Biochemical analysis of MeCP2

Robert John Klose

Thesis presented for the degree of Doctor of Philosophy
Wellcome Trust Centre for Cell Biology
University of Edinburgh
2005

Acknowledgments

First, and foremost, I would like to thank Adrian for taking me on as a PhD. student and for excellent scientific training under his mentoring. I would like thank all members of the Bird lab for guidance over the course of my PhD., but especially Skiry, with whom I have enjoyed discussing and debating science of all descriptions. In addition I would like to extend a special thanks to Irina Stancheva and Lars Schmiedeberg for their involvement in defining MeCP2 binding specificity, to Donald MacLeod who read the first draft of my thesis, and to my parents who have been supportive in my academic endeavours. Finally I would like to thank Emma for the opportunity to share our PhD. experiences together and for making life in Scotland an enjoyable experience.

The work in this thesis was supported by a Wellcome Trust Prize Studentship and an Overseas Research Scheme award from British Universities. I am indebted to these organizations for financially supporting my PhD. studies.

Table of Contents

TABLE OF CONTENTS **1**
FIGURE LIST **5**
ABSTRACT **7**
ABBREVIATIONS **9**

1. CHAPTER ONE- INTRODUCTION **12**

1.1 DNA METHYLATION 12
1.1.1 Methyl-cytosine, the fifth base in eukaryotic DNA *12*
1.1.2 DNA methylation in prokaryotes *12*
1.2 DNA METHYLATION IN EUKARYOTES 13
1.2.1 Cytosine methylation profiles in fungi *13*
1.2.2 Cytosine methylation in plants *14*
1.2.3 Cytosine methylation in lower eukaryotes *15*
1.2.4 Genomic methylation profiles in mammals *16*
1.2.5 Mammalian DNA methyltransferases *17*
1.2.6 DNA methylation in development, imprinting, and the X-chromosome *23*
1.2.7 DNA methylation and silencing of parasitic DNA elements *27*
1.2.8 Targeting DNA methylation by small RNA's in mammals? *28*
1.2.9 DNA methylation and cancer *30*
1.3 CHROMATIN REMODELING AND HISTONE MODIFICATION IN GENE EXPRESSION 31
1.3.1 Chromatin structure, histones, and histone modification nomenclature *31*
1.3.2 Active histone modification marks *32*
1.3.3 Repressive histone modification marks *35*
1.3.4 Does a histone code exist? *37*
1.3.5 Interplay between histone modification and DNA methylation *37*
1.3.6 ATP dependent chromatin remodeling factors and DNA methylation *39*
1.4 INTERPRETATION OF DNA METHYLATION 40
1.4.1 DNA methylation and silencing *40*
1.4.2 Identification of activities that bind methyl-CpG *40*
1.4.3 Isolation of a family of Methyl-CpG Binding proteins *41*
1.4.4 Structure of the MBD *41*
1.4.5 The ancestral MBD protein *43*
1.4.6 MeCP2 ..... *43*
1.4.7 MBD1 ..... *43*
1.4.8 MBD2 (and MBD3) *45*
1.4.9 MBD4 ..... *47*
1.4.10 Kaiso *48*
1.5 MECP2 49
1.5.1 Recognition of Methyl-CpG *49*
1.5.2 Transcriptional repression and protein partners *49*
1.5.3 Expression, evolution, and conservation *50*
1.5.4 MeCP2 and Rett Syndrome *51*
1.5.5 MeCP2 null mouse *53*
1.5.6 MeCP2 depleted Xenopus embryo *53*
1.5.7 MeCP2 target genes *54*
1.6 AIMS OF THIS THESIS 55

2. CHAPTER TWO- MATERIALS AND METHODS **56**

2.1 COMMON SOLUTIONS 56
2.2 BACTERIAL METHODS 56
2.2.1 Media *56*
2.2.2 Antibiotics *56*
2.2.3 Culture on plates *57*

2.2.4 *Liquid culture* 57
2.2.5 *Preparing chemically competent strains* 57
2.2.6 *Transformation* 58
2.3 ISOLATION OF EXOGENOUS DNA FROM BACTERIA 58
2.3.1 *Small scale plasmid DNA preparation* 58
2.3.2 *Sequencing grade plasmid DNA preparation* 59
2.3.3 *Large scale plasmid DNA preparation* 59
2.3.4 *Large scale preparation of Bacmid DNA* 59
2.4 MANIPULATION OF DNA AND CLONING 59
2.4.1 *Restriction endonuclease digestion of DNA* 59
2.4.2 *Dephosphorylation of DNA fragments* 60
2.4.3 *Methylation of DNA fragments* 60
2.4.4 *Oligonucleotides* 60
2.4.5 *Polymerase chain reaction* 60
2.4.6 *Purification of PCR fragments* 60
2.4.7 *Sequencing* 61
2.4.8 *Agarose gel extraction of DNA fragments* 61
2.4.9 *Ligation* 61
2.4.10 *Site directed mutagenesis* 61
2.4.11 *Radioactively labeling DNA fragments* 61
2.5 MAMMALIAN CELL CULTURE METHODS 61
2.5.1 *Mammalian cell culture* 61
2.5.2 *Cryogenic storage of mammalian cells* 62
2.5.3 *Transfection* 62
2.5.4 *Isolation of nuclei from tissue culture cells* 62
2.5.5 *Isolation of nuclei from brain tissue* 62
2.5.6 *Nuclear protein extracts from mammalian cells* 63
2.6 INSECT CELL AND BACULOVIRUS CULTURE METHODS 63
2.6.1 *Insect cell culture* 63
2.6.2 *Cryogenic storage of insect cells (SF9)* 63
2.6.3 *Generation of recombinant baculovirus genome* 64
2.6.4 *Transfection of SF9 cells and isolation of p1 baculovirus* 64
2.6.5 *Amplification of viral stock* 65
2.6.6 *Infection of SF9 cells for protein production* 65
2.7 XENOPUS LAEVIS METHODS 65
2.7.1 *Isolating Xenopus l. oocytes* 65
2.7.2 *Making Xenopus l. oocyte extract* 66
2.7.3 *Xenopus l. MeCP2 and Sin3a antibodies* 66
2.8 PROTEIN METHODS 66
2.8.1 *Measuring protein concentration* 66
2.8.2 *Dialysis* 66
2.8.3 *SDS-PAGE* 67
2.8.4 *Coomassie blue stain* 67
2.8.5 *Semi-Dry transfer to nitrocellulose membrane* 67
2.8.6 *Western blot* 68
2.8.7 *Immunoprecipitation* 68
2.8.8 *Sucrose gradient sedimentation* 68
2.8.9 *Kinase Assay* 69
2.8.10 *Mass Spectrometry* 69
2.9 EXPRESSION OF PROTEINS IN BACTERIA 69
2.9.1 *Cell preparation and induction of expression* 69
2.9.2 *Making non-denaturing extracts* 70
2.10 EXPRESSION OF PROTEINS IN INSECT CELLS 70
2.10.1 *Insect whole cell extracts* 70
2.11 PROTEIN PURIFICATION 70
2.11.1 *Chromatography media* 70
2.11.2 *Chromatography solutions* 71

2.11.3 Preparing disposable columns ... 72
2.11.4 Preparing FPLC based HR columns ... 72
2.11.5 Ion exchange chromatography ... 72
2.11.6 Affinity chromatography ... 72
2.11.7 Size exclusion chromatography ... 72
2.11.8 Purification of his tagged bacterial proteins ... 73
2.11.9 Purification of his tagged baculovirus expressed proteins ... 73
2.12 MANIPULATION OF PROTEIN ... 74
2.12.1 Dephosphorylation of baculovirus expressed MeCP2 ... 74
2.12.2 Nuclease treatment of purified rat MeCP2 ... 74
2.12.3 Denaturation of rat brain nuclear extract ... 74
2.13 DNA BINDING EXPERIMENTS ... 74
2.13.1 Generation of probes using single stranded oligonucleotides ... 74
2.13.2 Electrophoretic Mobility Shift Assay (EMSA) ... 75
2.13.3 Methyl-SELEX: Systematic Evolution of Ligand by Exponential Enrichment (SELEX) ... 75
2.13.4 End labeling DNA for foot-printing ... 77
2.13.5 Foot-printing ladder: Maxam-Gilbert G+A sequencing reaction ... 77
2.13.6 DNase1 foot-printing ... 78

3. CHAPTER THREE- PURIFICATION OF NATIVE MECP2 ... 79

3.1 INITIAL IDENTIFICATION OF MECP2 AND BIOCHEMICAL PROPERTIES ... 79
3.2 BIOCHEMICAL EVALUATION THE MECP2/SIN3A INTERACTION ... 80
3.2.1 MeCP2 co-immunoprecipitates with Sin3a ... 80
3.2.2 Absence of a stable MeCP2/Sin3a association in rat brain and cell line nuclear extracts ... 82
3.2.3 Xenopus laevis MeCP2 does not stably associate with Xenopus laevis Sin3a ... 82
3.3 PURIFICATION OF NATIVE MECP2 ... 85
3.3.1 MeCP2 elutes from size exclusion chromatography columns with a molecular weight of 400-500 kDa ... 85
3.3.2 The large molecular weight MeCP2 is stable over multiple biochemical purification steps ... 85
3.3.3 Large scale purification of native MeCP2 from rat brain nuclear extract ... 87
3.4 DISCUSSION ... 97
3.4.1 Mechanisms of MeCP2 mediated repression ... 97
3.4.2 MeCP2 does not stably associate with the Sin3a chromatin remodeling complex in nuclear extracts ... 97
3.4.3 The contribution of other MeCP2 co-factors in transcriptional repression ... 98
3.4.4 Distinct mechanisms of methylation mediated repression? ... 99

4. CHAPTER FOUR- BIOPHYSICAL ANALYSIS OF MECP2 ... 103

4.1 MECP2 HAS A LARGE ELUTION PROFILE BY SIZE EXCLUSION CHROMATOGRAPHY ... 103
4.2 BIOPHYSICAL ANALYSIS OF MECP2 ... 103
4.2.1 Elution profile of native MeCP2 is unaffected by nuclease treatment. ... 103
4.2.2 Producing preparative amounts of native MeCP2 for biophysical analysis ... 104
4.2.3 Native MeCP2 does not self associate ... 112
4.2.4 Biophysical analysis of MeCP2 using the Seigel and Monty equation ... 114
4.2.5 Mammalian and Xenopus MeCP2 are both monomeric elongated molecules ... 117
4.2.6 No individual domain of MeCP2 accounts for its elongated structure ... 119
4.3 DISCUSSION ... 121
4.3.1 Native MeCP2 is an elongated monomer ... 121
4.3.2 Is there a structural reason that MeCP2 is an elongated molecule? ... 121
4.3.3 Why might an elongated MeCP2 molecule be advantageous? ... 122

5. CHAPTER FIVE- MECP2 DNA BINDING SPECIFICITY ... 125

5.1 HOW DOES MECP2 RECOGNIZE METHYLATED DNA? ... 125
5.2 DNA BINDING PROPERTIES OF MECP2 ... 126

5.2.1 MeCP2 binds unique loci in vivo ... *126*
5.2.2 Enrichment of MeCP2 binding sites by Methyl-SELEX ... *127*
5.2.3 DNA fragments enriched by MeCP2 using Methyl-SELEX have an A/T rich sequence adjacent to the methyl-CpG ... *129*
5.2.4 MeCP2 requires an A/T rich sequence adjacent to the Methyl-CpG for efficient DNA binding ... *131*
5.2.5 The MeCP2 AT-Hook domain does not contribute to binding of an A/T run containing site ... *131*
5.2.6 MeCP2 methyl-CpG A/T run dependent binding resides in a MeCP2 fragment containing the MBD and flanking sequences. ... *135*
5.2.7 Binding to the promoter sequence of the MeCP2 target gene Bdnf is mediated by an methyl-CpG A/T run containing sequence ... *137*
5.3 DISCUSSION ... 137
5.3.1 MeCP2 binds unique loci which are not redundantly occupied by other MBP's ... *137*
5.3.2 Methyl-CpG with $[A/T]_{\geq 4}$ confers a high affinity directional binding site for MeCP2 *139*
5.3.3 Does MeCP2 bind to non-methylated DNA sequences? ... *140*
5.3.4 Do other MBP's have sequence specific DNA binding determinants? ... *141*
5.3.5 Mouse genetics suggest MBP's are not redundant ... *142*
5.3.6 Two modes of methyl-CpG recognition? ... *143*

6. CHAPTER SIX- PHOSPHORYLATION OF MECP2 ... **145**

6.1 MECP2 IS A PHOSPHOPROTEIN ... 145
6.2 PHOSPHORYLATION AND MECP2 ... 146
6.2.1 Phosphorylation of baculovirus produced MeCP2 does not affect DNA binding ... *146*
6.2.2 An activity that can phosphorylate MeCP2 is present in HeLa nuclear extracts ... *148*
6.2.3 There are two major activities in HeLa nuclei that phosphorylate MeCP2 ... *150*
6.2.4 Partial purification of MeCP2 kinase activity 1 ... *154*
6.2.5 CDK1 corresponds to the peak of MeCP2 kinase activity ... *156*
6.3 DISCUSSION ... 158
6.3.1 Phosphorylation of MeCP2 in baculovirus does not inhibit DNA binding. ... *158*
6.3.2 CDK1 phosphorylates Mecp2 ... *159*
6.3.3 The MeCP2 kinase complex, a work in progress... ... *159*
6.3.4 Are there other important MeCP2 kinases? ... *160*

7. CHAPTER SEVEN- CONCLUSIONS / FUTURE DIRECTIONS ... **162**

7.1 LEARNING MORE ABOUT MECP2 USING BIOCHEMISTRY ... 162
7.2 DETERMINING CO-FACTORS REQUIRED FOR MECP2 FUNCTION ... 163
7.3 THE EPIGENOME AND MECP2 TARGETING ... 164
7.4 MODULATING EPIGENETIC READOUT BY POST TRANSLATIONAL MODIFICATION ... 164
7.5 A SEARCH FOR THE CAUSE OF RETT SYNDROME ... 165

8. APPENDIX 1 ... **166**

9. REFRENCES ... **170**

10. PUBLICATIONS ... **195**

Figure List

FIGURE 1.1 - METHYLATED DNA BASES 13
FIGURE 1.2- CATALYTIC MECHANISMS OF 5-METHYL-CYTOSINE DNA METHYLTRANSFERASES (ADAPTED FROM (BESTOR, 2000) 18
FIGURE 1.3-DOMAIN ORGANIZATION OF THE MAMMALIAN DNA METHYLTRANSFERASE ENZYMES (ADAPTED FROM (BESTOR, 2000)) 20
FIGURE 1.4- IMPRINTING AT THE *H19/IGF2* LOCUS 25
FIGURE 1.5- THE METHYL-CPG BINDING PROTEIN FAMILY 42
FIGURE 1.6- STRUCTURE OF THE MBD DOMAIN OF MBD1 BOUND TO METHYLATED DNA (ADAPTED FROM (OHKI ET AL., 2001)) 42
FIGURE 1.7- ALIGNMENT OF MECP2 PROTEINS FROM VARIOUS ORGANISMS (THIS ALIGNMENT WAS A GIFT FROM SKIRMANTAS KRIAUCIONIS) 52
FIGURE 3.1- MAMMALIAN MECP2 DOES NOT STABLY ASSOCIATE WITH THE SIN3A CHROMATIN REMODELLING COMPLEX. 81
FIGURE 3.2- *XENOPUS LAEVIS* MECP2 DOES NOT STABLY ASSOCIATE WITH THE SIN3A CHROMATIN REMODELLING COMPLEX. 84
FIGURE 3.3- THE 400-500 KDA MECP2 IS STABLE OVER MULTIPLE CHROMATOGRAPHIC STEPS. 86
FIGURE 3.4- PREPARATIVE MECP2 PURIFICATION SCHEME. 88
FIGURE 3.5- SP-SEPHAROSE CHROMATOGRAPHY OF MECP2. 89
FIGURE 3.6- MONO-Q CHROMATOGRAPHY OF MECP2. 91
FIGURE 3.7- HEPARIN-SEPHAROSE CHROMATOGRAPHY OF MECP2 92
FIGURE 3.8- NI-NTA CHROMATOGRAPHY OF MECP2 94
FIGURE 3.9- SUPEROSE 12 SIZE EXCLUSION CHROMATOGRAPHY AND IDENTIFICATION OF MECP2 96
FIGURE 3.10- FLEXIBILITY AND DIVERSITY OF MECP2 FUNCTION *IN VIVO*? 100
FIGURE 4.1- NUCLEIC ACID DOES NOT CONTRIBUTE TO THE LARGE MOLECULAR WEIGHT OF NATIVE MECP2 104
FIGURE 4.2- GENERATION OF UNTAGGED MECP2 EXPRESSING BACULOVIRUS. 106
FIGURE 4.3- PRODUCTION OF BACULOVIRUS MECP2 107
FIGURE 4.4- BACULOVIRUS MECP2 IS A PHOSPHOPROTEIN 109
FIGURE 4.5- PRODUCTION OF UNTAGGED MECP2 IN BACTERIA. 111
FIGURE 4.6- NATIVE MECP2 DOES NOT SELF ASSOCIATE 113
FIGURE 4.7- MECP2 IS AN ELONGATED MONOMER. 115
FIGURE 4.8- BOTH HUMAN AND *XENOPUS* MECP2 ARE ELONGATED MOLECULES. 118
FIGURE 4.9- NO INDIVIDUAL DOMAIN OF MECP2 CONTRIBUTES TO ITS LARGE RADIUS. 120
FIGURE 4.10- IS MECP2 ELONGATED TO ACT IN A MODULAR FASHION? 124
FIGURE 5.1-METHYL-SELEX EFFICIENTLY SELECTS FOR HIGH AFFINITY MECP2 BINDING SEQUENCES. 128
FIGURE 5.2-HIGH AFFINITY MECP2 BINDING SITES ISOLATED BY METHYL-SELEX CONTAIN METHYL-CPG WITH AN ADJACENT RUN OF $[A/T]_{\geq 4}$ DNA. 130
FIGURE 5.3-MUTATION OF THE $[A/T]_{\geq 4}$ IN HIGH AFFINITY MECP2 BINDING SITE REDUCES AFFINITY.. 132
FIGURE 5.4- THE CONSERVED MECP2 AT-HOOK DOMAIN IS NOT REQUIRED FOR A/T RUN-DEPENDENT BINDING OF MECP2 *IN VITRO* OR *IN VIVO*. (EXPERIMENTS IN (C) AND (D) WERE CONTRIBUTED BY DR. LARS SCHMIEDEBERG) 134
FIGURE 5.5-A FRAGMENT OF MECP2 CONTAINING AMINO ACIDS 78-161 CONFERS METHYL-CPG $[A/T]_{\geq 4}$ DEPENDENT BINDING. 136
FIGURE 5.6 A NATURAL MECP2 TARGET GENE HAS A HIGH AFFINITY MECP2 BINDING SITE. 138
FIGURE 5.7- TWO MODES OF METHYL-CPG RECOGNITION? 144
FIGURE 6.1- PRODUCTION OF BACULOVIRUS MECP2 1-205 147
FIGURE 6.2- PHOSPHORYLATION OF BACULOVIRUS-PRODUCED MECP2 1-205 DOES NOT INHIBIT DNA BINDING 149
FIGURE 6.3- HELA CELLS HAVE MECP2 KINASE ACTIVITY 151
FIGURE 6.4- HELA CELL NUCLEAR EXTRACT CONTAINS TWO MECP2 KINASE ACTIVITIES 152
FIGURE 6.5- THE TWO NUCLEAR HELA MECP2 KINASE ACTIVITIES ARE BIOCHEMICALLY DISTINCT.. 153

FIGURE 6.6- PARTIAL PURIFICATION OF MECP2 KINASE ACTIVITY 1. 155
FIGURE 6.7- MECP2 KINASE ACTIVITY 1 CONTAINS CDK1/CYCLINB1 157

Abstract

DNA methylation is an epigenetic mark that has profound affects on the genome of higher eukaryotes. The majority of DNA methylation in mammalian genomes is found on cytosine bases in the context of symmetrical CpG dinucleotides. The main biological affect of DNA methylation is regulatory control of surrounding gene expression and chromatin structure. DNA methylation is unequivocally linked to repressed genes and regions of the genome that are permanently silenced and referred to as heterochromatin. The affect of DNA methylation is to a great extent realized through interpretation of the methyl-CpG signal by a family of methyl-binding proteins (MBP's). MBP's include: MeCP2, MBD1, MBD2, MBD3, MBD4, and Kaiso. With the exception of MBD4, which is involved in the repair of methyl-CpG deamination events, MBP's are all transcriptional repressors. The capacity of MBP molecules to modify chromatin structure and silence genes is in part dependent on associated co-factors that possess enzymatic activities directed towards chromatin.

This thesis deals predominantly with the MBP protein MeCP2. MeCP2 is a transcriptional repressor that recruits the Sin3a chromatin remodelling complex to methylated loci. Sin3a-associated histone deacetylases contribute to the ability of MeCP2 to repress transcription and modulate chromatin structure. The biomedical importance of normal MeCP2 function is highlighted by the discovery that inactivating mutations in MeCP2 cause the severe neurological disease Rett syndrome. By deleting the *Mecp2* gene, a mouse model of Rett syndrome has been generated and used to study the molecular and physiological outcome of MeCP2 deficiency. Inefficient regulation of neuronal gene expression may have a role in the etiology of Rett syndrome.

By studying the biochemical properties of MeCP2 this thesis addresses in detail three basic questions; (1) what are the native biochemical properties of MeCP2? (2) what specific DNA sequences does MeCP2 bind? and (3) what are the affects of post-translational modification on MeCP2? To address the composition of any

mammalian MeCP2 complexes, native MeCP2 was purified to near homogeneity from rat brain. Native MeCP2 is an elongated monomer that does not stably associate with other cofactors including Sin3a. Analysis of MeCP2 binding sites *in vivo* demonstrates that MeCP2 binds unique loci when compared to other MBP's. Using an unbiased *in vitro* DNA binding site evolution assay, Methyl-SELEX, MeCP2 was shown to require methyl-CpG sequences containing a flanking run of A/T rich DNA for high affinity binding. Finally, biochemical fractionation of nuclear proteins revealed activities that phosphorylate MeCP2, and the potential affects of this modification were explored. These biochemical studies shed new light on how MeCP2 may function to target genes for methylation-dependent repression and recruit co-factors that modulate chromatin structure and gene expression.

Abbreviations

aa	amino acid
ADP	adenosine diphosphate
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
C-	carboxyl
Ci	curie
Da	Dalton
dNTP	deoxyribonucleoside triphosphate
DMD	differentially methylated domain
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribuonuclease
DTT	dithiothreitol
EDTA	diaminoethanetetraacetic acid
EMSA	electrophoretic mobility shift assay
ES	embryonic stem

EST	expressed sequence tag
g	relative centrifugal force
h	hour
Hepes	N-2-hydroxyethylpeperazine-N'-2-ethanesulfonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
l	liter
M	molar
MBD	methyl CpG binding domain
mg	milligram
min	minutes
mol	mole
ng	nanogram
nm	nanometre
NMR	nuclear magnetic resonance
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction

pH	$-\log_{10}[H+]$
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
s	seconds
SAM	S-adenosyl-L-methionine
SELEX	systematic evolution of ligands by exponential enrichment
SDS	sodium dodecyl sulphate
TRD	transcriptional repression domain
Tris	2-amio-2(hydroxymethyl)-1,2,-propanediol
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
ug	microgram
UV	ultar violet
V	volts
°C	degree centigrade

1. Chapter one-
Introduction

1.1 DNA methylation

1.1.1 Methyl-cytosine, the fifth base in eukaryotic DNA

The complementary double stranded DNA helix consists of four basic deoxyribonucleic acid building blocks; deoxyadenosine, deoxyguanosine, deoxycytidine, and thymidine. These nucleotide building blocks are linked by phophodiester bonds to create the linear polymers that make up each individual strand of DNA. Biochemical examination of DNA identified a cytosine base pair variant which has a methyl group in the 5' position of the pyrimidine ring (Figure 1.1)(Hotchkiss, 1948; Johnson and Coghill, 1925). This methyl group is transferred to cytosine after DNA replication by DNA methyltransferase enzymes. Since the initial discoveries of 5-methyl cytosine it has become clear that this modification has profound implications for the biology of simple prokaryotes through to higher multi-cellular organisms.

1.1.2 DNA methylation in prokaryotes

DNA methylation in bacteria can occur on the N6 position of adenine and on the N4 and C5 positions of cytosine (Figure 1.1). DNA methylation in bacteria is part of a host restriction-modification system (Wilson and Murray, 1991), which is used to protect the organism from virus infection. The restriction modification system is composed of an endonuclease that attacks the invading viral genome, while the bacterial genome is protected by blockage of endonuclease sites with methylated DNA bases. Therefore, the bacteria can specifically target destruction of foreign DNA but not harm the coding and replication potential of their own genome. Other functions for DNA methylation in prokaryotes, including DNA repair and regulation of transcription, have also been proposed (Palmer and Marinus, 1994).

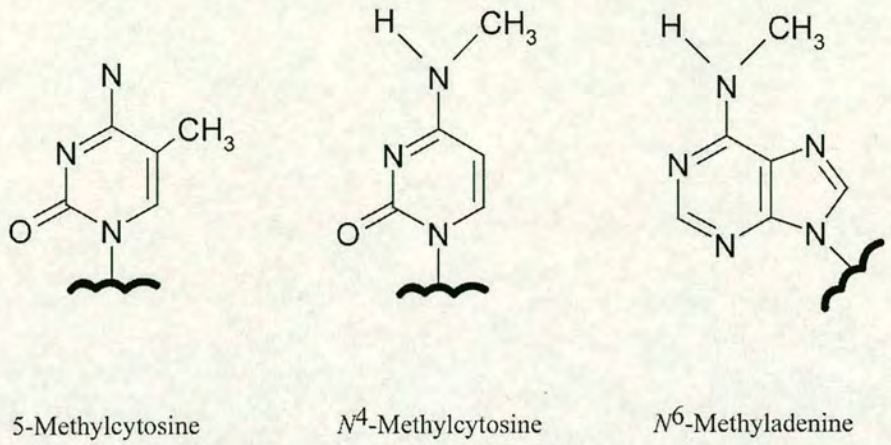


Figure 1.1 - Methylated DNA bases

Methylation of cytosine occurs on the amine group at the 4' position and the carbon at the 5' position in the DNA base. Adenine is methylated the amine group at the 6 position.

1.2 DNA methylation in eukaryotes

1.2.1 Cytosine methylation profiles in fungi

DNA methylation has not been observed in fission or budding yeast (Proffitt et al., 1984) despite exploratory efforts to identify this modification (Wilkinson et al., 1995). A DNA methyltransferase homologue, pmt1+ (Pinarbasi et al., 1996; Wilkinson et al., 1995), has been identified in fission yeast which has all the conserved domains found in functional methyltransferase enzymes, except a point mutation in the catalytic domain renders this enzyme non-functional. In contrast, DNA methylation has been observed in other species of fungi including *Candida albicans* (Russell et al., 1987) and *Neurospora crassa* (Bull and Wootton, 1984).

The DNA methylation system in *Neurospora crassa* has been thoroughly studied, (Selker et al., 2002) and appears to function in concert with a genome defence mechanism called repeat induced point mutation (RIP). RIP is induced by duplicated DNA sequences in the haploid genome of *Neurospora* during the sexual phase of its

life cycle, and results in mutations within the repeated DNA sequence (Selker et al., 2002). RIP functions in *Neurospora* to protect against invasion and exploitation of its genome by parasitic repeat elements (i.e. transposons). Induction of RIP directly causes accumulation of methylation on cytosine bases within the mutated region (Selker et al., 2002). One cytosine methyltransferase, Dim-2, is responsible for all DNA methylation in *Neurospora*, and cytosine methylation is not essential as a *Dim-2* null allele is viable (Kouzminova and Selker, 2001). Cytosine methylation in *Neurospora* affects most (>80%) of the cytosines in RIPed sequences and occurs in a sequence context independent manner. The complete genome sequence of *Neurospora crassa* allowed a genome wide survey of the DNA methylation profiles and most of the methylated sequences were found on relics of transposons that had been inactivated by RIP, and covered 2-3% of cytosine bases (Selker et al., 2003). DNA methylation has a very specific role in *Neurospora* for targeting regions RIPed by genome defence mechanisms. Presumably methylation acts as secondary silencing mechanism to keep RIPed sequences silenced, but evidence to support this suggestion has yet to be uncovered.

1.2.2 Cytosine methylation in plants

Cytosine methylation in plants is more widespread than in *Neurospora*, covering 5 - 25 % of total cytosine bases (Rangwala and Richards, 2004). The majority of DNA methylation is found concentrated in heterochromatic regions, including rDNA repeats, centromeric DNA, and parasitic / transposable elements. In contrast to the context independent profiles of methylation in *Neurospora*, cytosine methylation in plants is found in three contexts; CpG dinucleotides, CpNpG DNA, and CNN DNA (where N is A,T, or G) (Tariq and Paszkowski, 2004).

DNA methylation in *Arabidopsis* is mediated by at least four different cytosine methyltransferases: met1, drm1, drm2, and cmt3 (Tariq and Paszkowski, 2004). The met1 DNA methyltransferase is the *Arabidopsis* homologue of eukaryotic DNMT1 (Section 1.2.5) and is responsible for maintenance of CpG methylation during DNA replication (Saze et al., 2003). Drm1 and drm2 are homologous to the mammalian DNMT3 (Section 1.2.5) enzymes and have roles in setting up DNA methylation

profiles at CpG and CpNpG sites (Cao and Jacobsen, 2002). Cmt3 is responsible some locus specific non-CG methylation, and plays a role in histone methylation directed DNA methylation (Section 1.3.5). DNA methylation in *Arabidopsis* is associated with silenced regions of the genome, and may be actively interpreted (Berg et al., 2003).

1.2.3 Cytosine methylation in lower eukaryotes

A certain amount of variability exists between the DNA methylation profiles in fungi, and the same is true for multi-cellular eukaryotes. The nematode *Caenorhabditis elegans* (Simpson et al., 1986; Tweedie et al., 1997) does not contain DNA methylation nor a conventional DNA methyltransferase gene. A small amount of DNA methylation has been detected in *Drosophila melanogaster* (Gowher et al., 2000; Lyko et al., 2000), and the *Drosophila* genome encodes a DNA methyltransferase protein (DNMT2) which appears to be catalytically active (Kunert et al., 2003; Tang et al., 2003). The majority of DNA methylation in *Drosophila* occurs in the context of CpA dinucleotides, and interpretation of the this methylation appears to have a role in position effect variegation and normal chromosome segregation in *Drosophila* (Marhold et al., 2004b).

Most higher invertebrates contain moderate levels of methyl-cytosine (Bird, 2002; Tweedie et al., 1997) which is found almost exclusively on CpG dinucleotides. DNA methylation in invertebrates covers from 10 – 40 % (Tweedie et al., 1997) of the genome, consisting mostly of long stretches of densely methylated DNA compartments interspersed within completely unmethylated regions. High resolution mapping in *Ciona intestinalis* has shown that the majority of DNA methylation is found over the body of genes and not on intergenic or parasitic repeat sequences (Simmen et al., 1999). In invertebrates the absence of DNA methylation on parasitic repeat elements suggest the DNA methylation system does not function as a means of genome defence as has been suggested (Yoder et al., 1997b), but instead may have a role in modulating transcription (Simmen et al., 1999).

The invertebrate-vertebrate boundary delineates quite distinctly the shift in CpG methylation profile from, compartments of methylation that cover genes in the

invertebrate genome, to the extensive CpG methylation profile observed in vertebrates (Tweedie et al., 1997). Vertebrate genomes have CpG methylation virtually everywhere except for the CpG islands which are associated with the 5' end of most genes (Antequera et al., 1990). Many similarities are observed in the genomic methylation profiles of vertebrates, and this will be introduced in Section 1.2.4.

1.2.4 Genomic methylation profiles in mammals

DNA methylation is broadly distributed throughout the mammalian genome, but methyl-cytosine only accounts for about 1 % of total DNA bases. This number might seem surprisingly small but in the context of CpG dinucleotides, which is the main target sequence for DNA methylation in mammals, methylation is found on 70 - 80 % of cytosine bases.

Interestingly, the occurrence of methyl-CpG in the genome is actually a factor which drives loss of CpG dinucleotides, and this combination of bases is under-represented. The cause of CpG loss is a direct result of DNA methylation, as methyl-cytosine can spontaneously deaminate to thymine resulting in a G:T base pair mismatches. If repair of the mismatch does not occur before DNA replication, this can result in a C:G to T:A transition mutation in one of the resulting daughter strands. Therefore, simply by analyzing the genome sequence it is relatively easy to identify regions of the genome that are methylated and unmethylated based on CpG content. For example, regions of the genome which have higher than expected CpG content usually contain non-methylated CpG's, as is the case with CpG islands (see below). The coding regions of genes are usually under-represented with respect to CpG content, and the majority of CpG's contain DNA methylation. In addition to DNA methylation affecting the DNA composition of the genome it can also have profound affects on the coding potential of a gene. This is evident in human genetics as C:G to T:A transitions are a major cause of mutation in human disease (Cooper and Youssoufian, 1988). In other instances methylated CpG dinucleotides act as important mediators of gene regulation and constitute regulatory control elements, like those implicated in imprinted gene expression (Section 1.2.6). In these cases,

methyl-CpG is under selective pressure and must be maintained in order for normal gene regulation to occur.

Most non-methylated CpG dinucleotides are found in CpG islands at the 5' end of genes. In mammals, a CpG island is generally defined by a DNA sequence of over 200 base pairs, with greater than 50 % GC content, and an observed over expected ratio of CpG greater that 0. 6 (Gardiner-Garden and Frommer, 1987). Because CpG dinucleotides are generally under represented (as discussed above) an observed over expected value of 0.6 has generally been adopted as a cut off value to define a mammalian CpG island. Recent predictions from human genome sequencing efforts suggest that 29000 genes have CpG islands (Lander et al., 2001). In most tissues that have been analyzed, CpG islands remain non-methylated even if associated with silenced genes (Antequera et al., 1990). This suggests the genome has a mechanism to keep these regions free of DNA methylation, but the nature of this system remains undefined (Macleod et al., 1994). A defined biological role for CpG islands remains elusive, but one suggestion is that methylation free islands could act as a buffer to keep silenced regions of chromatin from impinging on the regulatory regions of genes.

1.2.5 Mammalian DNA methyltransferases

The process of DNA methylation in aqueous solution was initially considered to be chemically improbable (Bestor, 2000). The co-crystal structure of M. *HhaI* methyltransferase bound to an acceptor CpG (Klimasauskas et al., 1994) revealed an interesting mechanism for cytosine methylation. These studies showed that the cytosine base is pulled out of the DNA helix and inserted deep into the catalytic site of the enzyme. Once situated in the catalytic site, the DNA methylation reaction proceeds via thiol attack at the C6 position of the pyrimdine ring, resulting in formation of a covalent bond between the enzyme and the cytosine base (Figure 1.2). Enamine attack of the methyl donor, s-adenosyl-L-methionine (SAM), results in transfer of a methyl group to the C5 position of the pyrimidine ring (Figure 1.2). Abstraction of a proton from the C5 position allows the carbon-carbon double bond to reform and the base is released from the enzyme by β-elimination (Figure 1.2)

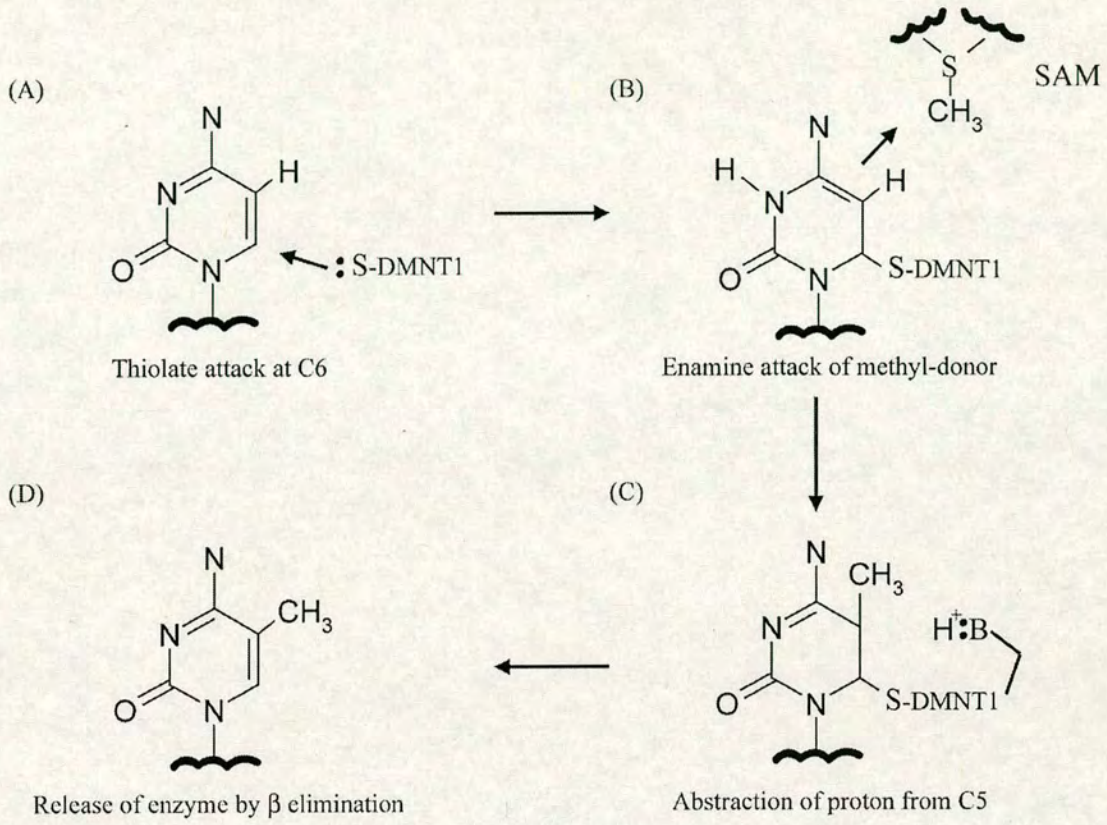


Figure 1.2- Catalytic mechanisms of 5-methyl-cytosine DNA methyltransferases (Adapted from (Bestor, 2000))

(A) The DNA methylation reaction proceeds via thiol attack at the C6 position of the pyrimdine ring, resulting in formation of a covalent bond between the enzyme and the cytosine base.
(B) Enamine attack of the methyl donor, s-adenosyl-L-methionine, results in transfer of a methyl group to the C5 position of the pyrimidine ring.
(C) Abstraction of a proton from the C5 position allows the carbon-carbon double bond to reform.
(D) The base is released from the enzyme by β-elimination.

(Bestor, 2000; Klimasauskas et al., 1994). The elegant mechanisms of DNA methylation, as revealed by structural studies of the bacterial methyltransferase M.*HhaI*, are likely to be conserved in mammalian DNA methyltransferase enzymes as the catalytic domains of these enzymes share striking sequence similarity.

Enzymatic activities capable of methylating CpG dinucleotides were identified biochemically in many different mammalian tissue sources (Bestor et al., 1988). The first mammalian enzyme identified at the cDNA level was DNA methyltransferase 1 (DNMT1), and its coding sequence revealed homology between the catalytic domains of the mammalian and bacterial DNA methyltransferase enzymes (Bestor et al., 1988). Since the initial identification of DNMT1, three other mammalian DNA methyltransferase enzymes have been identified, and include; DNMT2 (Yoder and Bestor, 1998), DNMT3a, and DNMT3b (Figure 1.3) (Okano et al., 1998a). Each of these enzymes shares homology to ten characteristic motifs found within DNA methyltransferases, six of which are highly conserved (Bestor, 2000; Okano et al., 1998a). Several functionally important motifs include, motif I and X that form the SAM binding region, motive IV that contains the thiol group required for enzymatic attack of the cytosine base, and motif IX which has roles in maintaining the structure of the enzyme that interacts with the major groove containing the cytosine base (Figure 1.3) (Bestor, 2000). Subsequent biochemical and genetic studies have been instrumental in delineating the role of each of these methyltransferase enzymes.

DNMT1 is the main enzyme required for maintenance of DNA methylation during DNA replication. As the replication fork passes over methylated CpG-dinucleotides, semi-conservative replication results in each newly replicated strand containing hemi-methylated CpG dinucleotides. DNMT1 can associate with the DNA polymerase clamp loader complex protein (PCNA)(Chuang et al., 1997) and can track with the replication fork to convert hemi-methylated DNA sites back into symmetrically dimethylated CpG's. Evidence *in vitro* indicates that DNMT1 has between 5 and 30 fold preference for hemi-methylated substrates (Yoder et al., 1997a) and converts hemi-methylated CpG in a processive manner (Vilkaitis et al., 2005). The efficient maintenance of DNA methylation patterns between cellular

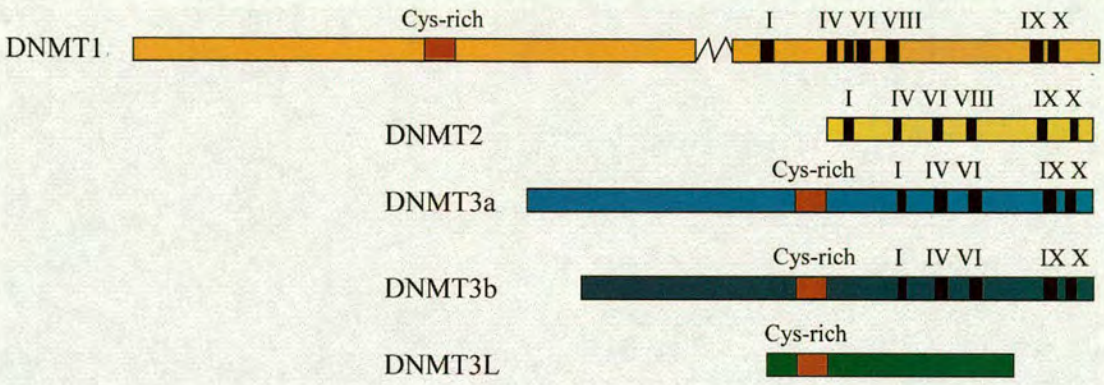


Figure 1.3-Domain organization of the mammalian DNA methyltransferase enzymes (Adapted from (Bestor, 2000)

The mammalian DNA methyltransferase enzymes (DNMT1, DNMT2, and DNMT3a/b) contain a series of conserved methyltransferase motifs (I-X) in the c-terminus of each protein. DNMT3L lacks the methyltransferase motifs but has a Cys-rich region which is shares with DNMT1 and DNMT3a/b.

replication events exemplifies why DNA methylation is considered a heritable epigenetic state (Bird, 2002). The importance of DNA methylation for normal development was realized when *Dnmt1* null alleles were introduced into mice. The mutant *Dnmt1* mice had a recessive lethal phenotype and null ES cells had an approximate three fold drop in the level of global DNA methylation (Li et al., 1992). Conversely, *Dnmt1* over-expression in mice causes a loss in the fidelity of the DNA methylation system resulting in DNA hypermethylation, loss on normal imprinting, and embryonic lethality (Biniszkiewicz et al., 2002). Mice carrying one null allele and one hypomorphic allele of *Dnmt1* survive and are viable despite reduction of DNA methylation levels to about 10 % of normal levels. These mice were runted at birth and died within 4 - 8 months developing T-cell lymphomas and displaying genomic instability (Gaudet et al., 2003). The evidence from mouse studies suggest even subtle affects in the expression profile of *Dnmt1* has pleotrophic affects on normal cellular function and development. Because of the difficulty in interpreting

the results of various mouse studies, controversy exists as to the exact cause of phenotypes resulting from *Dnmt1* manipulation, especially with regards to the catalytic activity of the protein and levels of DNA methylation. Interestingly DNMT1 interacts with transcriptional co-repressor molecules suggesting that functions in addition to the catalytic activity of DNMT1 are important for its function *in vivo* (Fuks et al., 2000; Rountree et al., 2000).

DNMT2 has proved the most enigmatic mammalian DNA methyltransferase. A surprising amount of effort has yielded very little useful understanding of how, or if, the DNMT2 protein functions *in vivo*. *Dnmt2* null ES cells have normal DNA methylation levels and are competent in silencing newly integrated viral DNAs (Okano et al., 1998b). Recombinant DNMT2 was generated using bacterial and baculovirus expression systems but no DNA methyltransferase activity was detected (Okano et al., 1998b). The crystal structure of DNMT2 revealed striking structural homology to bacterial methyltransferase M. *HhaI* (Dong et al., 2001), and DNMT2 was able to form denaturation resistant complexes with DNA suggesting that the protein forms covalent adducts with DNA in a similar manner to functional DNA methyltransferases (Figure 1.2). Recently, a detailed biochemical study of recombinant human DNMT2 has revealed very weak cytosine methyltransferase activity (Hermann et al., 2003) towards CpG in the context of ttnCGga(g/a) DNA. Additional *in vivo* analysis is required to determine the functional relevance of DNMT2 in human cells. Interestingly, homologues of DNMT2 have been identified in flies (Hung et al., 1999; Tweedie et al., 1999) and fission yeast (Section 1.2.1) suggesting that DNMT2 may have an evolutionarily conserved function.

