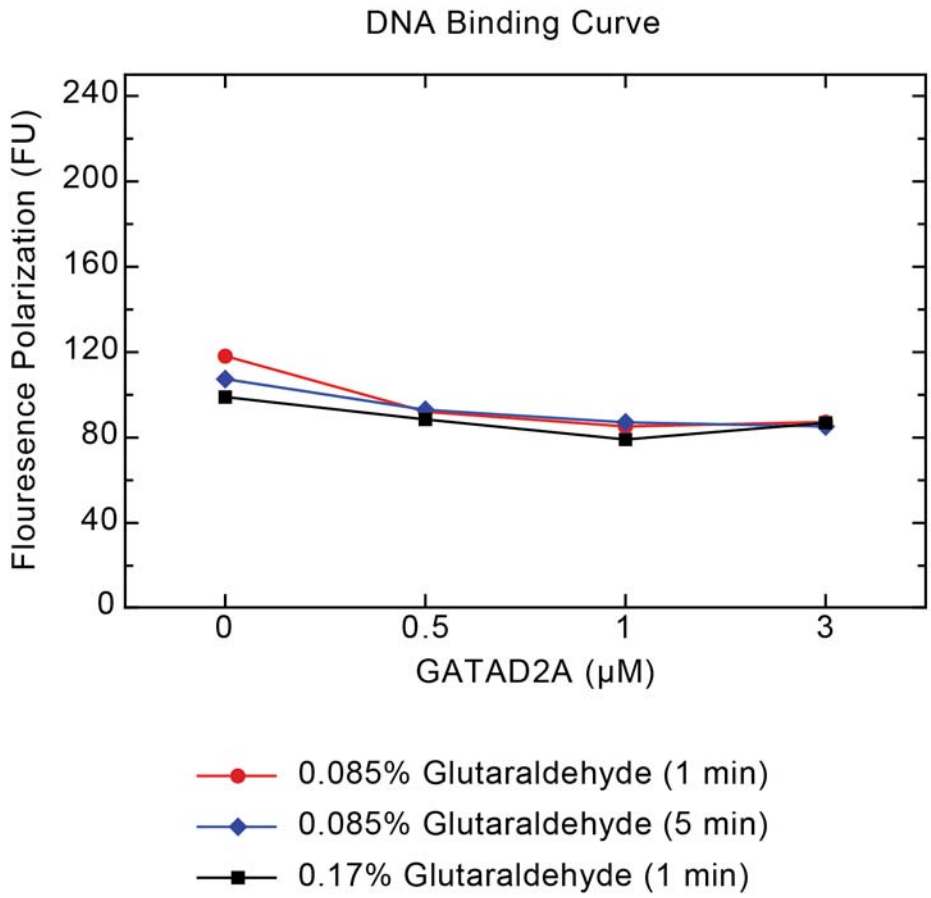

Figure 4.3 GATAD2A CR2 did not exhibit non-specific DNA binding activity
Plotted polarized fluorescence units (FU) detected with an increasing micromolar (μM) concentration of GST-CR2 protein at varying concentrations of glutaraldehyde crosslinking reagent.

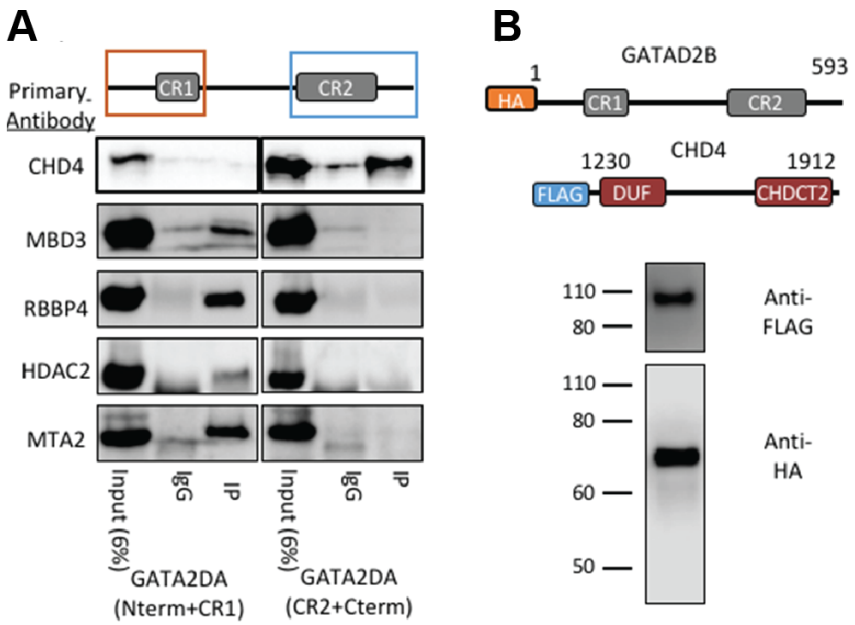

Figure 4.4 GATAD2A bridges the enzymatic components of the NuRD complex
A) The N-terminal half of GATAD2A interacts with most canonical NuRD components except CHD4 while C-terminal half interacts with CHD4. B) Likewise, immunoprecipitating the C-terminus of CHD4 is sufficient to detect GATAD2B.

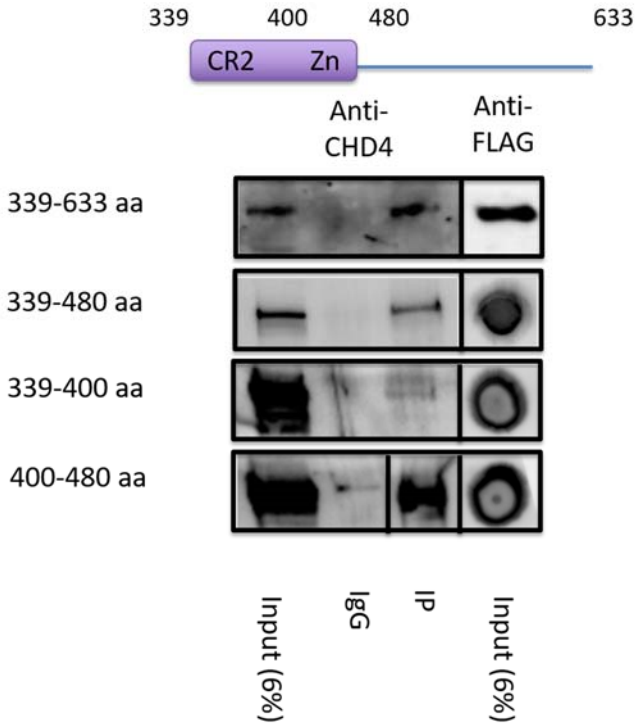

Figure 4.5 Zinc-finger domain of GATAD2A associates with CHD4
Removal of the amino acids c-terminal and n-terminal of the zinc-finger domain in the 339-480 aa and 400-480 aa constructs does not disrupt binding to CHD4. In the absence of the zinc-finger, the CR2 residues 339-400 do not pull down CHD4.

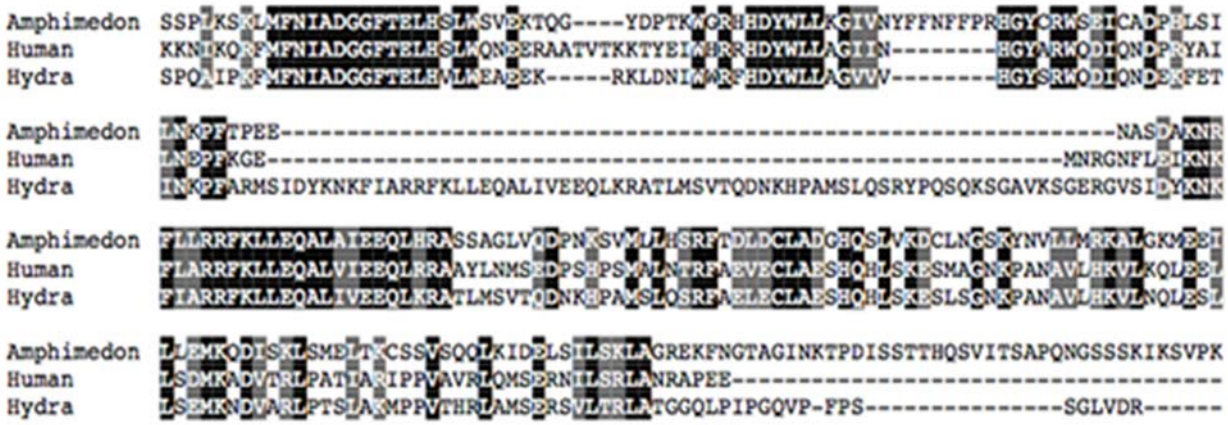

Figure 4.6 Evolutionary conservation of CHD4 C2 sequence
Alignment of human CHD4 amino acids 1686-1912 with orthologous proteins from early eukaryotes *Amphimedon queenslandica* and *Hydra vulgaris*. Identical residues are highlighted in black and similar residues are shaded in gray.