The DNMT3 methyltransferases, DNMT3a and DMNT3b, are encoded by separate genes but share wide ranging sequence homology at the amino acid level (Figure 1.3) (Bestor, 2000). Original biochemical studies identified several methyltransferase activities in cellular isolates, and in addition to DNMT1, the DNMT3 enzymes appear to account for these activities. In contrast to DNMT1, which is important for maintenance of pre-existing DNA methylation marks during DNA replication, the DNMT3 enzymes are largely thought to carry out the role of *de novo* DNA methylation in mammals (Bird, 1999; Okano et al., 1999; Xu et al., 1999). Cells

lacking the DNA methyltransferase 3a and 3b enzymes (Chen et al., 2003a) slowly lose genomic methylation, presumably because of imperfect fidelity in copying of DNA methylation profiles by DNMT1 during replication, and subsequent inability to add methylation to these missed methyl-CpG sites (Rhee et al., 2002). Reintroduction the DNMT3a enzyme into DNMT3a/b null cells restored wild type DNA methylation levels, verifying the role of DNMT3 enzymes in *de novo* methylation *in vivo* (Chen et al., 2003a). DNMT3a null mice are normal at birth but die at about 4 weeks later and DNMT3b null mice die at embryonic day 9. 5 due to multiple developmental defects (Okano et al., 1999). Therefore, *de novo* methylation is important for normal mouse development and survival. In support of the important role of DNMT3 enzymes in normal human development, mutations in DNMT3b have been linked to ICF syndrome (for immunodeficiency, centromere instability and facial anomalies)(Hansen et al., 1999; Xu et al., 1999). The ICF mutations in DMT3b are all in the C-terminal region of the protein which encodes the DNA methyltransferase domain, and at least one mutation is defective for *de novo* methyltransferase activity. In fitting with a loss of methyltransferase activity, ICF patients display a marked loss of DNA methylation over the human classical satellites 2 and 3 on chromosomes 1, 9, and 16 (Xu et al., 1999). The dynamic *de novo* methyltransferase function of the DNMT3a/b enzymes, in combination with the maintenance methyltransferase DNMT1 ensures fidelity of the DNA methylation system.

Recently, a third DNMT3 family member was discovered that shares homology to the cysteine rich region of the DNMT3 enzymes, but lacks the conserved methyltransferase motifs (Figure 1.3) (Bourc'his et al., 2001; Hata et al., 2002). Knockout studies in mice demonstrated that the DNMT3L protein was necessary for normal male germ cell development and silencing of transposable elements (Bourc'his and Bestor, 2004). DNMT3L also has an important role in setting up maternal DNA methylation imprints, and a maternal effect lethal phenotype is observed in the progeny of homozygous females (Bourc'his et al., 2001). The loss of DNA methylation in DNMT3L knockout mice lead to the speculation that DNMT3L may associate with the DNMT3 enzymes, a prediction that turned out to be correct

(Hata et al., 2002). Further studies have delineated an interesting role for DNMT3L in direct interaction with, and stimulation of the DNMT3a/b enzymes (Chedin et al., 2002; Suetake et al., 2004). Stimulation of DNMT3a/b enzymes by DNMT3L results from a binding induced conformational change in DNMT3a/b that causes increased DNA binding and catalysis (Gowher et al., 2005). An interesting but poorly studied aspect of DNMT3L function is interaction with histone deacetylase enzymes. Two reports have indicated that DNMT3L interacts with histone deacetylase enzymes to repress transcription, potentially linking active DNA methylation and chromatin remodelling (Aapola et al., 2002; Deplus et al., 2002). The contribution of DNMT3L to silencing of gene expression remains to be elucidated.

1.2.6 DNA methylation in development, imprinting, and the X-chromosome

One of the most interesting aspects of DNA methylation is its dynamic role during development in setting up allele specific gene expression profiles in mammals. DNA methylation is involved in two general forms of allelic silencing; firstly, DNA methylation contributes to imprinting of a group of genes which are expressed exclusively from either the maternal or paternal allele, and secondly DNA methylation contributes to the dosage compensation system which results in silencing one of the two X-chromosomes in female offspring.

In mice, the patterns of DNA methylation during development are well characterized. The DNA methylation marks in primordial germ cells are initially erased and then reform as the germ cells mature (Reik and Walter, 2001). The DNMT3 family of proteins are responsible for setting up these methylation marks (Bourc'his and Bestor, 2004; Bourc'his et al., 2001; Kaneda et al., 2004). After the egg is fertilized, the male pronucleus is rapidly demethylated, but the levels of methylation in the maternal genome remain high. From fertilization until the eight cell stage, passive demethylation of the maternal genome causes a global reduction in methylation levels until the maternal genome has roughly equal levels of methylation to the paternal genome (Mayer et al., 2000). Interestingly, there are certain imprinted genes which escape these demethylation events likely due to the function of a maternal store of DNMT1 (Delaval and Feil, 2004; Howell et al., 2001). From the eight cell

stage onwards, there is a gradual increase in the levels of DNA methylation until birth. DNA methylation marks are then maintained at high levels in most somatic cells with only minor alterations in methylation occurring in specific cells and loci.

During development, a subset of 'imprinted' loci are unaffected by global changes in DNA methylation and they retain information encoded in the maternal and paternal gamete. The requirement for imprinting was originally hypothesized during the early 1980's when pronuclear transfer studies using gynogenetic (only female chromosomes) and androgenetic (only male chromosomes) embryos highlighted the developmental requirement of a male and female gamete complement (Surani et al., 1984). The term 'imprint' was used to describe the differences in the male and female gametes that accounted for their ability to complement one another and produce a normal embryo. The insulin-like growth factor II gene (*Igf2)*, a fetal mitogen, was the first to fit the definition of an imprinted gene, being exclusively expressed from the paternal chromosome (DeChiara et al., 1991). The importance of imprinted *Igf2* gene expression is evident from gene targeting experiments where *Igf2* was deleted. In *Igf2* knockout mice, passage of the null allele through the female results in a normal mouse, but passage of the null allele thought the male germ-line resulted in growth deficiency (DeChiara et al., 1990). Further studies lead to the discovery that *Igf2* and *H19* genes were genetically linked on chromosome 7 and reciprocally imprinted in gene expression profile on the male and female chromosome (Zemel et al., 1992). From studying the *Igf2/H19* locus, the nature of the imprint responsible for control of allele specific expression was found to involve DNA methylation (Bartolomei et al., 1993; Ferguson-Smith et al., 1993). The general model of imprinting at the *Igf2/H19* locus depends on a region called the differentially methylated domain (DMD) which is located between the *Igf2* and *H19* genes (Figure 1.4). This CpG rich DMD is non-methylated on the maternal allele allowing the insulator protein CTCF to bind and block access of a downstream enhancer to the *Igf2* gene (Arney, 2003). Therefore, on the maternal allele the enhancer functions exclusively to drive H19 gene expression while *Igf2* remains silent. On the paternal allele the DMD is methylated which inhibits CTCF binding

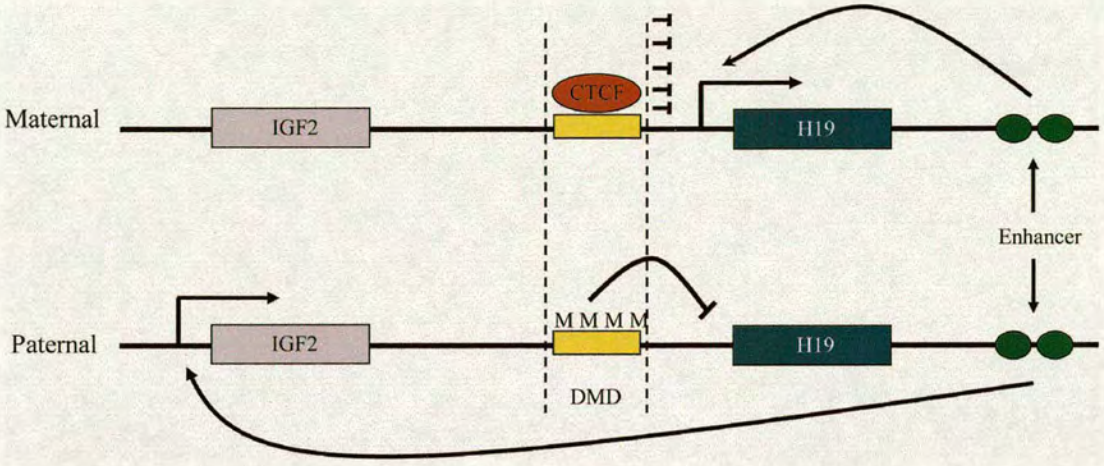


Figure 1.4- Imprinting at the *H19/IGF2* locus

The CpG rich DMD is non-methylated in the maternal allele where the insulator protein CTCF binds and blocks access of a downstream enhancer to the *Igf2* gene. Therefore the enhancer is free to drive *H19* gene expression while *Igf2* remains silent. On the paternal allele the DMD is methylated which inhibits CTCF binding. The downstream enhancer is able to act on *Igf2* allowing gene expression, while the *H19* gene remains silenced.

(Figure 1.4). The downstream enhancer is able to act on *Igf2* allowing gene expression, while the H19 gene remains silenced (Arney, 2003). The biological rational for reciprocal imprinting of the Igf2/H19 locus remains uncertain but it may be the result of competing maternal and paternal genomes (Reik and Walter, 2001). The *Igf2/H19* locus has been widely studied, but 60 other imprinted loci have been identified, and 100 are predicted to exist in the mouse genome (Scarano et al., 2005). In most cases, normal function of imprinted genes is associated with differentially methylated regions that also contain other epigenetic modifications (Fournier et al., 2002).

The importance of normal imprinting is highlighted by the association of several human diseases with loss of imprinting defects. For example, the autosomal dominant Angelman and Prader-Willi syndromes (Brannan and Bartolomei, 1999)

are caused by disruption of a cluster of imprinted genes on human chromosome 15. Normally, a sub-set of genes in this imprinted cluster are expressed exclusively from the maternal or paternal allele. These diseases usually result from a parent of origin deletion mutation which disrupts the normal balance of imprinted gene expression. Maternal defects result in Angelmans syndrome and paternal defects in Prader-Willi syndrome. Most mutations in Angelmans and Prader-Willi syndromes are the result of gene deletions, but direct defects in imprinting that correlate with perturbed DNA methylation accounts for about 5 % of cases (Kantor et al., 2004).

The female mammalian X chromosome undergoes radical changes in DNA methylation patterns associated with near complete silencing of genes on a single X chromosome. Early in development the female embryonic cells independently choose which of the two X chromosomes will be inactivated and this choice is propagated in future cell generations, making females mosaic in terms origin of X chromosome expression. The X inactivation choice is made independent of gametic imprints, and therefore selection of X chromosome expression is not affected by parent of origin. Some exceptions to this rule exist, for instance, there is imprinted X inactivation in the placenta of some organisms (Huynh and Lee, 2001). The molecular process of X inactivation in mice is regulated by competing expression between the *Xist* and *Tsix* non-coding RNA's in a region called the X inactivation center (XIC). On the active X (Xa), an insulator protein, CTCF, binds elements near the *Tsix* promoter resulting in stable *Tsix* expression and inhibition of *Xist* accumulation. Repression of *Xist* on the Xa is partly due to methylation of its CpG island (Panning and Jaenisch, 1996). Conversely, on the future inactive X (Xi) chromosome, *Tsix* expression is lost allowing accumulation of the *Xist* RNA, and the *Xist* promoter region remains free of DNA methylation. Curiously, *TSIX* does not appear to have the same role in regulating human X inactivation, suggesting alternative regulatory mechanisms might exists in humans (Migeon, 2003; Migeon et al., 2002). With time, *Xist* RNA is able to coat the entire X chromosome and this event correlates with the initiation of silencing and eventually loss of gene expression from that X chromosome. Coverage of the Xi by *Xist* RNA is sufficient to initiate silencing but not required for its continued maintenance (Wutz and Jaenisch,

2000). Other factors including various chromatin modulators follow *Xist* coverage and progress the silencing program. A unique feature of silencing on the Xi is dense methylation of the CpG islands in most genes. Usually CpG islands remain free of DNA methylation regardless of expression status, but on the Xi CpG islands are specifically targeted for DNA methylation, perhaps as a long term and heritable mechanism of maintaining dosage compensation. The exact role DNA methylation plays in silencing genes on the inactive X is unclear as embryos deficient for de *novo* methyltransferase activity allow reactivation of gene expression under some circumstances, but in other cases normal silencing is maintained (Sado et al., 2000; Sado et al., 2004). More recent evidence suggests an important role for polycomb proteins, histone modification, and chromatin modeling in the initiation stages of X chromosome inactivation (de Napoles et al., 2004; Fang et al., 2004; O'Neill et al., 2003; Plath et al., 2003). Multiple redundant silencing systems lead to normal X inactivation, and as a component of this silencing system DNA methylation is a distinct feature of the normal inactivated X chromosome.

1.2.7 DNA methylation and silencing of parasitic DNA elements

DNA methylation is integrally linked with genome defense in *Neurospora.* In mammals, DNA methylation protects against the deleterious affects of parasitic DNA elements in addition to well characterized roles in modulating gene expression. The human genome is littered with repetitive DNA (~35 %)(Yoder et al., 1997a), a large percentage of which corresponds to inactivated copies of parasitic elements. Most parasitic elements consist of short inderdispersed DNA elements (SINE's), long interdispersed DNA elements (LINE's), and LTR containing elements (Smit and Riggs, 1996). Activation of transposable elements and insertion into novel locations within the genome could have devastating affects on the normal coding potential of the genome. Most of the functional transposable elements in mammalian cells are densely methylated and silenced (Yoder et al., 1997b). DNA methylation likely contributes to the long term silencing of these elements through recruitment of chromatin modifying transcriptional repressors (Hata and Sakaki, 1997; Yu et al., 2001). In *Dnmt1* null mouse embryo's, loss of DNA methylation correlates with reactivation of a class of LTR elements called IAP's (Gaudet et al., 2004; Walsh et

al., 1998). Interestingly the LTR containing IAP elements do not undergo demethylation during the cleavage stages of mouse development (Lane et al., 2003), presumably to protect against activation of these functional transposons, during this sensitive time of development. A maternal store of DNMT1 (DNMT1o) or DNMT3 proteins likely mediate maintenance of IAP methylation at this critical point during mouse development (Howell et al., 2001). DNA methylation is also involved in the long term silencing of IAP elements and a general decrease in methylation of IAP elements in aged mice correlates with increased transcription of these elements (Barbot et al., 2002).

1.2.8 Targeting DNA methylation by small RNA's in mammals?

In plants, DNA methylation can be targeted by small double stranded RNA's to homologous DNA sequences as a means of silencing repetitive elements and modulating gene expression (Jones et al., 2001; Tariq and Paszkowski, 2004). RNAi induced DNA methylation occurs at most cytosines in the affected region, not just in context of CpG dinucleotides (Aufsatz et al., 2002). Recent studies in plants have identified a DNA methyltransferase (DRM)(Cao et al., 2003) and a class of RNA polymerase gene that are responsible for the RNAi mediated DNA methylation (Herr et al., 2005).

Short double stranded RNA in mammalian cells can lead to the active destruction of homologous single stranded RNA's via a post-transcriptional silencing mechanism (Matzke and Birchler, 2005; Sontheimer, 2005). Endogenous or exogenously produced double stranded RNA's generated by pairing of sense and antisense transcripts can act as a source of dsRNA that is processed by mammalian Dicer, an RNAse III related enzyme (Sontheimer, 2005). Dicer also processes microRNA's, which are shorter hairpin RNA's normally made in the cell that can target endogenous mRNA's for translational inhibition (Gebauer and Hentze, 2004). Substrates processed into small dsRNA's by Dicer are transferred to the RISC (RNA induced silencing complex) complex which mediates post transcriptional silencing. In plants and *S. pombe,* processing of double stranded RNA can result in transcriptional silencing affects in addition to post-transcriptional mRNA

degradation (Schramke and Allshire, 2003). In *S. pombe* the nuclear RNA induced initiation of transcriptional gene silencing (RITS) complex binds siRNA's and is targeted to the actively transcribing genomic locus corresponding to the bound siRNA and contributes to the transcriptional silencing program (Motamedi et al., 2004). This pathway in *S. pombe* is particularly important for silencing within the centromeric repeats (Motamedi et al., 2004; Verdel et al., 2004). Once RITS is located at the corresponding genomic loci it induces transcriptional silencing by modifying chromatin structure (Verdel et al., 2004).

Recently, DNA methylation directed by double stranded RNA has been observed in mammalian cells, suggesting that RNA mediated transcriptional silencing pathways, like that found in plants and fungi, might also exist in mammals. Two studies have successfully demonstrated short double stranded RNA (dsRNA) can target DNA methylation to corresponding DNA sequences. The first study used a combination of dsRNA's corresponding to the promoter of an endogenous gene and observed DNA methylation at CpG dinucleotides in the promoter DNA. Knock down of DNMT enzymes revealed that DNMT1 and DNMT3 are capable of mediating DNA methylation as a result of the dsRNA (Kawasaki and Taira, 2004). The second study confirmed that short double stranded RNA could silence an endogenous gene. In this study it appeared as it the nuclear envelope was a barrier for dsRNA entry into the nucleus, and effective silencing was only observed when the nucleus was made permeable to the siRNA (Morris et al., 2004). Other studies have not been able to target DNA methylation by dsRNA, suggesting that this process may be context dependent, or require unknown factors for efficacy (Park et al., 2004; Svoboda et al., 2004). In agreement with the existence of intact transcriptional silencing pathway in mammals, Dicer knockout ES cells display up-regulation of centromeric satellite repeat transcription, and have decreased levels of DNA methylation in the centromeric repeats (Kanellopoulou et al., 2005). More work is needed to delineate the functional properties of a nuclear RNAi silencing pathway in mammals, but evidence suggests that *de novo* DNA methylation is an active component.

1.2.9 DNA methylation and cancer

The initial link between levels of DNA methylation and cancer revealed that tumor samples were hypomethylated when compared to normal tissue of a similar origin (Feinberg and Vogelstein, 1983). Hypomethylation likely contributes to malignancy by allowing the re-expression of genes which are normally silenced by DNA methylation and by loss of genomic stability that comes from reduced methylation profiles (Feinberg and Tycko, 2004). DNA hypomethylation occurs in most cancer tissues and cells lines, but other regions of the malignant genome acquire distinct hypermethylation (Baylin et al., 1986; Toyota and Issa, 1999). Hypermethylation in promoter regions of genes is associated with silencing, and if a malignant cell can silence a gene that inhibits rapid cellular growth this is an obvious advantage for sustained proliferation. A common target of promoter methylation in cancer cells are tumor suppressor genes including *RB*, *p16*, and *MLH1* (Esteller, 2000). Inhibition of *RB* and *p16*, whose sole job is to control cell cycle events, allow the cancer cell to bypass checkpoints normally in place to preclude rapid and continual cell division. By silencing the mismatch repair protein MLH1, a cancer cell can increase the rate of mutation and accelerate the likelihood of acquiring new mutations that are advantageous to growth. Tumors in which the promoter region of *MLH1* is methylated and gene expression is silenced, display a reduced mismatch repair capacity (Herman et al., 1998). Lack of MLH1 mismatch repair activity has a profound affect on the repetitive microsatellite sequences in the genome, and this is likely an indicator of an overall inability to repair mismatch events that occur as DNA polymerase replicates DNA (Herman et al., 1998). The clinical relevance of DNA hypermethylation in cancer is exemplified by positive results in clinical trials treating cancer patients with low doses of the demethylating agent 5-Azacytidine (Esteller, 2005), and the recent approval of 5-azacytidine by the U. S. Food and Drug Administration for the treatment of myelodysplastic syndrome (Esteller, 2005; Leone et al., 2002).

A mouse model deficient in DNA methyltransferase activity has provided genetic evidence for an active role of DNA methylation in cancer progression. These studies have been carried out in the $Apc^{(Min/+)}$ mouse which was used as a model for

spontaneous colorectal cancer formation. $Apc^{(Min/+)}$ mice carry a germ-line mutation in one copy of the tumor suppressor gene, *Apc*, and are susceptible to polyp formation in the intestinal mucosa. In the $Apc^{(Min/+)}$ mouse, the level of CpG island hypermethylation was dramatically increased in the developing polyp, when compared to adjacent normal mucosa. Crossing mice with a hypomorphic allele of *Dnmt1* onto the $Apc^{(Min/+)}$ mouse background resulted in complete genetic suppression of polyp formation (Eads et al., 2002). Hypermethylation of CpG islands is normally observed in the developing polyp of $Apc^{(Min+)}$ mice but was absent in the *Dnmt1* hypomorph mice indicating that CpG island hypermethylation is a rate limiting step in polyp formation. These studies provide a direct link between CpG island hypermethylation and progression of colorectal cancer.

1.3 Chromatin remodeling and histone modification in gene expression

1.3.1 Chromatin structure, histones, and histone modification nomenclature

The basic protein unit responsible for compacting DNA into the relatively small confines of the nucleus is the histone octamer. The histone octamer consists of two dimers of histone H2A:H2B and H3:H4, around which approximately 147 base pairs of DNA is wrapped (Peterson and Laniel, 2004). Hierarchal nucleosome fibers form the basic octamer unit further increasing the potential for DNA compaction (Dorigo et al., 2004). In addition to the biophysical properties of chromatin, post translational modification of histones, and histone variants, can have profound affects on the activity of the genetic information wrapped around histone octamers. Over the past 5 – 10 years a large amount of information about histone modification and histone modification states has become available. Histone modifications include; biotinylation (Camporeale et al., 2004), ubiquitination, sumoylation, ADP-ribosylation, phosphorylation, acetylation, and methylation (Strahl and Allis, 2000). The nomenclature used to describe histone modification varies, so as a matter of clarity this section of the introduction uses the standardized BRNO nomenclature to describe histone modifications (Turner, 2005). The most thoroughly characterized histone modifications involved in regulating gene activity are acetylation and

methylation. This section of the introduction will mostly address these modifications as removal of histone acetylation and addition of certain repressive histone methylation marks is associated with DNA methylation.

1.3.2 Active histone modification marks

To activate gene expression in a genomic context there is a requirement to overcome the repressive chromatin environment that keeps genes silenced in the absence of activating factors. Most gene activation is associated with hyperacetylation of the lysine residues on core histone tails in the promoter region and 5' end of genes, both immediately prior to and during active transcription. In budding yeast, acetylation is associated with the 5' ends of genes, and in higher eukaryotes hyperacetylation is linked to promoter regions of genes (Roh et al., 2005; Roh et al., 2004). Recent genome wide chromatin mapping projects have demonstrated a tight correlation between the level of histone acetylation and the location of promoter and regulatory elements in genes (Bernstein et al., 2005; Roh et al., 2005). Moreover, the levels of histone acetylation are tightly correlated with the transcriptional state of the gene, and active genes are generally hyperacetylated (Schubeler et al., 2004). Acetylated lysines are found on multiple residues within the n-terminal tails of histone molecules, including; H2AK(5,9)ac, H2BK(5,12,15,20)ac, H3K(9,14,18,23)ac, and H4K(5,8,12,16).

Since the first histone acetyltransferase gene was cloned in *Tetrahymena* (Brownell et al., 1996), a series of acetyltransferase proteins containing the conserved histone acetyltransferase (HAT) domain have been identified in eukaryotes. In mammals, there are a surprisingly large number of histone acetyltransferase enzymes, which likely reflects their diverse substrate specificities and varying functions related to transcriptional activation (Ogryzko, 2001). In most cases, histone acetyltransferase enzymes do not have DNA binding capacity but instead rely on transcriptional activators and local chromatin modification for site specific recruitment and activity. Two of the most thoroughly studied acetyltransferase complexes are TFIID and PCAF/GCN5. In budding yeast, these two complexes have global affects on transcription, combining to affect the expression of about 70 % of genes (Lee et al.,

2000). A certain degree of specificity exists, as independently TFIID affects 30 % of genes and GCN5 (SAGA) 12 % of genes (Lee et al., 2000). The precise substrate specificity of the mammalian HAT enzymes *in vivo* (reviewed in (Kuo and Allis, 1998; Marmorstein, 2001)) is not entirely clear but individual enzymes do display amino acid preference within histones.

Histone acetylation can alter the charge properties of the N-terminal tails that protrude outwards from the core histone octamer, resulting in fewer positive charges. The loss of positive charge reduces the affinity of the n-terminal tails for negatively charged DNA molecules (Hong et al., 1993), which in turn may contribute to increased accessibility of DNA elements within chromatin containing hyperacetylated histones (Lee et al., 1993). In addition to the charge affects on histone tail/DNA interaction, another role of histone acetylation is to create docking sites for proteins that specifically recognize acetylated histone tails, which then act as a system to interpret the acetyl-lysine signal. An example of a domain that can directly recognize acetylated lysine residues in the N-terminal tales of histones is the bromodomain (Dhalluin et al., 1999; Jacobson et al., 2000; Zeng and Zhou, 2002). Structural analysis indicates that the bromodomain can interact with acetyl lysine but also requires interactions with flanking amino acid sequences providing individual bromodomains a certain amount of selectivity within proteins that contain multiple acetyl lysine residues (Mujtaba et al., 2002). Functional bromodomain proteins include histone acetyltransferases and chromatin remodeling enzymes (Hassan et al., 2002) suggesting that the bromodomain functions to interpret, and perhaps and propagate, active acetylated chromatin states (Cosgrove et al., 2004).

Methylation of specific arginine and lysine residues in histones H3 and H4 has also been linked to transcriptionally active genes. These methylation modifications occur on; H3K4, H3K36, H3K79, and H4R3. Unlike lysine acetylation, lysine methylation can occur in three different states mono, di, and tri methylation (ie H3K4mel, H3K4me2, and H3K4me3), depending on the enzyme which applies the modification (Zhang et al., 2003). Genome wide analysis has revealed that H3K4m2 is associated with the 5' and body of genes that are actively transcribed, or are in a poised state for transcriptional activation (Santos-Rosa et al., 2002; Schneider et al., 2004). H3K4m3

is a better indicator of genes actually in the process of transcription and is usually only associated with the 5' regions of the gene (Bernstein et al., 2005; Ng et al., 2003). Histone H3K36 methylation is intimately associated with the active process of transcription and is found over the body of actively transcribed genes (Bannister et al., 2005; Strahl et al., 2002; Xiao et al., 2003b). H3K79 methylation is directly linked to H3K4 methylation in the 'trans-regulatory' histone modification pathway (Briggs et al., 2002; Xiao et al., 2003b). This pathway requires ubiquitination of H2B123 which leads to methylation of both H3K79 and H3K4 in association with gene activation. Finally, H4R3 is methylated in response to active transcription by hormone receptors (Ma et al., 2001).

SET domain containing proteins are responsible for the majority of histone methylation and catalyze the formation of H3K4, H3K36, and H4K20 methylation (Xiao et al., 2003a). H4R3 methylation is catalyzed by PRMT1, a nuclear member of the protein methyltransferase family, and H3K79 methylation is mediated by an interesting protein, Dot1, that lacks a SET domain and seems to be unique in its capacity to modify H3K79. There are a growing number of SET domain proteins from yeast to humans that have active histone methyltransferase activity, and each uses roughly the same enzymatic mechanism (Xiao et al., 2003a). Although extensive similarity exists between SET domain containing proteins, each appears to have quite distinct specificity for both amino acid residue and methylation state. Several structural studies have identified intrinsic differences that contribute to this specificity (Xiao et al., 2003a).

Recently, the chromodomain was shown to interact with methylated lysine residues on histone tails, and may have an analogous roles to the acetyl-lysine recognition by the bromodomain (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). A large number of proteins associated with chromatin function have chromodomains, indicating that a diverse network of factors may be involved in interpreting histone methylation marks. For example, the yeast chromodomain containing protein Chd1, a component of the yeast SAGA histone acetyltransferase complex, specifically recognizes H3K4m2. This interaction suggests that histone methylation on H3K4 may recruit histone acetyltransferase activity via the chromodomain of Chd1,

efficiently coupling histone methylation and acetylation (Pray-Grant et al., 2005). Methylation of H3K4 also acts as a binding site for the ATPase dependent chromatin remodeling factor ISWI (Santos-Rosa et al., 2003; Zegerman et al., 2002), while inhibiting binding of the repressive nucleosome remodeling and deacetylation complex called NURD (Zegerman et al., 2002). Further investigation will uncover how histone methylation marks contribute to gene activation.

1.3.3 Repressive histone modification marks

A key step in gene silencing is the removal of histone modifications that are associated with the active state. When genes are silenced, the level of histone acetylation rapidly decreases in the promoter region of the gene. The rapid nature of this modification reversal is the result of a class of enzymes called histone deacetylases that actively remove acetyl groups from lysine residues (Katan-Khaykovich and Struhl, 2002). The identity of the genes responsible for histone deacetylation remained elusive until biochemical purification and cloning revealed an active deacetylase enzyme, named histone deacetylase 1 (HDAC1). HDAC1 turn out be homologous to the yeast transcriptional repressor Rpd3 confirming a role in gene silencing (Taunton et al., 1996). Bioinformatic analysis using HDAC1 amino acid sequence led to the identification of a family of enzymes that efficiently catalyze the removal of acetyl groups from lysine without co-factors (Taunton et al., 1996). Nicotinamide adenine dinucleotide (NAD)-dependent histone deacetylases have subsequently been identified which also efficiently remove acetyl groups from lysine (Imai et al., 2000). Histone deacetylase enzymes are found in a variety of co-repressor complexes in the nucleus and are recruited by transcriptional silencers to aid in the removal of active histone acetyl marks, during transcriptional repression (Jepsen and Rosenfeld, 2002).

An enzymatic activity that could remove methyl groups form lysine residues in histones was characterized in the early seventies (Paik and Kim, 1973), but studies analyzing the turnover of methylated histones suggested that methylation was not actively removed once transferred to lysine residues (Waterborg, 1993). Lysine specific demethylase 1 (LSD1) was recently identified, laying to rest the controversy

regarding the reversibility of histone methylation (Shi et al., 2004). Very little is known about the cellular roles of LSD1 but it does specifically remove H3K4m1 and H3K4m2 groups by an oxidation reaction involving FAD as a cofactor. Before the enzymatic activity of LSD1 was identified, it was found as a protein component of various cellular co-repressor complexes (Shi et al., 2004), that also contain histone deacetylase enzymes. Therefore, factors involved in reversing the active post-translational modification on histones appears to be actively coupled to the transcriptional repression systems affirming the importance of clearing active chromatin marks during gene silencing.

In addition to removing active histone modifications, silencing of gene expression also involves the addition of specific histone methylation modifications. Methylation of H3K9, H3K27, and H4K20 are all causatively linked to silenced chromatin. Methylation of H3K9 in mammals is found associated with the promoter regions of silenced genes (Kondo et al., 2004), repetitive DNA elements (Kondo et al., 2004), and regions of centromeric heterochromatin (Martens et al., 2005). H3K9m3 is concentrated in regions of centromeric heterochromatin and is part of a silencing pathway that results in H4K20m3 (Schotta et al., 2004). The exact function of the H3K9m3/H3K20m3 pathway remains unclear, but it does contribute to the normal silencing of aberrant centromeric repeat transcription (Schotta et al., 2004). Methylation of H3K27 is also linked to the repressed state (Peters et al., 2003). The best characterized functions for H3K27m are in silencing of transcription associated with the ultrabithorax genes in *Drosophila* (Cao et al., 2002), the *Hox* gene cluster in mammals (Cao and Zhang, 2004), and X-chromosome inactivation (Plath et al., 2003; Silva et al., 2003).

Like active histone modification marks, repressive histone methylation marks create binding sites for chromodomain containing proteins. The chromodomain containing heterochromatin protein 1(HP1), can recognize H3K9 methylation (Bannister et al., 2001; Jacobs et al., 2001; Lachner et al., 2001), and through interactions with H3K9 methyltransferase Suv39H1 propagates the silenced chromatin state (Yamamoto and Sonoda, 2003). The enhancer of zeste methyltransferase complex (Cao et al., 2002)

modifies H3K27, and this mark is recognized by the Polycomb protein which also has a chromodomain (Min et al., 2003).

1.3.4 Does a histone code exist?

Many modification states can occur on histone molecules in chromatin. Some modifications occur in combination, while others exclude subsequent modification at on the same molecule (Nishioka et al., 2002; Sarg et al., 2004). The interaction of histone modification states is exemplified by the H2BK123u/H3K79m/H3K4m trans regulatory pathway in budding yeast (Briggs et al., 2002; Sun and Allis, 2002). In this pathway H2BK123u is a prerequisite for H3K79 and H3K4 methylation, exemplifying how one modification can be required for the subsequent modification of other residues. Proteins that recognize different modification states on histone tails may be a mechanism to interpret these modifications and result in differing biological outcomes (de la Cruz et al., 2005; Turner, 2002). The diverse range and combinatorial affects of different histone modifications suggest these states may be an epigenetic system to code for biological readout. Therefore, the existence of a binary histone modification code that shares similarity to the genetic information stored in DNA code has been proposed (Fischle et al., 2003; Jenuwein and Allis, 2001; Turner, 2002). For many, the idea of a histone code is misleading as the information stored in histone modifications is not as rigid as protein coding information that is found in gene encoding DNA (Peterson and Laniel, 2004). Fine scale histone modification mapping and identification of proteins that interpret different combinations of histone modification will help to iron out the specificity of histone modifications and how their interplay results in biological outcome.

1.3.5 Interplay between histone modification and DNA methylation

In *Neurospora* all DNA methylation is directly linked to histone H3 K9 tri-methylation (H3K9me3). A mutation in the SET domain of the H3 trimethylase *dim-5* results in a complete loss of DNA methylation (Tamaru and Selker, 2001). DNA methylation is also completely dependent on HP1 recognition of H3K9me3 (Tamaru et al., 2003) suggesting HP1 is necessary to recruit the DNA methyltransferase activity of dim-2 (Section 1.2.1) to loci containing histone methylation.

In plants, interdependence between DNA methylation and histone H3K9me/H3K27me has also been identified. *Arabidopsis* with mutations in the Kryptonite (KYP) gene, which is a H3K9 methyltransferase, lose DNA methylation at CpNpGp sites (Jackson et al., 2002; Malagnac et al., 2002). The DNA methyltransferase (CMT3), which is responsible for CpNpGp methylation, is recruited by its chromodomain, to sites of H3K9me/K3K27me (Lindroth et al., 2004) to induce DNA methylation. This suggests that fungi and plants have evolved different methyl-histone recognition modules to induce DNA methylation at loci containing histone methylation. In Neurospora loss of DNA methylation does not affect histone methylation patterns, but in plants DNA methylation has the capacity to direct histone methylation marks (Soppe et al., 2002; Tariq et al., 2003). In plants CpG methylation mediated by the plant homologue of DNMT1 (MET1) affects the level of H3K9 methylation. Therefore plants appear to have evolved mechanisms by which DNA and histone methylation can influence each other.

In mammals the exact mechanisms by which the DNA and histone methylation pathways interact is less clear, but analogous to the situation in plants, histone methylation has an impact on DNA methylation and visa versa. In mammals, both methylation systems are physically linked with DNMT1 immunoprecipitating Suv39H1 (Fuks et al., 2003a) and DNMT3a/b associating with H3K9 methyltransferase activity (Fuks et al., 2003a; Geiman et al., 2004). There is no decisive evidence to indicate whether DNA methylation or histone methylation is the initiating signal in the silencing that employs both systems. In some instances, histone methylation appears to be the initiating signal, for example, the kinetics of silencing the $p16^{INK4a}$ gene in cancer cells indicates that histone methylation precedes DNA methylation (Bachman et al., 2003). A similar observation was made in the *Suv39h1* and *h2* double null ES cells that have reduced H3K9 methylation at pericentric heterochromatin, delocalized DNMT3b staining pattern, and reduced DNA methylation at satellite repeats (Lehnertz et al., 2003). In other instances, DNA methylation appears to be the initiating signal, as cells with reduced DNA methylation also have reduced histone H3K9 methylation levels (Espada et al.,

2004). In summary H3K9 methylation and DNA methylation in mammalian cells are associated but which is the recruiting or initiating signal may be context dependent.

1.3.6 ATP dependent chromatin remodeling factors and DNA methylation

The link between DNA methylation and ATP dependent chromatin remodeling activity was first identified in plants. A genetic screen for *Arabidopsis* mutants that have decreased DNA methylation levels revealed a causative gene, called *ddm1*, that has homology to yeast SWI2/SNF2 protein (Jeddeloh et al., 1999; Vongs et al., 1993). *In vitro* evidence suggests that the DDM1 protein can actively utilize ATP to remodel assembled chromatin templates (Brzeski and Jerzmanowski, 2003), and DDM1 may function *in vivo* to maintain DNA methylation levels by allowing access of DNA methyltransferase enzymes to their substrates in chromatin.

In mammals, efficient DNA methylation also relies on SWI2/SNF2-like ATPases. The mammalian LSH and ATRX (alpha thalassemia/mental retardation syndrome X-linked) proteins both contain SWI2/SNF2-like ATPase domains, and are required to maintain normal DNA methylation levels (Dennis et al., 2001; Gibbons et al., 2000). LSH null embryos have global losses of DNA methylation, with some tissues having ~ 40 % total reduction in methyl cytosine content (Dennis et al., 2001). Loss of DNA methylation in ATRX patients is slight and localized in rDNA arrays, a Y-specific satellite, and subtelomeric repeats (Gibbons et al., 2000). Whether LSH is a functional ATPase remains to be determined, but ATRX has been demonstrated *in vitro* to contain ATPase and helicase activity (Tang et al., 2004; Xue et al., 2003). Another mammalian SWI2/SNF2 protein hSNF2H interacts with the *de novo* methyltransferase DNMT3b (Geiman et al., 2004), but its affects on DNA methylation *in vivo* have yet to be investigated. The exact molecular role that ATPase containing chromatin remodelling factors play in maintaining DNA methylation remains poorly characterized, but chromatin remodelling activities likely permit DNA methyltransferase enzymes access to CpG dinucleotides within a chromatin context.

1.4 Interpretation of DNA methylation

1.4.1 DNA methylation and silencing

In some instances, transcription factor binding can be disrupted by the presence of DNA methylation within its recognition site. If the transcription factor is an activator, interference by DNA methylation can act as an indirect and passive silencing mechanism (Watt and Molloy, 1988). Passive mechanisms of repression only partially contribute to the affect of DNA methylation on silencing. If methylated and non-methylated reporter vectors are injected into rodent cells, initially gene expression from both constructs is equal, but within 24 - 48 hours the methylated construct becomes completely repressed, suggesting an active silencing program exists (Buschhausen et al., 1985; Buschhausen et al., 1987). From these experiments it was hypothesized that chromatin may be required for DNA methylation dependent silencing and the lag period required for efficient silencing was due to the gradual addition of nucleosomes to the exogenous reporter DNA. Subsequent studies indicated that methylated reporter vectors pre-assembled into chromatin and then injection into living cells were more rapidly silenced than the equivalent naked DNA, indicating that chromatin was a major contributor to the silencing phenomena (Buschhausen et al., 1987). Addition of nucleosomes to an unmethylated reporter vector did not silence its expression to the same extent, so it was presumed that other nuclear factors may be recognizing the methylated reporter vector and contributing to the robust silencing observed *in vivo* (Buschhausen et al., 1987; Kass et al., 1997). Subsequent studies using methylated reporter genes in transient transfections and *in vitro* transcription assays defined an indirect role of DNA methylation in silencing transcription (Boyes and Bird, 1991). Silencing of methylated reporter vectors *in vitro* was dependent on the level a methyl-CpG binding activity (MeCP1)(Meehan et al., 1989)(Section 1.4.2), suggesting that a protein component that recognizes DNA methylation was important in mediating silencing.

1.4.2 Identification of activities that bind methyl-CpG

The first indication that proteins might exist that bind methyl-CpG *in vivo* came from restriction endonuclease accessibility assays in isolated cellular nuclei (Antequera et al., 1989). In these assays, intact nuclei were refractory to endonuclease digestion at

methyl-CpG, but the same DNA in protein free isolates was easily digested. Two activities in were isolated from nuclear extracts that could specifically recognize sequences containing methyl-CpG *in vitro*, but not the equivalent sequence when it was non-methylated (Lewis et al., 1992; Meehan et al., 1989). The protein identity of the first activity, MeCP1, was not identified until several years later (Feng and Zhang, 2001), but the second activity, MeCP2, was very abundant in brain extracts facilitating conventional biochemical purification. A single polypeptide corresponding to the MeCP2 activity was micro-sequenced by Edman degradation and a corresponding cDNA isolated (Lewis et al., 1992).

1.4.3 Isolation of a family of Methyl-CpG Binding proteins

Mapping studies using southwestern blot analysis allowed localization of a methyl-CpG binding domain (MBD) within MeCP2 (Nan et al., 1993). The protein sequence of the MBD was used to search cDNA databases and a second methyl binding protein was identified and named PCM1 (protein containing methyl-CpG binding domain 1)(Cross et al., 1997). PCM1 specifically recognized methylated-CpG, and was later renamed MBD1 (Figure 1.5)(Hendrich and Bird, 1998). Using a similar bioinformatic approach three more mammalian MBD containing proteins were isolated and named MBD2, MBD3, and MBD4 (Hendrich and Bird, 1998). Mammalian MBD2 and MBD4 both bind methylated DNA specifically, but MBD3 lacks methyl-CpG dependent DNA binding capacity (Hendrich and Bird, 1998).

1.4.4 Structure of the MBD

The basic fold of the MBD consists of a wedge shaped α/β sandwich as determined by NMR structures of both the MeCP2 and MBD1 MBD domains (Ohki et al., 1999; Wakefield et al., 1999). The three dimensional structure of the MBD is made up of a four strand anti-parallel beta sheet linked to a single short alpha helix (Figure 1.6). An NMR structure of the MBD domain of MBD1 in complex with methylated DNA has provided an atomic level resolution the DNA / protein interface (Ohki et al., 2001). The beta sheet region of the MBD is composed of two short outer strands and two long inner strands. A nine amino acid loop between the inner strands of the beta sheet becomes structured upon DNA binding, and together with inner beta strands

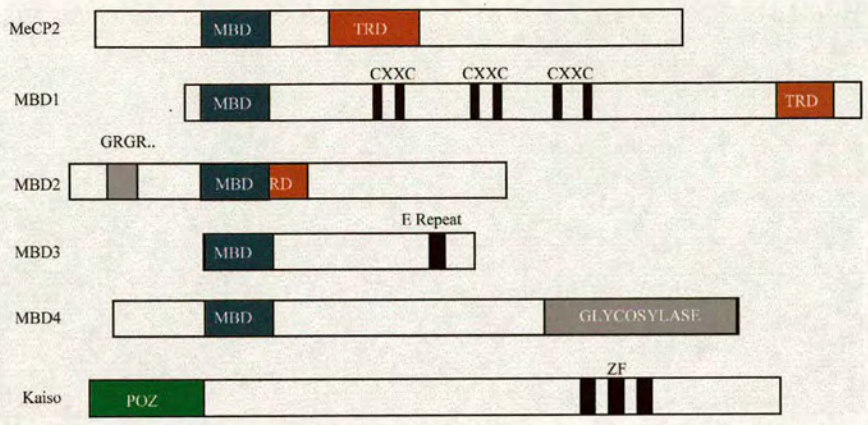


Figure 1.5- The methyl-CpG binding protein family

The DNA binding domains the methyl-CpG binding proteins are indicated by MBD, ZF, and CxxC domain. The MBD and ZF (Kaiso) are responsible for recognizing methylated DNA and the third CxxC domain of MBD1 binds non-methylated CpG. The transcriptional repression domains are in indicated by TRD (or POZ for Kaiso), and the DNA repair domain of MBD4 is indicated by its glycosylase domain

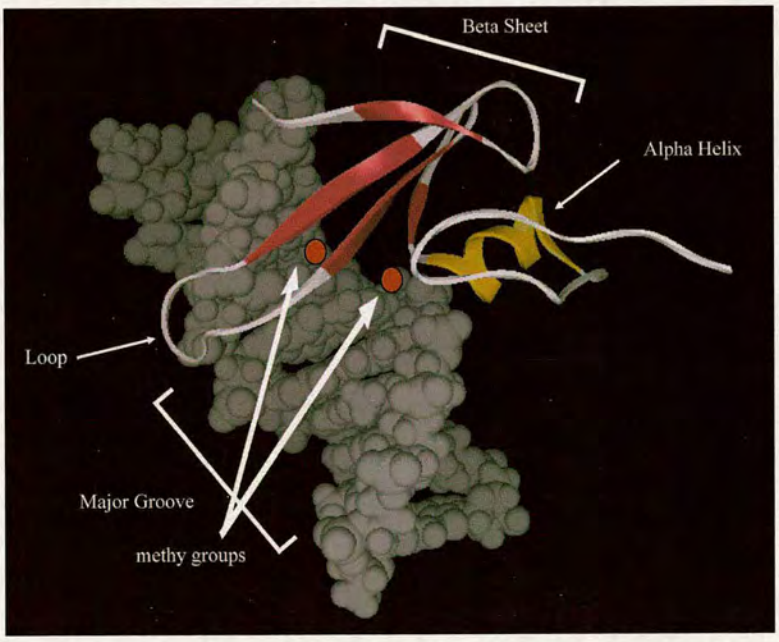


Figure 1.6- Structure of the MBD domain of MBD1 bound to methylated DNA (Adapted from (Ohki et al., 2001))

The long beta strands of the MBD bind deep in the major groove specifically recognizing one of the methylated bases. The loop joining to two long beta strands becomes structured upon DNA binding contributing to DNA recognition. The amino acids between the beta sheet and the alpha helix contribute to recognition of the second methyl-CpG while the alpha helix mostly contributes to DNA backbone contacts.

forms the bulk of the DNA biding face that lies in the major groove (Figure 1.6). The loop joining the beta sheet to the alpha helix projects into the major groove and makes important base contacts. The alpha helix mostly contributes contacts with the phosphate backbone (Figure 1.6). The methyl groups are directly recognized by hydrophobic amino acids in the beta sheet and also two residues in the loop between the beta sheet and the alpha helix.

1.4.5 The ancestral MBD protein

Bioinformatic analysis has successfully mapped the evolution of MBD containing proteins with the aid of newly available genome sequence information (Hendrich and Tweedie, 2003). MBD2 and MBD3 are likely the ancestral MBD proteins as they are the only vertebrate MBP'S that have homologues in invertebrate genomes (Hendrich and Tweedie, 2003). The MBD2/3 protein is encoded by a single gene in invertebrate genomes but diverges into two similar proteins in vertebrate genomes (Hendrich and Tweedie, 2003). The following sections will mostly deal with the MBP's found in vertebrates.

1.4.6 MeCP2

MeCP2 will be discussed in detail in the following section (Section 1.5)

1.4.7 MBD1

The MBD1 protein contains a N-terminal MBD and a potent transcriptional repression domain in the C-terminus (Figure 1.5). Between the MBD and TRD are several cysteine rich domains (CxxC) which potentially coordinate zinc ions. The *MBD1* gene gives rise to five splice variants in human cells (Fujita et al., 1999), and four splice variants in mouse cells (Jorgensen et al., 2004). MBD1 splice variants usually affect the number of CxxC domains included in the final protein product, or the sequence surrounding the TRD.

The MBD of MBD1 specifically binds methylated-CpG, but the CxxC3 domain is also a functional DNA binding domain (Jorgensen et al., 2004). The CxxC DNA binding domain was originally isolated in a screen for proteins that bind to non-methylated CpG dinucleotides (Voo et al., 2000). On its own the CxxC3 domain of

MBD1 has the capacity to bind non-methylated CpG dinucleotides, suggesting that certain MBD1 isoforms may be recruited to both methylated and non-methylated DNA sequences (Jorgensen et al., 2004). Interestingly, the CxxC3 domain of MBD1 shares homology with the CxxC domains of two other factors involved in chromatin methylation: HRX (a histone methyltransferase) and DNMT1 (a DNA methyltransferase)(Ayton et al., 2004).

The transcriptional repression domain of MBD1 is a very short (32 amino acid sequence) in the C-terminus of the protein that efficiently represses transcription at distance of up to 2 kilo-bases from an active promoter in a histone deacetylase dependent manner (Ng et al., 2000). The repression domain of MBD1 interacts with a co-repressor molecule called MCAF1 (Fujita et al., 2003b), and MBD1 actively recruits MCAF1 to methylated promoters *in vivo*. The MBD of MBD1 also interacts with CAF1 providing an indirect link between MBD1 and the CAF1 associated protein HP1 (Fujita et al., 2003c). The interaction of HP1 with MBD1/CAF1 might account for the observed recruitment of the HP1 protein partner, Suv39H1, to some MBD1 bound loci (Fujita et al., 2003c). A recent report has suggested that MBD1 interacts with methylpurine–DNA glycosylase activity to repress transcription, but the relevance of this interaction remains to be demonstrated (Watanabe et al., 2003). A study published last year has highlighted a novel role for MBD1 in copying histone H3K9 methylation marks to regions of DNA methylation during DNA replication (Sarraf and Stancheva, 2004). In this study MBD1 associated with histone H3K9 methyltransferase activity, and a yeast two-hybrid study mapped a direct interaction between a c-terminal region of MBD1 and the histone methyltransferase SETDB1. MBD1 and SETDB1 can form a complex with previously identified co-factor CAF1 during S-phase of the cell cycle, and this interaction is essential for deposition of H3K9me marks at MBD1 target loci. This interesting role of MBD1 during DNA replication indicates that eukaryotic MBD proteins may have diverse roles in interpreting the methyl-CpG signal in the mammalian cell.

An *Mbd1* null mouse has been generated and displays a very mild neurological phenotype (Zhao et al., 2003), but is viable, healthy, and fertile. Adult neuronal stem cells from *Mbd1* null mice also display reduced capacity to differentiate and have

genomic instability. The null mouse brain has reduced neurogenisis and the mice exhibited decreased spatial learning. The subtle defects in *Mbd1* null mice suggest a redundant role for MBD proteins in cellular function, or a specialized role for MBD1 under certain physiological circumstances not yet examined.

1.4.8 MBD2 (and MBD3)

The *Drosophila* MBD2/3 protein interacts with methylated cytosine in the context of a CpA dinucleotide (Marhold et al., 2004b). In vertebrates, the MBD2/3 protein is duplicated resulting in separate MBD2 and MBD3 proteins that share significant sequence similarity (Figure 1.5). In *Xenopus laevis* both MBD2 and MBD3 bind specifically to methyl-CpG dinucleotides (Wade et al., 1999), but mammalian MBD3 has an amino acid substitution that abolishes methyl-CpG specific binding (Hendrich and Bird, 1998). Mammalian MBD2 is a functional methyl-binding protein which specifically recognizes methyl-CpG (Hendrich and Bird, 1998). The repression domain of MBD2 overlaps with the MBD, and a GR_n repeat in the N-terminus has been reported to interact with RNA (Jeffery and Nakielny, 2004).

MBD2 is a potent HDAC dependent transcriptional repressor (Ng et al., 1999; Wade et al., 1999). Bioinformatic identification and subsequent characterization of MBD2 lead to the discovery that MBD2 is the methyl-CpG binding component of the MeCP1 activity observed in nuclear extracts (Ng et al., 1999; Zhang et al., 1999). Purification of MeCP1 from both *Xenopus laevis* and mammalian tissue culture cells showed that MBD2 and MBD3 stably associate with the NURD chromatin remodeling complex (Feng and Zhang, 2001; Wade et al., 1999). The core mammalian MeCP1 complex consists of; MBD2, MBD3, HDAC 1/2, MTA2, Mi-2, RbAP46/48, and p66/68. This large molecular weight complex preferentially deacetylates histones (via HDAC1/2) and remodels nucleosomes (via the Mi-2 ATPase) on methylated DNA (Feng and Zhang, 2001; Guschin et al., 2000). The p66/68 components of MeCP1 are involved in targeting MeCP1 and RbAP46/48 are likely a structural components (Brackertz et al., 2002; Feng et al., 2002). The exact role of MTA2 in MeCP1 remains unknown (Bowen et al., 2004). Interestingly the association of MBD2/3 with components of the NURD complex is conserved in

invertebrates, as the *Drosophila* MBD2/3 proteins co-fractionate with several factors in the *Drosophila* NURD complex (Ballestar et al., 2001; Barlow et al., 2001; Marhold et al., 2004a). In the majority of cases MBD2 has been associated with transcriptional repression, but one report suggests that MBD2 can potentiate gene expression though protein / protein interaction with RNA helicase A (RHA)(Fujita et al., 2003a). It will be interesting to determine what role MBD2 has in transcriptional activation.

Mbd3 null mice are embryonic lethal, but *Mbd2* null mice are fertile, viable, and healthy (Hendrich et al., 2001). Deletion of *Mbd3* in mice is presumably lethal due to disruption of the NURD co-repressor complex which has broad cellular roles in transcriptional repression (Bowen et al., 2004). Close examination of the *Mbd2* null mice has revealed subtle defects in maternal behavior, with *Mbd2* null mothers failing to care properly for their offspring (Hendrich et al., 2001). The reason for this behavior defect in the *Mbd2* null mice remains undiscovered. Because DNA methylation plays such an important role in polyp formation in the $Apc^{(min/+)}$ mouse model of colorectal cancer (introduced in Section 1.2.9), it was hypothesized that MBD2 might contribute to this affect by interpreting the methyl-CpG modification. *Mbd2* null mice were crossed onto the $Apc^{(Min/+)}$ background and a strong genetic interaction was observed, with MBD2 absence inhibiting polyp formation in a dose dependent manner (Sansom et al., 2003). The role of MBD2 in promoting polyp formation in the $Apc^{(min/+)}$ model is presumably related to its role in binding the promoter region of hypermethylated tumor suppressor genes, repressing transcription, and promoting tumorogenisis (introduced in Section 1.2.9). In agreement with its proposed role in malignancy in the intestine of the $Apc^{(min/+)}$ mice, MBD2 is found at the methylated promoter of the *p16* gene in cancer cells and contributes to its silencing (Magdinier and Wolffe, 2001). *Mbd2* null mice are also deficient in the normal cytokine gene expression with a stochastic relationship existing between the levels of MBD2 and expression of the *Il-4* gene in TH1 lineage T cells. Normally MBD2 inhibits *Il-4* expression by out-competing the activator GATA3 for promoter occupancy. MBD2 binds directly to the *Il-4* promoter to repress transcription and in its absence, GATA3 occupancy at *Il-4* promoter causes

ectopic gene expression (Hutchins et al., 2002). In addition to the gene expression defect in the null mouse, a more general defect in methylation dependent repression exists in MBD2 deficient cells as repression of methylated reporter vectors is partially alleviated compared to equivalent wild type cells (Hendrich et al., 2001). In depth analysis of the MBD2 null mice continues to reveal subtle defects in methylation dependent silencing (unpublished data Bird lab) suggesting that MBD2 has wide range of affects on normal methyl-CpG dependent repression.

1.4.9 MBD4

MBD4 is the only MBD protein which is not apparently associated with transcriptional repression, and instead MBD4 has a role in DNA repair. The DNA repair activity of MBD4 lies at the C-terminus of the protein, in a domain that has homology to DNA glycosylase repair proteins (Figure 1.5). In *in vitro* studies, the primary enzymatic activity of MBD4 is towards removal thymine or uracil within a mismatched CpG site (Bellacosa et al., 1999; Hendrich et al., 1999). This is an important repair process under normal cellular conditions due to the fact that methyl-cytosine can undergo spontaneous deamination resulting in a mismatched T:G base pairing within CpG (introduced in Section 1.2.4). In the case of methyl-cytosine deamination, if this mismatch is not repaired before DNA replication, the daughter strand can replicate normally giving rise to a cytosine to thymine transition mutation. A lack of complete repair of methyl-cytosine deamination likely contributes to the hypermutability of methyl-CpG bases in the mammalian genome (Duncan and Miller, 1980). In fact, cytosine to thymine transitions are the most frequent human disease causing point mutations, highlighting the importance of MBD4 function (Cooper and Youssoufian, 1988). The role of the MBD4 MBD in DNA repair is not entirely clear, but one possibility is that the MBD targets the protein to regions of methylated DNA where deamination events are most frequent. Another possibility is that the MBD binds deamination mismatches and hands the mismatched site to the glycosylase domain, which can then initiate the repair process.

MBD4 null mice are viable, healthy, and fertile, but show a three fold increase in the rate of CpG to TpG mutations (Millar et al., 2002; Wong et al., 2002). This increase

in mutation frequency is biologically relevant as crossing the MBD4 null mice onto the $Apc^{(Min/+)}$ mouse background resulted in an increase in tumor burden and reduction in life expectancy compared to normal littermates (Millar et al., 2002). Furthermore, sequencing of the *Apc* gene in the tumors of the MBD4 null mice revealed that loss of the second functional *Apc* allele was in large part due to CpG to TpG mutations that caused APC protein truncation. The affects of MBD4 seen in the mouse model of colorectal cancer may manifest in humans patients as subset of sporadic colon cancer patients have MBD4 truncation mutations, but whether this is a causative event in human tumorgenisis remains to be determined (Bader et al., 1999; Riccio et al., 1999; Yamada et al., 2002).

1.4.10 Kaiso

Kaiso is a methyl-CpG binding protein that utilizes zinc fingers instead of a MBD to recognize methyl-CpG (Prokhortchouk et al., 2001). In contrast to other MBP's that bind a symmetrical methyl-CpG, Kaiso recognizes methyl-CpG in the context of a CGCG. The Kaiso zinc fingers encompass a bi-modal DNA binding domain that also recognizes sequences which do not consist of methyl-CpG (Daniel et al., 2002). Kaiso is a transcriptional repressor, and the repression domain maps to the n-terminal POZ domain (Prokhortchouk et al., 2001). Kaiso was originally identified as a protein partner of the signaling molecule p120 catenin, suggesting Kaiso might respond to signals originating from outside of the cell (Prokhortchouk et al., 2001). Translocation of p120 catenin to the nucleus inhibits repression mediated by Kaiso, suggesting this may compose a biologically relevant signaling pathway (Kelly et al., 2004; Kim et al., 2004). Repression by Kaiso is mediated through its stable association with the NCoR complex (Yoon et al., 2003) which contains several repressor proteins, including a histone deacetylase enzyme (HDAC3). In HeLa cells the Kaiso/NCoR complex was required to repress the methylated MTA2 gene. Kaiso also represses methylated genes in *Xenopus laevis*, and when knocked down during development embryos prematurely activate a set of genes prior to the mid blastula transition (Ruzov et al., 2004). Interestingly, depletion of *Xenopus* Kaiso in developing embryos phenocopies depletion of DNMT1 (Stancheva and Meehan,

2000), suggesting a close link between Kaiso and methylation mediated repression in the *Xenopus* early embryo.

1.5 MeCP2

1.5.1 Recognition of Methyl-CpG

Full length MeCP2 specifically binds methylated DNA in southwestern blot assays, which provided a means by which to identify and purify the protein in rat brain nuclear extract (Lewis et al., 1992). Recombinant MeCP2 produced in bacteria retained the methyl-CpG dependent binding activity, allowing localization of the domain responsible for methyl-CpG binding (Nan et al., 1993). Analysis of the methyl-binding domain (MBD) demonstrated that MeCP2 has high affinity for methylated CpG dinucleotides (~10^{-9} M), covers roughly 12 base pairs by DNAseI footprint analysis, and interacts with the major groove of DNA (Nan et al., 1993). Binding of MeCP2 to DNA containing more than one methyl-CpG is not co-operative and therefore MeCP2 does not likely self association on DNA. Recognition of methyl-CpG in chromatin is uninhibited by the presence of nucleosomes and MeCP2 retains the capacity to interact with the methyl-CpG containing major groove DNA exposed on the surface of a nucleosome (Chandler et al., 1999). In mini-chromosome binding assays MeCP2 specifically binds methylated chromatin while ignoring the corresponding unmethylated template (Nan et al., 1997a).

1.5.2 Transcriptional repression and protein partners

MeCP2 can efficiently repress transcription from methylated promoters by *in vitro* transcription assays (Jones et al., 1998; Nan et al., 1997a). Transcriptional repression *in vitro* is mediated through MeCP2 association with the basal transcription factor TFIIB, and likely functions through interrupting the initiation stages of transcription (Kaludov and Wolffe, 2000). The MeCP2 repression domain (TRD) was mapped by reporter assays *in vivo* and consists of a roughly 100 amino long sequence just downstream of the MBD (Jones et al., 1998; Nan et al., 1997a). The TRD domain also encompasses a nuclear localization domain which is required for normal sub-cellular targeting of MeCP2 (Nan et al., 1996). Repression by MeCP2 is not only limited to promoter proximal regions as effective repression occurs at a distance of

up to 2 kilo-bases from the start site of transcription (Nan et al., 1998). In addition to direct affects on transcriptional initiation MeCP2 can also actively repress transcription in a histone deacetylase dependent fashion. HDAC sensitive repression by MeCP2 has been attributed to its association with the mammalian Sin3a co-repressor complex (Jones et al., 1998; Nan et al., 1998). The Sin3 co-repressor complex is a very large multi-protein complex, which is targeted to genomic loci through transient interactions with transcriptional repressor proteins (Silverstein and Ekwall, 2005). Core components of the mammalian Sin3a co-repressor complex include and Sin3a, HDAC 1/2, and RbAP46/48 (Silverstein and Ekwall, 2005). Most Sin3a complexes share the five core components listed above, but various other proteins stably associate with the Sin3a complex to form a multitude of sub-complexes (Kuzmichev et al., 2002). Sin3a complexes use their associated histone deacetylase activity to remove active acetyl marks from lysine residues and remodel chromatin contributing to the silencing of gene expression. MeCP2 also associates with H3K9 methyltransferase activity, complementing the removal of acetylation marks with the active deposition of repressive histone methylation marks (Fuks et al., 2003b). In addition to association with the Sin3a chromatin remodeling complex MeCP2 interacts with various other factors including; Suv39H1, c-Ski, DNMT1, Co-Rest, LANA, PU1, Splicing factors, Brm, and RNA (Buschdorf and Stratling, 2004; Harikrishnan et al., 2005; Jeffery and Nakielny, 2004; Kimura and Shiota, 2003; Kokura et al., 2001; Krithivas et al., 2002; Lunyak et al., 2002; Suzuki et al., 2003). The relative contribution of these other factors to the normal function of MeCP2 remain to be investigated.

1.5.3 Expression, evolution, and conservation

The MeCP2 protein is expressed from one locus, but differential splicing gives rise to two isoforms (MeCP2 α and β) that have unique amino acid composition at the N-terminus (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004). MeCP2 (α) is the most highly expressed of the two isoforms in most mouse tissues, but both isoforms are more equally expressed in humans (Mnatzakanian et al., 2004). The role of alternative MeCP2 splice variants in normal MeCP2 function remains unknown. MeCP2 is ubiquitously expressed throughout human and mouse tissues but is most

abundant in the brain. The human MeCP2 mRNA codes for a highly basic 486 amino acid protein, which is evolutionarily conserved having homologues in fish, amphibians, birds and mammals (Figure 1.5). In all cases, the protein domain arrangement of MeCP2 is similar, with the MBD and TRD clustering towards the N-terminus of the protein and a long C-terminal region of unknown function (Figure 1.5 and Figure 1.7)). The MBD and TRD domains function in both mammals and amphibians to bind methylated DNA and repress transcription respectively. The conservation of MeCP2 function is exemplified by the observation that human MeCP2 supports embryonic development in *Xenopus laevis* when the endogenous *Xenopus* MeCP2 protein is depleted (Stancheva et al., 2003).

1.5.4 MeCP2 and Rett Syndrome

Rett syndrome is a relatively frequent form of severe neurological disease that occurs approximately one in every 10000-22000 female births (Kriaucionis and Bird, 2003). Rett syndrome is characterized by a normal period of development during the first year life, after which a rapid regression in development occurs with patients exhibiting; loss of acquired speech and motor skills, microcephaly, seizures, autism, ataxia, intermittent hyperventilation and stereotypic hand movements (reviewed in, (Kriaucionis and Bird, 2003) Many patients survive into adulthood, but few procreate.

Genetic linkage analysis revealed a region on Xq28 that was responsible for a rare familial case of Rett syndrome (Amir et al., 2000). Sequencing of the *MECP2* gene, on the Xq28 region, identified a mutation that segregated with the disease (Amir et al., 1999). Rett syndrome is a dominant disease and usually arises from a spontaneous mutation in the *MeCP2* gene. Large scale screening programs have revealed numerous types of Rett syndrome mutations in all regions of the MeCP2 protein (Figure 1.7). Interestingly, there is not an extensive genotype / phenotype correlation between the type of MeCP2 mutation and disease prognosis. Some mutations that occur in the MBD and that affect DNA binding are phenotypically similar to mutations found outside of mapped functional domains (Kriaucionis and Bird, 2003). Molecular studies are slowly beginning to expand the understanding of

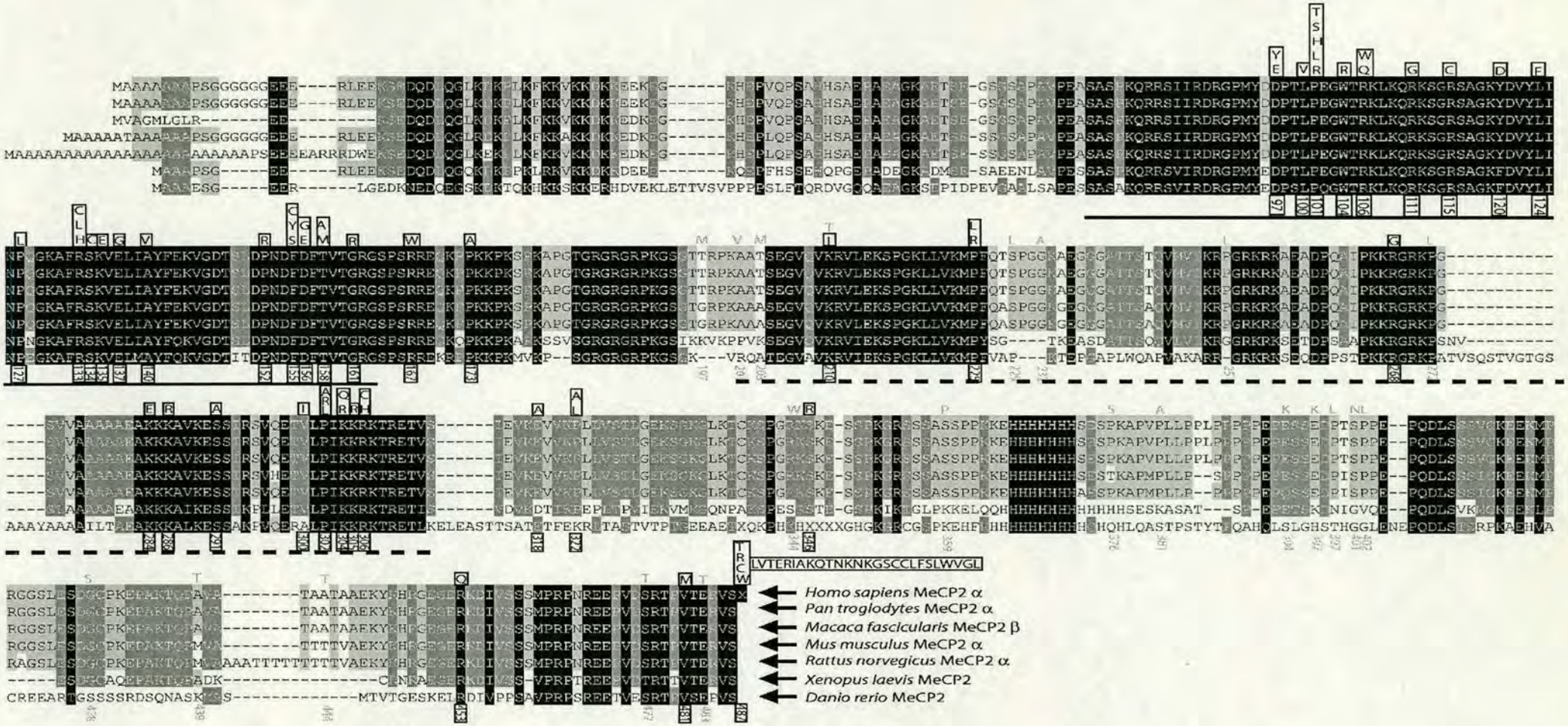


Figure 1.7- Alignment of MeCP2 proteins from various organisms (This alignment was a gift from Skirmantas Kriaucionis)

Where available, the amino acid sequence of the major isoform of MeCP2 (α) was used in a cross species alignment of amino acid sequence. The solid black line under the alignment indicates the MBD domain and the dotted black line the transcriptional repression domain. Rett syndrome mutations are indicated above the alignment as black text with a black box. Non-disease causing polymorphisms in MeCP2 are indicated in grey. The amino acid number corresponding to the human sequence is indicated below the marked Rett syndrome mutations and normal polymorphisms.

how Rett mutations affect MeCP2 function (Free et al., 2001; Kudo et al., 2001; Kudo et al., 2003; Yusufzai and Wolffe, 2000).

1.5.5 MeCP2 null mouse

MeCP2 null male ES cells were unable to contribute to germ line transmission, and it was originally believed that loss of MeCP2 in mice was lethal (Tate et al., 1996). Subsequently, using a conditional knockout allele strategy, MeCP2 null animals were achieved by two groups independently (Chen et al., 2001; Guy et al., 2001), and the original conclusion regarding embryonic lethality was revised. No phenotype was observed in MeCP2 null animals until about 3 to 8 weeks of age, at which point they developed stiff uncoordinated gait, and a loss of spontaneous movement. Most animals eventually developed hind limb clasping, irregular breathing, and on average died at 54 days after birth (Chen et al., 2001; Guy et al., 2001). Female mice heterozygous for loss of MeCP2 displayed a similar phenotype except that the onset of symptoms occurred between 3 to 9 months and these mice were able to raise normal litters (Chen et al., 2001; Guy et al., 2001). The similarity in time of disease onset and symptoms between human Rett patients and the MeCP2 heterozygous mice, makes this a reasonable genetic model to study Rett syndrome. Another male *Mecp2* allele resulting in truncation of the MeCP2 protein at amino acid 308 displays a similar phenotype to the null animals, except that the onset of symptoms occurs later (Shahbazian et al., 2002).

1.5.6 MeCP2 depleted *Xenopus* embryo

Xenopus laevis is tractable experimental system to study early development. Depletion of xMeCP2 in the in the developing *Xenopus* embryo resulted in a cessation of development at the neurula stage (Stancheva et al., 2003). Development could be rescued by re-expressing human MeCP2, suggesting a significant amount of functional overlap between MeCP2 orthologues. Interestingly, re-introduction of a mutant form of MeCP2 (R168X) found in human Rett syndrome patients resulted in a partial rescue of development in the MeCP2 depleted embryo. The mutant embryos exhibited a bent morphology and hyperactive spastic behavior (Stancheva et al., 2003). In contrast to the normal early developmental pattern in the MeCP2 null mice, differences appear to exist in function of the amphibian protein as MeCP2 deficiency results in a clear early developmental defect.

1.5.7 MeCP2 target genes

A series of studies have analyzed the function of MeCP2 in regulating specific genes in cancer cell lines, but few MeCP2-regulated genes have been identified in normal cells. Micro-array analysis comparing the gene expression pattern in normal wild type and MeCP2 null mouse cells has revealed very few changes in gene expression (Tudor et al., 2002). Similar gene expression analysis by micro-array using cell lines or tissue from human MeCP2 patients have given variable results (Ballestar et al., 2005; Colantuoni et al., 2001; Traynor et al., 2002). The surprising observation from all the large scale gene expression studies using MeCP2 null tissue is the apparent lack of mis-regulated genes. Given the wide genomic distribution of methyl-CpG, it was predicted that MeCP2 would have pleotrophic and global affects on gene expression, but this is not the case.

Recently, the mouse *Bdnf* gene was identified as a direct target for MeCP2-mediated repression in mouse neurons (Chen et al., 2003b; Martinowich et al., 2003). MeCP2 binds the promoter proximal region of the gene and maintains silencing of *Bdnf* in isolated mouse brain neurons. Activation of the *Bdnf* gene through calcium signaling pathways results in MeCP2 phosphorylation and subsequent loss of binding to the *Bdnf* promoter. Because MeCP2 may function to keep *Bdnf* silent in the absence of activators, the level of *Bdnf* re-expression observed in the MeCP2 null cells was small. This type of gene expression change would not have been identified in genome wide micro-array from tissues with a complex cellular make-up. I will be important to look at gene expression patterns in isolated cell types from the brain to identify genes, like *Bdnf*, that are specifically regulated by MeCP2.

Xenopus laevis embryos depleted of xMeCP2 fail to develop past the neurula stage. Developmental defects in the depleted embryos appear to stem from inappropriate silencing of the transcriptional regulator Hairy2a (Stancheva et al., 2003). In MeCP2 depleted embryos enhanced expression of the *Hairy2a* gene results in a decrease in the number of neuronal precursor cells, which likely contributes to the phenotypic cessation of growth during neurulation. MeCP2 works in a similar way at the *Hairy2a* gene as it does at the *Bdnf* gene. MeCP2 binds the promoter proximal region of the repressed Hairy2a gene and maintains the silenced state. Upon developmental activation of Hairy2a, MeCP2 leaves the promoter and gene activation can proceed

unimpeded. It appears from studies in both mice and frogs that MeCP2 may have a role in silencing inducible genes. It will be interesting to establish whether other inducible neurological genes are regulated in this manner by MeCP2.

Initial studies in cells from Rett patients revealed no changes in the expression of imprinted genes (Balmer et al., 2002). Recently mis-regulation of the imprinted *DLX5/6* (Horike et al., 2005) and *UBE3A* (Makedonski et al., 2005; Samaco et al., 2005) genes in MeCP2 patients and mice was observed. In these studies, other imprinted genes were not mis-regulated, suggesting *DLX5/6* (Horike et al., 2005) and *UBE3A* may be unique target genes for MeCP2. A more thorough analysis of the growing list of imprinted genes will be required to understand the contribution of MeCP2 to normal imprinted gene expression.

1.6 Aims of this thesis

The flurry of interest in MBD proteins over the last decade has revealed many new and exciting discoveries regarding the diverse properties of methyl-CpG binding proteins. To try and understand in more detail the function of MeCP2 the work described in this thesis uses a biochemical approach to address three basic questions regarding MeCP2 function; (1) what are the native biochemical properties of MeCP2? (2) what specific DNA sequences does MeCP2 bind? and (3) what are the affects of post-translational modification on MeCP2?

Chapters 3 and 4 address question (1) though biochemical purification of native MeCP2 from rat brain and recombinant sources. This investigation clarifies the nature of the biochemical association between MeCP2 and Sin3a and also uncovers some interesting biophysical properties of MeCP2. Chapter 5 addresses question (2) and uncovers an unexpected DNA binding specificity of MeCP2. These data suggest that MeCP2 is targeted to specific methyl-CpG sites *in vivo*. Chapter 6 addresses question (3) and uncovers a biochemical activity that is capable of phosphorylating MeCP2. Together these biochemical studies reveal new attributes of the MeCP2 molecule which may be relevant to normal function *in vivo* and Rett syndrome.

2. Chapter two-
Materials and Methods

2.1 Common solutions

Phosphate buffered saline (PBS) - 140 mM NaCl, 3 mM KCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4.

Tris-glycine SDS (TGS) (1 X) - 25 mM Tris, 250 mM Glycine, 0.1 % SDS.

Tris-buffered saline (TBS) (1 X) - 50 mM Tris-HCl pH 8.0, 150 mM NaCl

Tris-acetate EDTA (TAE) (1 X) - 40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA and pH adjusted to 8.0.

Tris-borate EDTA (TBE)(1X)- 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA and pH adjusted to 8.0.

2.2 Bacterial methods

2.2.1 Media

Luria-Bertani (L.B.) Media – 10 g / l Bacto tryptone, 5 g / l Yeast extract, 10 g / l NaCl, and adjusted to pH 7.0. L.B. was then autoclaved and stored at room temperature until required.

L.B. Agar – Formulated as per L.B. except that 15 g / l of bacto-agar was included. L.B. agar was autoclaved and stored at room temperature until required.

2.2.2 Antibiotics

Ampicillin – Ampicillin was diluted in distilled water to a stock concentration of 50 mg / ml, sterilized through a 0.2 μm filter, and stored at – 20 °C . Ampicillin was thawed and added to LB or LB agar at a working concentration of 50 μg / ml.

Kanamycin – Kanamycin was diluted in distilled water to a stock concentration of 50 mg / ml, sterilized through a 0.2 μm filter, and stored at – 20 °C. Kanamycin was thawed and added to LB or LB agar at a working concentration of 50 μg / ml.

Gentamicin – Gentamicin was diluted in distilled water to a stock concentration of 10 mg / ml, sterilized through a 0.2 μm filter, and stored at – 20 °C. Gentamicin was thawed and added to LB or LB agar at a working concentration of 7 μg / ml.

Chloramphenicol – Chloramphenicol was diluted in ethanol to a stock concentration of 34 mg / ml and stored at -20 °C. Chloramphenicol was added to LB or LB agar at a working concentration of 34 μg / ml.

Tetracycline - Tetracycline was diluted in ethanol to a stock concentration of 10 mg / ml and stored at -20 °C. Tetracycline was added to LB or LB agar at a working concentration of 10 μg / ml.

2.2.3 Culture on plates

L.B. agar was taken either immediately out of the autoclave or re-dissolved in a microwave and allowed to cool to 50 °C before antibiotic was added. The L.B. agar mix was poured into round 10 cm dishes and allowed to cool at room temperature until solidified. The plates were then stored inverted at 4 °C. Prior to use plates were warmed to 37 °C. Bacteria were grown by streaking a single colony on the plate, or by diluting liquid culture and spreading evenly over the surface of the plate using a sterile bent glass rod. Plates were inverted and incubated overnight at 37 °C.

2.2.4 Liquid culture

Antibiotic was added to L.B. media and mixed thoroughly. For small cultures, a single bacterial colony was scraped from the plate using a Gilson pipette tip and dropped into a 20 ml falcon tube. The culture was grown shaking at 250 rpm overnight at 37 °C. For larger cultures small overnight cultures were grown as above and the following afternoon diluted 1:500 – 1:1000 in fresh L.B. containing antibiotics and grown at 250 rpm overnight at 37 °C.

2.2.5 Preparing chemically competent strains

Bacterial cells were streaked from a glycerol stock onto LB agar plates and incubated overnight at 37 °C. A single colony was inoculated in 5 ml of LB overnight shaking at 37 °C. The 5 ml culture was then inoculated into a 500 ml flask of LB and incubated shaking at 37 °C until cell density reached an absorbance ob 0.5 at 650 nm. The cells were collected by centrifugation at 2400 rpm 15 min 4 °C and placed on ice

for 10 min. Using sterile technique the cells were resuspend in 165 ml of sterile filtered buffer A (100 mM RbCl, 50 mM MnCl2, 30 mM K-Acetate, 10 mM CaCl2, 15 % glycerol) pH 5.8. Cells were incubated on ice 45 min, and then recovered by centrifugation at 2400 rpm for 15 min 4 °C. The cells were resuspended in 40 ml of sterile filtered buffer B (10 mM MOPS, 10 mM RbCl2, 75 mM CaCl2, 15 % glycerol) pH 6.8 and incubated on ice for 15 min. Aliquots of competent cells (200 µL) were snap frozen in liquid nitrogen and stored at – 80 °C. Competency was evaluated by transforming 0 ng, 10 ng, or 100 ng of super-coiled plasmid DNA into aliquots of the newly made competent cell stock, and tranformants per µg derived by counting viable plasmid selected colonies arising the next day on LB agar plates. Competent cells were typically 10^6 to 10^8 colonies per µg of plasmid DNA.

2.2.6 Transformation

Competent cells were defrosted on ice and DNA was swirled into the competent cell mix with at pipette tip. The transformation reaction was then allowed to rest on ice for 20 min. The cells were heat shocked for 1 min and 30 s at 42 °C in a water bath, immediately followed by at 2 minute resting period on ice. 800 µl L.B. was then added to the transformation mixture and it was incubated at 37 °C shaking for 1 hr. If the DNA was a ligation reaction the cells were collected at 6000 g for 1 min and then re-suspended in 150 µl of LB before plating on selective LB agar. If the transformation DNA was supercoiled plasmid 35 µl of the transformation reaction was directly plated on selective LB agar.

2.3 Isolation of exogenous DNA from bacteria

2.3.1 Small scale plasmid DNA preparation

Analytical digest quality DNA was isolated by growing a single colony overnight in 5 ml L.B. containing antibiotic selection. 1 ml was transferred a 1.5 ml Eppendorf tube and the remainder saved at 4 °C for subsequent sequencing grade DNA preparation. The 1 ml sample was centrifuged at room temperature for 40 s, and the supernatant removed and discarded. The cell pellet was resuspended in 100 µl of STET (8 % Sucrose, 5 % Triton x-100, 50 mM Tris-HCl pH 8.0, 50 mM EDTA) supplemented with 25 µl of 10 mg / ml lysozyme. Stock Lysozyme was prepared in 10 mM Tris-HCl pH 8.0 with 1mM EDTA and stored at – 20 °C. The sample was boiled for 45 s,

and then centrifuged for 15 min at 13000 rpm. The cell debris was removed with a sterile tooth pick and discarded. 125 µl of isopropanol was added to the supernatant to precipitate the DNA, and the DNA was collected by centrifugation at room temperature for 10 min. The supernatant was removed and discarded, followed by resuspension of the DNA pellet in a 25 µl volume of Tris-HCl pH 8.0 supplemented with 80-100 µg / ml of DNAse free pancreatic RNAse.

2.3.2 Sequencing grade plasmid DNA preparation

The remaining bacterial culture from the analytical DNA preparation (Section 2.3.1) was used to purify plasmid DNA for sequencing. Sequencing grade DNA was made using the Qiagen mini preparation DNA kit according to the manufactures recommendations.

2.3.3 Large scale plasmid DNA preparation

Large scale preparation of plasmid DNA was done according to the Qiagen Maxi Prep Kit instruction manual. The precipitated DNA was resuspended in 10 mM Tris pH 8.0 and stored at -20 °C.

2.3.4 Large scale preparation of Bacmid DNA

Large scale preparation of Bacmid DNA was done according to the Qiagen Midi Prep Kit with the exception that Bacmid DNA was eluted from the resin by first heating the elution buffer to 65 °C. The Bacmid DNA pellet was resuspended in 10 mM Tris pH 8.0 and stored at - 20 °C.

2.4 Manipulation of DNA and cloning

2.4.1 Restriction endonuclease digestion of DNA

Restriction digest were generally done in 20 µl or 50 µl volumes in the appropriate buffers supplied with restriction enzyme (NEB). The reactions were supplemented with 100 ng / µl bovine serum albumin (BSA) if the product literature recommending its addition. Most digests were carried out at 37 °C overnight, or at the temperature given in the product literature. Efficiency and completeness of digestion was verified by agarose gel electrophoresis in 1 X TAE or TBE and visualization by ethidium bromide staining.

2.4.2 Dephosphorylation of DNA fragments

The 5' phosphate of DNA molecules was removed with calf intestinal alkaline phosphatase (CIP) (NEB). DNA was diluted to 50 ng / µl in NEB buffer 3 and 0.5 units of CIP per µg of total DNA was added to the reaction. The reaction was allowed to proceed for 1 hr at 37 °C and the DNA re-purified on a gel extraction column (Qiagen).

2.4.3 Methylation of DNA fragments

Methylation reactions using M.*HpaII*, M.*HhaI*, or M.*SssI* methyltransferases (NEB) were carried out according to manufactures instructions, with the exception that methylation reactions were carried out overnight at 37 °C follow by the addition of fresh enzyme and SAM in the morning followed by an additional 1 – 2 hour incubation time.

2.4.4 Oligonucleotides

All oligonucleotides were supplied by Sigma-Genosys and purified by desalting or poly-acrylamide gel electrophoresis. Stock oligonucleotides were diluted in H_2O to 1 ug / µl and stored at -20 C.

2.4.5 Polymerase chain reaction

Polymerase chain reactions were carried out in 20 µl or 50 µl reaction volumes generally with 50 ng of each primer per reaction. If template DNA was plasmid 5 ng of DNA was used and if genomic DNA 50 – 100 ng was used. Cycling parameters varied depending on DNA source and primer properties. If the PCR products were subsequently to be cloned for protein expression PFU polymerase (Stratagene) was used to amplify the DNA. If the PCR was for analytical purposed Red Hot Taq polymerase (Abgene) was used. In each case the reaction buffer and salt solution provided with these commercial polymerases were used.

2.4.6 Purification of PCR fragments

PCR fragments were purified with the Qiagen PCR purification kit according to the manufacturer recommendations.

2.4.7 Sequencing

Sequencing reactions were assembled using the Big Dye terminator V3 kit (Roche). Reactions were assembled in 10 µl volumes containing 2 µl of Big Dye sequencing mix, 20 ng of sequencing primer, and 200 - 400 ng of template DNA. The sequencing reactions were done using the following program on an MJ thermo-cycler : 96 °C for 1 minute followed by 25 cycles at 96 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 4 min. Sequences were analyzed on an ABI 3730 sequencer by the ICAPB sequencing facility at the University of Edinburgh.

2.4.8 Agarose gel extraction of DNA fragments

DNA was extracted from agarose gels using the Qiagen gel extraction kit according to the manufacturer's instructions and the DNA eluted in 30 µl.

2.4.9 Ligation

Ligation reactions were carried out in a 10 µl reaction volume using 6 Weiss units of T4 DNA ligase (NEB) according to the manufactures recommendations.

2.4.10 Site directed mutagenesis

Site directed mutagenesis was carried out with the QuikchangeXL mutagenesis kit (Stratagene) according to the manufactures recommendations. All mutagenesis primers were PAGE purified by Sigma-Genosys.

2.4.11 Radioactively labeling DNA fragments

50 – 100 ng of template DNA was end labeled in a 30 µl reaction volume with 20 units of T4 polynucleotide kinase (NEB) and 1 µl of 10 mCi / ml of gamma-^{32}P dATP for 1 hour at 37 °C. Unincorporated isotope was removed by purification of the labeled DNA on a Qiagen gel purification column, and the purified DNA was eluted in 75 - 150 µl of Tris.HCl pH 8.0.

2.5 Mammalian cell culture methods

2.5.1 Mammalian cell culture

Mammalian cells were grown in defined growth media (DMEM or Alpha MEM (Invitrogen)) with 10 % donor calf serum, and 100 µg / ml of a penicillin / streptomycin antibiotic mix (Invitrogen). The cells were maintained at 37 °C in 5 %

CO_2. Adherent cells were dislodged at 90 % confluence by washing with a dilute solution of trypsin / EDTA and replating at a 1 : 5 ratio. Suspension cell culture was maintained in a similar manner except cells were diluted and replated omitting the trypsinization.

2.5.2 Cryogenic storage of mammalian cells

Cells were collected by centrifugation at 1500 g for 5 min and supplemented with 1 ml of media containing 10 % DMSO per 5 X 10^6 cells. Cells were stored on ice for 1 hour followed by storage at - 80 °C overnight. The following day the cells were transferred to liquid nitrogen for permanent storage.

2.5.3 Transfection

Cells were transfected with lipofectamine (Invitrogen) according to the manufacture recommendations. The next day the cells were washed with fresh media and allowed to recover for 48-72 hours before cells were harvested for further manipulation.

2.5.4 Isolation of nuclei from tissue culture cells

Adherent tissue culture cells were scraped into 1 X PBS with the aid of a rubber police man. The cells we collected by centrifugation at 1500 g for 5 min at room temperature. The cells were washed once with PBS and resuspended in 10 volumes of buffer A (10 mM Hepes pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, and complete protease inhibitors (Roche)). The cells were incubated on ice for 10 min and recovered by centrifugation at 1500 g for 5 min. The cells were then resuspended in 3 volumes of buffer A and nuclei released by 20 – 40 plunges of a tight fitting dounce. Nuclei were recovered by centrifugation at 1500 rpm for 5 min.

2.5.5 Isolation of nuclei from brain tissue

Rat brains (obtained from Pel-Freez Biologicals) were ground to a fine powder in liquid nitrogen with a mortar and pestle. The brain powder was diluted 5 volumes to 1 in ice-cold buffer A (10 mM Hepes (pH 7.5), 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 M sucrose, 10 % glycerol, and complete protease inhibitors (Roche)) followed by homogenization in a 60 ml Dounce (Braun) on a Potter S (Braun) motorized homogenizer (five strokes at 1100 rpm). The homogenate was layered onto a 10 ml cushion of buffer A and centrifuged in pre-chilled SW28

rotor at 24,000 rpm in a Beckman XL100 ultracentrifuge for 40 min at 3 °C to recover the nuclei.

2.5.6 Nuclear protein extracts from mammalian cells

Recovered nuclei were resuspended in 1 volume of buffer C containing 5 mM Hepes (pH 7.9), 26 % glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, and complete protease inhibitors (Roche Applied Science) supplemented with 400 mM NaCl. The extraction was allowed to proceed for 1 h on ice with occasional agitation, and then the nuclei were pelleted at 13,000 rpm for 20 min at 4 °C. The supernatant was taken as the nuclear extract. Nuclear extract was either used immediately or snap frozen in liquid nitrogen and stored at – 80 °C

2.6 Insect cell and baculovirus culture methods

2.6.1 Insect cell culture

All baculovirus stock generation and expression was done using SF9 cells (Invitrogen) which are derived from the papal ovarian tissue of the fall army worm *Spodoptera frugiperda*. Cells were grown in serum free SF900 II medium containing L-glutamine (Gibco) supplemented with penicillin and streptomycin (100 µg / ml), and incubated at 27 °C. Cells were grown as a monolayer in tissue culture flasks and split when confluent by dislodging the cells physically. Cells were diluted 1:10 to 1:20 in fresh SF900 II medium and either re-plated in new tissue culture flasks or inoculated as 50 ml cultures in a 200 ml spinner culture flask at 80 - 90 rpm. SF9 cells grown as suspension cultures in spinner flasks were maintained at densities of less than 2.5 X 10^6 cells per ml. Cells were passaged by centrifuging the cells at 1500 rpm for 5 min to isolate the cells, and the old media was removed and replaced with fresh SF900 II media at 0.5 X 10^6 cells per ml.

2.6.2 Cryogenic storage of insect cells (SF9)

SF9 cells were cryogenically stored by centrifuging the cells at 1500 rpm for 5 min and then resuspended the collected cells at 1 X 10^7 cells per ml in SF900 II media supplemented with 10 % DMSO. Cells were stored on ice for one hour and then placed at -70 °C overnight. In the morning the cells were transferred for permanent storage in liquid nitrogen. To recover stocks from liquid nitrogen the cells were rapidly defrosted at 37 °C and immediately diluted in fresh SF900 II media and

placed in tissue culture flasks. Flasks were incubated at 27 °C and dead cells removed and replaced with fresh media the following day.

2.6.3 Generation of recombinant baculovirus genome

The Fastbac system (Invitrogen) was used to generate recombinant baculovirus genome. The virus was generated according to instructions of the manufacturer and described briefly here. 10 ng of Fastbac plasmid DNA was added to 100 µl of competent DH10Bac cells which had been defrosted on ice. Cells were incubated on ice for 30 min. Cells were then heat shocked for 45 seconds at 42 °C, and immediately chilled on ice for 2 min. 900 µl of L.B. broth was added and the tubes were shaken for 4 hours at 37 °C. The cells were then plated as 15 µl of the transformation mix or 100 µl each of a 10^{-1}, 10^{-2}, and 10^{-3} dilution. Cells were grown on plates containing kanamycin, gentamicin, tetracycline, X-gal, and IPTG for 48 hours at 37 °C. White colonies should contain transposed baculovirus genome , but as a second screening 10 white colonies were re-streaked onto the same selective media and incubated overnight at 37 °C. Then a single white colony from the streaked plate was selected and grown in liquid culture overnight under the same selective pressure. In the morning cells were pelleted and the baculovirus genome recovered by Midi prep (Qiagen). To elute the DNA from the Midi prep column the elution buffer was heated to 65 °C before applying to column. To verify the transposition of the protein coding sequence into the baculovirus genome, diagnostic PCR was carried out using M13 forward and reverse primers that flank the insertion site, and additional internal gene specific primers. To facilitate amplification of the large diagnostic PCR fragments a PCR reaction with an initial denaturation of 94 °C for 3 min followed by 34 cycles of 94 °C for 30 second, 55 °C for 45 seconds, and 72 °C for 5 min was carried out. A final extension of 72 °C for 7 min was used to extend PCR products. Diagnostic PCR reactions consisted of 100 ng of bacmid DNA and 50 ng of diagnostic primer in a 50 µl reaction volume.