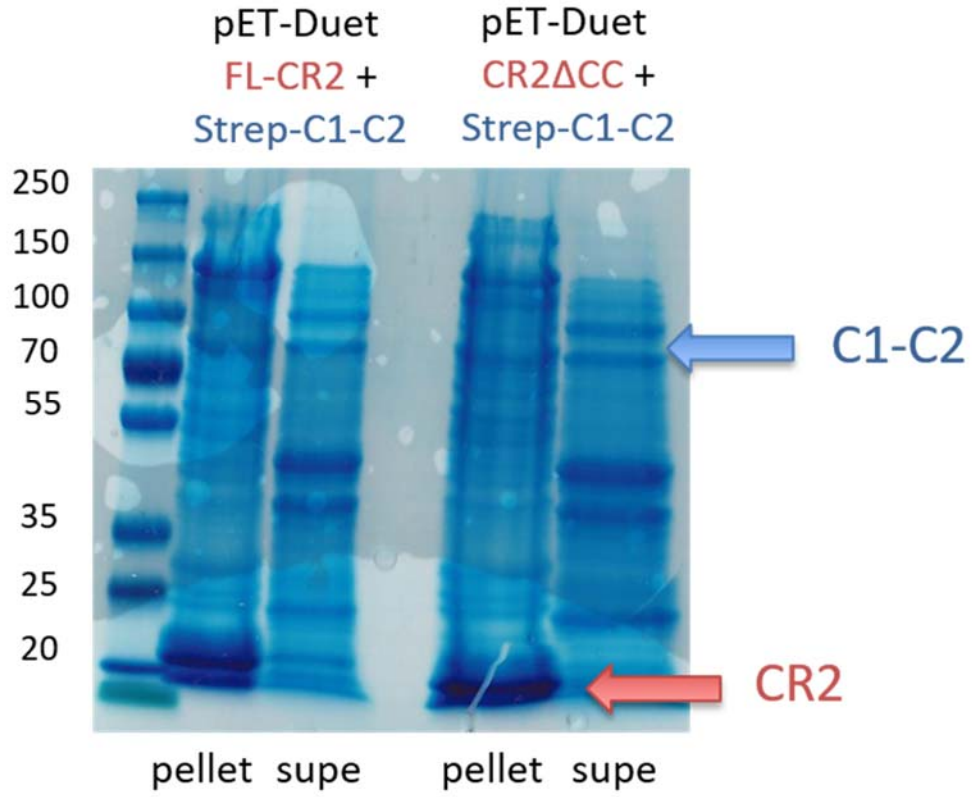

Figure 4.7 GATAD2A CR2 and CHD4 co-expression
Coomassie-stained SDS-PAGE gel of the insoluble (pellet) and soluble (supe) fractions of whole cell lysates after expressing CR2 and c-terminal CHD4 (C1-C2) constructs from the same bicistronic petDuet vector. Expected molecular weights are 17 kDa (CR2) and 80 kDa (CHD4).

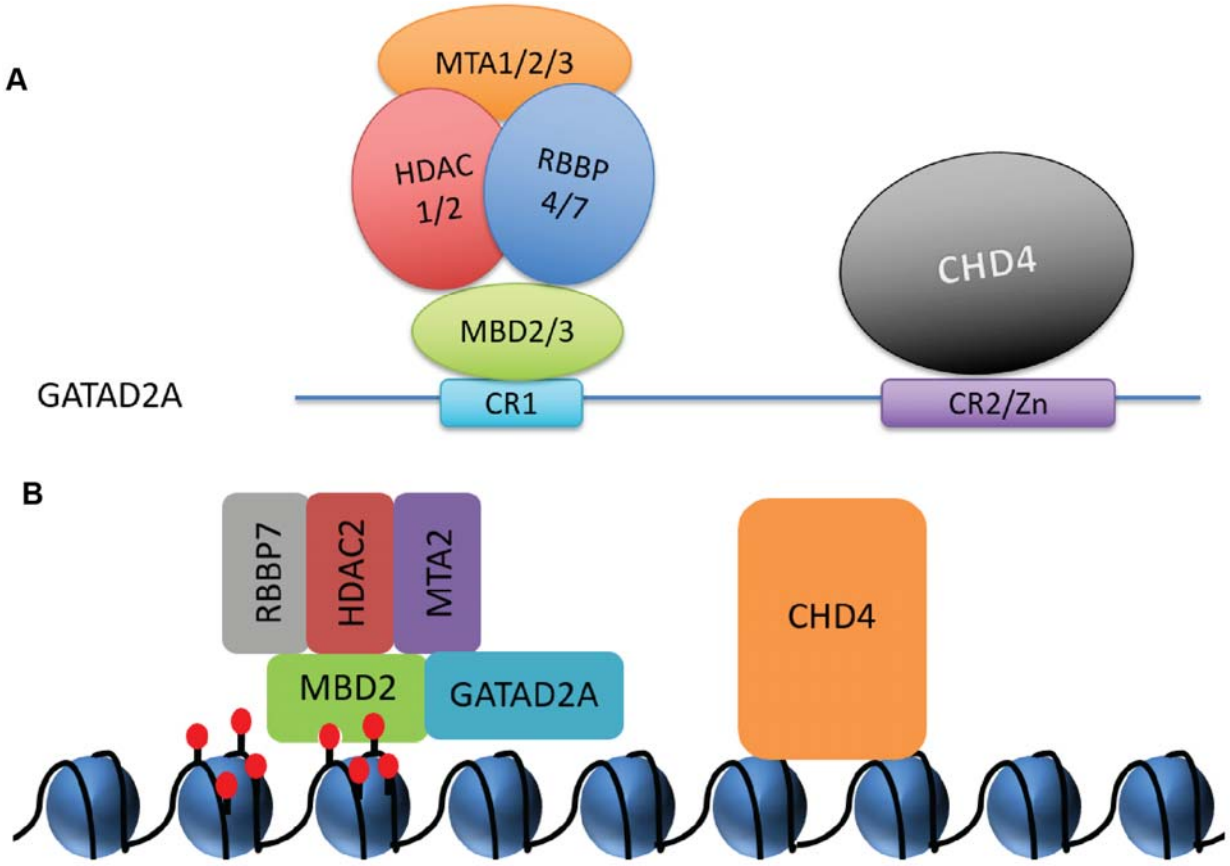

Figure 4.8 Models of CHD4 Recruitment to NuRD
A) GATAD2A bridges the histone deacetylase and nucleosome remodeling activities of NuRD associating with the HDCC through the CR1 interaction with MBD2 and with CHD4 through an interaction with the zinc-finger of CR2.
B) The zinc-finger domain of GATAD2A may contribute to the localization of CHD4 at methylated CpG-islands through the CR1 domain interaction with MBD2.

Functional Characterization of MTA2 Zinc Finger Domain

Zinc-fingers are classically considered nucleic acid binding domains [7–11], however, many have also been found to be protein-protein interaction domains [12–14], including histone modification readers [15,16]. MTA2 is a component of the histone deacetylase core complex (HDCC) which we have shown interacts with the NuRD complex through contact with the MBD2 intrinsically disordered region (IDR). Residues of the MBD2 IDR critical to the HDCC complex interaction were identified, however, the identity of HDCC domains involved in the interface remain elusive [17]. The zinc-finger domain (ZnF) of MTA2 has not been functionally characterized, therefore we sought to examine its possible function as a domain for interaction with the MBD2 IDR, histone tails, and DNA. Here, we present findings that suggest the MTA2 ZnF in isolation does not associate with the MBD2 IDR, unmodified or modified histone tail peptides, and may have non-specific DNA binding activity.

Materials and Methods

Cloning, Expression, and Purification of the MTA2 Zinc-Finger Domain

The MTA2 zinc-finger (178-255 aa) was cloned into the bacterial expression vector pGEX4T-1 (GE Healthcare Biosciences), which was engineered for the expression of recombinant proteins with an N-terminal tandem fusion tag of glutathione S-transferase (GST). The resulting GST-MTA2 fusion was overexpressed in the Rosetta II (DE3) (Invitrogen) *E. coli* strain and grown in Luria Bertani (LB) medium containing 50 µg/ml of ampicillin to an optical density of 0.5 -0.6 at 600 nm, and then induced with 0.2 mM isopropyl-β-D-thiogalactopyranooside for 2 h at 37 °C. Cells were harvested by centrifugation and lysed by sonication in 30 ml of the B-PER reagent (Thermo Scientific)

supplemented with EDTA-free protease inhibitor tablets (Thermo Scientific). The fusion protein was isolated from the soluble fraction by GST affinity chromatography (GSTrap FF, GE Healthcare) and eluted with reduced glutathione buffer (25 mM glutathione, 50 mM Tris pH 8.8, 200 mM NaCl).