2.6.4 Transfection of SF9 cells and isolation of p1 baculovirus

To generate infectious baculovirus particles SF9 cells were seeded at a density of 9 X 10^{5} in six well plates. In one tube 1 µg of baculovirus genome was diluted in 100 µl with un-supplemented Graces Medium (Invitrogen) and in a second tube 6 µl of Cellfectin (Invitrogen) was diluted in 100 µl Graces Medium. The two tubes were

mixed and allowed to incubate for 45 min at room temperature. While the DNA / Cellfectin reagents were incubating the cells were washed once with 2 ml of Graces medium and replaced with 800 µl of Graces medium. After 45 min the DNA / Cellfectin mixture was added to the plates and allowed to incubate at 27 °C for 5 hours. After 5 hours the transfection mix was removed and 2 ml of SF900 II medium was added. Cells were incubated at 27 °C for 72 hours and the supernatant collected as the viral p1 stock. The p1 stock was stored at – 70 °C for long term storage. The cells from the transfection were collected and lysed in 100 µl of 2 X SDS loading buffer. The lysate was analyzed by western blotting for the expressed protein as a verification of baculovirus genome transfection.

2.6.5 Amplification of viral stock

To amplify the P1 viral stock a 20 ml culture of SF9 cells was grown to 2 X 10^6 cells per ml and 1 ml of p1 viral stock was added to the SF9 culture which was incubated spinning for 48 hours. The cells were centrifuged for 5 min at 1500 rpm and the supernatant collected as p2 stock and stored at 4 °C protected from light. Several aliquots of the p2 stock were frozen at - 70 °C for long term storage. The cell pellet was used to make whole cell extract from which small scale purifications were done to test solubility and functionality of the expressed protein.

2.6.6 Infection of SF9 cells for protein production

To generate preparative amounts of protein using Baculovirus, 500 ml spinner cultures were set up and grown to a density of 2 X 10^6 cells per ml at 27 °C. 25 - 30 ml of baculovirus amplified from the p2 stock was used to infect the cells for 48 hours. Cells were collected by centrifugation at 1500 rpm for 5 min and the supernatant removed and discarded. The cell pellets were stored at - 70 °C until protein preparation.

2.7 Xenopus laevis methods

2.7.1 Isolating *Xenopus l.* oocytes

Female *Xenopus laevis* were sacrificed by Dr. Donald MacLeod and the oocytes dissected immediately before oocyte extract was prepared.

2.7.2 Making *Xenopus l.* oocyte extract

Xenopus laevis oocytes were isolated and washed several times in OR-2 (5 mM Hepes pH 7.9, 1 mM $Na_2(PO_4)$, 82.5 mM NaCl, 2.5 mM KCl, and 1 mM MgCl2) and collagenase Type II (Sigma Chemical) was added to 0.75 mg/ml. The oocytes were incubated on a shaking platform for 60 – 90 min until the oocytes were dispersed. Immature oocytes and follicle cells were removed by constant washing with OR-2. Oocytes were then washed several times in extractions buffer (20 mM Hepes pH 7.5, 5 mM KCl, 1.5 mM MgCl2, 1 mM EGTA, 10 % glycerol, 10 mM β-glycerophosphate, 0.5 mM DTT, 1 mM PMSF, and complete protease inhibitor (Roche)) and placed in 6 ml volumes into SW40 ultracentrifuge tubes and the remainder of the volume made up with extraction buffer. The oocyctes were centrifuged in a SW40 rotor at 38 000 rpm for 1 hr at 4 °C. The clarified extract was used immediately for biochemical fractionation.

2.7.3 *Xenopus l.* MeCP2 and Sin3a antibodies

Xenopus laevis MeCP2 and Sin3a antibodies were a generous gift of Peter L Jones. The specificity of the MeCP2 antibodies was verified by western blot using recombinant *Xenopus laevis* protein as bait. Sin3a antibodies gave an immunoreactive band at the predicted size in oocyte extract.

2.8 Protein methods

2.8.1 Measuring protein concentration

Protein concentration was measured using the BioRad bradford reagent kit. Briefly, the stock reagent was diluted 1 in 4 with water and mixed thoroughly. Increasing volumes of protein were then added to 1 ml of bradford reagent. The reagent was allowed to incubate with the protein for five min and the absorbance at 595 nm recorded. Known concentrations of BSA were added to bradford reagent in parallel to generate a standard curve by which the concentration of the test protein sample could be determined.

2.8.2 Dialysis

Protein was added to sealed dialysis tubing or cassettes (Pierce) with an appropriate molecular weight cut off range (as recommended by manufacturer). Dialysis was

allowed to proceed overnight in the appropriate buffer at approximately 300 - 500 fold excess of dialysis buffer. The following morning the dialysis buffer was replaced with an equivalent volume of fresh dialysis buffer and allowed to proceed for an additional 3 - 4 hours before the protein sample was collected and either stored at - 70 °C or used immediately.

2.8.3 SDS-PAGE

Proteins were diluted in loading buffer (8 % SDS, 100 mM Tris (pH 6.8), 100 mM DTT, 10% glycerol, and 0.1% brophenol blue) and boiled for 5 min at 100 °C. 0.75 mm thick gels were assembled in either a Bio-Rad mini protean II apparatus or a large Bio-Rad XL casting apparatus according to the manufacturer's recommendation. Gels consisted of a stacking gel buffered at pH 6.8 and a separating gel pH 8.8. All gels used an acrylamide to bis-acrylamide ratio of 29:1 with 0.001% SDS and the acrylamide was polymerized by the addition of ammonium persulfate and TEMED as a catalyst. Gels were run in TGS at 25 mA per gel small gel or 45 mA per large gel at 4 °C until the loading dye migrated off the bottom of the gel.

2.8.4 Coomassie blue stain

Coomassie blue stain consisted of 50 % methanol with 10 % glacial acetic acid and 0.1 % Brilliant blue R250 dye. SDS PAGE gels were incubated with fresh Coomassie blue stain rocking for one hour. Excess stain was removed with H_2O and the gel was immersed in de-stain solution consisting of 30 % methanol and 10% acetic acid. The de-stain was allowed to proceed for several hours until the background staining was no longer visible. Gels were then soaked overnight in H_2O and imaged the following morning.

2.8.5 Semi-Dry transfer to nitrocellulose membrane

Semi-Dry transfer was carried out on a Bio-Rad electro-blot system according to the manufacture recommendations. Briefly, three layers of 0.3 mm whatman paper was soaked in semi-dry transfer buffer (39 mM glycine, 48mM Tris, 0.037% SDS, and 20 % methanol). A single layer of nitrocellulose filter paper was wetted in transfer buffer and placed on top of the three layers of whatman paper. The SDS PAGE gel was carefully placed on top of the nitrocellulose membrane and an additional three layers of moist 0.3 mm Whatman paper was placed on top. Bubbles were removed by

rolling a 25 ml pipette several times over the assembled gel sandwich. The top electrode was place over of the gel sandwich and the transferred with 0.2 amps for between 1 and 1.25 hours. After transfer the gel sandwich was disassembled and the SDS PAGE gel Coomassie stained to ensure efficient transfer and the nitrocellulose membrane blocked for at least one hour in 5 % milk 1 X TBS solution.

2.8.6 Western blot

For most antibodies the following western blot procedure was used. After a 1 hour block in a 5 % milk 1 X TBS solution, fresh blocking solution containing a 1:500 to 1:1000 dilution of the primary antibody was applied. The primary antibody was incubated for one hour at room temperature or overnight at 4 °C followed by three consecutive ten minute washes in a 5 % milk / 1 X TBS solution. The secondary antibody was applied at a dilution of 1:5000 in a 5 % milk / 1 X TBS solution for one hour at room temperature. Two consecutive ten minute washes in a 5 % milk / 1 X TBS solution were followed by one ten minute wash in 1 X TBS. The nitrocellulose membrane was then placed in a freshly prepared equal volume mixture of ECL solution 1 (2.5 mM luminal, 0.396 mM p-coumeric acid, and 100 mM Tris-HCl pH 8.5) and solution 2 (5.6 mM H_2O_2, 100 mM Tris-HCl pH 8.5) for one minute. The nitrocellulose blot was wrapped in a single layer of ceran wrap and exposed to ECL hyper-film (Amersham Biosciences).

2.8.7 Immunoprecipitation

Antibodies were incubated with 500 µg of rat brain nuclear extract diluted in binding buffer (20 mM Hepes (pH 7.9), 0.15 M NaCl, 10% glycerol, 0.2 mM EDTA, 0.01% Triton X-100) at 4 °C for 4 h. 20 µl of protein A conjugated Sepharose beads (Amersham Biosciences) were added to the immunoprecipitation and incubated for 1 h at 4 °C rocking. Beads were washed four times with binding buffer and eluted with 1 X SDS PAGE loading buffer.

2.8.8 Sucrose gradient sedimentation

Sucrose gradients were formed at 4 °C in 13 ml SW40 tubes using a manual two chamber gradient former. Chamber one was loaded with buffer A (300 mM KCl, 20 mM Hepes pH 7.9, 2 mM EDTA, 10 % Glycerol, 10 mM beta-mecaptoethanol) containing 5 % sucrose and chamber 2 with buffer A containing 20% sucrose. The sample containing 25 – 50 ug of standard proteins BSA, Appoferritin, B-amylase, and

ADH were loaded onto the 5 – 20 % sucrose gradient and centrifuged at 40 000 rpm in a SW40 rotor for 19 hours at 4 °C. 500 µl fractions were collected manually from the top of the gradient using a peristaltic pump fitted with a capillary tube. Each fraction was TCA precipitated and roughly an equal portion of the precipitated protein was used for an SDS PAGE gel followed by Coomassie stain to identify marker proteins or a western blot to identify MeCP2.

2.8.9 Kinase Assay

Kinase assays were carried out by adding 1 – 3 ul of crude protein extract, or protein fraction, to a 20 µl reaction in buffer HB5 (60 mM β-glycerophosphate, 15 mM ρ-nitrophenylphosphate, 25 mM MOPS (pH7.2), 5 mM EGTA, 15 mM $MgCl_2$, 1mM DTT, 0.1 mM Na_3VO_4, 1% Triton X-100) supplemented with 15 uM ATP, 60 ng / ul recombinant MeCP2 (or control protein), and 5 nl of 10 mCi / ml of gamma-^{32}P dATP. The reaction was allowed to proceed for 45 min at 30 °C and stopped by the addition of 1 X SDS PAGE loading buffer. Samples were separated by SDS PAGE and Coomassie stained prior to drying the gel onto Whatman paper for 1 hour at 80 °C. The dried gels were visualized by exposure to a phosphor-imager screen (Amersham) for 1 hour or a Bio-Max MS auto-radiography film (Kodak) for 2 hours.

2.8.10 Mass Spectrometry

To identify purified proteins, excised bands were in-gel trypsinized, and peptides were eluted from the gel slice. Mass spectrometry analysis was carried out on an Applied Biosystems Voyager DE-STR matrix-assisted laser desorption ionization time-of-flight instrument using α-cyano-4-hydroxycinnamic acid matrix. Spectra were analyzed in MS-Fit and then submitted to Protein Prospector (prospector.ucsf.edu/) for peptide matching by the Edinburgh Protein Interaction Center.

2.9 Expression of proteins in bacteria

2.9.1 Cell preparation and induction of expression

A 25 ml starter culture was initiated with a fresh colony from an LB agar plate. The starter culture was grown overnight shaking at 37 °C under the appropriate antibiotic selection. The next morning the 25 ml culture was diluted in 1 – 2 l of LB containing antibiotic selection and grown shaking at 37 °C until an optical density of 0.6 was at 600 nm was obtained. At this point the culture was transferred to 30 °C and 1 mM

IPTG was added to induce protein production. The culture was induced at 30 °C shaking for 3 hours and the bacteria were collected by centrifugation at 6000 g for 10 min at 4 °C. At this point the bacterial were either used to make extract or stored at -80 °C for future use.

2.9.2 Making non-denaturing extracts.

The cell pellet from a 2 l culture of bacteria was resuspended in 30 ml of lysis buffer (20mM Tris.HCl pH 8.0, 500mM NaCl, 0.1% NP 40) containing complete protease inhibitors (Roche) in a 50 ml falcon tube. The resuspended lysate was sonicated on ice at output 4 - 5 at duty cycle 30 % (on a Branson Sonifier 250) for 3 min, followed by a 3 min rest period on ice, and then a second 3 min sonication. The insoluble material after sonication was pelleted at 12000 g for 20 min at 4 °C. The supernatant was collected as whole cell extract.

2.10 Expression of proteins in insect cells

2.10.1 Insect whole cell extracts

The cell pellet was resuspended 1: 20 in lysis buffer (50 mM NaH_2 PO_4 pH 7.9, 500 mM NaCl, 10 mM Beta-glycerphosphate, 15 % Glycerol, 0.01 % NP-40, and complete protease inhibitors). The cells were then added to a 40 ml dounce and incubated on ice for 30 min over which time frame 40 plunges of the dounce were applied. The lysate was then centrifuged at 14500 rpm for 20 min at 4 °C and the supernatant taken as soluble whole cell extract.

2.11 Protein purification

2.11.1 Chromatography media

All chromatography media was purchased from Amersham biosciences except for BioRex 70 cation exchange resin which was purchased from Bio-Rad, and fractogel TMAE which was purchased from EMD biosciences.

Resin	Application
Chelating Sepharose	affinity chromatography (bound with Ni)
Bio-Rex70	cation exchange resin (50-100 mesh)

Sp-Sepharose	cation exchange resin
Mono S	cation exchange resin (mono dispersed beads)
Mono Q	anion exchange resin (mono dispersed beads)
Fractogel TMAE	anion exchange resin (1 ml analytical and 180 ml preparative)
Superose 12	Size exclusion chromatography (24 ml analytical)
Superose 6	Size exclusion chromatography (24 ml analytical)
Sephacryl S-300	Size exclusions chromatography (360 ml preparative)
Phenyl-Sepharose	Hydrophobic interaction chromatography (2 ml analytical)

2.11.2 Chromatography solutions

Anion exchange buffer contains 20 mM TrisHCl (pH7.9), 0.2 mM EDTA, 1 mM DTT, 10 % Glycerol, supplemented with 100 mM NaCl (AE 100) or 1000 mM NaCl (AE1000).

Cation exchange buffer contains 20 mM Hepes (pH 7.6), 0.2 mM EDTA, 1 mM DTT, 10% Glycerol, supplemented with 100 mM NaCl (CE100), or 150 mM NaCl (CE150), or 200 mM NaCl (CE200), or 1000 mM NaCl (CE1000).

Buffers for his-tag purification contain 50 mM NaH2PO4, 300 mM NaCl, 10 % Glycerol (pH 8.0), supplemented with 20 mM imidazole (N20), 100 mM imidazole (N100), or 250 mM imidazole (N250).

Size exclusion chromatography buffers were made in 20 mM HepesKOH (pH 7.9), 3 mM MgCl2, 10 % glycerol, supplemented with 150 mM KCl (GF150) or 500 mM KCl (GF500).

Buffers for hydrophobic interaction contain 0.2 mM EDTA, 50 mM Tris pH 7.9, 1 mM DTT, and 10 % glycerol supplemented with 0.7 M $(NH_4)_2SO_4$ (PE700) or no $(NH_4)_2SO_4$ (PE0).

2.11.3 Preparing disposable columns

Disposable 10 ml or 20 ml poly-prep columns were purchased from Bio-Rad. Columns containing ion exchange resins were prepared by washing the resin in low salt buffer and then applying the resin to the column. The resin was allowed to settle and then washed with 5 column volumes of high salt buffer to fully charge the resin and then was re-equilibrated with low salt wash buffer.

2.11.4 Preparing FPLC based HR columns

Empty FPLC columns were purchased from Amersham biosciences. The columns were filled with resin according to the manufacturer's recommendations. The resins were stored in 20 % ethanol, and regenerated for use by washing with water followed by high salt buffer and low salt buffer to charge the resin.

2.11.5 Ion exchange chromatography

Protein samples were dialyzed into the appropriate CE or AE buffer and applied to the prepared resin either in a disposable column of a FPLC based column. The sample was then washed with 20 – 40 column volumes of the dialysis buffer followed by consecutive step elutions or a linear gradient elution of increasing ionic concentration. Factions were collected during all steps and stored at – 80 °C or used immediately.

2.11.6 Affinity chromatography

Heparin-Sepharose was used in same manner as described in Section 2.11.5.

2.11.7 Size exclusion chromatography

Proteins were applied directly to a gel filtration column equilibrated according to the manufactures recommendations. Molecular weight size standards (Sigma) were run over the column and their elution volumes monitored by absorbance at 280 nm. Protein samples were applied to the column with volumes not exceeding that recommended by the manufacturer and eluted with a constant buffer flow rate of 0.2 ml / minute collecting fractions within the inclusion limit. Samples were stored at – 80 °C or used immediately

2.11.8 Purification of his tagged bacterial proteins

Whole cell lysate was transferred to a 50 ml falcon tube conataining 1 ml of Ni-Nta beads pre-washed in lysis buffer. The lysate was incubated for 1 hour at 4 °C on a spinning fly to allow proteins containing a histidine run to bind the matrix. The lysate bead mixture was poured directly onto a disposable 10 ml poly-prep column (Bio-Rad) allowing the unbound material to flow through. The bound material was then washed with 20 – 40 column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH adjusted to 8.0). The protein was eluted 5 times with 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH adjusted to 8.0). A 3 µl aliquot of each elution was run on SDS PAGE and analysed by Coomassie staining. The protein was dialysed overnight into dialysis buffer (20 mM Hepes pH 7.9, 200 mM NaCl, 0.5 mM EDTA, 0.1 mM DTT, and 10 % glycerol). The concentration of the purified protein was determined by Bradford assay and visualized by SDS PAGE and Coomassie staining for purity. Purified proteins were stored short term at - 20 °C and long term at - 80 °C. If further purification was required other chromatographic methods indicated in Section 2.11 were used where indicated.

2.11.9 Purification of his tagged baculovirus expressed proteins

Whole cell lysate was transferred to a 50 ml falcon tube conataining 1 ml of Ni-Nta beads prewashed in lysis buffer. The lysate was incubated for 1 hour at 4 °C on a spinning fly to allow proteins conataining histide runs to bind the matrix. The lysate bead mixture was poured directly onto a disposable 10 ml poly-prep column (Bio-Rad) allowing the unbound material to flow through. The bound material was then washed with 20 – 40 ml of column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH adjusted to 8.0). The protein was eluted 5 times with 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH adjusted to 8.0). A 3 µl aliquot of each elution was run by SDS PAGE and analysed by Coomassie staining. The protein was dialysed overnight into dialysis buffer (20 mM Hepes pH 7.9, 200 mM NaCl, 0.5 mM EDTA, 0.1 mM DTT, and 10 % glycerol). The concentration of the purified protein was determined by Bradford assay and visualized by SDS PAGE and Coomassie staining for purity. Purified proteins were stored short term at - 20 °C and long term at - 80 °C. If further purification was

required other chromatographic methods indicated in Section 2.11 were used where indicated.

2.12 Manipulation of protein

2.12.1 Dephosphorylation of baculovirus expressed MeCP2

Purified baculovirus protein (Section 2.11.9) was incubated with lambda phosphatase buffer (50 mM Tris-HCl, 5 mM dithiothreitol, 0.1 mM Na_2EDTA, 0.01 % Brij 35, pH 7.5) and 400 units of lambda phosphatase (NEB) for 1 hour at 30 °C. To verify specificity of the dephosphorylation reaction mock dephosphorylation and dephosphorylation reactions carried out with 50 mM EDTA as an inhibitor were done in parallel. Changes in relative mobility were determined by SDS PAGE and Coomassie staining. To obtain pure dephophorylated MeCP2 large dephosphorylation reactions were carried out and the protein re-purified by ion exchange chromatography (Section 2.11.5)

2.12.2 Nuclease treatment of purified rat MeCP2

Purified rat brain MeCP2 (15 µl) was added to 240 µl of 50 mM Tris.HCl pH 8.0 and 2 mM $MgCl_2$ containing 90 units of Benzonase (Merck). The reaction was incubated at 4 °C for 3 hours according to the manufacturer's recommendations, followed by size exclusion chromatography on a Superose 12 column in GF 500.

2.12.3 Denaturation of rat brain nuclear extract

Concentrated urea was added to 200 µg of rat brain extract to a final concentration of 4 M. The sample was heated for 10 min at 65 °C and then applied to a Superose 12 column in GF 150.

2.13 DNA binding experiments

2.13.1 Generation of probes using single stranded oligonucleotides

Equal molar amounts of a long single stranded oligonucleotide (30 – 75 bases) and a complementary shorter reverse primer were incubated in 10 mM tris pH 8.0 with 50 mM NaCl for 10 min in a 100 °C hot block. The hot block was removed from the heat source and allowed to cool to room temperature. The annealing reaction was

used as a template to generate double stranded DNA by incubating for 1 hour at 37 °C with 6 units of klenow polymerase and 5 µl of 2 mM dNTP's.

2.13.2 Electrophoretic Mobility Shift Assay (EMSA)

Probes used for EMSA were generated either by PCR amplification from plasmid DNA followed by purification on a Qiagen PCR purification column, or by extension of single stranded of oligonucleotides using complementary oligonucleotides and Klenow polymerase extension (Section 2.13.1) followed by purification on Qiagen gel extraction column. EMSA reactions were assembled at room temperature by combining 500 ng of competitor DNA (poly(dG-dC).poly(dG-dC)) and protein sample in binding buffer (32 % ficoll, 80 mM hepes pH 7.9, 150 mM KCl, 4mM EDTA, and 2 mM DTT). This reaction was allowed to incubate 10 min at room temperature and then the labelled probe containing bromophenol blue was added and the reaction continued for a further 25 min at room temperature. The binding reaction was loaded on to a 6 % non denaturing polyacrylamide that had been pre run for 1 hour at 240 V. The samples were electrophoresed at 240 V until the bromophenol blue dye front had migrated 9.5 cm into the gel. The gel was transferred to 2 layers of 0.2 mm Whatman paper and dried at 80 °C for 1 hour. Dried gels were exposed to a phosphor-imager screen overnight and developed the next day.

2.13.3 Methyl-SELEX: Systematic Evolution of Ligand by Exponential Enrichment (SELEX)

A 74 base oligonucleotide was generated that has M13 forward and M13 reverse primer binding sites at the 5' and 3' ends respectively. The central core of the oligonucleotide consisted of an *HpaII* methyltransferase / restriction endonuclease site flanked by 18 base pairs of random sequence on either side (gttttcccagtcactac-(N_{18})-CCGG-(N_{18})-gtcatagctgtttcctg) and the primer was purified by Sigma-Genosys through polyacrylamide gel electrophoresis. The single stranded oligonucleotide was used to generate a double stranded DNA template by mixing 5 µg of the long single stranded oligonucleotides library with an equamolar amount of the M13R primer. The primers were incubated in 10 mM tris pH 8.0 with 50 mM NaCl for 10 min in a 100 °C hot block. The hot block was removed from the heat source and allowed to cool to room temperature. The annealing reaction was used as a template to generate double stranded DNA by incubating for 1 hour at 37 °C with 6 units of klenow polymerase

and 5 µl of 2 mM dNTP's. The resulting double stranded DNA was purified using a Qiagen gel extraction column and eluted in 50 µl of elution buffer. The purified DNA was methylated with 20 units of M.*HpaII* methyltransferase overnight and purified on a Qiagen gel extraction column, and eluted in 50 µl. The efficiency of methylation was tested by digesting 50 ng of the methylated probe with *HpaII* restriction enzyme and resolved on a 10 % polyacrylamide gel. Efficiently methylated probe was completely refractory to *HpaII* digest as observed by ethidium bromide staining of the polyacrylamide gel. The concentration of the purified methylated methyl-SELEX DNA was quantified by comparing varying amounts of low mass DNA ladder from invitrogen on a 10 % polyacrylamide gel by ethidium bromide staining. A 50 ng aliquot of the purified methylated methyl-SELEX template was end labelled using 30 units of T4 polynucleotide kinase (NEB) and 1 µl of ~ 0.5 MBq / ml gamma-^{32}P dATP. The labelled methyl-SELEX DNA was purified on a Qiagen gel extraction column, and eluted in 75 µl of DNA. Then, 1.6 ng of labelled methylated methyl-SELEX DNA template was used in a 20 µl EMSA reaction (as described in section 2.13.2) and the binding buffer consisted of 10 mM Tris pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 8 % Ficol, 1mM DTT, and 0.1 % NP40. 44 ng, 89 ng, or 133 ng of recombinant MeCP2 1-205 was used as the shifting protein. The EMSA was run on a 6 % polyacrylamide gel for approximately 2 hours at 240 V until the loading dye had migrated ~ 10 cm into the gel. After drying the EMSA gel to 0.2 mM Whatman paper at 80 °C for one hour under vacuum the blot was exposed to KODAK BioMax MS film with a MS intensifier screen at - 80 °C for 1 hr and developed. The resulting film was used as a guide to cut out specific bands that were shifted by the lowest amount of MeCP2 protein. The gel slices were incubated in 100 µl H_2O at 100 °C for 10 min. 5 µl of the DNA eluted from the gel slice was subject to PCR amplification in a reaction containing 1.5 mM $MgCl_2$, 50 ng of the M13 forward / reverse primer, and 0.5 units of Taq polymerase (Abgene). The PCR reaction consisted of 95 °C for 4 min, followed by 24 cycles of 95 °C for 40 seconds, 60 °C for 40 seconds and 72 °C for 10 seconds. An 8 minute extension at 72 °C was followed by addition of 100 ng of M13 forward / reverse primer, 2 µl of 2 mM dNTP's, and 0.5 units taq polymerase. A final cycle for 94 °C for 4 min, 64 °C for 2 min, and 72 °C for 10 min was used to fully extend all PCR products (this final cycle was essential to obtain all double stranded product). The PCR product was purified

with a Qiagen gel extraction kit and the resulting purified PCR product were methylated with M.*HpaII* methyltransferase as described above. This fragment was then used for a subsequent round of selection by repeating the EMSA and re-isolating shifted DNA by PCR as described above. The selected fragments were enriched over 8 rounds of EMSA and the final PCR product was cloned into the pGEMTeasy TA cloning system (Promega). Sequences were analyzed by chain termination sequencing using the SP6 primer located in the pGEMTeasy vector (Promega).

2.13.4 End labeling DNA for foot-printing

A single stranded DNA fragment corresponding to the foot-printing probe was first labelled. Labelling was done in a 30 µl reaction containing 500 ng of single stranded template DNA, 1.5 µl of p32 dATP, and 3 µl of T4 polynucleotide kinase. Un-incorporated nucleotide was removed by purification of the single stranded DNA on a Qiagen nucleotide removal column and eluted in 60 µl of tris pH 8.0. An equal molar amount of an oligonucleotide corresponding to the reverse complement of the 3' end of the labelled single stranded DNA was added and incubated in a 100 °C heat block for 10 min in Tris.HCl pH 8.0 with 50 mM NaCl. The block was removed from the heat source and allowed to cool to room temperature. The annealing reaction was directly used in a klenow extension reaction containing 6 units of klenow polymerase and 5 µl of 2 mM dNTP's. The extension reactions was purified on a Qiagen gel extraction column and eluted in a 60 µl volume of Tris.HCl pH 8.0. Half of the reaction was methylated overnight with M.*Sss*I and the other half mock methylated. The final product was purified on a Qiagen gel extraction column and the relative incorporation of gamma-^{32}P dATP determined in a scintillation counter adjusting the final volume to make a stock of 10 000 cpm in 10 mM Tris pH 8.0.

2.13.5 Foot-printing ladder: Maxam-Gilbert G+A sequencing reaction

Probe (approximately 8 µl) corresponding to 80 000 cpm was added to a reaction containing 1 ug of *E. coli* genomic DNA in a total volume of 10 µl. 1 µl of a 4 % solution of formic acid was added and the mixture and incubated at 37 °C for 25 min. During this incubation a fresh stock of 10 % piperidine (1 M) was freshly made. 150 µl of the 10 % piperidine solution was added directly to the formic acid reaction and the tube incubated at 90 °C for 30 min. This reaction was then placed on ice for 5 min

followed by addition of 1 ml of n-butanol. After vortexing, the sample was centrifuged for 2 min at maximum speed in a micro-centrifuge to pellet the DNA. The DNA pellet was resuspended in 150 µl of 1% SDS and an additional 1 ml of n-butanol was added followed by vigorous vortexing and centrifugation at maximum speed for 2 min in a micro-centrifuge. The supernatant was removed and extracted twice more in 0.5 ml of n-butanol with centrifugation between each step. To remove the remaining traces of n-butanol the pellet was dried under vacuum for 10 min. The resulting pellet was resuspended in 40 µl of loading buffer (98% formamide, 10 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylen Cyanol) and stored at -20 until required. 10 µl of this sequencing reaction was sufficient per lane on a typical sequencing gel.

2.13.6 DNase1 foot-printing

DNA binding reactions were set up in the same manner as an EMSA reaction (Section 2.13.2). Briefly reactions were assembled in binding buffer containing 0-100 ng of poly(dG-dC).poly(dG-dC), 500 ng of BSA, and 140 ng of MeCP2 1-205 protein in a 18 µl reaction volume. Reactions were incubated at room temperature for 10 min, and then 2 µl of probe at 10000 cpm labelled at one end was added to each reaction and it was allowed to go for another 25 min at room temperature. 0.025 µg of DNase 1 from bovine pancrease (Boringher Manheim) diluted in DNase dilution buffer (50 mM NaCl, 20 mM Tris pH 7.6, 5 mM $MgCl_2$, and 3 mM $CaCl_2$) was added to the reaction in a 4 µl volume. The DNase 1 digestion was allowed to continue for 2 - 4 min at room temperature. Then, 50 µl of 8 mM EDTA was added to stop the digestion. To recover the DNA fragments 1 ug of *E. coli* genomic DNA was added to the reaction followed by phenol chloroform extraction and ethanol precipitation. Air dried DNA pellets were resuspended in 10 µl of loading buffer (98 % formamide, 10 mM EDTA, 0.05 % Bromophenol Blue, 0.05 % Xylen Cyanol) and incubated at 100 °C for 2 min prior to loading on a 6 % sequencing gel that had been pre-run at 1400 V until the temperature reached approximately 50 °C. The DNA fragments were electrophoresed at 1400 V until the Xylene Cyanol had migrated into the gel 17 cm. The gels were dried on 2 layers of 0.2 mm Whatman paper and exposed to a phosphor-imager screen overnight before developing.

3. Chapter three-
Purification of Native MeCP2

3.1 Initial identification of MeCP2 and biochemical properties

The identification of two biochemical activities in mammalian cells that specifically recognize methyl-CpG indicated that DNA methylation could be actively recognised and perhaps even interpreted (Lewis et al., 1992; Meehan et al., 1989). The identity of the first activity, named MeCP1, was elusive and it was not until recently that MBD2 was determined to be the methyl-CpG binding component of MeCP1 (Feng and Zhang, 2001; Ng et al., 1999). The second biochemical activity, MeCP2, was the first to be purified and have its coding sequence identified.

MeCP2 was originally purified from rat brain by nuclear protein extraction followed by conventional biochemical fractionation. Using south-western analysis to follow methyl-CpG binding activity, a relatively pure fraction consisting of only one main polypeptide was achieved (Lewis et al., 1992). A partial polypeptide sequence of MeCP2 was obtained by Edman degradation, and degenerate oligonucleotides probes were used to isolate a full length MeCP2 cDNA. For many years MeCP2 was the only methyl binding activity for which the polypeptide sequence was known (Lewis et al., 1992) demonstrating how this brute force biochemical approach was instrumental in advancing the understanding of factors that interpret methyl CpG.

Two important functional domains have been mapped in MeCP2: the methyl-CpG binding domain (MBD) and the transcriptional repression domain (TRD)(Nan et al., 1997a; Nan et al., 1993). The MBD was used in bioinformatic studies to identify other methyl CpG binding proteins (MBP's) (Cross et al., 1997; Hendrich and Bird, 1998; Nan et al., 1993). The TRD of MeCP2 represses transcription when bound to methylated CpG's near promoter sequences and also from promoter proximal regions (Nan et al., 1998; Yu et al., 2000). The versatile repressive capabilities of MeCP2 rely on its tight association with chromatin, and interaction with proteins that modulate chromatin structure (Meehan et al., 1992; Nan et al., 1997a). One chromatin remodelling machine that MeCP2 interacts with is the Sin3a co-repressor complex. The chromatin remodelling and histone modifying capabilities of this large

multi-protein complex contribute to HDAC-dependent transcriptional repression by MeCP2 (Jones et al., 1998; Nan et al., 1998). In addition to Sin3a, various other factors have been shown to bind MeCP2 (Section 1.5.2) but the relative contribution of these proteins to MeCP2 function remains a matter of uncertainty (Fuks et al., 2003b; Kimura and Shiota, 2003; Kokura et al., 2001; Krithivas et al., 2002; Lunyak et al., 2002; Suzuki et al., 2003).

Recent advances in protein biochemistry have simplified the purification of proteins and protein complexes from crude cellular extracts (Forler et al., 2003). With the aid of mass spectrometry and extensive genome information very small amounts of purified protein can be identified. This is achieved through peptide mass fingerprint analyses combined with bioinfomatic comparison to theoretical peptide fingerprints in publicly curated databases (Steen and Mann, 2004).

To try and understand the relative biochemical contribution of the MeCP2 protein to the Sin3a complex and to identify other factors that stably associated with MeCP2, MeCP2 was purified by conventional biochemistry. In contrast to previous mammalian MeCP2 isolations (Lewis et al., 1992; Nan et al., 1997a) this analysis was done under milder conditions with the aim of maintaining and evaluating stable protein / protein interactions.

3.2 Biochemical evaluation the MeCP2/Sin3a interaction

3.2.1 MeCP2 co-immunoprecipitates with Sin3a

Mammalian Sin3a co-immunoprecipicates with MeCP2 in experiments using MeCP2 and Sin3a specific antibodies (Nan et al., 1998). *In vitro* analysis using recombinant proteins showed that MeCP2 binds directly to Sin3a. To verify this interaction *in vivo* a co-immunoprecipitation from rat brain nuclear extract with MeCP2 specific antibodies was done and the immunoprecipitated material probed with Sin3a antibodies (Figure 3.1A). Sin3a was immunoprecipitated with MeCP2 antibodies but not an irrelevant antibody (Gal4), in agreement with previously published

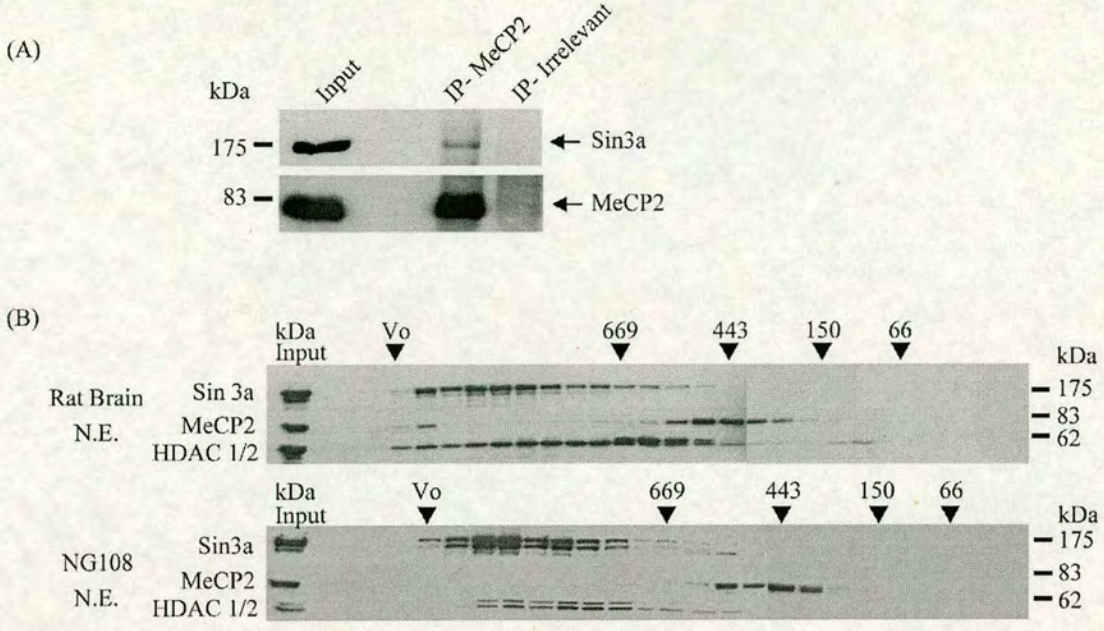


Figure 3.1- Mammalian MeCP2 does not stably associate with the Sin3a chromatin remodelling complex.

(A) An MeCP2 specific antibody or an irrelevant antibody (anti-gal4) were used to immunoprecipitate proteins from rat brain nuclear extract. The immunoprecipitates were probed by western blot with an anti-Sin3a antibody. The MeCP2 antibody specifically co-immunoprecipitates Sin3a signal from rat brain nuclear extracts, but an irrelevant antibody does not.
(B) Rat brain and NG108 nuclear extracts were separated by Superose 6 size exclusion chromatography and fractions probed by western blot with antibodies against Sin3a, MeCP2, and HDAC1/2. Native molecular weight markers are indicated as arrows above the western blot. MeCP2 elutes between 400-500 kDa devoid of the majority of Sin3a and HDAC 1/2 signal.

observations (Nan et al., 1998). A relatively weak signal for Sin3a is observed in MeCP2 immunoprecipitates suggesting that MeCP2 associates with a small amount of cellular Sin3a.

3.2.2 Absence of a stable MeCP2/Sin3a association in rat brain and cell line nuclear extracts

To address, in a more qualitative manner, the biochemical association between MeCP2 and Sin3a, brain extracts were analyzed by Superose 6 size exclusion chromatography. MeCP2 and components of the Sin3a co-repressor complex (Sin3a and HDAC 1/2) were identified in the size exclusion fractions by SDS PAGE and western blot analysis (Figure 3.1B). As previously reported (Kuzmichev et al., 2002), Sin3a complexes eluted between the void volume and the 443 kDa marker in fractions also containing HDAC 1/2 western blot signal. Interestingly very little Sin3a signal was observed in fractions containing MeCP2 western signal (Figure 3.1B). This observation is in agreement with the small amount of Sin3a associated with MeCP2 by co-immunoprecipitation (Figure 3.1A). A similar elution profile for Sin3a and MeCP2 was also observed in nuclear extracts from the neural-glial fusion cell line NG108 (Figure 3.1B) and several other transformed human cells (data not shown).

3.2.3 *Xenopus laevis* MeCP2 does not stably associate with *Xenopus laevis* Sin3a

Studies of MeCP2 in the African clawed frog (*Xenopus laevis)*, also identified an interaction between xMeCP2 and xSin3a in oocyte extract (Jones et al., 1998). In these experiments, association between xMeCP2 and xSin3a was evaluated by co-immunoprecipitation and western blot analysis using antibodies raised against the *Xenopus* proteins. In agreement with the observations in mammalian extracts these co-immunoprecipitations suggested that only small amounts of xSin3a/xMeCP2 are stably associated. In the same study, a rudimentary biochemical fractionation of oocyte extract was combined with size exclusion chromatography analysis to evaluate the relative amount of xMeCP2 and xSin3a that associate. In these experiments the authors observed that the majority of xMeCP2 was associated with xSin3a, in contrast to the small amount of xSin3a immunoprecipitated with MeCP2 antibodies. The different results with these two assays could be explained by disruption of the

xSin3a/xMeCP2 association during antibody binding in the immunoprecipitation experiments, but two more aspects of this study necessitate re-evaluation. Firstly, xMeCP2 and xSin3a do not precisely co-elute from the size exclusion chromatography column bringing into question the homogeneity of the proposed stable xMeCP2/xSin3 complex. Secondly, and most importantly, the xSin3a western blot signal in the biochemical fractionation migrated just above the xMeCP2 protein. This is much smaller than would be expected for xSin3a which is a 1275 amino acid protein. In other organisms Sin3a migrates by SDS PAGE at roughly 175 kDa. It is difficult to reconcile the abnormal migration of the proposed xSin3a molecule in this study.

To address these inconsistencies newly generated xMeCP2 and xSin3a antibodies were obtained from Peter L Jones. The specificity of the xMeCP2 antibodies were verified by western blot analysis of recombinant xMeCP2 (data not shown). In *Xenopus laevis* oocyte extracts the xMeCP2 antibodies recognized a band just under 80 kDa by SDS PAGE, suggesting endogenous xMeCP2 was recognized by this reagent. The xSin3a antibodies reacted with three bands at approximately 175 kDa, and looked similar to the signal observed in mammalian Sin3a western analysis. Multiple xSin3a bands may arise in oocyte extract due to Sin3a splice isoforms or proteolyic degradation the full length protein. To partially purify MeCP2, oocyte extract was applied to Bio-Rex 70 cation exchange resin and washed with low salt buffer (Figure 3.2A). xMeCP2 and xSin3a eluted from the Bio-Rex70 column with 500 mM NaCl and the presence of each protein was verified by western blot analysis (Figure 3.2B). The eluted proteins from the Bio-Rex 70 column were then applied to a Superose 6 size exclusion column. The size excluded fractions were TCA precipitated and run on an SDS PAGE gel followed by western blot analysis (Figure 3.2C). xMeCP2 eluted between 400-500 kDa and xSin3a eluted over several fractions in the high molecular weight range, indicating the xSin3a complex is intact

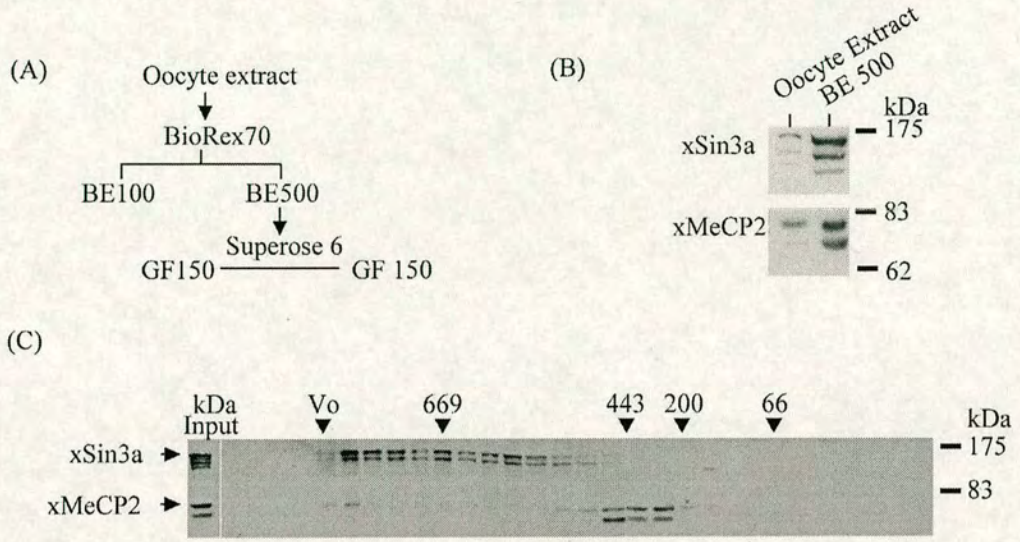


Figure 3.2- *Xenopus laevis* MeCP2 does not stably associate with the Sin3a chromatin remodelling complex.

(A) *Xenopus laevis* oocyte extract was separated by Bio-Rex 70 ion exchange chromatography followed by Superose 6 size exclusion chromatography.
(B) The 500 mM NaCl elution from the Bio-Rex 70 cation exchange column contains xMeCP2 and xSin3a as analysed by western blot analysis.
(C) An aliquot of the 500 mM NaCl elution from the Bio-Rex 70 cation exchange column was separated by Superose 6 size exclusion chromatography and the fractions analysed by western blot analysis using xMeCP2 and xSin3a antibodies. Native molecular weight markers are indicated as arrows above the western blot. xMeCP2 was found in fractions ranging from 400-500 kDa devoid of the majority of xSin3a.

in these experiments (Figure 3.2C). These results suggest that very little xMeCP2 is stably associated with xSin3a. This observation is analogous to the results in Figure 3.2C using mammalian sources and is consistent with poor co-immunoprecipitation of Sin3a with MeCP2 antibodies. In light of these observations, the original finding that *Xenopus laevis* xMeCP2/xSin3a exist as a stable complex is in doubt.

3.3 Purification of Native MeCP2

3.3.1 MeCP2 elutes from size exclusion chromatography columns with a molecular weight of 400-500 kDa

Although MeCP2 does not form a stable complex with Sin3a, MeCP2 eluted from Superose 6 size exclusion column with a relative molecular weigh between 400-500 kDa (Figure 3.2C). This MeCP2 elution profile is nearly ten times larger than its predicted size by amino acid composition (52.4 kDa). To determine if native MeCP2 has a large size exclusion profile due to unidentified stably associated proteins, MeCP2 was purified from rat brain.