Fluorescent Polarization DNA Binding Studies

5'-end 6-FAM labeled DNA oligos were ordered from Integrated DNA Technologies, and annealed and HPLC purified. Oligos and protein were each buffer exchanged into sample buffer [10mM HEPES (pH 7.5), 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT]. The protein and DNA concentrations were measured by UV absorbance and diluted to 5 uM and 70 nM, respectively, in assay buffer [10mM HEPES (pH 7.5), 50 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 0.005% TWEEN 20]. In a black, flat bottom, half volume 96 well plate (COSTAR), protein was serially diluted 1:2 across each row, except for the protein-free control wells. DNA was added to a final concentration of 10 nM in every well, and samples in each row were mixed with a multi-channel pipette. The protein-DNA complexes were incubated for 10 minutes at room temperature and then crosslinked with glutaraldehyde of varying concentrations and incubation times. Plates were read on a CLARIOstar plate reader (BMG Labtech), by focus adjusting on a protein-free control well and measuring fluorescent anisotropy at 495 nm.

NMR Binding Experiments

GST-MTA2 fusion protein (GST-MTA2_ZnF) was overexpressed in the Rosetta II (DE3) (Invitrogen) *E. coli* strain and grown in minimal (M9) media supplemented with ^{15}N ammonium chloride and containing 50 µg/ml of ampicillin. Cells were grown to an

optical density of 0.5 -0.6 at 600 nm and then induced with 0.2 mM isopropyl-β-D-thiogalactopyranooside for 4 h at 37 °C. Cells were harvested by centrifugation and lysed by sonication in 30 ml of the B-PER reagent (Thermo Scientific) supplemented with EDTA-free protease inhibitor tablets (Thermo Scientific). The fusion protein was isolated from the soluble fraction by GST affinity chromatography (GSTrap FF, GE Healthcare) and eluted with reduced glutathione buffer (25 mM glutathione, 50 mM Tris pH 8.8, 200 mM NaCl). The GST-tag was removed by thrombin cleavage overnight at 4 °C. Thrombin was removed by gravity affinity chromatography with a benzamidine sepharose resin (Thermo Scientific) and MTA2 ZnF protein was isolated via gel filtration chromatography with a Superdex 75 10/300 GL column (GE Healthcare). Fractions containing MTA2 ZnF protein were further purified via ion exchange chromatography over a Source 15S 4.6/100 column (GE Healthcare), then combined and buffer exchanged into NMR buffer [10 mM sodium phosphate pH 6.5, 10% D_2O]. Final protein concentration was determined by UV absorption at 280 nm.

Seventeen base pair oligonucleotides (forward: GAGGCGCTCGGCGGCAG; reverse: CTGCCGCCGAGCGCCTC) were purchased (IDT Technologies), annealed, and purified by ion exchange chromatography over a Source 15Q 4.6/100 column (GE Healthcare). Fractions containing double-stranded DNA were then combined and buffer exchanged into NMR buffer. Final DNA concentration was determined by UV absorption at 260 nm. One, two and three molar equivalents of DNA were titrated into a 200 μM ^{15}N labeled MTA2 ZnF protein sample in NMR buffer.

^{15}N labeled MBD2 intrinsically disordered region (IDR) protein was expressed and purified as previously described [1]>. One or two molar equivalents of unlabeled

MTA2 ZnF protein in NMR Buffer was added to the ^{15}N labeled MBD2 IDR sample. HSQC spectra were collected at 25°C with a Bruker Avance III NMR spectrometer equipped with a cryogenic probe and operated at a ^{1}H frequency of 700 MHz.

Results

We purified a GST-tagged fusion of MTA2 zinc-finger (GST-MTA2_ZnF) (Figure 4.1) that our collaborators in the Strahl lab screened for histone binding activity with an EPITITAN™ modified histone tail peptide array (EpiCypher), which yielded zero hits. We also ran a preliminary fluorescence polarization DNA binding assay, in the presence of the protein-DNA cross-linker glutaraldehyde, and did not detect a non-specific interaction with DNA (Figure 4.2). To test for interaction with the intrinsically disordered region (IDR) of MBD2 we ran HSQC titration experiments with ^{15}N labeled MBD2 IDR and unlabeled MTA2 ZnF protein. The MBD2 IDR alone has a distinct pattern of chemical shifts concentrated between 7.5 and 8.75 ppm in the 1H dimension, seen in the left panel of Figure 4.3. The chemical shifts did not change in the presence of increasing concentrations of MTA2 ZnF protein, demonstrated by the overlying blue and green spectra in the right panel of Figure 4.3. An additional HSQC experiment was conducted with ^{15}N labeled MTA2 ZnF protein in the presence of increasing concentration of double stranded DNA. Data could not be collected on the samples, however, as the addition of one, two or three molar equivalent of DNA immediately formed a white precipitate and signal was no longer obtainable from the NMR sample over the noise. Increasing the salt concentration did not prevent precipitate formation (data not shown).

Discussion and Future Directions

The absence of hits from the modified histone tail peptide array and the lack of observable chemical shift perturbations in the MBD2 IDR spectrum upon exposure to the MTA2 ZnF protein leads us to conclude that they do not likely interact within the NuRD complex. It is possible that larger interaction interfaces are required to observe the MTA2 ZnF contribution to the interaction of the histone deacetylase core complex (HDCC) with MBD2 or histones, and future experiments could be designed to test this. Co-expression and purification of HDAC, RBBP4/8, and MTA constructs with and without the zinc-finger domain would allow for further examination of the role of the MTA2 ZnF in complex formation by testing the interaction with MBD2 full-length protein and IDR domain. Such interactions could be examined with analytical size exclusion chromatography, light scattering, in vitro pull-downs or HSQC titration experiments. Taking measurements of binding affinity with histone proteins or assembled nucleosomes may also inform if the MTA2 ZnF measurably adds to the association of the HDCC with chromatin.

Alternatively, the precipitation of MTA2 ZnF protein in the presence of equal and excess concentrations of double-stranded DNA might indicate non-specific DNA binding and this possibility should also be further explored. Sequence specificity could be examined with SELEX experiments or DNA footprinting assays focused on sequences near loci thought to be regulated by the NuRD complex. Enriched sequences should then be confirmed and characterized by biochemical or biophysical techniques, such as electromobility shift assay (EMSA), fluorescence polarization, isothermal calorimetry or

HSQC titration experiments, towards ultimately solving a solution NMR structure of a relevant complex.

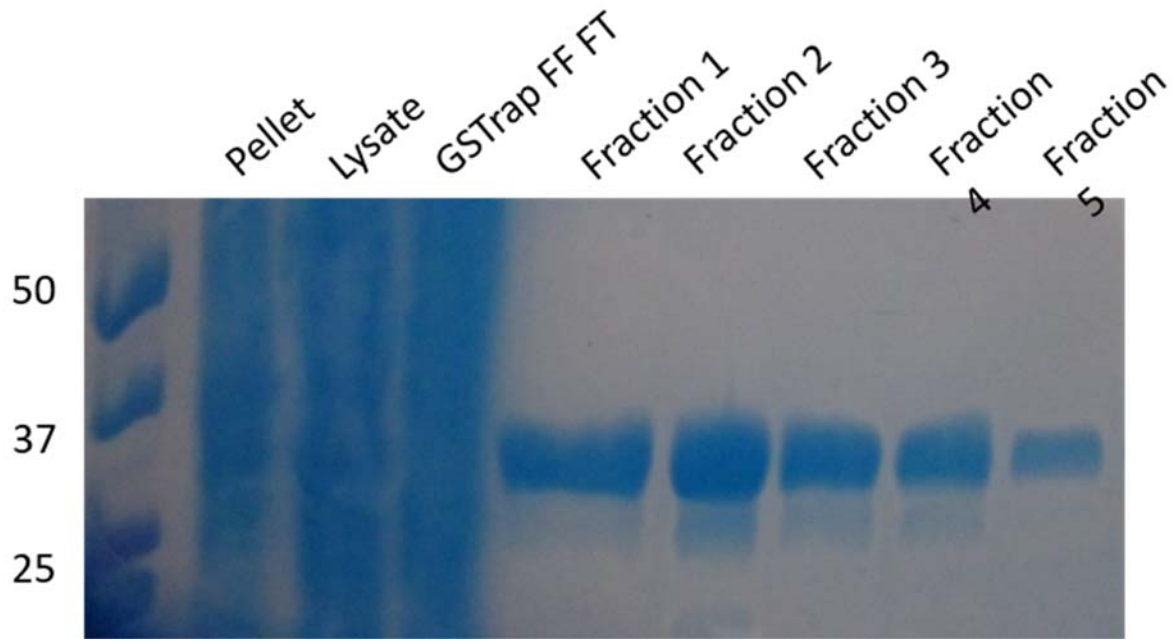

Figure 4.9 GST-fusion MTA2 zinc-finger purification
Coomassie-stained SDS-PAGE gel of the insoluble (pellet) and soluble (lysate) fractions of bacterial whole cell lysates following expression of the GST-tagged MTA2 zinc-finger (GST-MTA2_ZnF) domain. Subsequent lanes include the flow-through (FT) and elution fractions from a GSTrap FF column extracting the GST-MTA2 fusion protein (expected molecular weight 35 kDa).