3.3.2 The large molecular weight MeCP2 is stable over multiple biochemical purification steps

To biochemically purify MeCP2 and any stably associated factors, it was necessary to verify the stability of MeCP2 over a series of conventional chromatography steps. To obtain nuclear extract, rat brain nuclei were isolated by ultracentrifugation of brain homogenate over a sucrose cushion, followed by extraction of nuclear proteins with 400 mM NaCl. A three step purification scheme was devised (Figure 3.3A) including; Sp-Sepharose cation exchange, followed by a Mono-Q anion exchange, and finally a Ni-Nta affinity step. At each step of the purification fractions containing MeCP2 were identified by western blot analysis (data not shown). The nuclear extract was diluted to 100 mM NaCl and applied to a 1 ml Sp-Sepharose column and eluted with a linear gradient of NaCl from 100 mM to 1 M. The peak MeCP2 signal, as determined by western blot, was between 450-500 mM NaCl. The MeCP2 containing fractions were combined and dialysed into buffer containing 100 mM NaCl. The active fractions were then applied to a 1 ml Mono-Q column and MeCP2 was found in the flow through. The flow through

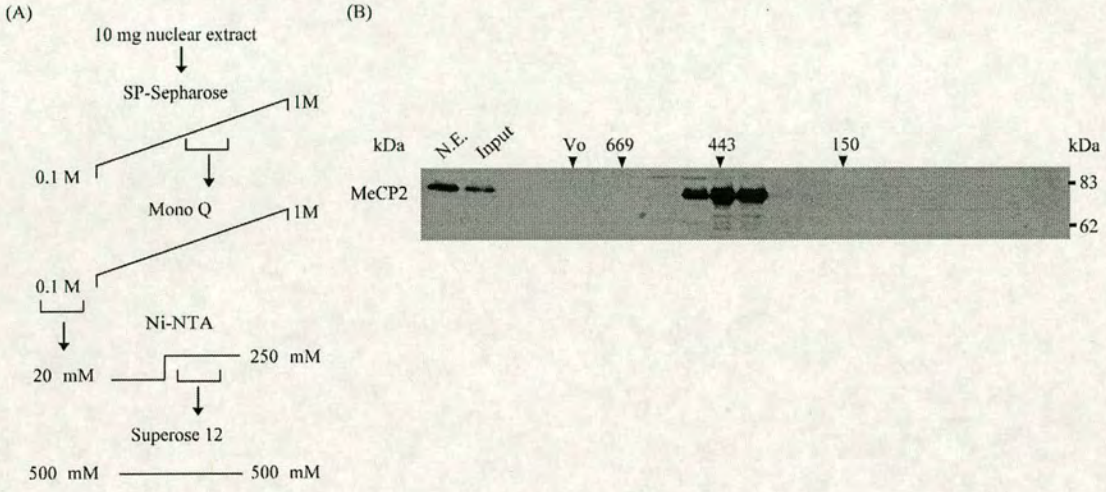


Figure 3.3- The 400-500 kDa MeCP2 is stable over multiple chromatographic steps.

(A) A four step pilot purification was devised to ensure the large molecular weight MeCP2 was not lost during multiple chromatographic steps. This purification involved Sp-Sepharose, Mono Q, Ni-Nta, and Superose 12 chromatographic steps. MeCP2 was followed by western blot analysis using MeCP2 specific antibodies. (B) After the first three chromatographic steps MeCP2 containing fractions were separated by size exclusion chromatography and the fractions tested by western blot using MeCP2 specific antibodies. Native molecular weight markers are indicated as arrows above the western blot. MeCP2 remained as a 400-500 kDa species over multiple purification steps.

was applied to a 0.5 ml Ni-Nta affinity resin. Ni-Nta resin was chosen at this step as MeCP2 has a run of seven histidines in the C-terminus which bind with high affinity to Ni-Nta resin. The Ni-Nta resin was washed extensively and eluted with 250 mM imidazole. MeCP2 eluted in the 250 mM fraction as determined by western blot. To evaluate whether MeCP2 still had a high molecular weight after the purification this fraction was applied to a Superose 12 size exclusion column and fractions collected within the inclusion limit (Figure 3.3B). MeCP2 eluted from the Superose 12 column with a relative molecular weight between 400-500 kDa, consistent with the native molecular weight of MeCP2 observed in unfractionated rat brain nuclear extract. This initial pilot purification demonstrates that the large native molecular weight MeCP2 species is maintained over several conventional biochemical purification steps and is amenable to large scale preparative biochemical purification.

3.3.3 Large scale purification of native MeCP2 from rat brain nuclear extract

To facilitate preparative purification of MeCP2 from rat brain nuclear extract large amounts of rat brain nuclei were isolated. Using approximately 400 rat brains (obtained from Pel-freez biological) as starting material, batches of 25 brains were pulverized into a fine powder under liquid nitrogen in a large mortar and pestle. The brain powder was mechanically homogenized and the nuclei purified by ultracentrifugation through a sucrose cushion. The resulting nuclei proved to be of high quality and were suitable for salt extraction of nuclear proteins. Using the pilot purification as a template, a modified purification scheme was devised including a Heparin affinity step to facilitate enhanced purification (Figure 3.4). Rat brain extract (140 mg) was dialysed into cation exchange buffer with 100 mM NaCl (CE 100) and applied to an 8 ml Sp-Sepharose column. The protein was eluted with a linear gradient from 100 mM NaCl (CE 100) to 1 M NaCl (CE 1000) and every second fraction was analyzed by western blot with MeCP2 antibodies. At the same time antibodies recognizing Sin3a and HDAC 1/2 were used to verify the initial observations that MeCP2 was not stably associated with the Sin3a chromatin remodeling complex (Figure 3.5A). Figure 3.5B shows the relative purification by total protein as measured by absorbance at 280 nm. In the first step, MeCP2

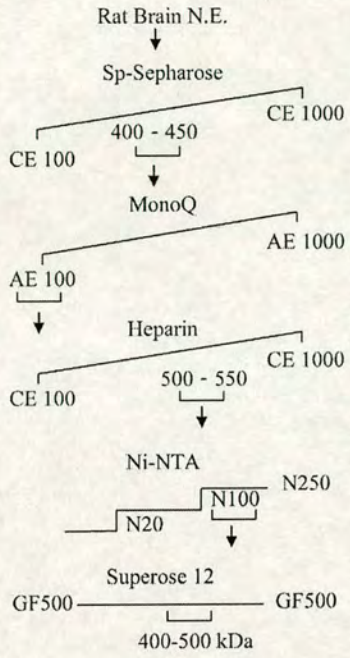


Figure 3.4- Preparative MeCP2 purification scheme.

A five step preparative purification scheme was devised to purify MeCP2 from 140 mg of rat brain nuclear extract. The purification included Sp-Sepharose, Mono-Q, Heparin-Sepharose, Ni-Nta, and Superose 12 chromatography steps.

eluted in a tight peak between 450-500 mM NaCl. Sin3a eluted over two peaks, one in the flow through/wash phase and a second major peak between 100mM NaCl and 450 mM NaCl. HDAC 1/2 eluted over a wide range of fractions and was found both in the flow through/wash phase and the NaCl elutions from 100mM to 500mM. The existence of several Sin3a and HDAC1/2 containing complexes fits with the wide elution profiles of these three proteins (Kuzmichev et al., 2002). Importantly, the majority of MeCP2 eluted in one major peak which was distinct from the peak Sin3a/HDAC 1/2 containing fractions (Figure 3.5A).

The MeCP2 containing fractions from the Sp-Sepharose column were combined and the protein dialyzed into anion exchange buffer containing 100mM NaCl (AE100). The dialyzed protein, 8.8 mg as determined by Bradford assay, was applied to a 1ml

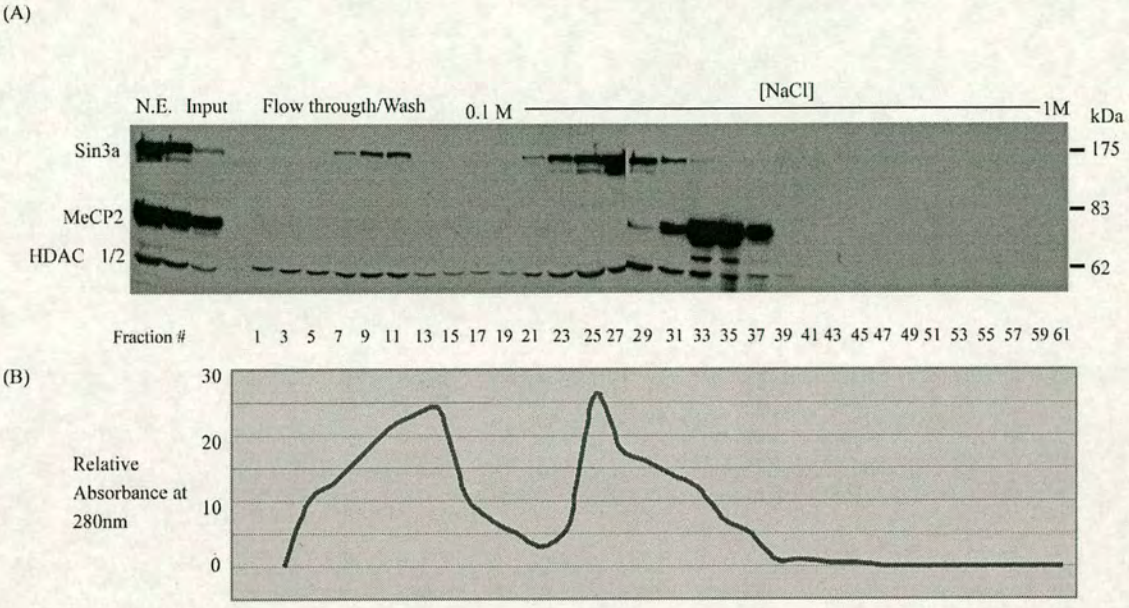


Figure 3.5- Sp-Sepharose chromatography of MeCP2.

(A) Rat brain nuclear extracts were separated on a Sp-Sepharose cation exchange column and the fractions probed by western blot with antibodies against Sin3a, MeCP2, and HDAC 1/2. The stage of the chromatography run (input, wash/flow-through, elution) is indicated above the western blot and the fraction number is marked below. MeCP2 bound tightly to the cation exchange resin (fractions 29-37) and these fractions are mostly devoid of Sin3a.
(B) The absorbance at 280 nm over the Sp-Sepharose purification was plotted to indicate the relative amount of total protein in each fraction.

Mono-Q column and eluted with a linear gradient of anion exchange buffer from 100 mM NaCl (AE100) to 1 M NaCl (AE 1000) (Figure 3.6A). Most of the MeCP2 was identified by western blot in the flow through/wash phase with a small amount of signal eluting at around 350-400 mM NaCl. This step of the purification revealed a complete separation of MeCP2 containing fractions from the residual Sin3a/HDAC1/2 that was brought through from the Sp-Sepharose cation exchange step. The remaining Sin3a and HDAC1/2 eluted from the Mono-Q column between 400-450 mM NaCl, completely devoid of detectable MeCP2 (Figure 3.6A). Therefore, the first two purification steps completely separate MeCP2 from the Sin3a/HDAC complexes, verifying the observation that in crude extracts most cellular MeCP2 does not exist in a stable complex with the Sin3a chromatin remodeling complex. Although MeCP2 does not bind to the Mono-Q resin, this proved to be a very useful purification step as is shown in Figure 3.6B. The flow through/wash phase is distinct from the major peak of protein as determined by absorbance at 280 nm (Figure 3.6B).

The MeCP2 containing fractions from the flow through/wash phase of the Mono-Q purification were combined and dialyzed into buffer containing 100 mM NaCl (CE100). The dialyzed protein, 4 mg as determined by Bradford assay, was applied to a 1ml Heparin affinity column and eluted with a linear gradient from 100 mM NaCl (CE100) to 1 M NaCl (CE 1000) (Figure 3.7A). The Heparin affinity step was added to the large scale preparative purification to enhance the overall purification of MeCP2. A Heparin affinity resin has two advantages which make it ideal as a 'polishing step' during the purification of MeCP2. First, Heparin mimics DNA in structure and has been shown to associate with DNA binding factors (Gadgil and Jarrett, 1999), and secondly the resin also contains a cation exchange moiety which was shown in the pilot purification to bind MeCP2 efficiently. Increased binding of MeCP2 to the Heparin Sepharose column is demonstrated by the requirement of higher ionic strength conditions (500-500 mM NaCl) to elute MeCP2 (Figure 3.7A) when compared to a cation exchange resin alone (450-500mM NaCl) (Figure 3.5A). Because Sin3a and HDAC 1/2 were completely separated from the MeCP2 containing fractions during the Mono-Q step, they were not assayed in subsequent

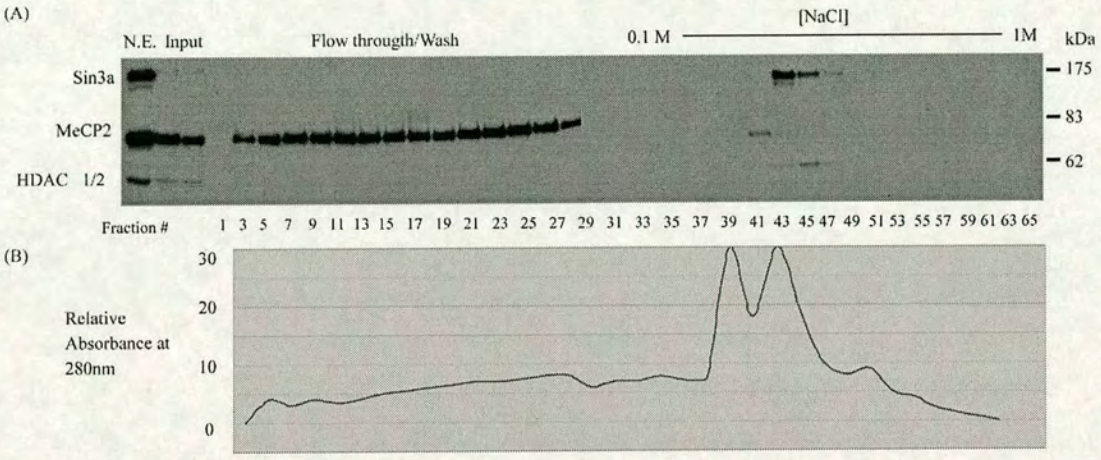


Figure 3.6- Mono-Q chromatography of MeCP2.

(A) The MeCP2 containing fractions from the Sp-Sepharose purification were dialysed and separated on a Mono-Q anion exchange column probing each fraction by western blot with antibodies against Sin3a, MeCP2, and HDAC 1/2. The stage of the chromatography run (input, wash/flow-through, and elution) is indicated above the western blot and the fraction number is marked below. MeCP2 did not bind the Mono-Q anion exchange resin (fractions 3-27), and completely separated MeCP2 away from the remaining Sin3a and HDAC 1/2.
(B) The absorbance at 280 nm over the Mono-Q purification was plotted to indicate the relative amount of total protein in each fraction.

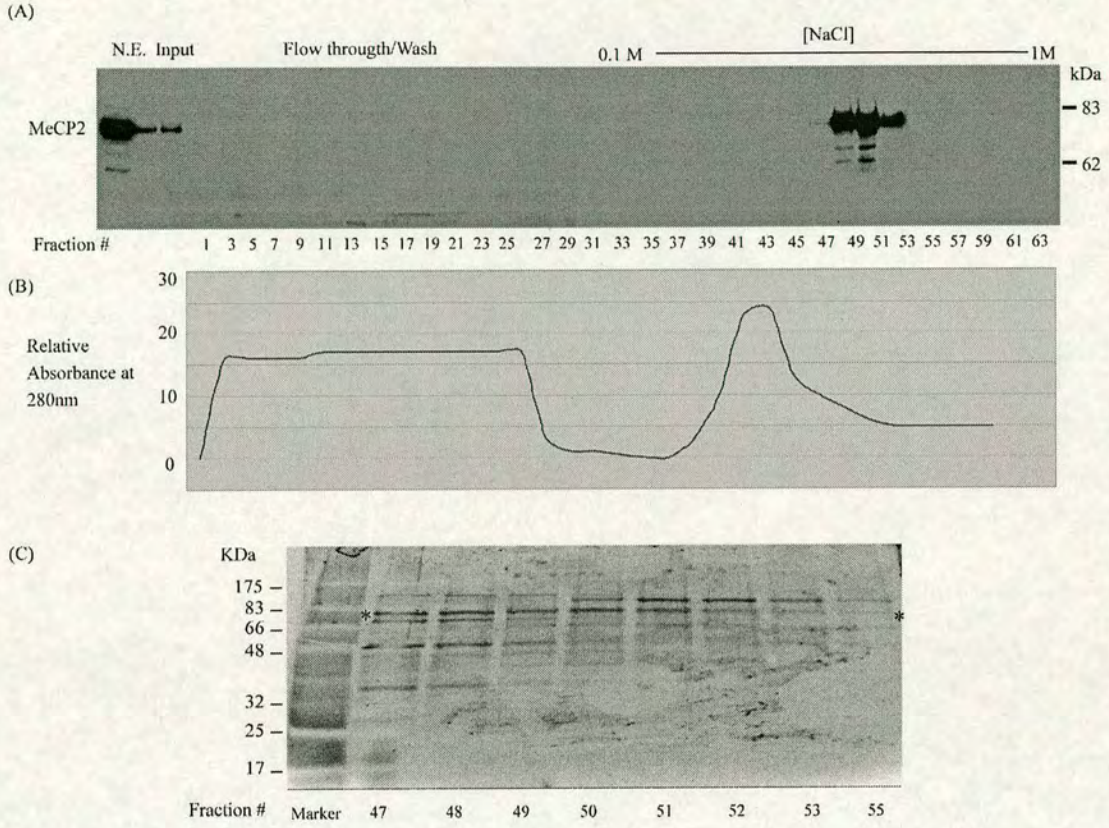


Figure 3.7- Heparin-Sepharose chromatography of MeCP2.

(A) The MeCP2 containing fractions from the Mono-Q purification were dialysed and separated on a Heparin-Sepharose affinity column and the fractions probed by western blot with antibodies against MeCP2. The stage of the chromatography run (input, wash/flow-through, and elution) is indicated above the western blot and the fraction number is marked below. MeCP2 bound tightly to the Heparin affinity resin (fractions 49-53).
(B) The absorbance at 280 nm over the Heparin purification was plotted to indicate the relative amount of total protein in each fraction.
(C) An aliquot of fractions 47-53 was separated by SDS PAGE and Sypro-ruby stained with MeCP2 indicated by an asterisk, and the molecular weight markers in kDa displayed left.

western blot analysis. The increased affinity of MeCP2 for the Heparin column proved an efficient purification step as indicated by the absorbance trace at 280 nm (Figure 3.7B). On the Heparin column the majority of the contaminating proteins from the Mono-Q step elute outside of the MeCP2 containing fractions, as is evident from the total protein absorbance trace at 280 nm (Figure 3.7B). To evaluate the progress of the purification, fractions 47 – 55 containing MeCP2 were separated by SDS PAGE and stained for total protein using Sypro-Ruby stain (Figure 3.7C). Sypro-Ruby has advantages over traditional silver stain as it is more sensitive and less biased towards preferential interactions with specific amino acids in the protein (White et al., 2004). Fractions 47-55 are highly enriched for MeCP2 (as indicated by an asterisk) and contain various other proteins of different intensities and unknown identity.

The MeCP2 containing fractions were combined and dialyzed into cation exchange buffer containing 100 mM NaCl (CE100) omitting EDTA and DTT. The dialyzed protein, 0.408 mg as determined by Bradford assay, was applied to a 0.5 ml Ni-Nta affinity column and washed extensively with Ni-Nta wash buffer containg 20 mM imidazole (N20). The column was then eluted in batch with Ni-Nta elution buffer containing 100 mM imidazole (N100) followed by Ni-Nta elution buffer containing 250 mM imidazole (Figure 3.8A). The MeCP2 containing fractions were again monitored by western blot with MeCP2 antibodies. MeCP2 protein was lost at this step in the flow through and washes but the vast majority of MeCP2 eluted during the 100 mM imidazole elution step (Figure 3.8A). The Ni-Nta purification step proved to be very efficient in isolating MeCP2 as indicted by analysis of the peak MeCP2 fractions on a Sypro-Ruby stained gel (Figure 3.8B). The peak MeCP2 containing fractions from the Ni-Nta elution consist of only three major polypeptides, with the asterisk indicating MeCP2. The purity of the MeCP2-containing sample has improved in comparison to the more crude MeCP2-containing fractions from Heparin affinity step (Figure 3.7C). To evaluate the purification to this

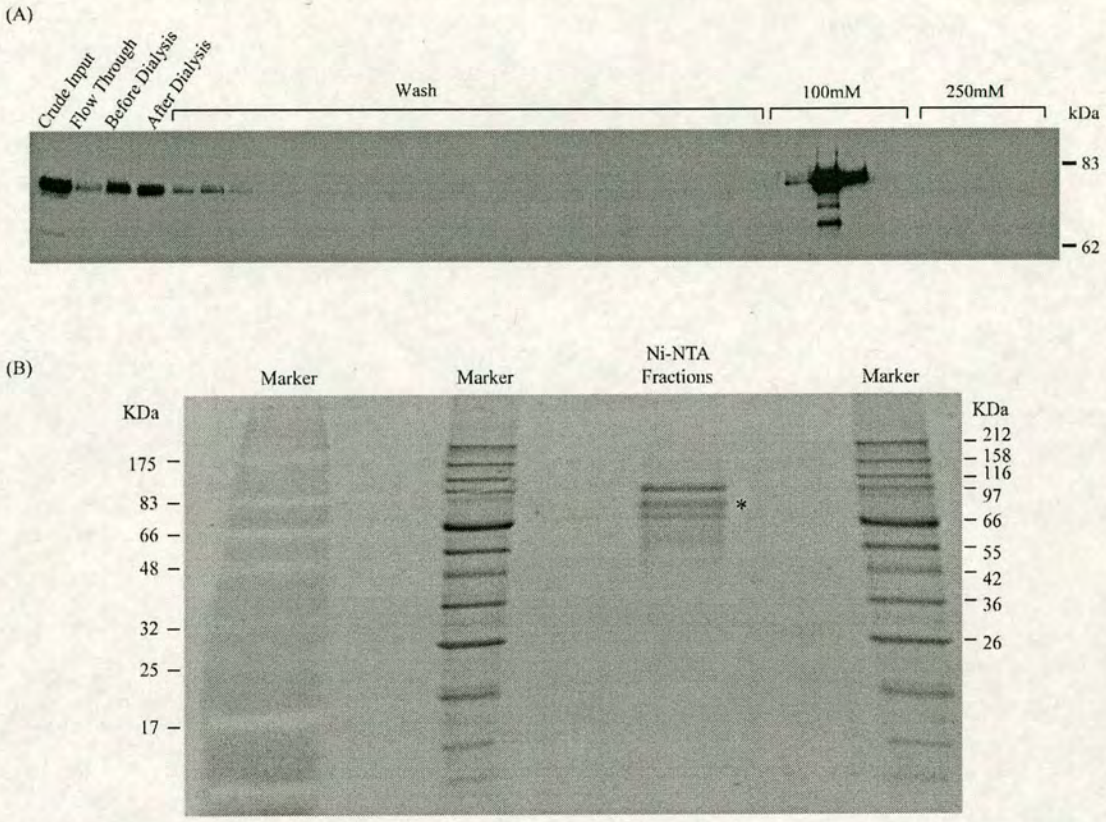


Figure 3.8- Ni-Nta chromatography of MeCP2.

(A) The MeCP2 containing fractions from the Heparin-Sepharose column separated on a Ni-Nta affinity column and the fractions probed by western blot with antibodies against MeCP2. The stage of the chromatography run (input, wash/flow-through, and elution) is indicated above the western blot. MeCP2 bound tightly to the Ni-Nta affinity resin and eluted with 100 mM imidazole.
(B) The 100 mM imidazole elutions were combined and an aliquot was separated by SDS PAGE and Sypro-ruby stained. The Ni-Nta purification resulted in three polypeptides with MeCP2 indicated by an asterisk. The molecular weight in kDa is displayed by two different molecular weight markers left and right

point, aliquots of each purification step were separated by SDS PAGE and Sypro-Ruby stained. To the left of the protein gel is a table indicating the total amount of protein at each purification step (Figure 3.9A). MeCP2 is enriched throughout the purification.

To determine whether MeCP2 remained in a high molecular weight fraction over the purification and whether this large molecular weight fraction resulted from stably associated proteins, the Ni-Nta fractions (0.148 mg of protein) containing MeCP2 were applied to a Superose 12 size exclusion column and eluted in gel filtration buffer containing 500 mM NaCl (GF500) (Figure 3.9B). It was found to be essential at this step to elute the small amount of total protein from the Superose 12 size exclusion column using high salt to inhibit non specific retention on the column matrix. Importantly the purified MeCP2 activity was the same native molecular weight (400-500 kDa) as was observed in the input nuclear extract (Figure 3.9B). This suggests that native MeCP2 has remained intact over the purification. A portion of each size exclusion fraction was separated by SDS PAGE and stained with Sypro-Ruby stain. Surprisingly, all three proteins eluted in different peak fractions on the size exclusion column suggesting that they are not part of a multi-protein complex (Figure 3.9B). The smallest polypeptide eluted in fractions devoid of MeCP2, and this protein has not successfully been identified by mass spectrometry. The middle band and the upper band were excised from the gel and peptides isolated by in-gel tryptic digestion and time of flight mass spectrometry. Peptide fingerprinting revealed the upper band as topoisomerase I and the middle band is MeCP2. The absorbance at 280 nm plotted below the Sypro-ruby stained gel indicates that MeCP2 and topoisomerase I account for the majority of protein in these fractions (Figure 3.9C). To verify that MeCP2 does not physically interact with topoisomerase I, MeCP2 was immunoprecipitated from nuclear extracts and topoisomerase I protein identified with anti-topoisomerase I antibodies. Figure 3.9D demonstrates that topoisomerase I does not associate with MeCP2 in rat brain nuclear extracts. Although MeCP2 does not stably associate with other proteins it still has a large native molecular weight as determined by

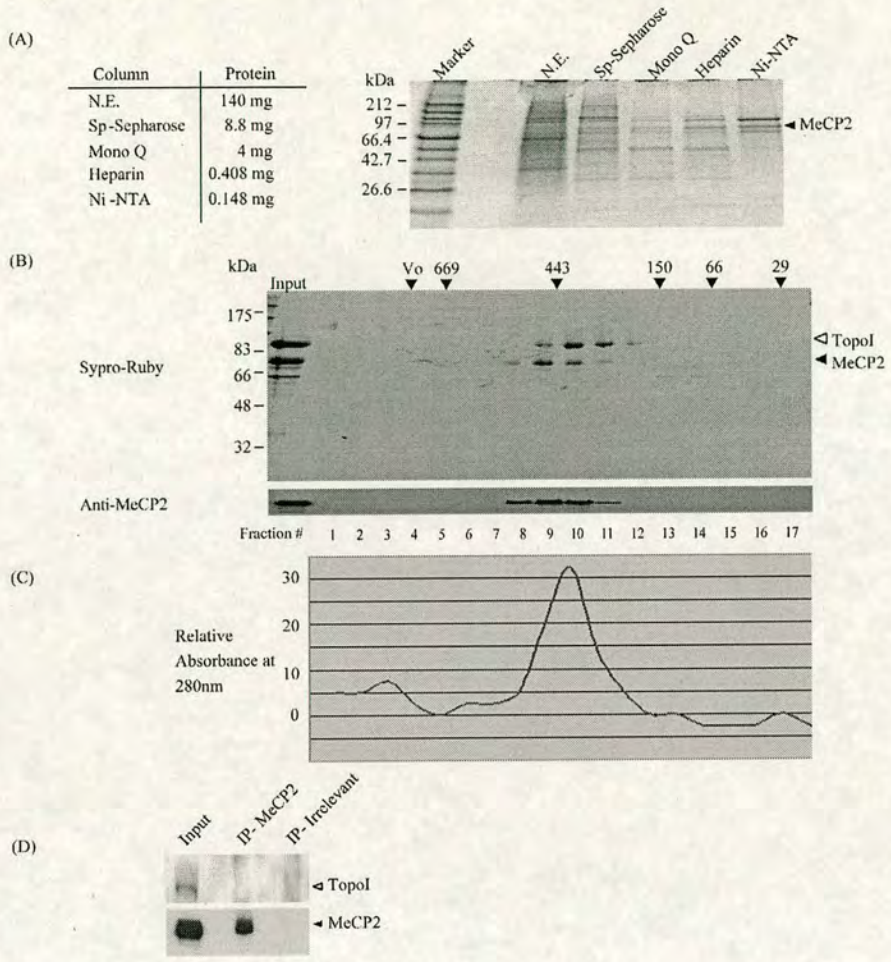


Figure 3.9- Superose 12 size exclusion chromatography and identification of MeCP2.

(A) Aliquots of each purification stage were separated by SDS PAGE and Sypro-ruby stained. Above the Sypro-ruby stained gel is the purification step, to the left is molecular weight markers in kDa, and to the right MeCP2 is indicated by an arrow. The table to the left indicates the total protein concentration at each step in milligrams as determined by Bradford assay. MeCP2 is consistently enriched and purified through the first four purification steps.
(B) An aliquot of the Ni-Nta fraction was separated by Superose 12 size exclusion chromatography and the fractions analysed by western blot analysis using MeCP2 specific antibodies (bottom). MeCP2 eluted in fractions 8-12. The same Superose 12 fractions were separated by SDS-PAGE and analysed by Sypro-ruby staining (top). The two largest proteins were identified by mass spectrometry to be topoisomereaseI and MeCP2 as indicated (right) with arrows.
(C) The absorbance at 280 nm over the Superose 12 size exclusion step was plotted to indicate the relative amount of total protein in each fraction.
(D) MeCP2 was immunoprecipitated from rat brain nuclear extract with MeCP2 specific antibodies and probed by western blot with anti-topoisomerase antibodies. MeCP2 failed to immunoprecipitate topoisomeraseI specific signal suggesting it is a contaminant in the purification.

size exclusion chromatography. The biochemical and biophysical explanation for this observation is determined and discussed in Chapter four.

3.4 Discussion

3.4.1 Mechanisms of MeCP2 mediated repression

DNA binding transcription factors mediate their effects independently or through association with other protein co-factors, and in some cases, through a combination of both mechanisms (Ogbourne and Antalis, 1998; Thiel et al., 2004). In *in vitro* transcription assays, MeCP2 represses transcription (Nan et al., 1997a). This function is likely mediated through interaction with TFIIB, a component of the basal transcriptional machinery (Kaludov and Wolffe, 2000). In addition to interfering with basal transcription, many transcriptional repressors utilize large protein complexes that contain enzymatic activities capable of modulating chromatin structure to repress transcription (Bird and Wolffe, 1999; Lusser and Kadonaga, 2003; Peterson and Laniel, 2004). Using a candidate approach, MeCP2 was found to interact with the mammalian Sin3a protein by co-immunoprecipitation (Nan et al., 1998). The mammalian Sin3a protein was identified by its homology with the yeast repressor sin3 (Ayer et al., 1995; Schreiber-Agus et al., 1995) and several groups biochemically purified Sin3a associated with a host of protein co-factors. Histone deacetylases, HDAC1 and HDAC2, were identified as stable components of the Sin3a chromatin remodeling complex (Alland et al., 1997; Laherty et al., 1997; Zhang et al., 1997) and contribute to the ability of Sin3a to modulate gene expression in a chromatin context. Importantly, MeCP2 was shown to associate with an intact Sin3a co-repressor complex as both HDAC1 and HDAC2 bound MeCP2 in an enzymatically active state. The contention that MeCP2 existed as a stable component of the Sin3a co-repressor complex was the result of work studying the *Xenopus laevis* xMeCP2 protein. A combination of the observations from studies in mammalian and *Xenopus* sources has led to the belief that MeCP2 is intimately associated with Sin3a and that this complex is the sole mediator of active transcriptional repression by MeCP2.

3.4.2 MeCP2 does not stably associate with the Sin3a chromatin remodeling complex in nuclear extracts

Using an unbiased approach, the work described in this chapter evaluates the MeCP2/Sin3a association in mammalian nuclear extracts. By repeating co-

immunoprecipitation experiments in rat brain nuclear extracts it was shown that very little cellular Sin3a is associated with MeCP2. Size exclusion chromatography was also used to analyze the MeCP2/Sin3a interaction in nuclear extracts and it became obvious from these experiments that MeCP2 was not found in fractions containing the Sin3a co-repressor complex. This observation fits with the fact that no group has identified MeCP2 as a component of a Sin3a chromatin remodeling complex despite several independent purifications from different cellular sources (Ahringer, 2000). To try and resolve the contradiction regarding the nature of the MeCP2/Sin3a interaction in mammals and amphibians newly generated xMeCP2 and xSin3a antibody reagents were obtained. By analyzing xMeCP2 and xSin3a in oocyte extracts, it was determined that these two molecules do not stably associate. Our observations are in agreement with those from Ryan et al who separated oocyte extract by glycerol gradient fractionation and found peak xMeCP2 and xSin3a fractions were distinct (Ryan et al., 1999). In light of recent evidence in both mammalian and amphibian systems it is apparent that MeCP2 is not a stable component of the Sin3a complex, but does weakly associate with Sin3a in a manner that can be observed by co-immunoprecipitation.

Purification of MeCP2 from rat brain failed to identify other stably associated co-factors but left questions as to the cause of its larger than expected apparent molecular weight by size exclusion chromatography. MeCP2 is a 52.4 kDa protein but elutes from size exclusion column with a relative molecular weight of nearly ten times its predicted size. Biochemical and biophysical analysis of MeCP2 in Chapter four addresses this discrepancy.

3.4.3 The contribution of other MeCP2 co-factors in transcriptional repression

Although MeCP2 does not form stable interactions with Sin3a in DNA free extracts, there is evidence that they interact on DNA. Chromatin immunoprecipitation assays show certain methylated genes recruit Sin3a and/HDAC1 associated proteins in an MeCP2 dependent fashion (Chen et al., 2003b; Horike et al., 2005; Martinowich et al., 2003). Interestingly, a series of other protein factors have also been shown to interact with MeCP2 including; Suv39H1, c-Ski, DNMT1, Co-Rest, LANA, PU1, Brm, Splicing factors, and RNA (Buschdorf and Stratling, 2004; Harikrishnan et al., 2005; Jeffery and Nakielny, 2004; Kimura and Shiota, 2003; Kokura et al., 2001;

Krithivas et al., 2002; Lunyak et al., 2002; Suzuki et al., 2003). The relative contribution of these factors to MeCP2-mediated repression and cellular function is poorly understood and in most cases remains to be thoroughly investigated. The realization that MeCP2 is not a stable biochemical component of the Sin3a chromatin remodeling complex leaves open the possibility that other factors contribute to the normal function of MeCP2. This is exemplified by the recent observation that MeCP2 associates with histone methyltransferase activity, and that cell lines over expressing MeCP2 contain increased histone methylation at loci containing DNA methylation (Fuks et al., 2003b). Further investigation of alternative MeCP2 co-factors may help to elucidate the repressive capabilities and cellular functions of MeCP2.

3.4.4 Distinct mechanisms of methylation mediated repression?

As depicted in Figure 3.10A, MeCP2 is the only MBD protein that has been isolated principally devoid of stably associated co-factors. Other MBP's are all stably associated with co-factors; MBD1 stably associates with SetDB1 (Sarraf and Stancheva, 2004), Kaiso is a component of NCoR, and MBD2 is part of the MeCP1 complex (Feng and Zhang, 2001; Yoon et al., 2003). The fact that MeCP2 is isolated independent of associated co-factors suggests a possible mechanistic difference in the way MeCP2 functions to read DNA methylation, and perhaps increased flexibility once bound to DNA to repress transcription. Because MeCP2 is not part of a stable complex, it can bind methylated sites and then repress transcription either independently, (Option A, Figure 3.10B), or by subsequent recruitment of co-factors like the Sin3a chromatin remodeling complex, (Option B, (Figure 3.10B). Recruitment of MeCP2 to regulatory regions free of pre-determined co-factors lends flexibility as MeCP2 can repress transcription in a context-specific, or even, gene specific-manner, perhaps recruiting different co-factors under unique situations; (Option C (Figure 3.10B). This starkly contrasts with the rigid nature of repression directed by other complex-bound MBD proteins. MBD1, Kaiso, and MBD2 containing complexes arrive at methylated CpG's ready to modify chromatin and repress transcription immediately, using a generalized and predefined set of repressive tools. The protein composition of complex bound MBP's may relate to the context in

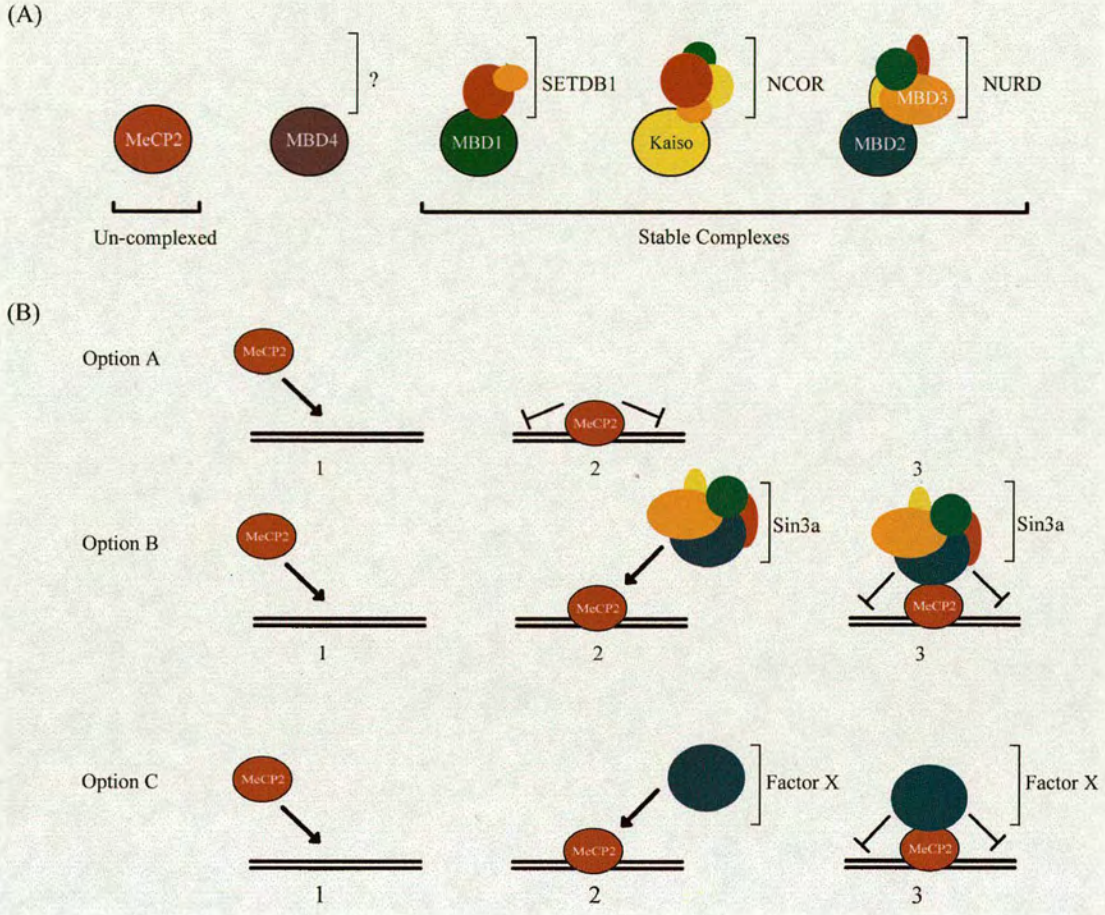


Figure 3.10- Flexibility and diversity of MeCP2 function *in vivo*?

(A) MeCP2 is the only MBD containing protein so far that has been purified in the absence of stably associated co-factors.
(B) A schematic demonstrating three of the potential ways MeCP2 may function to repress transcription *in vivo*. Option A, MeCP2 binds methylated DNA and uses intrinsic repressive capabilities. Option B, MeCP2 binds methylated DNA and recruits the Sin3a chromatin remodelling complex to repress transcription and remodel surrounding chromatin. Option C, MeCP2 binds methylated DNA and recruits other factors to mediate repression in a gene specific or context dependent manner.

which each represses transcription, and the fact that MeCP1 related complexes are found in lower organisms suggest that they may have evolved this way (Marhold et al., 2004a; Wade et al., 1999). MBD1 is recruited by CAF1 to active DNA replication forks coupling histone H3 lysine 9 methylation via SETDB1 to target genes with DNA methylation (Sarraf and Stancheva, 2004). Similarily, Kaiso/NCoR and MeCP1 (MBD2) may recruit stably associated co-factors to genes which the complex is specifically tailored to repress (Hendrich et al., 2001; Hutchins et al., 2002; Ruzov et al., 2004). Although MeCP2 has more flexibility in co-repressor recruitment, co-factor bound MBD proteins have at least one specific advantage in that they do not rely on a two step mechanism, binding then recruitment, to impart silencing.

Interestingly, MeCP2 has different biochemical attributes than other MBP's. Mild extraction conditions (300-350 mM NaCl) recover MBD1 and Kaiso (Helle Jorgensen, personal communication), and also MBD2 (Meehan et al., 1992) from nuclei. The ease by which complexed MBD proteins can be isolated from nuclei suggests they are not tightly associated with chromatin. This fluidity may reflect a requirement for complex-bound MBP's to constantly associate and dissociate with a variety of sites that they target for repression to maintain and re-enforce the repressed state. In contrast, MeCP2 is tightly bound to chromatin and requires more harsh conditions for efficient release from isolated nuclei (>400 mM NaCl)(Meehan et al., 1992). MeCP2 might use a different strategy than complexed MBP's, by binding tightly to specific loci and re-inforcing silencing by recruiting and then releasing co-factors. Tight binding, coupled to potentially promiscuous recruitment of co-factors might suit the functional requirement of MeCP2 to exploit different co-factors to repress different classes of genes (Figure 3.10). If the diverse set of co-factors that MeCP2 can utilize to repress transcription in turn permits MeCP2 to affect a range of target genes this could partially account for the phenotypic severity of MeCP2 null mice in comparison to other MBP null mice. Certain functions and cofactors that MeCP2 utilizes may be required to silence a defined sub-set of MeCP2 target genes which cannot be modulated by complexed MBP's. In agreement with this contention, transcriptional profiling in MeCP2 null mice have revealed little evidence to support the notion that MeCP2 is a global methylation dependent repressor of transcription (Tudor et al., 2002), as one might expect of a methyl-CpG binding protein. However, MeCP2 has been shown to have gene-specific roles in transcriptional repression of the

Bdnf, *Ube3a*, and *Dlx-5/6* genes (Chen et al., 2003b; Horike et al., 2005; Martinowich et al., 2003). Functional analysis of MeCP2 target genes and evaluation of redundancy between MBD proteins will begin to address the individual functions of MBP proteins.

4. Chapter four-
Biophysical analysis of MeCP2

4.1 MeCP2 has a large elution profile by size exclusion chromatography

When methyl-CpG binding activities were originally identified, MeCP1 was analysed by size exclusion chromatography and eluted in a high molecular weight fraction (Meehan et al., 1989). Purification and identification of the proteins in MeCP1 revealed a multi-protein complex containing MBD2 as the active methyl-CpG binding protein (Feng and Zhang, 2001; Ng et al., 1999). In Chapter 3 a purification of native rat brain MeCP2 showed that MeCP2 does not stably associate with other proteins, yet has a large apparent molecular weight by size exclusion chromatography. Because MeCP2 is predicted to encode a 52.4 kDa protein, it was hard to reconcile this large elution profile (between 400-500 kDa) based on the predicted molecular weight of MeCP2 alone. In this chapter three possibilities were adressed to explain this discrepancy; 1- MeCP2 has a DNA or RNA component which was not detected in the biochemical purification, 2- MeCP2 self associates and forms a large molecular weight complex, or 3-MeCP2 is an abnormally shaped molecule.

4.2 Biophysical analysis of MeCP2

4.2.1 Elution profile of native MeCP2 is unaffected by nuclease treatment.

To test whether RNA or DNA components contribute to the large relative molecular weight of MeCP2, an aliquot of purified rat brain MeCP2 protein was treated with Benzonase (Moreno et al., 1991). Benzonase is a non-specific nuclease that will efficiently degrade RNA and DNA even in complex mixtures. After treatment with Benzonase the protein sample was separated by size exclusion chromatography on a Superose 6 column and the fractions analyzed by western blot with MeCP2 specific antibodies (Figure 4.1). Treatment with nuclease had no affect on the mobility of the purified material and the elution profile mirrored that of the untreated sample (Figure 3.1C). This observation suggests that an RNA or DNA does not contribute to the

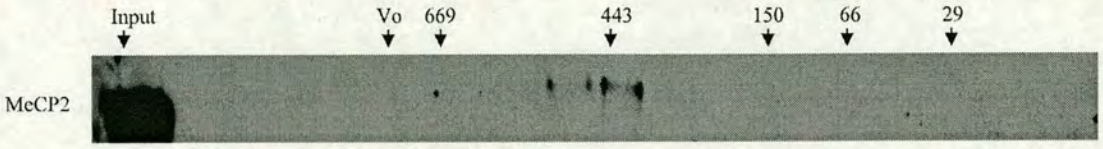


Figure 4.1- Nucleic acid does not contribute to the large molecular weight of native MeCP2.

MeCP2 purified from rat brain was treated with Benzonase and separated by size exclusion chromatography on a Superose 12 column. MeCP2 was identified by western blot analysis using MeCP2 specific antibodies. Native molecular weight markers are displayed above the blot in kDa.

abnormal size exclusion profile of native MeCP2. In light of a recent report that MeCP2 associates with various forms of RNA, it is informative that RNA does not contribute greatly to the native size exclusion profile of MeCP2 (Jeffery and Nakielny, 2004).

4.2.2 Producing preparative amounts of native MeCP2 for biophysical analysis

Although native MeCP2 from rat brain was efficiently purified in micro-gram quantities, much larger amounts of native MeCP2 protein were required to carry out biophysical analysis and determine the cause of the aberrant size exclusion profile of MeCP2. To produce preparative amounts of MeCP2 material, both a baculovirus expression system and a prokaryotic expression system were utilized. In both systems MeCP2 was expressed untagged and purified by a combination of affinity chromatography and conventional biochemistry.