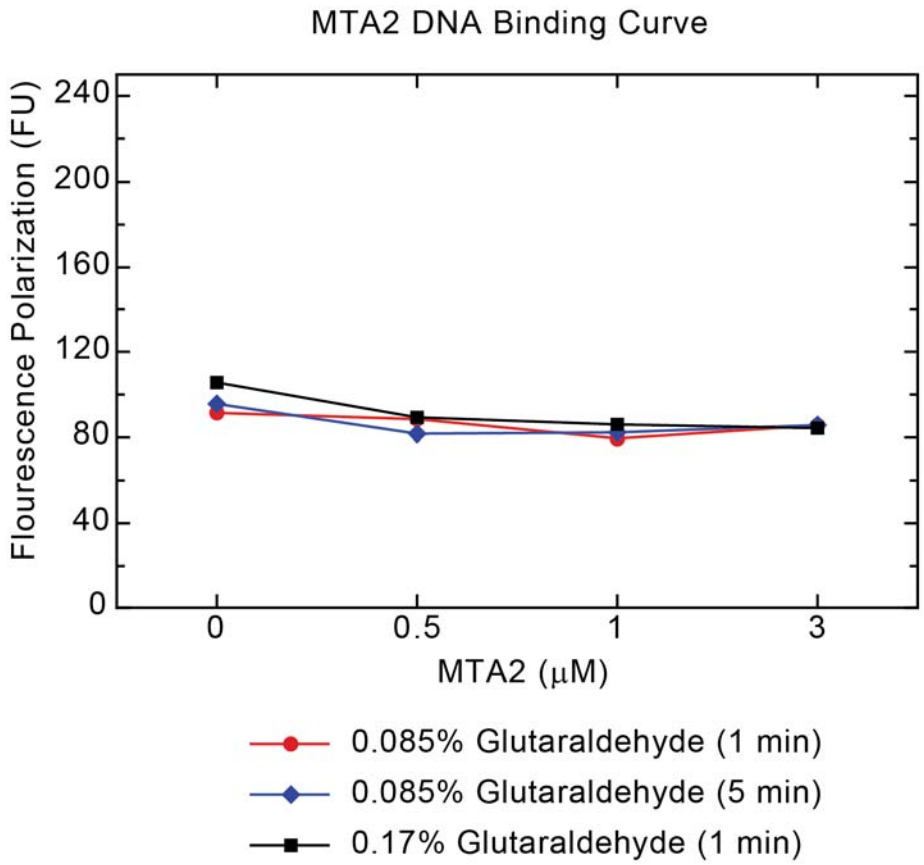

Figure 4.10 MTA2 zinc-finger did not exhibit DNA binding activity by fluorescence polarization
Plotted polarized fluorescence units (FU) detected with an increasing micromolar (μM) concentration of GST-MTA2_ZnF protein at varying concentrations of glutaraldehyde crosslinking reagent.