Baculovirus expression of mammalian proteins has proved very convenient for obtaining full length proteins that are difficult to express in bacterial systems. Historically, full length MeCP2 produced in bacteria as a GST-fusion suffered from proteolytic degradation (Xinsheng Nan, personal communication), so baculovirus expression was chosen to overcome this problem. Untagged human MeCP2 was cloned into the fastbac1 baculovirus shuttle vector from Invitrogen (Figure 4.2). The

Invitrogen system takes advantage of site specific recombination to produce a recombinant baculovirus genome in bacteria (Luckow et al., 1993). This system relies on cloning the gene of interest, in this case MeCP2, into a shuttle vector followed by transformation into bacterial cells containing a recombination proficient viral genome (Figure 4.2). Once recombination is induced in bacteria, only the recombined viral genome is maintained through selection for a co-integrated marker gene (gentR, Figure 4.2). The viral DNA is then purified and transfected into SF9 insect cells which produce recombinant protein while also assembling and budding infectious recombinant viral particles. Recombinant viral particles can be isolated from the transfected SF9 cells and subsequently used as a vehicle for further SF9 cell infection and recombinant protein expression.

Recombinant baculovirus DNA was transfected into SF9 cells and viral particles were collected (P1 viral stock). To ensure the recombinant virus was producing MeCP2, the transfected cells were harvested and protein expression assayed by western blot using MeCP2 specific antibodies. Strong MeCP2 western blot signal was observed in the transfected cell extracts and no signal was observed in the untransfected SF9 line (Figure 4.3A). The P1 viral stock was amplified and used for large scale infection of SF9 cells in spinner culture. SF9 insect cells (~ 10^9 cells) were infected with amplified MeCP2 virus and whole cell extract obtained (Figure 4.3B). The clarified extract was applied to Ni-Nta resin and eluted in 250 mM imidazole. A strong band at approximately 80 kDa, corresponding to MeCP2, was observed in the first two elutions from the Ni-Nta column (Figure 4.3B). MeCP2 containing fractions were pooled, applied to a Sp-Sepharose cation exchange column, and eluted with a linear gradient of increasing NaCl concentration from 0.2 M to 1 M. The majority of MeCP2 eluted over four fractions (Figure 4.3B) which were dialyzed into storage buffer and the concentration determined by a combination of Bradford assay and visual quantification by Coomassie staining of SDS-PAGE gels. The production of MeCP2 in baculovirus proved efficient, as purified MeCP2 protein

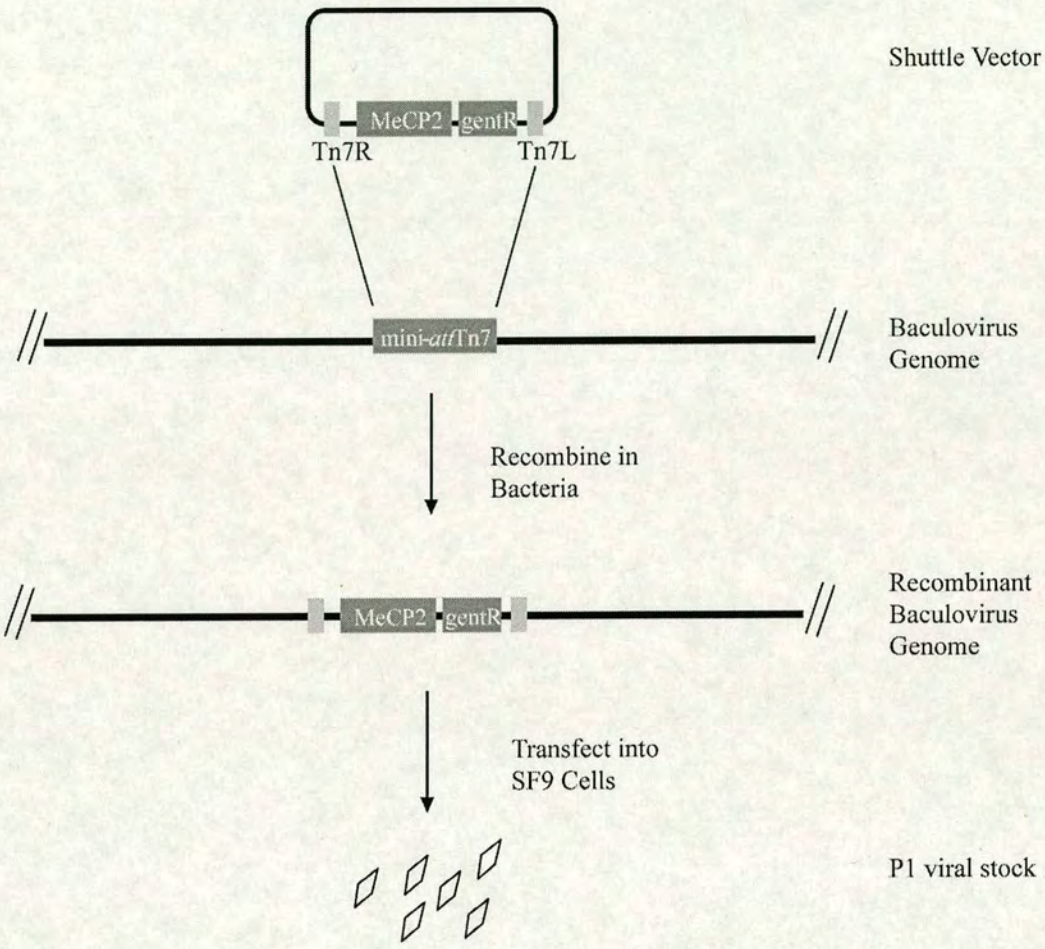


Figure 4.2- Generation of untagged MeCP2 expressing baculovirus.

MeCP2 expressing baculovirus was generated by cloning MeCP2 into a baculovirus shuttle vector. The shuttle vector has recombination sites (Tn7R/L) flanking MeCP2 and a selectable marker (gentR). The shuttle vector was transformed into bacteria that contain a baculovirus genome with a mini*att*Tn7 recombination site. Bacteria containing the recombined MeCP2 baculovirus genome were selected for by the incorporated antibiotic resistance cassette (gentR). The baculovirus genome was purified from bacteria and transfected into SF9 cells to produce P1 viral stock and verify MeCP2 expression.

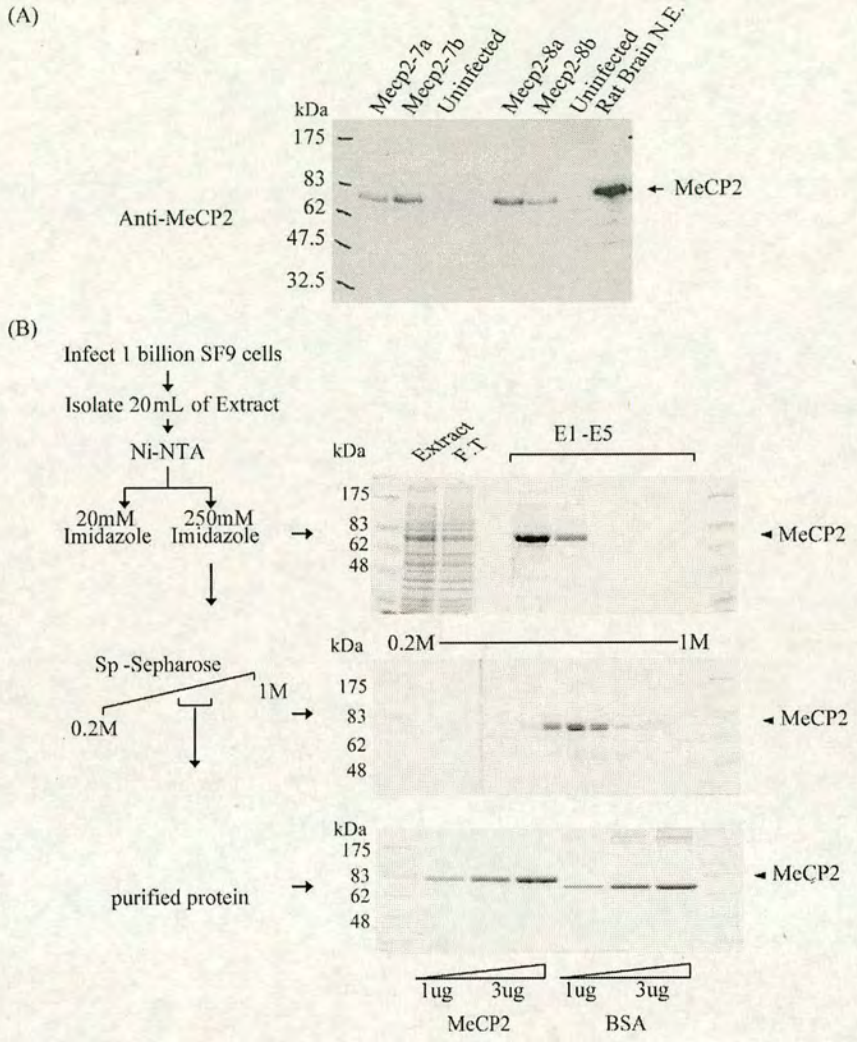


Figure 4.3- Production of baculovirus MeCP2.

(A) Western blot analysis for MeCP2 in whole cell extracts from transfected SF9 cells. Recombinant MeCP2 baculovirus genome isolates (7 and 8) were transfected into SF9 cells in duplicate (a and b).
(B) Amplified viral stock was used to infect one billion SF9 cells and 20 mL of clarified whole cell extract was applied to a Ni-Nta column. Unbound proteins were washed away and MeCP2 was eluted with 250 mM imidazole. Elutions (E1-5) were analysed by Coomassie stained SDS PAGE. MeCP2 containing fractions E1 and E2 were applied to a Sp-Sepharose cation exchange column and eluted with a linear gradient of NaCl from 0.2 M to 1M. MeCP2 eluted between 450-550 mM NaCl as determined by Coomassie stained SDS PAGE. MeCP2 containing fractions were dialysed into storage buffer and the protein concentration determined by Bradford and Coomassie stained SDS PAGE.

(1.3mg) was virtually all full length. Therefore, expression of MeCP2 in baculovirus overcomes some of the problems associated with production of full length MeCP2 in bacteria (Figure 4.3B).

When applied to a Superose 12 size exclusion column, the baculovirus-produced MeCP2 protein showed the same elution profile as the native rat brain material, suggesting that this may be a useful source of protein for biophysical studies (Figure 4.4A). During the purification of MeCP2 from SF9 cells, the presence of a MeCP2 doublet was evident on SDS PAGE gels (Figure 4.4A). Because insect cells contain a repertoire of eukaryotic kinases, it was predicted that this mobility difference might exist due to phosphorylation of the MeCP2 protein. To test this possibility, baculovirus produced protein was treated with lambda phosphatase and separated by SDS PAGE (Figure 4.4B). The phosphatase treated protein had increased mobility when compared to untreated material indicating loss of phosphorylation (Figure 4.4B). As a control, the de-phosphorylation reaction was carried out in the presence of phosphatase inhibitors which did not affect the mobility of treated MeCP2 (Figure 4.4B). Because native MeCP2 purified from rat brain did not migrate as doublet, this modification could arise through different kinase substrate preference in insect cells. Inappropriate phosphorylation of MeCP2 in insect cells could interfere with normal protein function, therefore the baculovirus produced material was held in reserve so that MeCP2 produced in bacteria could be investigated further.

Intact, full-length GST-MeCP2 protein is difficult to purify from bacteria. For this reason it would be advantageous to have a bacterial expression system in which MeCP2 could be isolated as a full length protein devoid of tags. To this end, the His/S-tag was removed from the commercial prokaryotic expression vector pet30b, and MeCP2 was cloned in-frame with the translational start site of the vector. To overcome any expression problems that might arise because of differential eukaryotic and prokaryotic codon usage, the MeCP2 expressing plasmid was transformed into *E. coli* BL21 cells that express rare human tRNA codons. Protein was produced by induction of MeCP2 expression at 30 °C for 3 hours, and whole cell bacterial lysate was made by sonication followed by removal of insoluble material by

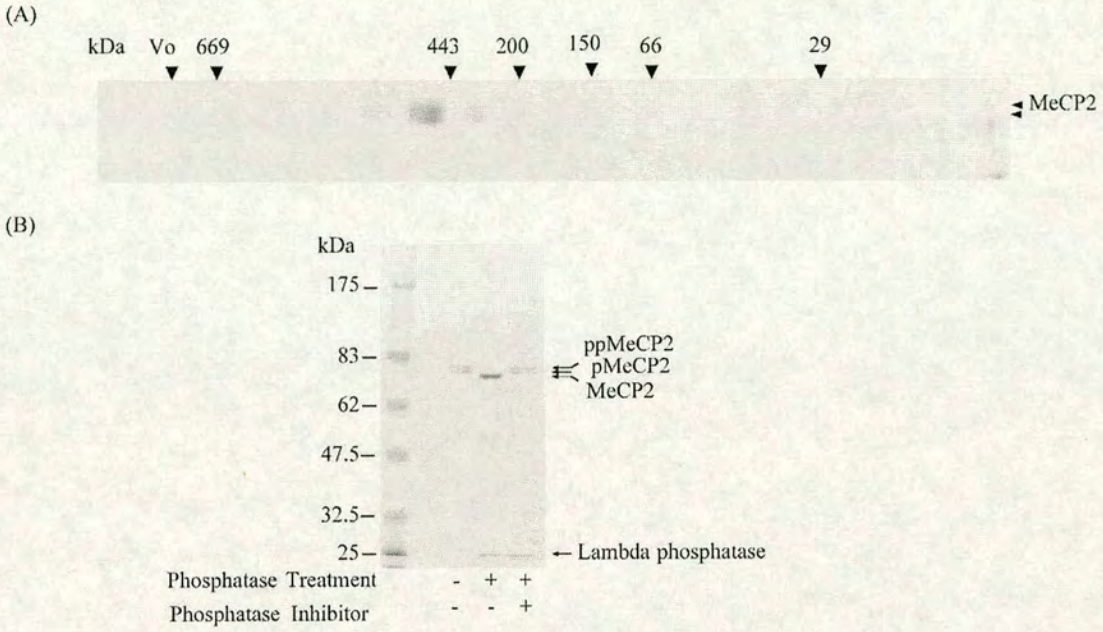


Figure 4.4- Baculovirus MeCP2 is a phosphoprotein

(C) Baculovirus produced MeCP2 was separated by Superose 12 size exclusion chromatography and analysed by Coomassie stained SDS PAGE. Baculovirus produced MeCP2 has the same apparent molecular weight as native rat brain MeCP2. (D) Baculovirus produced MeCP2 is a phosphoprotein. MeCP2 was purified as a doublet from SF9 cells and this doublet is lost by treating the protein with lambda phosphatase as observed by Coomassie stained SDS PAGE. Phosphatase inhibitors alleviate the increase in mobility observed by phosphatase treatment

centrifugation. Clarified lysate was applied to a Ni-Nta column followed by extensive washing of the bound MeCP2 protein. MeCP2 was eluted from the Ni-Nta column with 250 mM imidazole (Figure 4.5A). The MeCP2 containing fraction from the Ni-Nta column was applied to a Sp-Sepharose cation exchange column and MeCP2 was eluted with a linear NaCl gradient from 0.2 M to 1 M. MeCP2 eluted between 400-450 mM NaCl and these fractions were then directly applied to a Sephacryl S-300 HR 26/60 column. MeCP2 eluted from the Sephacryl S-300 HR 26/60 column between 400-500 kDa. The MeCP2 containing fractions from the Sephacryl S-300 column were dialyzed into buffer containing 200 mM NaCl and applied to a Mono S cation exchange column. A linear gradient from 0.2 M to 1 M NaCl was then applied to the Mono S column and MeCP2 eluted between 400 and 450 mM NaCl. The MeCP2 containing fractions were pooled and dialyzed into storage buffer. MeCP2 purity was >95% when analysed by Coomassie stained SDS-PAGE (Figure 4.5A). After extensive purification the MeCP2 yield approached milligram-per-liter quantities, indicating the efficiency of this expression regime. To determine whether the bacterially produced, untagged, full length MeCP2 behaved like the endogenous rat brain protein, purified bacterial MeCP2 was separated on a Superose-12 size exclusion column (Figure 4.5B). As was previously observed for endogenous rat brain MeCP2, the bacterially produced MeCP2 protein eluted between 400-500 kDa. As the bacterial protein had the same size exclusion profile as the endogenous protein, it was necessary to verify that the protein was functional. Since MeCP2 is a methyl-CpG binding protein, its ability to bind methylated DNA by EMSA was tested (Figure 4.5C). MeCP2 was incubated with a probe containing multiple-CpG dinucleotides that were fully methylated using the prokaryotic methyltransferase M.*SssI*. Bacterially produced MeCP2 specifically shifted the methylated probe and not its unmethylated counterpart (Figure 4.5C). The methyl-specific EMSA could be competed with excess unlabelled methylated DNA but not unmethylated DNA, verifying the specificity of the observed shift (Figure 4.5C). Although the MeCP2 EMSA showed specificity for methylated DNA, the MeCP2 DNA complex migrated only a short distance into the gel matrix and appeared to consist of one major complex (Figure 4.5C). This result is unexpected given that the probe contains multiple methyl-CpG's. Theoretically, MeCP2 should associate with

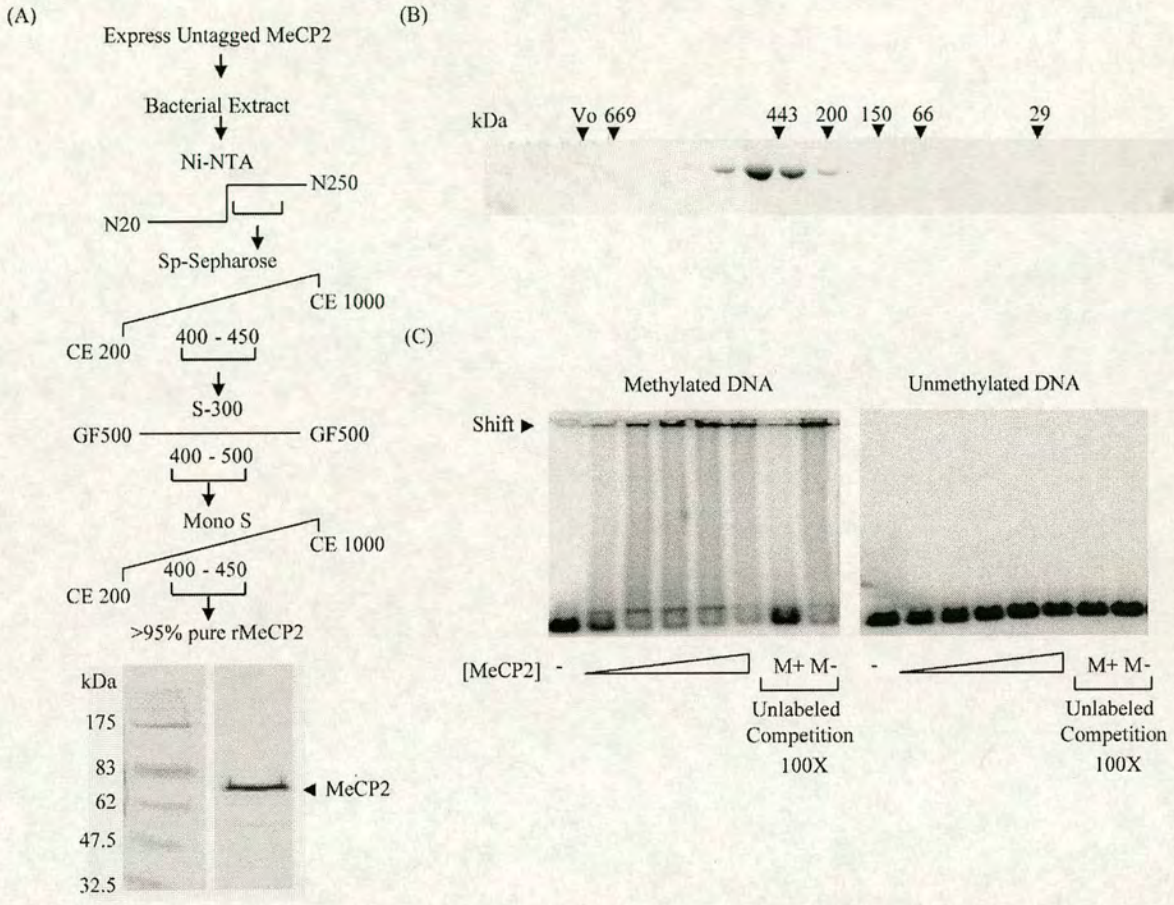


Figure 4.5- Production of untagged MeCP2 in bacteria.

(A) A four step purification scheme was devised to purify untagged MeCP2 from bacteria. Purified MeCP2 was determined to be >95% pure by Coomassie stained SDS PAGE.
(B) Recombinant MeCP2 was separated by Superose 12 size exclusion chromatography and analysed by Coomassie stained SDS PAGE. MeCP2 produced in bacteria has the same apparent molecular weight as native rat brain protein.
(C) Recombinant MeCP2 specifically shifts fully a methylated CG11 probe (Meehan et al., 1989) by EMSA on a 1.5 % agarose gel. Methylated probe was shifted by increasing concentrations of MeCP2 (left panel), and could only be competed by unlabelled methylated DNA. MeCP2 did not shift unmethylated DNA (right panel).

the probe in an incremental and concentration dependent manner as it occupies individual methyl-CpG sites, which would result in multiple DNA proteins complexes of differing apparent molecular weight. The results in Figure 4.5C do not behave according to this rational, and the reason for this is currently unknown. Despite the peculiarities in the EMSA appearance, MeCP2 produced in bacteria appears to be an excellent source of material with which to study the biophysical properties of native MeCP2.

4.2.3 Native MeCP2 does not self associate

Bacterially produced recombinant MeCP2 was used to test the hypothesis that MeCP2 may self-associate. If two molecules in solution are physically associated, it is possible to covalently link the two proteins using cross-linking reagents, but molecules that are not associated do not cross-link. When cross-linked, physically associated molecules exhibit a reduced mobility when analyzed by SDS-PAGE and can be differentiated from unassociated molecules by this feature. MeCP2 was treated with increasing concentrations of the chemical cross-linking reagent EGS (Figure 4.6A). As a control the same experiment was carried out using BSA, a monomeric protein, and ADH a tetrameric complex. BSA was unaffected by increasing EGS concentration but ADH cross-linked into higher molecular weight dimeric, trimeric and tetrameric species (Figure 4.6A). MeCP2 displayed no cross-linked species suggesting that it does not self associate. However, an increase in mobility of MeCP2 was observed with increasing EGS concentration probably due to intramolecular cross-links (Figure 4.6A). Therefore, chemical cross-linking data indicates that full length recombinant MeCP2 does not self associate. In agreement with the chemical cross-linking experiments, two differently sized molecules of MeCP2 also do not associate in solution. A fragment of containing 1-318 amino acids of MeCP2 was purified from bacteria and mixed with full length MeCP2 (Figure 4.1B). Size exclusion chromatography of the mixed full length and 1-318 proteins resulted in two independent elution profiles, which corresponded to the each protein run on the same column individually (Figure 4.6C). Although one cannot rule out the possibility that deleting amino acids 319-486 interferes with potential MeCP2 self association, these data are in fitting with EGS cross-linking studies indicating

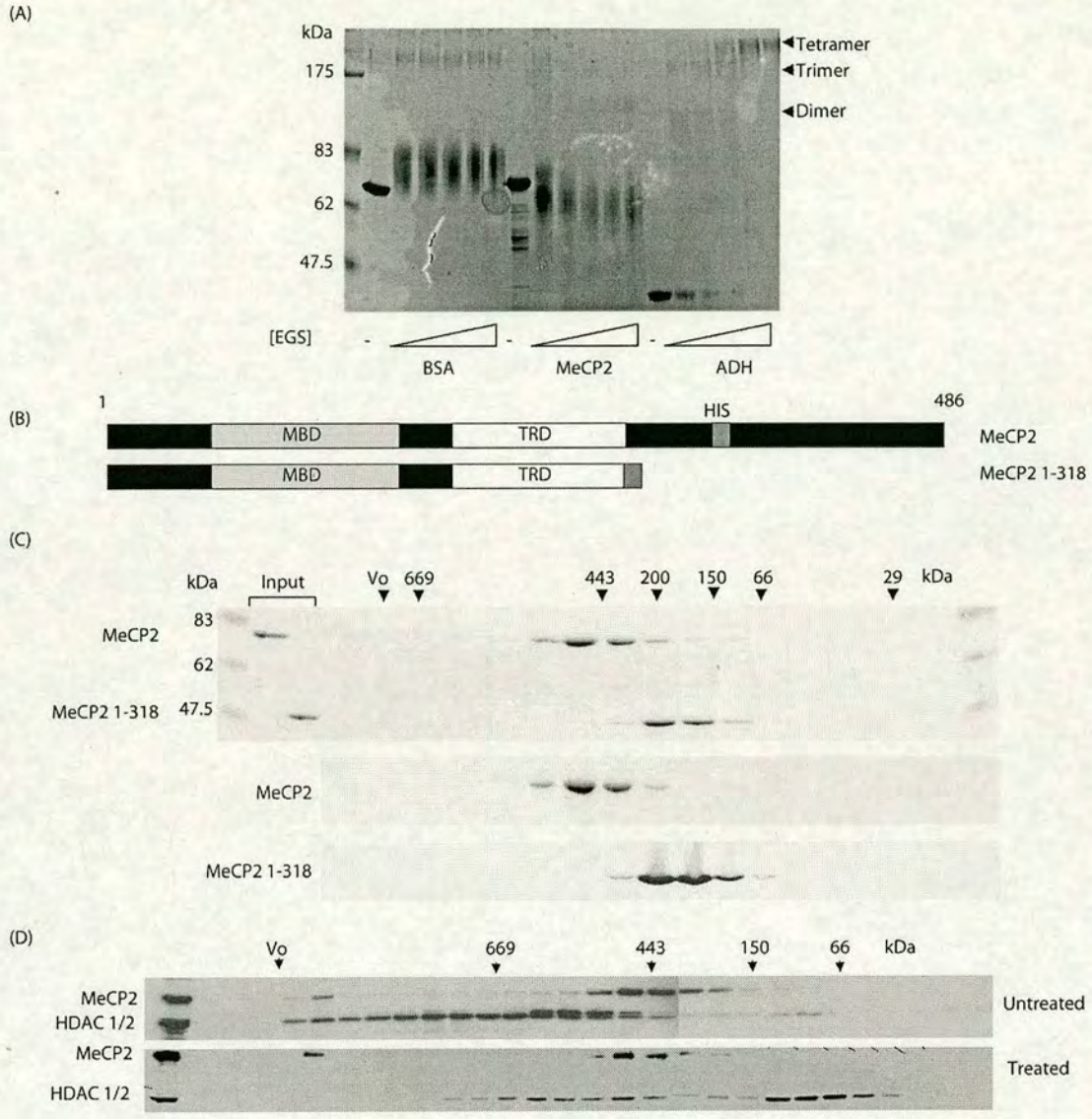


Figure 4.6- Native MeCP2 does not self associate.

(A) Full length MeCP2 does not self associate as determined by chemical crosslinking. MeCP2, BSA, and ADH were untreated (-) or treated with increasing concentrations of EGS and analysed by Coomassie stained SDS-PAGE.
(B) Diagram showing untagged MeCP2 and MeCP2 1-318, which were produced in bacteria and purified.
(C) Full length MeCP2 and MeCP2 1-318 were mixed and applied to a Superose 12 size exclusion column with the molecular weight markers indicated above each Coomassie stained SDS PAGE gel. Each protein eluted independently (upper panel) and with the same apparent molecular weigh as each protein separated individually (lower panels).
(D) Native MeCP2 does not self associate *in vivo*. Rat brain nuclear extract was left untreated (top panel) or treated (bottom panel) with 6M urea and separated by Superose 12 size exclusion chromatography. MeCP2 and HDAC 1/2 were identified by western blot analysis

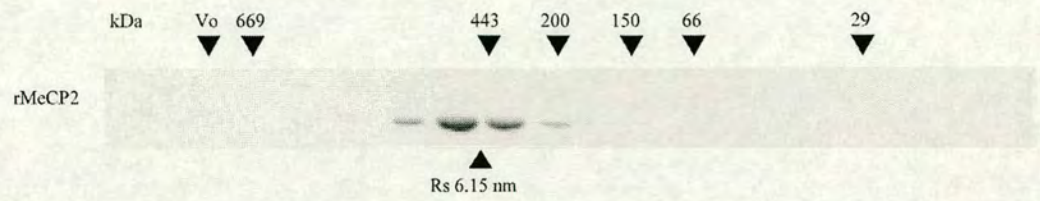
that full length MeCP2 does not self associate. Therefore, MeCP2 does not self associate *in vitro*.

To verify that the endogenous rat brain MeCP2 does not self associate, nuclear extract was treated with 6M urea at 65 °C for 10 minutes and then applied to a size exclusion column. Denaturing conditions did not significantly affect the native mobility of MeCP2 but did eliminate most of the high molecular weight HDAC 1/2 complexes (Figure 4.6D). This suggests that, like the recombinant MeCP2, native MeCP2 from rat brain does not self associate.

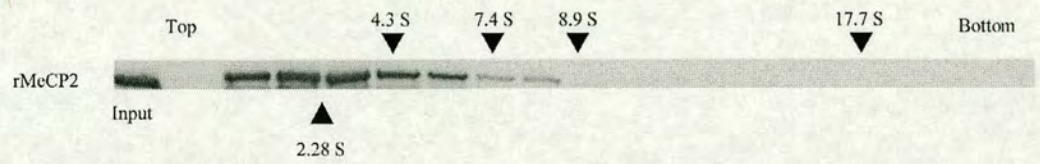
4.2.4 Biophysical analysis of MeCP2 using the Seigel and Monty equation

In the mid-sixties, a very elegant series of experiments by Siegel and Monty determined that a combination of the stokes radius (radius of protein) and the sedimentation coefficient could be used to accurately determine the native molecular weight of both globular and non-globular proteins and protein complexes (Siegel and Monty, 1966). Their analysis utilized a combination of size exclusion chromatography (which determines stokes radius) and density gradient sedimentation (which evaluates sedimentation coefficient) to analyze the native properties of proteins, and protein complexes. The power of the Siegel and Monty method is realized when studying the native molecular weight of non-globular molecules as it is essential to account for the radius (size exclusion) and sedimentation co-efficient (sedimentation) to accurately calculate mass. Without this biophysical information non-globular molecules will appear to have a very large molecular weight by size exclusion chromatography (radius) and a disproportionately small sedimentation co-efficient when compared to protein size markers which are globular proteins. The slow sedimentation of non-globular molecules is attributed to drag as the molecule sediments though the density gradient. By applying the equation of Seigel and Monty, size exclusion and sedimentation anomalies are cancelled and the native molecular weight of a non-globular molecule can be precisely determined.

(A)



(B)



(C)

Hydrodynamic Properties of rMeCP2

Theoretical Mass	Sedimentation Coefficient	Stokes Radius	Frictional Ratio	Derived Mass
Da	s 20,w	nm	f/f_o	Da
52,437	~2.28	~6.15	2.41	57,880

(D)

Protein	Frictional ratio
Albumin	130
Cytochome C	1.19
Fibrinogen	2.33
Myosin	3.62

Figure 4.7- MeCP2 is an elongated monomer.

(A) Native recombinant MeCP2 was separated by Superose 12 size exclusion chromatography. The molecular weight of standard proteins are indicated above and the radius of MeCP2 indicated below the Coomassie stained SDS PAGE.
(B) Native recombinant MeCP2 was separated through a 5-20% sucrose gradient by ultracentrifugation and fractions analysed by western blot analysis. The sedimentation co-efficient of standard proteins are indicated above, and the sedimentation co-efficient of MeCP2 below.
(C) Hydrodynamic analysis of MeCP2 was carried out using the Siegel and Monty equation. Native MeCP2 has a derived mass corresponding to a monomeric MeCP2 molecule, and frictional ratio calculations indicate that MeCP2 it is an elongated molecule.
(D) The frictional ratio of other globular (albumen and cytochrome C) and elongated (fibrinogen and myosin) molecules.

To facilitate Siegel and Monty analysis native MeCP2 produced in bacterial was separated on a size exclusion column and sedimented through a sucrose gradient. The Superose 12 size exclusion profile was the same as observed previously (Figure 4.5B), and the radius of the molecule was extrapolated from the known radius of the calibrated molecular weight standards. The radius of native MeCP2 is 6.15 nM (Figure 4.7A). To determine the sedimentation coefficient, native MeCP2 and sedimentation markers were run through a 5-20% sucrose gradient by ultracentrifugation. The MeCP2 signal from the gradient fractions was analyzed by densitometry and the peak value used to extrapolate the sedimentation coefficient from the included sedimentation co-efficient standards. The sedimentation coefficient of native MeCP2 is 2.28 S (Figure 4.7B). To determine the molecular weight of MeCP2 the stokes radius and sedimentation co-efficient were plugged into the Siegel and Monty formula as follows,

$$M_r = 6\pi\eta_{20,w} \bullet s_{20,w} \bullet R_S \bullet N / (1 - \rho_{20,w}\nu)$$
$$M_r = 57.880 \text{ kDa}$$

where:

R_S is the Stoke's radius (nm) (extrapolated from gel filtration profile of MeCP2)
$R_{S(MeCP2)} = 6.15$ nm

$s_{20,w}$ is the sedimentation velocity (S x 10^{-13}) (extrapolated form the sucrose gradient)
$s_{20,w(MeCP2)} = 2.28$ S

$\eta_{20,w}$ is the viscosity of water at 20 °C (0.01002 $g \cdot s^{-1}\ cm^{-1}$)
N = Avogadro's number (6.022 x $10^{23} \cdot mol^{-1}$)
$\rho_{20,w}$ is the density of water at 20 °C (0.9981 $g \cdot cm^3$)
ν is the partial specific volume (used 0.725 cm^3/g).

By combining the radius and sedimentation co-efficient the native mass of MeCP2 was determined to be 57.880 kDa, which is very close to the predicted monomeric MeCP2 molecular weight of 53.437 kDa by amino acid sequence (Figure 4.7C). In agreement with the previous experiments in this chapter, MeCP2 does not self associate and the Siegel and Monty equation derives a mass corresponding to a single MeCP2 molecule.

4.2.5 Mammalian and *Xenopus* MeCP2 are both monomeric elongated molecules

MeCP2 has a large radius and small sedimentation coefficient suggesting that is a non-globular molecule. The method of Siegel and Monty was used to determined the frictional ratio (f/f_0), which is a measure of shape,

$$f/f_0 = 6\pi\eta_{20,w} \cdot R_S / 6\pi\eta_{20,w} \cdot (3\nu M_r / 4\pi N)^{1/3}$$
$$= 2.41$$

where:

R_S is the Stoke's radius (nm) (extrapolated from gel filtration profile of MeCP2)
$R_{S(MeCP2)} = 6.15$ nm

$M_r = 57.880$ kDa (calculated in Section 4.2.4)

$\eta_{20,w}$ is the viscosity of water at 20 °C (0.01002 $g \cdot s^{-1}\ cm^{-1}$)
N = Avogadro's number ($6.022 \times 10^{23} \cdot mol^{-1}$)
ν is the partial specific volume (used 0.725 cm^3/g).

MeCP2 has a frictional ratio of 2.41 (Figure 4.7C), and for the purposes of comparison the frictional ratio of several common proteins in shown in Figure 4.7D. Globular molecules like albumen and cytochrome C have frictional ratios just above one, whereas elongated molecules like fibrinogen and myosin have frictional ratios above two. It has also been reported that a chicken protein, called ARBP, that shares homology with the MBD of MeCP2 has an elongated shape (von Kries et al., 1994). The significance of an elongated MeCP2 molecule remains to be evaluated.

The size exclusion chromatography elution profile was determined for endogenous *Xenopus laevis* oocyte MeCP2 in Chapter 3. Like its mammalian counterpart, the *Xenopus* protein was found in a large molecular weight fraction devoid of Sin3a

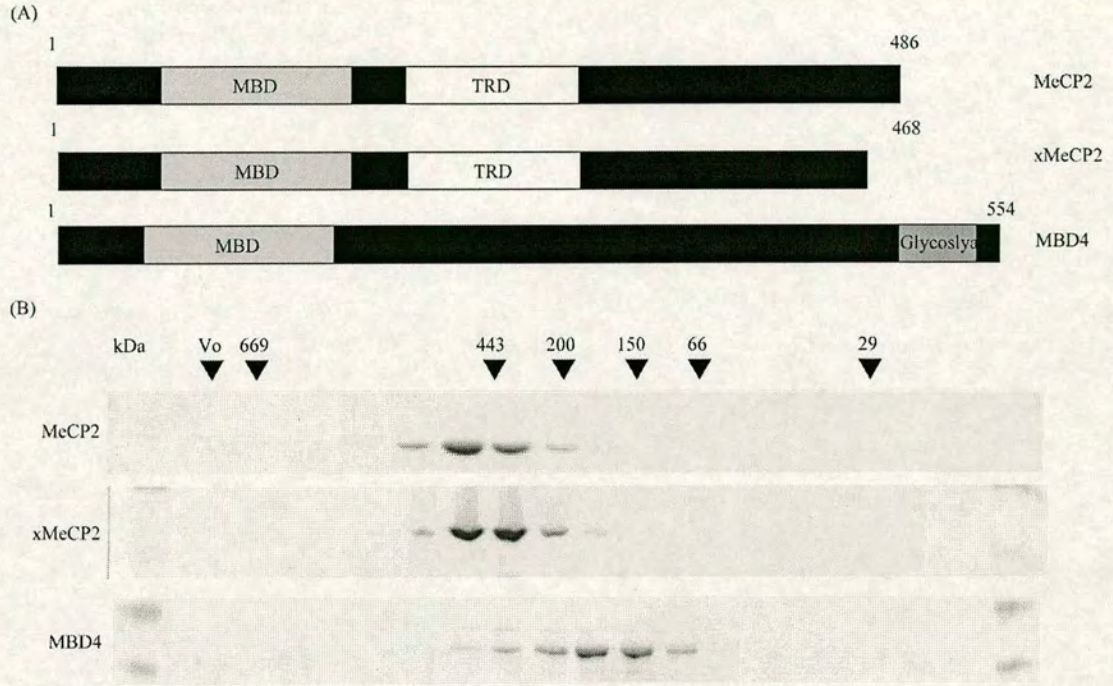


Figure 4.8- Both human and *Xenopus* MeCP2 are elongated molecules.

(A) Diagram showing human MeCP2, *Xenopus* MeCP2, and MBD4 structures. These proteins were expressed in bacteria and purified.
(B) The three proteins purified in (A) were separated by size Superose 12 size exclusion chromatography and fractions analysed by Coomassie stained SDS PAGE. Both human and *Xenopus* MeCP2 have large radii as exemplified by large apparent molecular weight by size exclusion chromatography. MBD4, although larger by amino acid sequence, elutes with smaller relative molecular weight in comparison to MeCP2.

(Figure 3.2C). To address whether the high molecular weight elution profile of *Xenopus* MeCP2 was also the result of a highly elongated structure, xMeCP2 was produced as an untagged recombinant protein in bacteria (Figure 4.8A), and separated on a Superose 12 size exclusion column. Recombinant xMeCP2 had an elution profile similar to that of mammalian MeCP2 (Figure 4.8B), suggesting that endogenous MeCP2 in *Xenopus* oocyte extracts is also an elongated monomeric molecule. Therefore both mammalian and amphibian MeCP2 appear to be highly elongated molecules. Interestingly when mammalian MBD4 was produced in bacteria and separated by size exclusion it had a much smaller radius despite containing an additional sixty eight amino acids (Figure 4.8B). Therefore, elongated structure is not an intrinsic characteristic of MBD containing proteins.

4.2.6 No individual domain of MeCP2 accounts for its elongated structure

To try and address which domain(s) contribute to the large radius of MeCP2, a series of recombinant MeCP2 deletion mutants were generated (Figure 4.9A). MeCP2 protein fragments were expressed and purified from bacteria under the same conditions as full length MeCP2. The purified proteins were then separated by size exclusion chromatography on a Superose 12 column. Figure 4.9B indicates how much larger the observed protein radius was than would be expected based on predicted amino acid sequence. The wild type protein has a radius about 9.5 times larger than would be predicted from amino acid sequence alone. The deletion proteins 1-373 and 1-318 both elute with radii more than 5 times greater that would be predicted. When smaller fragments encompassing individual domains of MeCP2 were analyzed the values were only 2.5-3.1 times larger than predicted. The general trend observed from these experiments is that smaller MeCP2 fragments have a less pronounced disparity in predicted radius compared to observed radius. Therefore, the elongated structure of MeCP2 is not determined by a single domain, but rather by large portions of MeCP2 protein. Interestingly, MeCP2 is about 11% proline by amino acid composition which may contribute to the large radius.

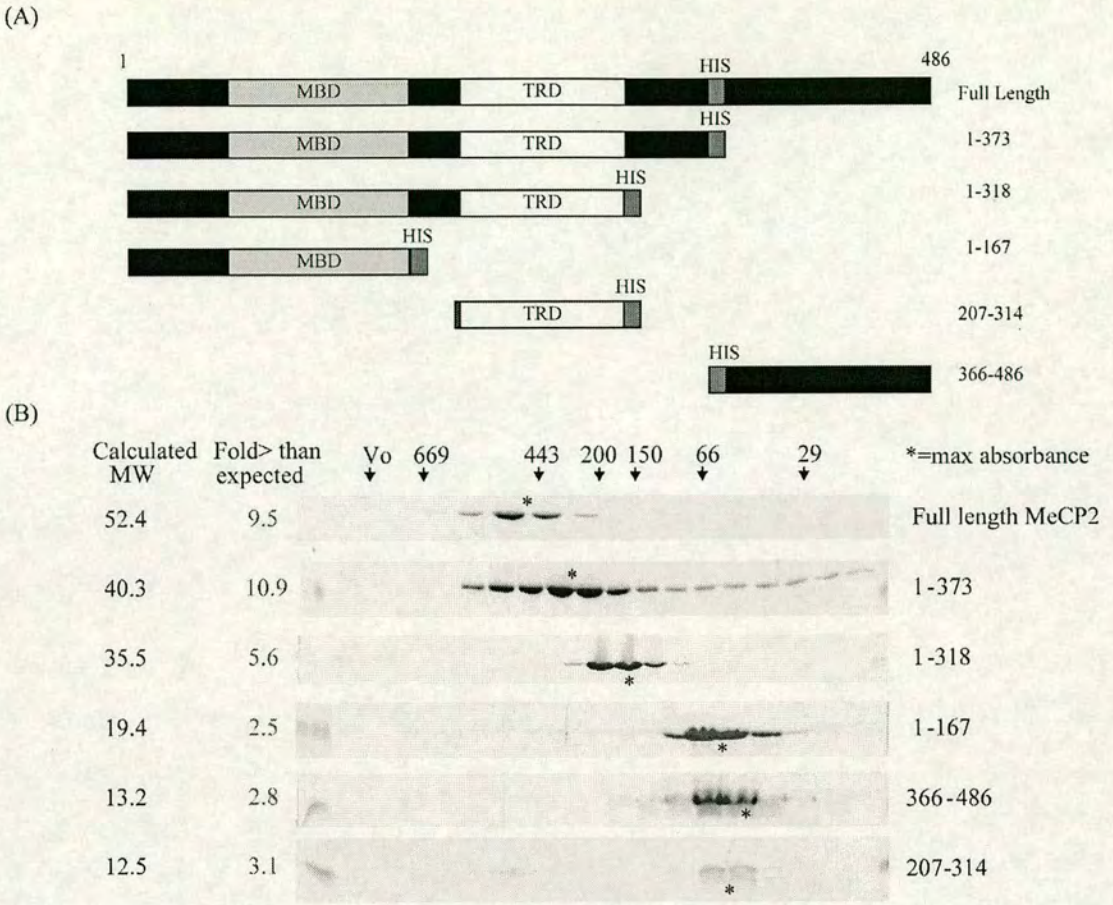


Figure 4.9- No individual domain of MeCP2 contributes to its large radius.

(A) Diagram showing a series of deletion mutants which lack MeCP2 domains. These proteins were expressed in bacteria and purified.
(B) The proteins illustrated in (A) were separated by Superose 12 size exclusion chromatography and analysed by Coomassie stained SDS PAGE. Above each panel the relative molecular weight of each molecule is indicated and the predicted molecular weight by amino acid sequence is indicated far left. Immediately to the left of each panel is the fold difference in observed versus predicted radius of each molecule. An asterisk indicates the peak absorbance of each elution profile and to the right is the construct identity.

4.3 Discussion

4.3.1 Native MeCP2 is an elongated monomer

Very little data is available about the biophysical properties of native MBD proteins. Purified native MeCP2 from rat brain was found in a large molecular weight fraction although it was not stably associated with other proteins. To address this discrepancy three possible explanations were tested; 1-MeCP2 associates with an RNA or DNA component, 2-MeCP2 stably self associates, 3-MeCP2 is a non globular molecule. By treating purified rat brain MeCP2 with nucleases no change in the size exclusion profile was observed, suggesting DNA or RNA are not involved. In order to study the second two possibilities, preparative amounts of recombinant MeCP2 were produced and shown to contain the same properties as the MeCP2 molecule purified from rat brain. No self association of MeCP2 was observed under several experimental conditions, and biophysical analysis using the Siegel and Monty analysis demonstrated that MeCP2 is an elongated monomer.

4.3.2 Is there a structural reason that MeCP2 is an elongated molecule?

Biophysical analysis demonstrates that mammalian and *Xenopus* MeCP2 are both highly elongated molecules. The biophysical properties of the other MBD proteins remain to be tested but preliminary evidence suggests that MBD4 does not have the same elongated structure as MeCP2. Efforts to map a domain required for the elongated shape of MeCP2 proved inconclusive, and the majority of the protein is required for the observed radius. A possible explanation for the large radius could be the relatively significant contribution of proline in MeCP2, which accounts for 11% of amino acids. Several other molecules including P-57 (Masure et al., 1986) and myosin light chain kinase (Crouch et al., 1981) have large radii and a disproportional contribution of the amino acid proline in their sequence. The peptide bond that forms in proline is inherently rigid and often causes in kinks in the protein backbone (Barlow and Thornton, 1988) and will act as a helix breaker. The frequent occurrence of the amino acid proline might be a structural factor that contributes to the large radius of native MeCP2 by perturbing normal globular protein folding. With the exception of the MBD, simple secondary structure prediction analysis of MeCP2 reveals no obvious regions that might contribute to structure of the molecule and

domain mapping failed to reveal any highly elongated sub-domain (Figure 4.9B). Protein crystallization and three dimensional structure analysis may be the only useful way to fully understand the amino acid determinants that result in an elongated MeCP2. Currently, crystallization trials are underway to try and elucidate the structural determinants that lead to elongation of MeCP2.

4.3.3 Why might an elongated MeCP2 molecule be advantageous?

The shape of a molecule usually serves a functional or structural requirement, for example, the rod like structure of myosin is functionally related to its structural role in muscle contraction (Holmes, 1997). While myosin has a distinct shape to suit its biomechanical functions, the yeast transcription factor TFIIIB is highly elongated and postulated to act as a bridge linking together proteins necessary for initiation of RNA pol III transcription (Klekamp and Weil, 1987). Therefore, an elongated MeCP2 molecule may serve some structural advantage related to function *in vivo*. One possibility is that elongated MeCP2 functions in a modular fashion with each of its domains functioning independently from each other (Figure 4.10A). For example, the MBD would be responsible for binding methylated DNA and other distinct protein domains, the TRD, N-terminus and the C-terminus, might carry out their function in a manner structurally independent of the MBD (Figure 4.9A). Therefore, MeCP2 might bind DNA with the N and C termini extended outward as a platform to mediate interactions with other proteins and/or potentially to exclude other factors from interacting with surrounding DNA sequence (Figure 4.10B). Recent evidence suggests that MeCP2 is involved in chromatin looping and uncharacterized domains of MeCP2 may be involved in this function (Horike et al., 2005). Since the MBD and TRD can both function independently when expressed as separate domains (Jones et al., 1998; Nan et al., 1997a; Yu et al., 2000), the suggestion that MeCP2 functions in a modular fashion might not be completely unfounded. Interestingly, Rett syndrome mutations exist along the length of the MeCP2 molecule, and regardless of whether these mutations affect mapped functional domains, they result in relatively similar disease phenotypes. Since many Rett mutations leave methyl-CpG binding and repression modules intact, these mutations could interfere with other structurally distinct functional domains in MeCP2 (Figure 4.10C). With the recent discovery of MeCP2 regulated genes (Chen et al., 2003b; Horike et al., 2005; Martinowich et al.,

2003) it will be of functional and structural importance to dissect what domains of MeCP2 are required for normal function, and analyze whether an elongated MeCP2 molecule contributes to these functions.

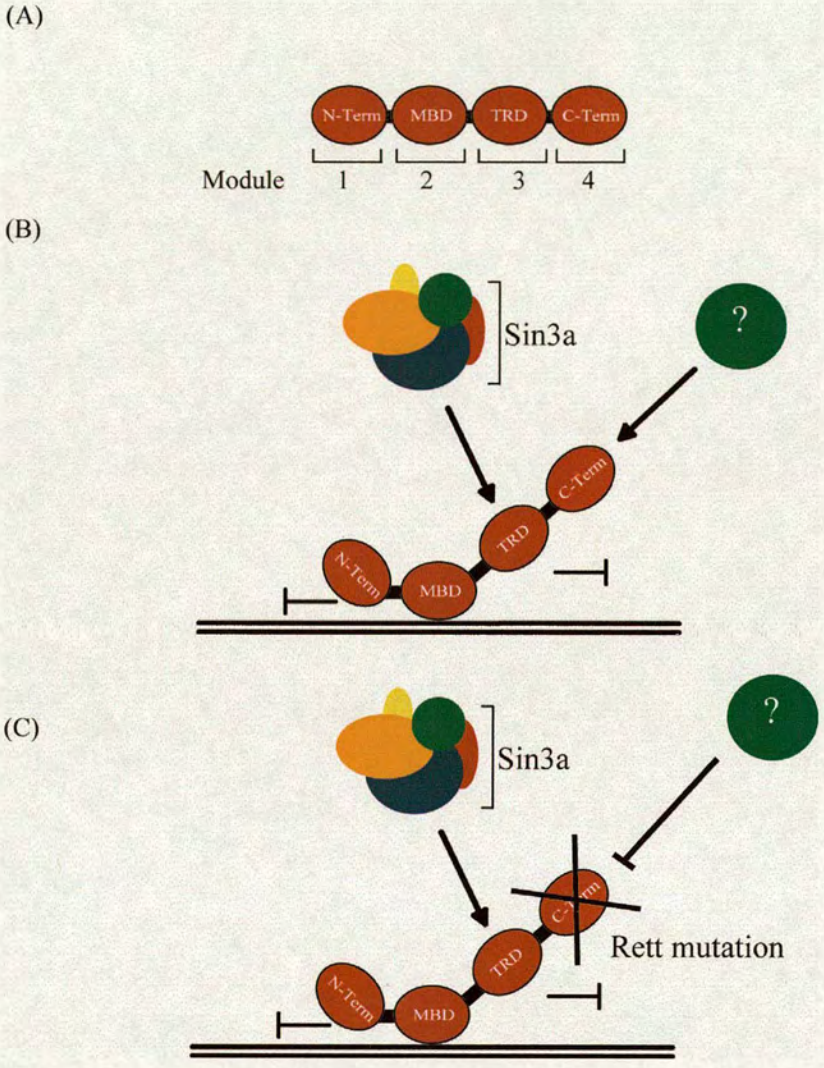


Figure 4.10- Is MeCP2 elongated to act in a modular fashion?

(A) MeCP2 is a highly elongated molecule suggesting that multiple modular domains may exist. Within the context of an elongated MeCP2 molecule each individual domain could function independently.
(B and C) As explained in the text, an elongated modular MeCP2 molecule could act as a versatile landing pad for co-factor recruitment and/or to sterically hinder binding of other transcription factors to surrounding DNA (B). Because many patients with a classical Rett syndrome phenotype have mutations that occur outside of known functional domains, other modular domains could be important for normal MeCP2 function (C).

5. Chapter five-
MeCP2 DNA Binding Specificity

5.1 How does MeCP2 recognize Methylated DNA?

DNA methylation in vertebrates is recognized and interpreted by methyl-CpG binding proteins (MBP's) that repress transcription and translate epigenetic information found on DNA to surrounding chromatin. Little is known about how MBP's are targeted to specific methylated loci *in vivo*, but substantial *in vitro* evidence has defined the specificity of MeCP2 for methylated DNA (Free et al., 2001; Meehan et al., 1992; Nan et al., 1993; Yusufzai and Wolffe, 2000). The MBD domain structure of MeCP2 and MBD1 in the absence of methylated DNA have been solved by NMR (Ohki et al., 1999; Wakefield et al., 1999) More recently, the structure of the MBD1 MBD in complex with a methyl-CpG containing DNA fragment provided molecular insight into how methyl-CpG is recognized (Ohki et al., 2001). The core structure of the MBD is formed by an α/β sandwich, and distinct residues in the MBD contact both the major groove and each individual methyl-cytosine (Ohki et al., 2001). The NMR structure of a chicken protein, named ARBP1, which has homology to the MBD domain of MeCP2 has been solved in the absence of DNA and shown to contain the same structural fold (Brunner et al., 2000; Heitmann et al., 2003). Initial reports characterized ARBP1 as a matrix attachment region (MAR) binding protein. ARBP1 binds the sequence GGTGT, but also has affinity for methyl-CpG (Buhrmester et al., 1995; von Kries et al., 1991a; Weitzel et al., 1997a). The biologically relevant binding specificity of ARBP1 remains to be determined, but only methyl-CpG binding specificity has been thoroughly demonstrated for MeCP2. The biomedical importance of MBD function was realized with discovery that the human neurological disease Rett syndrome is caused by mutations in the MeCP2 gene (Amir et al., 1999). Mutational profiling of Rett syndrome patients has identified a significant portion of point mutations that inactivate the MBD of MeCP2, again pointing to the biological importance of active interpretation of the DNA methylation signal (Kriaucionis and Bird, 2003).

Recently, several studies have used chromatin immunoprecipitation (ChIP) to map MBP binding loci *in vivo*. Two studies have used cancer cell lines to identify genes that MeCP2 targets in malignant sources. In a study by Ballestar et al (2003) the binding profiles of MBP proteins were compared by hybridizing ChIP DNA to CpG island micro-arrays. This study showed that some methylated CpG islands were targeted by individual MBD proteins, but a large proportion were targeted by more than one MBP. A recent report using a chromatin immunoprecipitation-and-clone (ChIP-and-Clone) approach identified the DLX 5/6 locus in mouse brain as a target for MeCP2-mediated repression (Horike et al., 2005). This chapter investigates the binding properties of MeCP2 *in vivo* and *in vitro,* and identifies a new DNA binding requirement for high affinity recognition of methyl-CpG containing DNA by MeCP2.

5.2 DNA binding properties of MeCP2

5.2.1 MeCP2 binds unique loci *in vivo*

Methyl-CpG is distributed ubiquitously throughout the mammalian genome (Bird, 2002). MBD proteins can specifically recognize symmetrically methylated CpG, but how each individual MBD family member contributes to distinguishing this signal in a genomic context remains unclear. Our collaborator Dr. Irina Stancheva has used MeCP2 specific antibodies in ChIP-and-clone experiments to identify methylated loci that MeCP2 occupies *in vivo* in the human primary fibroblast line MRC5. In these studies MeCP2 bound to loci which were not extensively bound by other MBP molecules. MeCP2 binding *in vivo* was dependent on DNA methylation as pre-treatment with 5-aza-cytidine caused loss of MeCP2 ChIP signal at normally bound loci. Because methyl-binding proteins recognize methylated CpG it was expected that overlap might exist between DNA's occupied by MBP's. In contrast to this expectation, it appears as if MeCP2 binds unique loci that generally do not contain the other MBD protein relatives. Given that all MBP's recognize methyl-CpG it was of interest to determine whether it was possible for other MBP family members to occupy MeCP2 binding sequences in the absence of MeCP2 protein. To investigate this possibility, Dr. Stancheva carried out depletion studies of MeCP2 and found that MBD2 could efficiently occupy unbound MeCP2 sites. Using similar experimental approach Dr. Stancheva isolated MBD2 binding loci *in vivo* and found that depletion

of MBD2 did not result in occupancy of MBD2 binding sites by MeCP2. These data suggest that MeCP2 may be targeted to specific methyl-CpG's *in vivo*.

5.2.2 Enrichment of MeCP2 binding sites by Methyl-SELEX

ChIP-and-clone experiments suggest that MBP's occupy distinct loci *in vivo*. MeCP2 binds nearly exclusively to loci devoid of other MBP's and cannot take over sites vacated by MBD2 in depletion experiments. It is conceivable that the specific MeCP2 occupancy observed *in vivo* may be due to a requirement for additional DNA binding determinants outside of methyl-CpG alone. To test this hypothesis an experiment was designed to isolate high affinity DNA binding sites for MeCP2, using a modified form of the *in vitro* selection technique SELEX (systematic evolution of ligands by exponential enrichment). Traditional SELEX has been used in many different systems to isolate aptamers that are high affinity ligands for a given bait molecule (Klug and Famulok, 1994). A DNA aptamer has been successfully used to isolate DNA sequences that have a high affinity for DNA binding factors (He et al., 1996). Here a modification of traditional SELEX (Methyl-SELEX) was devised to determine whether MeCP2 has additional sequence binding specificity outside of its known binding requirement of methyl-CpG.

Methyl-SELEX involved generating double stranded DNA fragments that have a fixed central CpG in the context of a *Hpa*II methyltransferase/restriction endonuclease site, flanked by random DNA sequence (Figure 5.1A). The initial pool of DNA fragments was methylated using M.*Hpa*II methyltransferase and tested for complete methylation by loss of sensitivity to cleavage by *Hpa*II restriction endonuclease. The N-terminal half of human MeCP2 (amino acids 1 - 205) was used for Methy-SELEX, as it forms a discrete DNA-protein complex more reproducibly than the full-length protein (1 - 486). The protein was mixed with the starting DNA and the protein DNA complex was recovered following an electrophoretic mobility shift assay (EMSA). DNA from the complex was amplified, re-methylated and once more bound to MeCP2 (1-205). After 8 cycles, MeCP2 binding efficiency was increased approximately five fold, demonstrating the efficiency of this technique for isolating DNA fragments that have an enriched affinity for MeCP2 (Figure 5.1B).

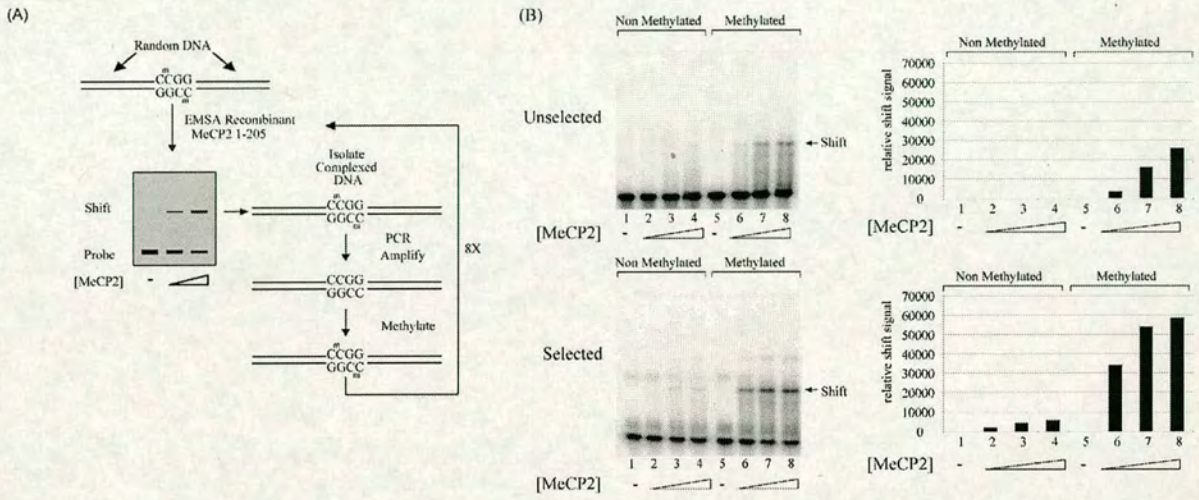


Figure 5.1-Methyl-SELEX efficiently selects for high affinity MeCP2 binding sequences.

(A) A four step Methyl-SELEX enrichment as detailed in the Results text involves evolution of high affinity MeCP2 binding sites by multiple EMSA enrichment steps. (B) Unselected and selected DNA from Methyl-SELEX were end-labelled with ^{32}P and used as probes in an EMSA. Methyl-SELEX selected material showed an increase in EMSA signal in comparison to that observed with unselected material. The methylation status of each probe is indicated above and MeCP2 protein concentration is shown below. (–) indicates free probe. The intensity of each signal was quantified and plotted as relative shift signal in the right hand panel.

5.2.3 DNA fragments enriched by MeCP2 using Methyl-SELEX have an A/T rich sequence adjacent to the methyl-CpG

The sequence of 88 selected fragments and 86 unselected fragments was determined (5.3A)(Appendix 1). Analysis of these sequences revealed the frequent presence of an A/T run of 4 or more bases ($[A/T]_{\geq 4}$) close to the methyl-CpG site in the enriched sample. When sequences were aligned with the A/T run to the right of the methyl-CpG (Figure 5.2A)(Appendix 1), an $[A/T]_{\geq 4}$ motif appeared to cluster in two vertical stripes located 1-3 base pairs or 6-9 base pairs from the methylated CCGG site. To determine whether the DNA fragments in the selected pool that contained $[A/T]_{\geq 4}$ had an increased affinity for MeCP2, these fragments were synthesized as individual double stranded DNA sequences. Five selected fragments S1-S5 (Figure 5.2B) were chosen for this analysis. Two fragments (S1 and S2) that had an $[A/T]_{\geq 4}$ within three bases of the methyl-CpG, two fragments (S3 and S4) that had an $[A/T]_{\geq 4}$ within eight bases of the methyl-CpG, and a single fragment (S5) that had no $[A/T]_{\geq 4}$ adjacent to the methyl-CpG. As a control two fragments from the unselected pool were chosen (U1 and U2). Fragments S1-5 and U1-2 were used in an EMSA with the central CpG methylated (M+) or unmethylated (M-). Fragments S1-S4 containing an $[A/T]_{\geq 4}$ adjacent to the methyl-CpG demonstrated efficient methylation dependent binding (Figure 5.2C). Fragment S5 which was isolated from the selected pool of DNA's bound poorly to the methylated probe, suggesting that this is a low affinity binding site that has been retained though the enrichment. Fragments U6 and U7 were isolated from the unselected pool and showed little detectable binding even to the methylated fragment (Figure 5.2C). To determine if the presence of the A/T run causes directional binding of MeCP2 to a symmetrical methyl-CpG, the S1 probe was bound to MeCP2 and subjected to DNAseI foot-printing analysis. The high affinity S1 fragment resulted in a foot-print which covers the A/T run and causes an asymmetrical protection of the CpG in the direction of the A/T run containing sequence (Figure 5.2D). Therefore, DNaseI analysis suggests that the presence of an A/T run confers direction binding of MeCP2 on methyl-CpG containing sequences.

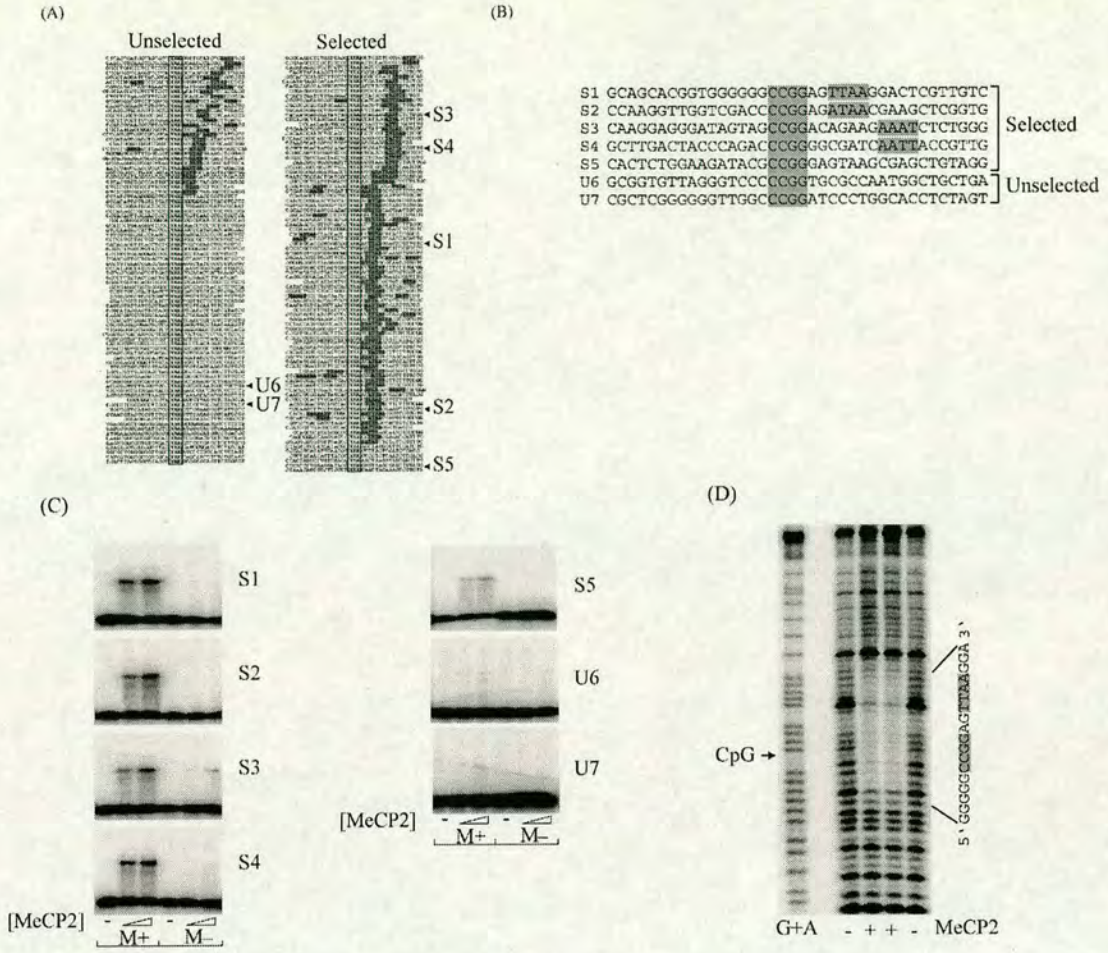


Figure 5.2-High affinity MeCP2 binding sites isolated by Methyl-SELEX contain methyl-CpG with an adjacent run of $[A/T]_{\geq 4}$ DNA.

(A) A representative sample of the random unselected and MeCP2 selected DNA sequences from the Methyl-SELEX experiment were cloned, sequenced, and aligned. The central fixed CCGG site is indicated by an open box and runs of four or more A/T's are highlighted with a shaded box.
(B) Five selected (S1-S5) and two unselected sequences (U6 and U7); see arrows in part A) were chosen for further analysis.
(C) EMSA analysis of MeCP2 selected probes S1-S4 show efficient binding, but selected probe S5 that lacks $[A/T]_{\geq 4}$ binds poorly. No detectable binding was observed to unselected probes U6 and U7. M+ and M- indicate presence or absence of methyl-CpG in the EMSA fragment. The concentration of MeCP2 is indicated below with (–) indicating free probe
(D) DNaseI foot-printing analysis on the ^{32}P end labelled S1 fragment demonstrates asymmetrical protection of the methyl-CpG towards the A/T run (large bracket with protected DNA sequence to the right). Location of the methyl-CpG in the foot-printing probe is indicated by an arrow to the left. G+A indicates the Maxam/Gilbert sequencing reaction on the same DNA fragment as size marker. Samples were incubated with (+) or without (-) MeCP2 protein.

5.2.4 MeCP2 requires an A/T rich sequence adjacent to the Methyl-CpG for efficient DNA binding

To test the requirement of $[A/T]_{\geq 4}$ for efficient MeCP2 binding, fragments S1 and S3 were mutated (Figure 5.3A) by the addition of cytosine or guanine at two positions within the A/T rich run. In both cases altering the base composition from an A/T run to non-A/T run containing probe drastically reduced binding efficiency (Figure 5.3B). All probes used to this point in the study were methylated in the context of a fixed central *Hpa*II methyltransferase site (CCGG). To ensure that the extra bases surrounding the central CpG did not contribute to the enhanced MeCP2 binding an artificial probe (A1) was generated in which the central CG of probe S1 was placed in the context of GCGC and could be specifically methylated with *Hha*I methyltransferase (Figure 5.3A). Mutation of the *Hpa*II site to a *Hha*I methyltransferase site did not affect the high affinity binding (Figure 5.3B). If an $[A/T]_{\geq 4}$ adjacent to the methyl-CpG is a requirement for efficient MeCP2 binding it should be possible to engineer a high affinity MeCP2 binding site using these criteria. To test this hypothesis a probe was generated (A2) that contains a central CpG with an adjacent A/T run of four bases (Figure 5.3A) and is surrounded by random DNA sequence. In EMSA assays the engineered DNA sequence bound MeCP2 as efficiently as the Methyl-SELEX fragments containing an $[A/T]_{\geq 4}$ (Figure 5.3B). To demonstrate that the $[A/T]_{\geq 4}$ contributes to enhanced binding, this element was mutated (A2M). The engineered probe with the mutated $[A/T]_{\geq 4}$ had a drastically reduced affinity for MeCP2. Together these observations demonstrate that the efficient DNA binding observed in the Methyl-SELEX fragments is due to the $[A/T]_{\geq 4}$ sequence that lies adjacent the methyl-CpG.

5.2.5 The MeCP2 AT-Hook domain does not contribute to binding of an A/T run containing site

It was noted in early studies that MeCP2 contains a domain that is shared with proteins that bind the minor groove of A/T DNA (Lewis et al., 1992; Nan et al., 1993). The effect of this domain on MeCP2 function remains unknown, but close sequence similarity with an AT-hook domain of HMGA1 (formerly HMG I/Y) and

(A)

```
S1   GCAGCACGGTGGGGGGCCGGAGTTAAGGACTCGTTGTC
S1M  GCAGCACGGTGGGGGGCCGGAGTGCAGGACTCGTTGTC

S3   CAAGGAGGGATAGTAGCCGGACAGAAGAAATCTCTGGG
S3M  CAAGGAGGGATAGTAGCCGGACAGAAGACTCCTCTGGG

A1   GCAGCACGGTGGGGGGGCGCAGTTAAGGACTCGTTGTC
A1M  GCAGCACGGTGGGGGGGCGCAGTGCAGGACTCGTTGTC

A2   ACGTACATGCCTCATGCCGGACTTAACTGCAGCTACTC
A2M  ACGTACATGCCTCATGCCGGACTGCACTGCAGCTACTC
```

(B)

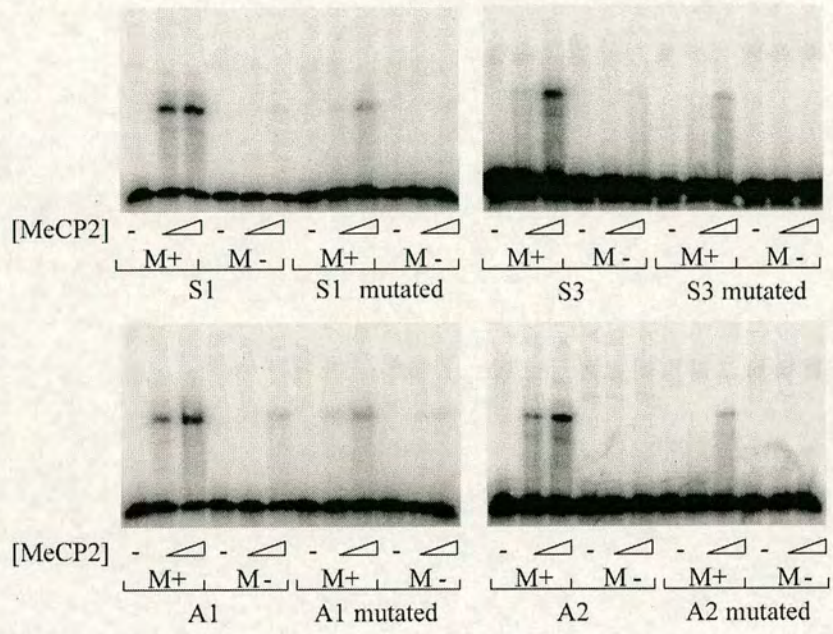


Figure 5.3-Mutation of the $[A/T]_{\geq 4}$ in high affinity MeCP2 binding site reduces affinity.

(A) Selected fragments S1 and S3 were mutated in the A/T run as indicated by the shading (S1M and S3M). In fragment A1, the HpaII has been altered to a HhaI site (GCGC), and A1M combines this with a mutated A/T run. Fragments A2 and A2M are artificially engineered MeCP2 binding sites consisting of random DNA sequence in which is embedded a methyl-CpG with an adjacent A/T run (A2) or mutated A/T run (A2M).
(B) EMSA analysis shows that mutation of the A/T run (S1M and S3M) reduces the affinity of MeCP2 for S1 and S3 as indicated by reduction of EMSA signal in the S1 M and S3 M lanes. A1 and A2 both bind tightly to MeCP2 and binding is dependent on an adjacent A/T run (A1M and A2M).

across MeCP2 orthologues is striking (Aravind and Landsman, 1998) (Figure 5.4A). Given the requirement for an $[A/T]_{\geq 4}$ at high affinity MeCP2 binding sites, experiments were performed to analysed whether the AT-hook might be responsible for the binding specificity of MeCP2. To facilitate this analysis two conserved arginine amino acids within the MeCP2 AT-hook were mutated to glycine (Figure 5.4A). Equivalent mutations within the AT-hook of HMGA1 significantly inhibited binding to heterochromatin in photobleaching experiments (Harrer et al., 2004). When wild type and mutant MeCP2 proteins were assayed by EMSA for binding to the S1 and S1-mutated DNA fragments, loss of the AT-hook domain had no effect on the affinity or specificity of MeCP2 for probes that contained $[A/T]_{\geq 4}$ (Figure 5.4B).

It seemed possible that the AT-hook affects MeCP2 binding *in vivo*, but escapes detection by *in vitro* EMSA. To examine the *in vivo* dynamics of the association between MeCP2 and mouse heterochromatin a collaborator, Dr. Lars Schmiedeberg, employed fluorescence recovery after photo-bleaching (FRAP). In mouse cells MeCP2 has a distinct sub-nuclear localization profile at pericentric heterochromatin, which contains the major satellite DNA (Figure 5.4C). Mouse satellite DNA accounts for about 7 – 10 % of the mouse genome and comprises an A/T-rich repeated sequence in which CpG dinucleotides are highly methylated (Hörz and Altenburger, 1981; Manuelidis, 1981). Based on the observations reported above, the combination of concentrated methyl-CpGs adjacent to runs of A/T should provide an optimal binding location for MeCP2. For FRAP analysis, wild type and AT-hook mutant MeCP2 were fused to GFP and transfected into mouse cells. Both the wild type and the AT-hook mutant protein localized normally to DAPI bright spots, which correspond to heterochromatic foci (Figure 5.4C). To test the kinetics of MeCP2 and the MeCP2 AT-hook mutant in heterochromatic foci, NIH3T3 and $p53^{n/n}$ mouse fibroblasts were photobleached and the recovery time of GFP measured. Wild type MeCP2 had a recovery half time of about 25 s, with no significant difference ($p>0.0001$) between wild type and AT-hook mutant in both cell lines tested (Figure 5.4D). This recovery time is comparable to linker histone H1 (Lever et al., 2000), but almost 10 times slower than other structural non-histone chromatin proteins, for example HP1 (Cheutin et al., 2003; Schmiedeberg et al., 2004). From both the *in*

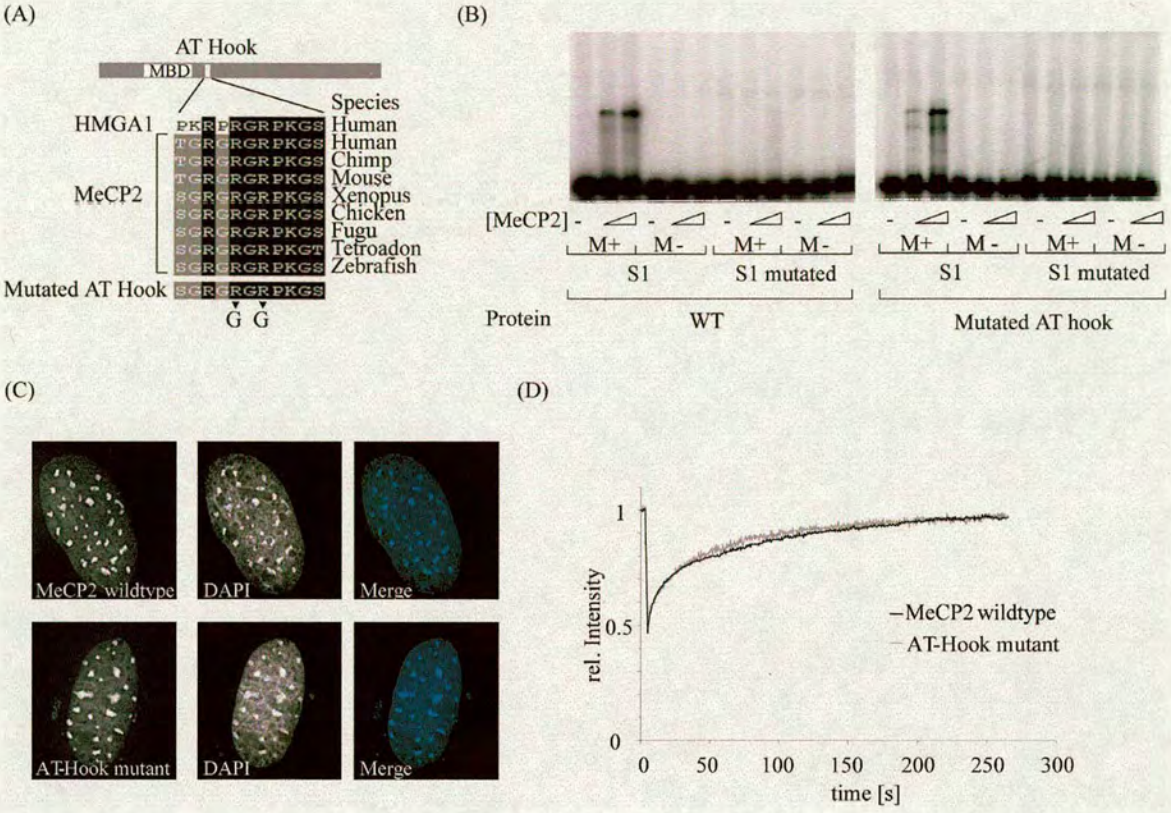


Figure 5.4- The conserved MeCP2 AT-hook domain is not required for A/T run-dependent binding of MeCP2 *in vitro* or *in vivo*.
(Experiments in (C) and (D) were contributed by Dr. Lars Schmiedeberg)

(A) Alignment of the conserved AT-hook domains of MeCP2 with an AT-hook found in HMGA1. A mutant MeCP2 protein was generated by replacing two essential arginines in the MeCP2 AT-hook with glycine.
(B) Mutation of the AT-hook of MeCP2 does not affect methyl-CpG A/T run dependent binding to the S1 probe as indicated by EMSA using ^{32}P labelled probes.
(C) GFP-MeCP2 and GFP-MeCP2 (AT-hook mutant) both localized to DAPI stained foci in living cells.
(D) Fluorescence recovery after photobleaching at heterochromatic sites is indistinguishable between wild-type and mutant MeCP2

vitro EMSA and *in vivo* FRAP analysis the AT-hook is not essential for the recognition of high affinity MeCP2 binding sites in which methyl-CpG is followed by an A/T run.

5.2.6 MeCP2 methyl-CpG A/T run dependent binding resides in a MeCP2 fragment containing the MBD and flanking sequences.

The MeCP2 AT-Hook does not contribute to the observed requirement for an $[A/T]_{\geq 4}$ adjacent to methyl CpG for efficient MeCP2 binding. Therefore, to map the MeCP2 domain responsible for high affinity binding, a series of deletion mutants were expressed as recombinant proteins (Figure 5.5A) and used in EMSA on probes S1 and S1 mutated. MeCP2 1-167, which completely lacks the AT-Hook, has the same selectivity for a methyl-CpG with an $[A/T]_{\geq 4}$ as fragment 1-205 that contains the intact AT-Hook domain (Figure 5.5A). This observation reaffirms the initial demonstration that the AT-Hook is not an obligate component of the observed specificity. Further deletion from both the N and C termini of MeCP2 toward the MBD (fragments 77-167 and 78-161) retained high affinity binding for the $[A/T]_{\geq 4}$ containing probe (Figure 5.5A), excluding these regions as specificity determinants. A further N-terminal deletion of MeCP2, resulting in a fragment containing amino acids 90-161, completely abolishes DNA binding. Since a MeCP2 fragment containing amino acids 78-161 bound the S1 probe, this suggests that high affinity MeCP2 binding requires amino acids between 78 and 90 in addition to the conserved MBD domain (Figure 5.5A). To verify that binding to methyl-CpG with an adjacent $[A/T]_{\geq 4}$ is a unique feature of the MeCP2 protein, a fragment of MBD2 covering the MBD, and having roughly the same amino acids as MeCP2 90-161 was used in an EMSA. Figure 5.5B demonstrates the MBD2 protein has no sequence specificity with respect to the presence of an $[A/T]_{\geq 4}$ adjacent to the methyl-CpG. Therefore MeCP2, but not MBD2, appears to require sequence determinants outside of methyl-CpG alone for efficient DNA binding.

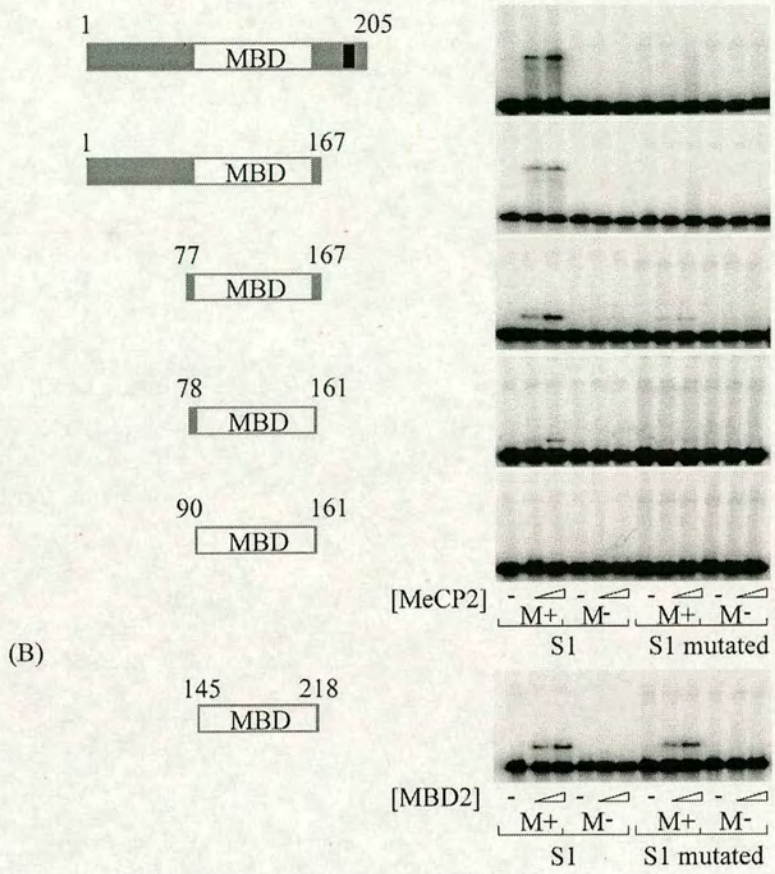


Figure 5.5-A fragment of MeCP2 containing amino acids 78-161 confers methyl-CpG $[A/T]_{\geq 4}$ dependent binding.

(A) Deletion fragments of MeCP2 were used in EMSA with S1 labelled probe. A fragment containing amino acids 78-161 has methyl-CpG $[A/T]_{\geq 4}$ dependent binding but this activity is lost in a fragment containing only amino acids 90-161.
(B) MBD2a (145-218) was used in an EMSA on the same labelled S1 probe and showed equivalent binding to both the S1 and S1 mutated probe, indicating MBD2 does not require an $[A/T]_{\geq 4}$ for efficient binding.

5.2.7 Binding to the promoter sequence of the MeCP2 target gene *Bdnf* is mediated by an methyl-CpG A/T run containing sequence

MeCP2 has been shown to bind an inducible promoter region of the rodent *Bdnf* gene and contribute to maintenance of the silenced state (Figure 5.6A)(Chen et al., 2003b; Martinowich et al., 2003). Bisulfite genomic sequencing and competitive EMSA analysis suggested that MeCP2 bound preferentially to a methyl-CpG at -148 base pairs from the transcription start site (Chen et al., 2003b; Martinowich et al., 2003). Analysis of the sequence surrounding -148 revealed an A/T run of four base pairs adjacent to the CpG (Figure 5.6B). To test the efficiency of MeCP2 binding to this sequence, we generated a probe (-148) corresponding to the MeCP2 binding site and a control probe of the same length containing one CpG from a sequence within *Bdnf* exon 1 (Figure 5.6B). The -148 sequence bound MeCP2 by EMSA, as predicted, and the control fragment showed no obvious binding (Figure 5.6C). Mutation of the $[A/T]_{\geq 4}$ moiety of –148 significantly reduced the MeCP2 dependent EMSA. The persistence of residual MeCP2 binding is likely the result of an additional A/T rich segment further downstream of the CpG (Figure 5.6C). Interestingly, the sequence corresponding to the *Bdnf* site (5'CGGAATT3') occurred six times in the Methyl-SELEX sample of 88 sequences (Figure 5.6A). To determine if MeCP2 utilizes the $[A/T]_{\geq 4}$ in the -148 *Bdnf* site for directional binding, DNaseI foot-printing analysis was performed. As was observed with artificially selected S1 fragment (Figure 5.3D), MeCP2 binding was not centred on the methyl-CpG site, but extended asymmetrically to cover the AATT run (Figure 5.6D). These data suggest that MeCP2 binding at this regulatory region is orientated with respect to the methyl-CpG through interaction with the adjacent $[A/T]_{\geq 4}$.

5.3 Discussion

5.3.1 MeCP2 binds unique loci which are not redundantly occupied by other MBP's

Studies analysing the specificity of MBD containing proteins for methylated DNA have mostly used artificial templates which contained a high density of CpG dinucleotides (Fraga et al., 2003; Meehan et al., 1992; Nan et al., 1993) and

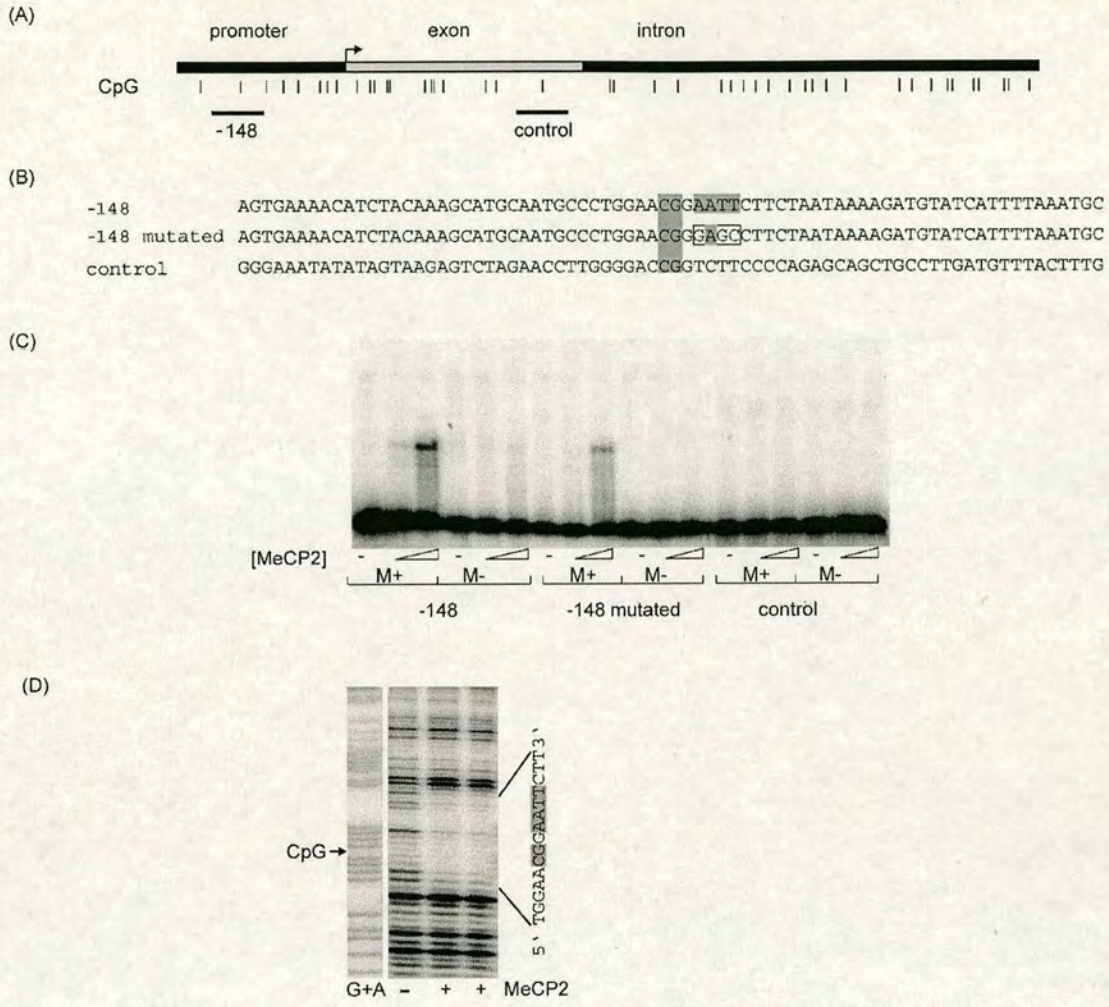


Figure 5.6- A natural MeCP2 target gene has a high affinity MeCP2 binding site.

(A) Analysis of the mouse *Bdnf* promoter IV region showing the MeCP2 binding site at -148 bases from the transcription start.
(B) Probes are shown corresponding to the -148 sequence, -148 with a mutated A/T run, and a CpG-containing fragment from exon 1 of the *Bdnf* gene.
(C) EMSA shows that efficient MeCP2 binding to the –148 probe is reduced when the A/T run in mutated. The exon 1 control probe shows little detectable binding. (D) DNaseI foot-printing analysis on ^{32}P labelled -148 *Bdnf* fragment in the presence (+) or absence (-) of MeCP2. The footprint due to the addition of MeCP2 is bracketed, with the protected DNA sequence to the right. G+A indicates the Maxam/Gilbert sequencing reaction on the same DNA fragment as size marker. Methyl-CpG is indicated by an arrow (left).

fragments containing methylated promoters regions of tumour suppressor genes (Fraga et al., 2003). The EMSA probes used in these studies show the MBD specifically binds to methylated DNA. The relative contribution of MBD proteins to occupancy of methyl-CpG dinucleotides *in vivo* is unclear. Previous studies using chromatin immunoprecipitation to profile MBD bound loci were carried out in cancer cell lines and thus addressed the targeting of MBD proteins in malignant sources (Ballestar et al., 2003; Koch and Stratling, 2004; Sarraf and Stancheva, 2004). Recently, a ChIP and clone approach was used to isolate MeCP2 bound loci from mouse brain, which led to the isolation of *DLX5/DLX6* genes as targets of MeCP2 mediated repression (Horike et al., 2005). Since cancer tissue and transformed cell lines have perturbed methylation levels and patterns (Klug and Famulok, 1994; Paz et al., 2003), our collaborator Dr. Stancheva used the primary human cell line MRC5 to analyse the binding characteristics of MeCP2. In Dr. Stancheva's study, MeCP2 antibodies were used to chromatin immunoprecipitate and subsequently clone genomic fragments bound by MeCP2. Surprisingly, MeCP2 bound to unique loci that were mostly unoccupied by the other MBD proteins. Interestingly, MBD2- bound loci identified by ChIP-and-clone, also lacked MeCP2 occupancy. The fact that very little MBP co-occupancy occurs at the isolated MeCP2 and MBD2 loci suggests that MBD containing proteins recognize methyl-CpG containing loci in a non-redundant fashion. MBD2 was able to occupy MeCP2 binding sites when MeCP2 was depleted suggesting that normal binding of MeCP2 at its target loci excludes binding of other MBP's, in particular MBD2.