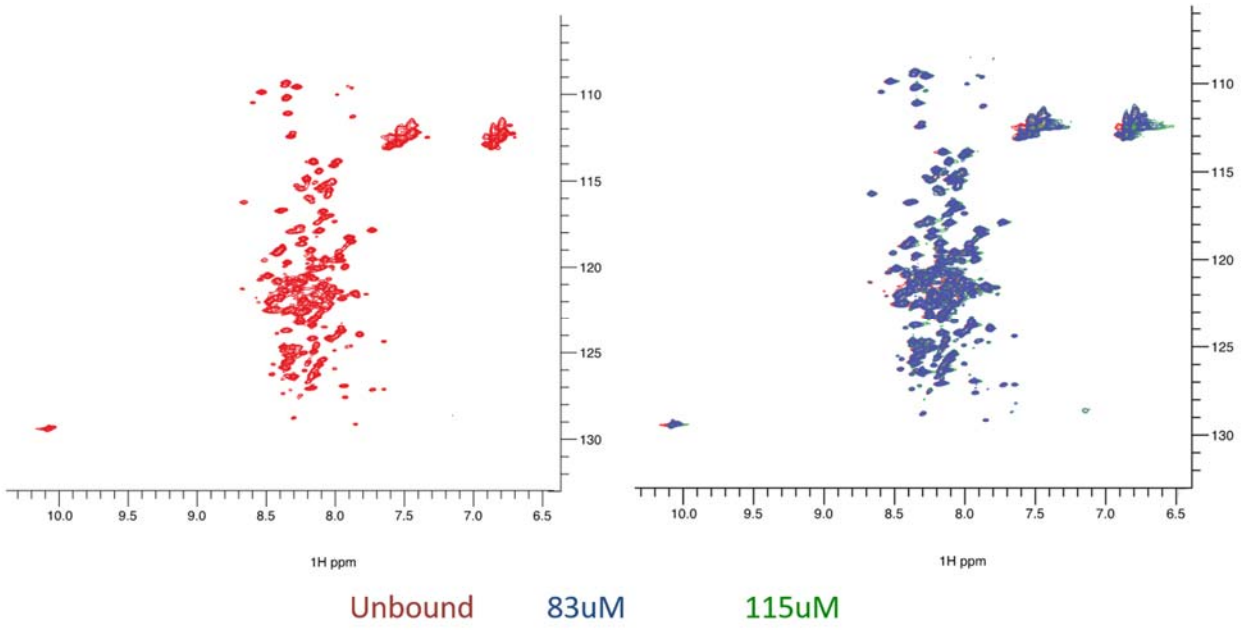

Figure 4.11 MTA2 zinc-finger did not exhibit binding to the MBD2 IDR via HSQC
HSQC spectra of ^{15}N labeled MBD2 intrinsically disordered region (IRD) protein free in solution (red). The left panel displays overlaid spectra of the IDR incubated with one (blue) or two (green) molar equivalents of unlabeled MTA2 ZnF protein.

REFERENCES

1. Gnanapragasam, M. N. *et al.* p66α–MBD2 coiledcoil interaction and recruitment of Mi2 are critical for globin gene silencing by the MBD2–NuRD complex. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7487–7492 (2011).

2. Ramirez, J., Dege, C., Kutateladze, T. G. & Hagman, J. MBD2 and Multiple Domains of CHD4 Are Required for Transcriptional Repression by Mi-2/NuRD Complexes. *Mol. Cell. Biol.* **32,** 5078–5088 (2012).

3. Stothard, P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28,** 1102–1104 (2000).

4. Brackertz, M., Gong, Z., Leers, J. & Renkawitz, R. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res.* **34,** 397–406 (2006).

5. Kloet, S. L. *et al.* Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.* **282,** 1774–1785 (2015).

6. Sun, S., Yang, X., Wang, Y. & Shen, X. In Vivo Analysis of Protein–Protein Interactions with Bioluminescence Resonance Energy Transfer (BRET): Progress and Prospects. *Int. J. Mol. Sci.* **17,** 1704 (2016).

7. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* **356,** 408–414 (1992).

8. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA Recognition by Cys2Hi2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 183–212 (2000).

9. Omichinski, J. G. *et al.* NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261,** 438–46 (1993).

10. Lu, D., Searles, M. A. & Klug, A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426,** 96–100 (2003).

11. Lee, B. M. *et al.* Induced fit and "lock and key" recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J. Mol. Biol.* **357,** 275–91 (2006).

12. Fox, A. H. *et al.* Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. *EMBO J.* **18,** 2812–22 (1999).

13. Westman, B. J., Perdomo, J., Matthews, J. M., Crossley, M. & Mackay, J. P. Structural studies on a protein-binding zinc-finger domain of Eos reveal both similarities and differences to classical zinc fingers. *Biochemistry* **43,** 13318–27 (2004).

14. Simpson, R. J. Y. *et al.* Classic zinc finger from friend of GATA mediates an interaction with the coiled-coil of transforming acidic coiled-coil 3. *J. Biol. Chem.* **279,** 39789–39797 (2004).

15. Liu, Y. *et al.* Family-wide Characterization of Histone Binding Abilities of Human CW Domain-containing Proteins. *J. Biol. Chem.* **291,** 9000–9013 (2016).

16. He, F. *et al.* Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* **18,** 1127–39 (2010).

17. Desai, M. A. *et al.* An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex. *Nucleic Acids Res.* **43,** 3100–3113 (2015).