5.3.2 Methyl-CpG with $[A/T]_{\geq 4}$ confers a high affinity directional binding site for MeCP2

Binding of MeCP2 to unique loci in ChIP-and-clone experiments led to the hypothesis that MeCP2 might have additional binding specificity outside methyl-CpG. A modified form of SELEX (Klug and Famulok, 1994), Methyl-SELEX, isolated high affinity MeCP2 binding sites which contain methyl-CpG with an adjacent run of four or more nucleotides of A/T rich DNA. This extra DNA binding specificity suggests that MeCP2 may recognize a sub-set of genomic methyl-CpG containing sites. Under normal conditions in MRC5 cells, MeCP2 can out-compete MBD2 for binding to

specific methyl-CpG sites, suggesting that $[A/T]_{\geq 4}$ may confer sequence specific and high affinity binding for MeCP2 to these sites. High affinity MeCP2 binding sites are non symmetrical and DNaseI footprint analysis indicates that MeCP2 binding covers the $[A/T]_{\geq 4}$ and is asymmetrical with respect to the methyl-CpG. A MeCP2 binding site in the promoter region of the *Bdnf* gene (Chen et al., 2003b; Klose and Bird, 2003; Martinowich et al., 2003) also showed directional MeCP2 binding and A/T run protection in DNaseI assays. It is particularly attractive to hypothesize that the direction in which MeCP2 binds a regulatory region may have an influence on how other protein co-factors are oriented with respect to elements necessary for gene regulation. Interestingly, other transcription factors, including E2F4 (Zheng et al., 1999) and SKN-1(Kophengnavong et al., 1999), bind a cognate sequence and rely on an adjacent run of A/T rich DNA for optimal DNA binding. This new insight into the binding specificity of MeCP2 will help to identify binding regions for MeCP2 in newly identified target genes, and it will be of interest to determine the consequence of directional MeCP2 binding.

5.3.3 Does MeCP2 bind to non-methylated DNA sequences?

Is MeCP2 primarily targeted to the sites identified here or is it potentially able to bind other sequences, including some that are non-methylated? The evidence that binding of MeCP2 is dependent on DNA methylation is based on multiple independent studies. In addition to many *in vitro* studies that define the preference for methylated CpG sites (Fraga et al., 2003; Meehan et al., 1992; Nan et al., 1993), ChIP analyses have consistently shown that MeCP2 associates with a specific DNA sequence *in vivo* when it is methylated, but not to the same sequence at the same genetic locus when it is hypo-methylated (El-Osta et al., 2002; Ghoshal et al., 2002; Gregory et al., 2001; Lorincz et al., 2001; Nan et al., 1996; Nguyen et al., 2001; Rietveld et al., 2002). In the present study, it was shown that 12 randomly selected MeCP2 binding sites all lose their association with MeCP2 when DNA methylation is reduced by 5-azacytidine treatment. Taken together, these studies leave little doubt that MeCP2 functions as a methyl-CpG binding protein *in vivo*.

In vitro studies showed that MeCP2 can associate with non-methylated fragments of genomic DNA and mouse satellite DNA, although DNA methylation, when present, significantly altered the pattern and increased the affinity of these interactions (Lewis

et al., 1992; Weitzel et al., 1997b). Analysis of binding sites for chicken MeCP2 (originally designated ARBP1; (von Kries et al., 1991b) identified a common GGTGT motif, mutation of which reduced the affinity for MeCP2. *In vitro* binding site selection with mammalian MeCP2 did not lead to enrichment of the GGTGT motif in the tight binding population of DNA fragments. Interestingly, earlier studies proposed that ARBP1 binding to GGTGT is enhanced by a flanking A/T run, as two molecules that intercalate with A/T DNA interfered with binding to MeCP2 (Buhrmester et al., 1995). These findings might be explained if GGTGT represents a weak mimic of methyl-CpG whose affinity for MeCP2 can be strengthened by a flanking run of A/T DNA. Whether this DNA methylation-independent binding has biological relevance remains to be seen.

Most surprising in this context is a report (Georgel et al., 2003) that MeCP2 can complex with non-methylated DNA assembled *in vitro* into nucleosomal arrays, leading to compaction of the arrays as assayed by electron microscopy. In contrast a prior study indicated that MeCP2 exclusively targets nucleosomal arrays that contain methylated DNA, but ignores non-methylated arrays in the same chromatin solution (Nan et al., 1997b). Given this discrepancy and the reported inability of others to find MeCP2 bound to non-methylated loci *in vivo* (see above), it will be essential to show that the association of MeCP2 with nucleosomes that carry non-methylated DNA is biologically relevant.

5.3.4 Do other MBP's have sequence specific DNA binding determinants?

MBD proteins were initially identified based on their ability to recognize methyl-CpG containing sequences, and little effort has been spent investigating whether sequence determinants in addition to a methyl-CpG contribute to their DNA binding properties. Evidence is provided in this chapter that demonstrates MeCP2 binds with a high affinity to sequences with a methyl-CpG and an adjacent run of $[A/T]_{\geq 4}$ DNA. These observations raise the important question of whether other MBP's utilize sequences in addition to methyl-CpG for high affinity binding. Currently, studies are underway to address this possibility utilizing the methyl-SELEX procedure, but some evidence is already emerging that suggests that MBD1 and Kaiso have additional targeting determinants. MBD1 can utilize a zinc co-coordinating CxxC domain to bind to

nonmethylated CpG dinucleotides (Jorgensen et al., 2004). Whether the MBD and the CxxC domain of MBD1 collaborate to recognize methyl-CpG in a specific context remains to be determined, but this suggests that there may be some interplay between methyl-CpG and nonmethyl-CpG DNA recognition by MBD1. Kiaso, a zinc finger DNA binding transcription factor binds specifically to methyl-CpG containing sequences. For Kaiso, DNA binding specificity has been partially determined, and it was shown to prefer methyl-CpG in the context of CGCG. It will be interesting to determine whether the additional DNA binding properties of MBD1 and Kaiso contribute to their binding specificities *in vivo*.

To date, MBD2 has not revealed any sequence specificity outside methyl-CpG. Bioinformatic analysis suggests that MBD2 is the ancestral methyl binding domain protein because it is the only MBD protein found in lower eukaryotes (Hendrich and Tweedie, 2003). No alternative DNA binding domains have been identified in MBD2, nor has sequence preference outside a single symmetrically methylated CpG dinucleotide been observed. It is possible that MBD2 first evolved the capacity to bind methyl-CpG dinucleotides in a context independent manner, and perhaps it has maintained this function where other MBD containing proteins have evolved and diverged in their DNA binding preferences. The data in this chapter supports this idea, as un-occupied MeCP2 loci were a good substrate for MBD2 but not MBD1. The sequence independent binding properties of MBD2 raise an interesting possibility with regard to the evolution of the MBD domain. If the ancestral MBD (MBD2) protein arose to recognize methyl CpG, in any context, it is possible to envisage that expansion of the MBD protein family in vertebrates created specialised MBD proteins that acquired additional DNA sequence specificity as seen for MeCP2. With identification of more MBP binding loci, and genes that these proteins regulate, it will begin to become more clear as to how the DNA methylation signal is interpreted by proteins that recognize methyl-CpG.

5.3.5 Mouse genetics suggest MBP's are not redundant

Null mice have been generated for all identified mammalian MBP's that are involved in transcriptional repression; MeCP2 (Chen et al., 2001; Guy et al., 2001), MBD1 (Zhao et al., 2003), MBD2 (Hendrich et al., 2001), MBD3 (Hendrich et al., 2001), and Kiaso (Selfridge et al unpublished). The phenotypes of these null mice are mild, with

the exception of MeCP2 which is a model for the debilitating human disease Rett syndrome. Interestingly, when MeCP2 and MBD2 null mice were crossed the phenotype of the resulting double null progeny was not more severe than the MeCP2 single null mice (Guy et al., 2001), perhaps indicating that MeCP2 is responsible for regulating a non-overlapping set of genes. This contention is further supported by the fact that a MeCP2, MBD2 and Kaiso triple null animals do not display synergy in phenotype, but again have the same symptoms as the MeCP2 null mouse (Selfridge et al unpublished). If MeCP2 affects a specific sub-set of methylated target genes, which can not be regulated by other MBP's, this could explain why its deficiency causes such a severe and non-overlapping phenotype when compared to other MBP null mice.

It is tempting to hypothesize from these genetic observations that MBP's are functioning in non-redundant pathways through intrinsic DNA recognition determinants. One could envisage that MeCP2, MBD1, and Kaiso are targeted to specific loci by intrinsic recognition of methyl-CpG in the context of a defined DNA sequence. In contrast, MBD2 might have a wider and more probabilistic affect on methyl-CpG dependent silencing due to a less discriminatory role in recognizing methylated loci (Hutchins et al., 2002).

5.3.6 Two modes of methyl-CpG recognition?

If DNA binding specificity has a role in division of work between MBP's one can envisage two modes of methyl-CpG dependent repression. The first mode would be global, in which a protein would have the capacity, like MBD2, to recognize methyl-CpG and repress transcription without context dependent requirements. Therefore, MBD2 with its tightly associated chromatin remodeling complex NURD may act as the global minder of methylation dependent silencing (Figure 5.7). The second mode would be a more defined and targeted means of methyl-CpG repression in which sequence specific determinants outside methyl-CpG alone are required to target the MBP's to genes they regulate. MeCP2, MBD1 and Kaiso might fit this role, as each has DNA binding preference, or activity, in addition to a single methylated CpG. As introduced in Chapter 1, site specific recruitment of MBP's may also explain their differential usage of co-repressor molecules. For instance, MeCP2 may require the services of Sin3a to repress a subset of methylated genes to which it is specifically

targeted by DNA specificity (Figure 5.7). Likewise, MBD1 and Kaiso would utilize their distinct co-repressor molecules to repress an independent subset of target genes. Therefore, within this division of labor, each MBP would be tailored to efficiently repress their respective target genes, and MBD2 would have a role in maintaining the global fidelity of methylation-mediated repression. With more defined understanding of DNA recognition by MBP's and their functions *in vivo* it will become more clear how the methyl-CpG signal is interpreted by MBP's.

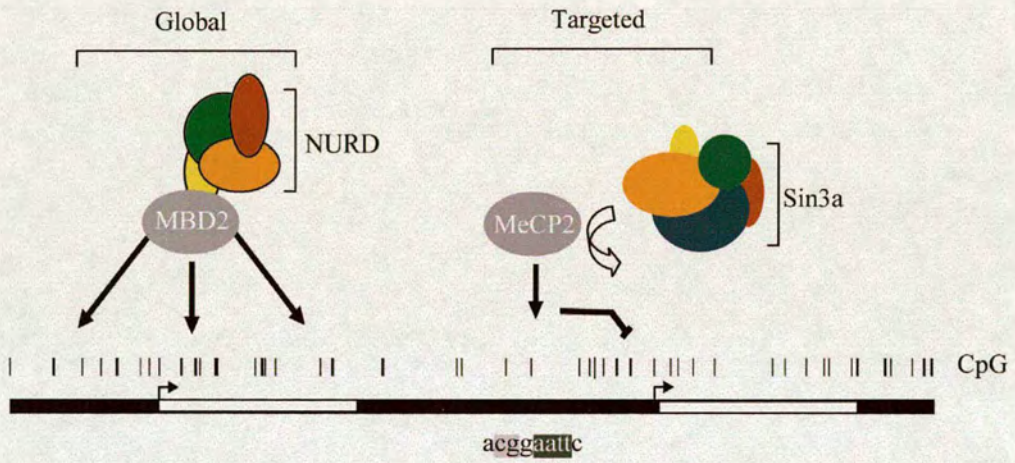


Figure 5.7- Two modes of methyl-CpG recognition?

MBD2 appears to recognize methyl-CpG independent of DNA context, and therefore may fit the role of global repressor of methylated loci (Left). On the other hand, factors like MeCP2 that have sequence specificity in addition to methyl-CpG may be targeted to specific methyl-CpG containing sequences (Right) and use specially tailored repression mechanisms to modulate the expression of a sub-set of specific target genes.

6. Chapter Six-
Phosphorylation of MeCP2

6.1 MeCP2 is a phosphoprotein

MeCP2 is involved in silencing the *Bdnf* gene in rodent neurons (Chen et al., 2003b; Martinowich et al., 2003). *Bdnf* silencing is in part maintained through binding of MeCP2 to methylated DNA upstream of the promoter, and recruitment of the Sin3a chromatin remodelling complex. One study suggested that MeCP2 was lost from the *Bdnf* promoter during gene activation due to a reduction of DNA methylation in the promoter region, but the changes in methylation were subtle (Martinowich et al., 2003). The second also study showed that *Bdnf* expression induced by calcium signalling pathway resulted in MeCP2 loss from the promoter concomitant with transcriptional activation (Chen et al., 2003b). Interestingly, in this study very little change in the DNA methylation pattern was observed at the *Bdnf* promoter region, before or after gene induction, indicating mechanisms other than loss of methylation profile are causative in removal of MeCP2 from the promoter (Chen et al., 2003b). The authors also showed that a more slowly migrating form of MeCP2 was observed by western blot analysis when calcium signalling pathways were activated. The slower migrating form of MeCP2 was eliminated by treating the protein extracts with recombinant phosphatase, demonstrating that the change in mobility was due to phosphorylation. This suggests that phosphorylation may have a role in liberating MeCP2 from the *Bdnf* gene during gene activation. In agreement with phosphorylation affecting MeCP2 occupancy of the *Bdnf* promoter during gene activation, south-western blot analysis showed that phosphorylated MeCP2 had reduced methyl-CpG binding capacity (Chen et al., 2003b). The identification if *Bdnf* as a MeCP2 target gene has led to the discovery that MeCP2 can act as dynamic interpreter of DNA methylation through modulation of its DNA binding capacity. Recently MeCP2 was identified in a proteomic screen for phosphoproteins in HeLa cell nuclear extract (Beausoleil et al., 2004), but little is know about the exact phosphorylation sites or kinases that mediate modification of MeCP2.

In this Chapter the DNA binding capacity of phosphorylated MeCP2 produced in baculovirus was analysed and a kinase activity that modifies MeCP2 was partially purified from HeLa cell nuclear extract.

6.2 Phosphorylation and MeCP2

6.2.1 Phosphorylation of baculovirus produced MeCP2 does not affect DNA binding

Recombinant MeCP2 produced in baculovirus is a phosphoprotein (Section 4.2.2). In light of the recent observation that phosphorylation of MeCP2 affects DNA binding (Chen et al., 2003b; Martinowich et al., 2003), the capacity of Baculovirus produced MeCP2 to bind methylated DNA was analyzed. Full length MeCP2 protein does not perform well in EMSA, therefore a protein encompassing amino acids 1-205 of MeCP2, that works well in EMSA, was produced in baculovirus. MeCP2 1-205 viral stock was amplified and used for large scale infection of SF9 cells (~ 10^9 cells)(Figure 6.1A). The clarified extract was applied to Ni-Nta resin and eluted in 250 mM imidazole. A strong Coomassie stained band at approximately 32.5 kDa on SDS PAGE, corresponding to MeCP2 1-205, was observed in the first two elutions from the Ni-Nta column (Figure 6.1A). MeCP2 containing fractions were pooled, applied to a Mono-S cation exchange column, and eluted with a linear gradient of increasing NaCl concentration from 0.2 M to 1 M. The majority of MeCP2 eluted over three fractions (Figure 6.1A), which were dialyzed into storage buffer and the concentration determined by a combination of Bradford assay and visual quantification by Coomassie stained SDS-PAGE (Figure 6.1A).

To verify that MeCP2 1-205 was also modified by phosphorylation in SF9 cells, purified MeCP2 1-205 was treated with lambda phosphatase, in the presence or absence of phosphatase inhibitors, and then analyzed for a change in electrophoretic mobility. As was observed with full length MeCP2, treatment with phosphatase caused an increase in the mobility of MeCP2 1-205 and this was ablated by phosphatase inhibitors (Figure 6.1B). Although the change in mobility is very modest upon phosphatase treatment the increase in mobility is convincingly demonstrated on a high resolution gel in Figure 6.2A. In contrast to full length MeCP2 (Section 4.2.2),

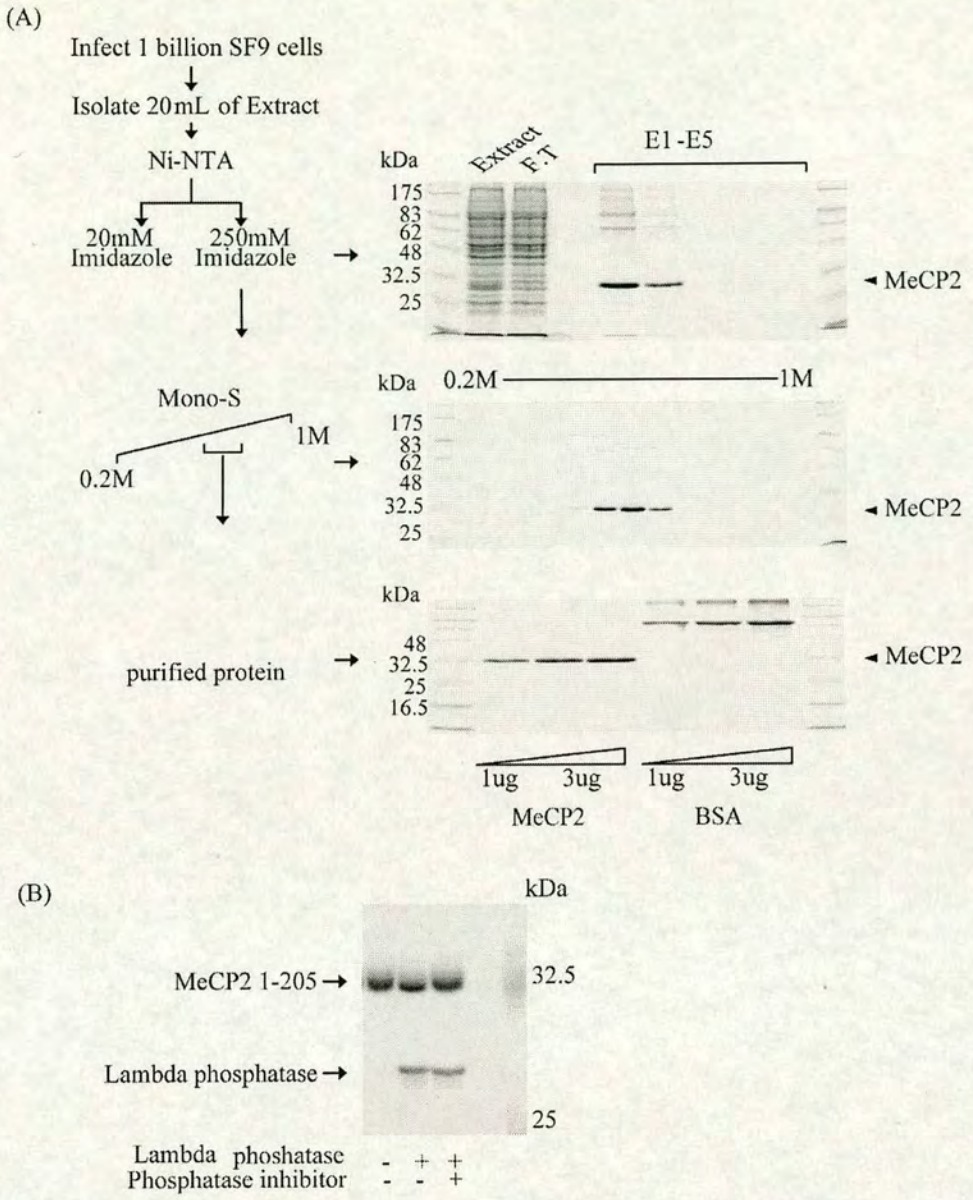


Figure 6.1- Production of baculovirus MeCP2 1-205.

(A) Amplified viral stock was used to infect 10^9 SF9 cells and 20 ml of clarified whole cell extract was applied to a Ni-Nta column. MeCP2 eluted from the Ni-Nta column with 250 mM imidazole. Elutions (E1-5) were analysed by Coomassie stained SDS PAGE. MeCP2 containing fractions E1 and E2 were applied to a Mono-S cation exchange column and eluted with a linear gradient of NaCl from 0.2 M to 1M. MeCP2 eluted between 450-550 mM NaCl as determined by Coomassie stained SDS PAGE. MeCP2 containing fractions were dialysed into storage buffer and the protein concentration determined by Bradford and Coomassie stained SDS PAGE.
(B) Baculovirus produced MeCP2 1-205 is a phosphoprotein. Purified MeCP2 1-205 was treated with lambda phosphatase and this resulted in an increase in electrophoretic mobility which was alleviated by phosphatase inhibitors.

MeCP2 1-205 exhibited a less pronounced shift in mobility indicating that perhaps fewer residues are modified in this truncated molecule.

To accurately determine whether the DNA binding capacity of the phosphorylated Baculovirus MeCP2 1-205 protein was different from the unmodified protein it was first necessary to remove phosphorylation from an aliquot of protein sample. To this end, phosphorylated MeCP2 1-205 was incubated with lambda phosphatase or mock treated with lambda phosphatase, and then, purified away from the phosphatase protein by cation exchange chromatography (Figure 6.2A). Purified phosphorylated and dephosphorylated proteins were quantified by Bradford analysis and visualization by Coomassie stained SDS PAGE. Efficient dephosphorylation was evident from the increased mobility of the lambda phosphatase treated material (Figure 6.2A). The DNA binding efficiently of the modified MeCP2 1-205 protein was compared to the unmodified protein in EMSA using probe S1 (Section 5.2.3). MeCP2 1-205 bound specifically to methylated DNA irrespective of its phosphorylation state, and had similar DNA binding properties to the corresponding protein produced in bacteria (Figure 6.2B). These data suggest that phosphorylation of MeCP2 in SF9 cells is not sufficient to inhibit DNA binding, and therefore the modifications in Baculovirus produced MeCP2 may not correspond to the phosphorylation modifications that inhibit DNA binding in mammalian cells.

6.2.2 An activity that can phosphorylate MeCP2 is present in HeLa nuclear extracts

A recent report identified MeCP2 as a phosphoprotein in HeLa cell nuclear extract (Beausoleil et al., 2004). It would be of interest to identify kinases that affect MeCP2 in mammalian cells and use this information to understand what affect phosphorylation has on normal MeCP2 function. To identify nuclear MeCP2 kinase activity an *in vitro* assay was devised to monitor incorporation of phosphate groups into MeCP2. The basis of the assay is very simple and involves producing full length recombinant MeCP2 in bacteria and then mixing it with nuclear extract. By adding a ^{32}P labeled phosphate donor (^{32}P gamma-ATP), potential kinase activity can be monitored by incorporation of isotopic phosphate into MeCP2. Since MeCP2 is

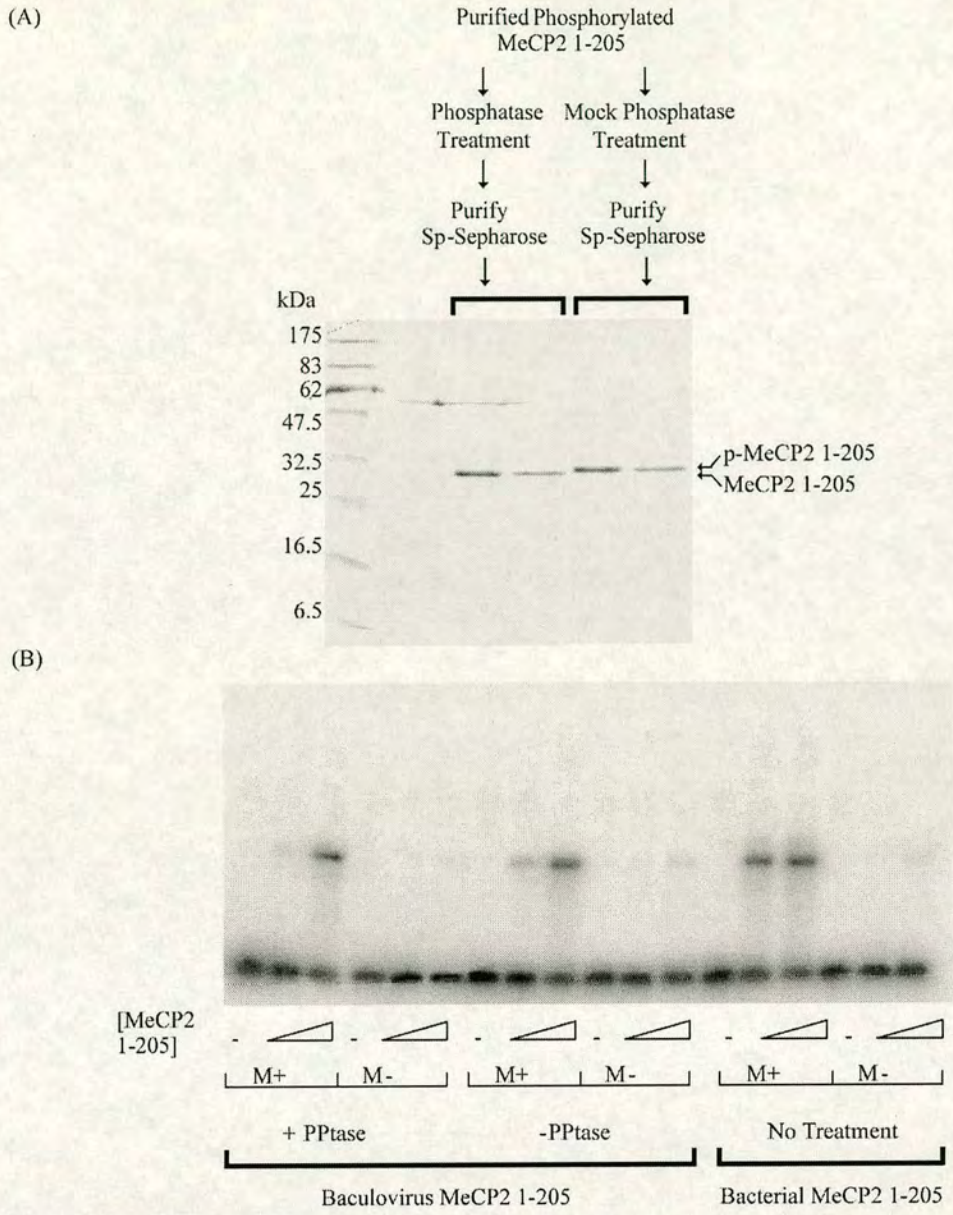


Figure 6.2- Phosphorylation of baculovirus-produced MeCP2 1-205 does not inhibit DNA binding

(A) Preparative amounts of phosphorylated and dephosphorylated MeCP2 were made by treating purified baculovirus MeCP2 1-205 with lambda phosphatase (dephosphorylated) or a mock treatment (phosphorylated). The MeCP2 molecules were re-purified by Sp-Sepharose cation exchange chromatography and visualized by Coomassie stained SDS PAGE. The phosphorylated form of MeCP2 has a slower mobility compared to the dephosphorylated protein.
(B) The phosphorylated and dephosphorylated MeCP2 1-205 samples were used in an EMSA experiment. The phosphorylation status of the MeCP2 molecule did not affect DNA binding and both fragments bound specifically to methylated DNA with a similar affinity to MeCP2 1-205 produced in bacteria

phosphorylated in HeLa cells, HeLa nuclear extract was used as a source potential kinase activity, and an *in vitro* kinase assay was carried out. The reaction was carried out using various concentrations of HeLa nuclear extract in the presence of phosphatase inhibitors so that newly incorporated phosphate groups would not be removed. The *in vitro* kinase reaction products were then separated by SDS-PAGE and visualized by phosphor-imaging. Labeled phosphate was efficiently incorporated into recombinant MeCP2 indicating there is a kinase activity in HeLa nuclei that can phosphorylate MeCP2 (Figure 6.3). To verify that the ^{32}P labeled band in the *in vitro* kinase assay was MeCP2, a parallel experiment was carried out in which the MeCP2 protein was immunoprecipitated before SDS PAGE analysis. The resulting signal on the phosphor-imager corresponded to the immunoprecipitated MeCP2 protein, verifying that MeCP2 is the target of phosphorylation in these assays (Figure 6.3).

6.2.3 There are two major activities in HeLa nuclei that phosphorylate MeCP2

To gain a better understanding of the biochemical nature of the MeCP2 kinase activity, HeLa nuclear extract was subjected to biochemical fractionation (Figure 6.4A). The majority of the kinase activity extracted from HeLa nuclei remained soluble after dialysis into low salt anion exchange buffer and the soluble extract was applied to a 1 ml fractogel-TMAE anion exchange column. The bound protein was eluted with increasing salt concentration, and the majority of the MeCP2 kinase activity eluted from the column between 250 and 400 mM NaCl as indicated by an *in vitro* kinase assay (Figure 6.4B). The active fractions were pooled and applied to a Superose 12 size exclusion column. *In vitro* kinase reactions were carried out on the Superose 12 fractions and two peaks of activity were resolved, one corresponding to ~ 300 kDa and the second corresponding to ~ 45 kDa (Figure 6.4C). Therefore, HeLa nuclei have two major MeCP2 kinases that bind an anion exchange column have different apparent molecular weights by size exclusion chromatography.

To verify that the MeCP2 kinase activities that elute in the same fractions on the anion exchange column are biochemically distinct, HeLa nuclear extract was applied to a 1 ml Sp-Sepharose cation exchange column and eluted with increasing salt

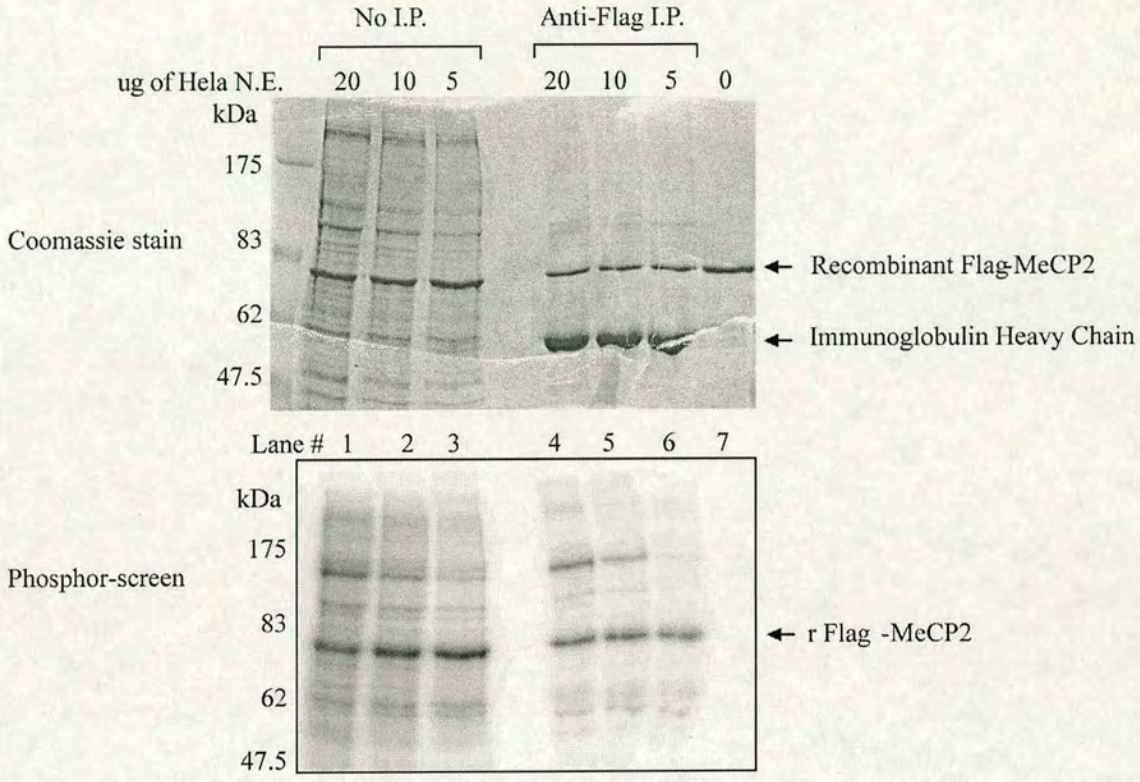


Figure 6.3- HeLa cells have MeCP2 kinase activity

An *in vitro* kinase assay was designed to monitor kinase activity directed towards MeCP2. Varying amounts of HeLa nuclear extracts (20, 10, and 5 ug) were mixed with recombinant MeCP2, and ^{32}P labelled ATP to monitor MeCP2 phosphorylation (Lanes 1, 2, and 3). The top gel is a Coomassie stained gel of the proteins in the reaction, and the bottom gel is a phosphor-imager screen analysing incorporated isotope. MeCP2 was heavily labelled by addition of HeLa nuclear extract in the *in vitro* kinase assay. To verify that the phosphorylated 80 kDa band was MeCP2 the same reaction was carried out, but before analysis by SDS PAGE the recombinant MeCP2 protein was immunoprecipitated by virtue of an N-terminal flag-tag. The 80 kDa protein with incorporated ^{32}P immunoprecipitated with flag antibody (Lanes 4, 5, and 6) verifying that it is recombinant MeCP2.

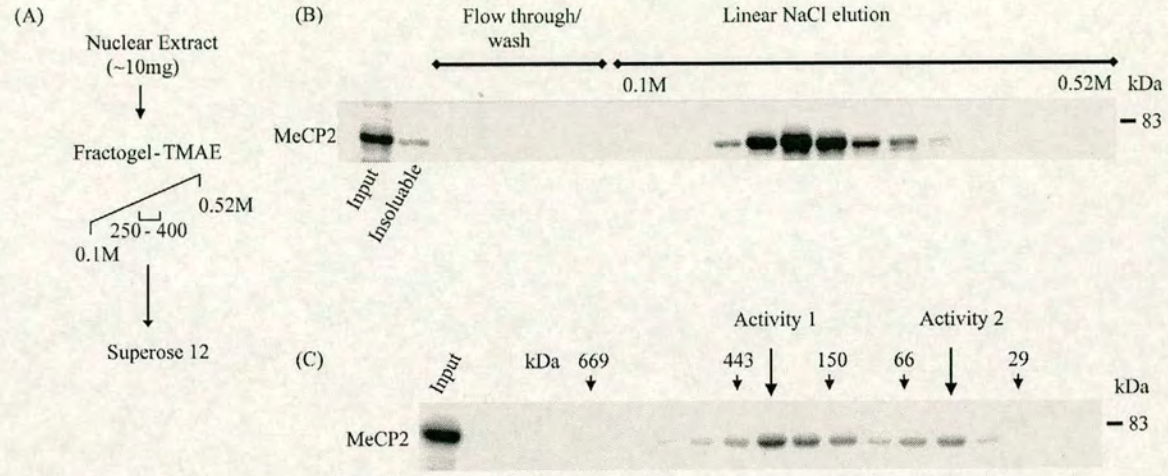


Figure 6.4- HeLa cell nuclear extract contains two MeCP2 kinase activities

(A) To biochemically characterize the MeCP2 kinase activity 10 mg of HeLa nuclear extract was applied to a 1 ml Fractogel-TMAE anion exchange column and eluted with increasing NaCl concentration.
(B) The fractions collected from the anion exchange column were used in an *in vitro* kinase reaction. The majority of the MeCP2 kinase activity eluted between 250 and 400 mM NaCl.
(C) The fractions containing MeCP2 specific kinase activity from the anion exchange column were applied to a Superose 12 size exclusion chromatography column. The size excluded fractions were used in an *in vitro* kinase assay, which resolved two MeCP2 kinase activities, the first with a relative molecular weight of ~ 300 kDa and the second with a relative molecular weight of ~ 60 kDa.

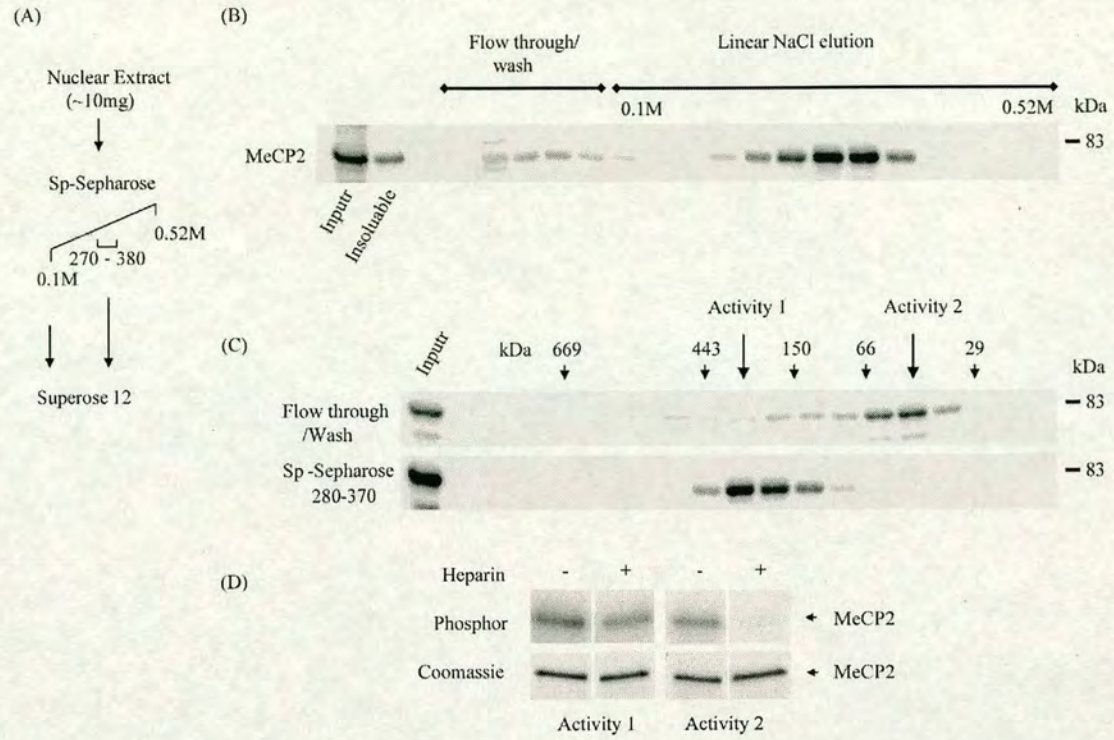


Figure 6.5- The two nuclear HeLa MeCP2 kinase activities are biochemically distinct

(A) HeLa cell nuclear extract was applied to a Sp-Sepharose cation exchange column and eluted with increasing NaCl concentration.
(B) The fractions collected from the cation exchange column were used in an *in vitro* kinase reaction. MeCP2 kinase activity was observed in both the flow through / wash phase of the elution and the 280 and 370 mM NaCl elution.
(C) Both the flow through / wash elution and the 280 – 370 mM NaCl elutions were applied independently to a Superose 12 size exclusion column. The flow through / wash phase corresponded to the ~ 60 kDa MeCP2 kinase activity (activity 2), and the 280 – 370 mM elution corresponded to the ~ 300 kDa activity (activity 1).
(D) To determine if kinase activity one or two were Casein Kinase II (CKII), the peak fractions from the Superose 12 size exclusion step were used for *in vitro* kinase assays in the presence or absence of heparin, which inhibits CKII activity. MeCP2 kinase activity 1 is not affected by heparin, but activity 2 was completely abolished by the addition of heparin to the reactions. Therefore activity 2 likely corresponds to CKII.

concentration in an attempt to separate the two activities (Figure 6.5A). Again, the majority of the kinase activity remained soluble when dialysed into low salt cation exchange buffer, but in contrast to the anion exchange column, some of the activity eluted in the flow through / wash stage and the remainder between 280 and 370 mM NaCl (Figure 6.5B). To determine if these two elution profiles corresponded to a separation of the MeCP2 kinase activities, each was applied to a Superose 12 column independently. The flow through / wash activity eluted from the size exclusion column at ~ 45 kDa and the 280 – 370 mM elution corresponded to the ~ 300 kDa activity (Figure 6.5C). Therefore, cation exchange chromatography is efficient in separating the two MeCP2 kinase activities.

One of the most abundant nuclear kinases in the HeLa cells is Casein kinase II (CKII). CKII has relatively loose substrate specificity and often phosphorylates a multitude of substrates *in vitro* that it may not significantly modify *in vivo* (Meggio and Pinna, 2003). Fortunately, CKII activity is readily inhibited by heparin, so its presence can be easily identified in crude extracts by testing the sensitivity of the kinase activity to heparin (O'Farrell et al., 1999). To determine if either of the two MeCP2 kinase activities was due to CKII, an *in vitro* kinase assay was carried out in the presence of a CKII inhibiting concentration of heparin. The ~ 300 kDa (activity 1) or the ~ 45 kDa (activity 2) activity was used as an enzyme source. Activity 1 is unaffected by heparin, but activity 2 was completely abolished by addition of heparin, suggesting activity 2 contains CKII (Figure 6.5D).

6.2.4 Partial purification of MeCP2 kinase activity 1

The speed and sensitivity of the *in vitro* kinase assay makes it ideal for following kinase activity during biochemical purifications. Several other studies have successfully used this approach to purify and identify kinases that modify a specific target molecule (Friedl et al., 2003; Hong et al., 2004). MeCP2 kinase activity 2 likely corresponds to nuclear CKII and was not nearly as robust as activity 1. Therefore activity 1 was chosen for further purification and analysis. Large amounts (~ 1.3 grams) of HeLa nuclear extract were prepared using a 350 mM NaCl extraction followed by dialysis into AE100. The dialysed extract was applied to a 180 ml Fractogel-

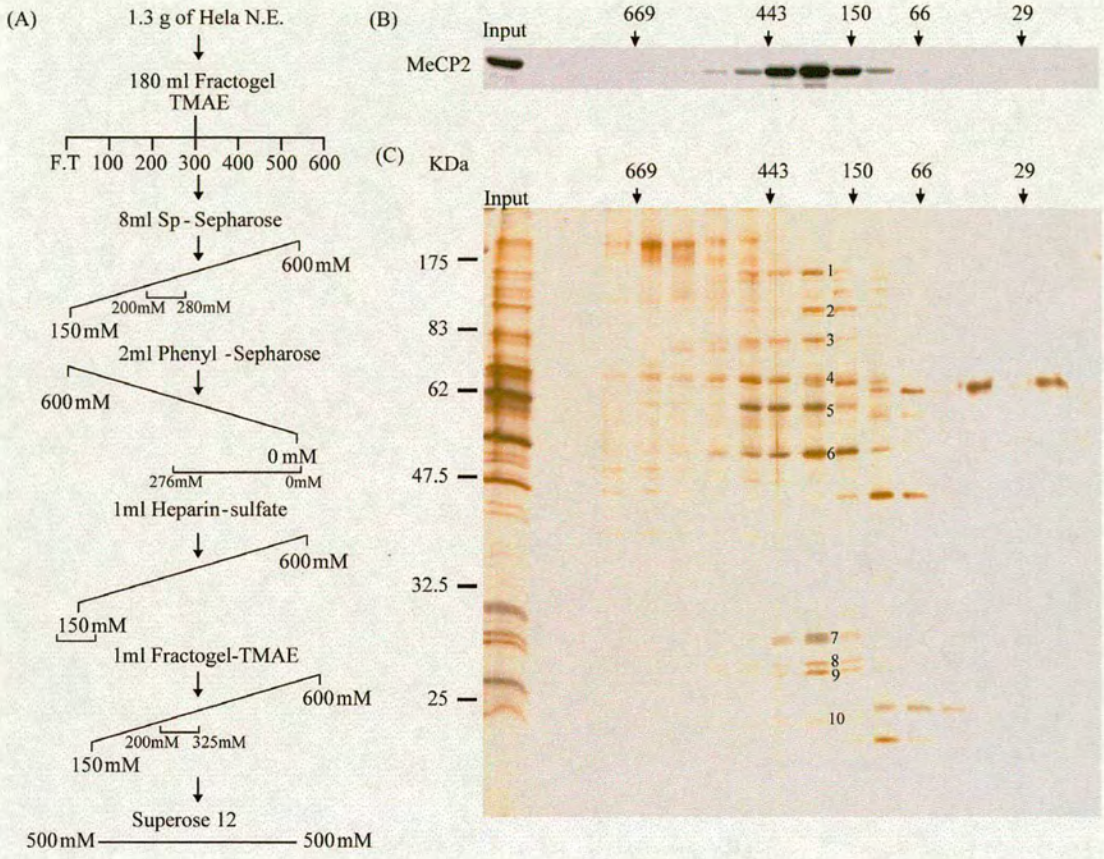


Figure 6.6- Partial purification of MeCP2 kinase activity 1.

(A) A schematic illustrating the purification steps and elution conditions used to purify MeCP2 kinase activity 1. At each step of the purification the activity was followed by *in vitro* kinase activity against recombinant MeCP2.
(B) The fractions collected from the final Superose 12 purification step were used in an *in vitro* kinase assay. MeCP2 kinase activity 1 was purified intact and retained its ~ 300 kDa relative molecular weight (Compare with Figure 6.5).
(C) An aliquot of the fractions used in (B) were separated on a 10 % SDS PAGE gel and analysed by silver stain. The peak fraction containing MeCP2 kinase activity consists of ten major polypeptides, four of which (6,7,8, and 9) co-migrate with the MeCP2 kinase activity.

TMAE column and batch eluted in steps of 100 mM NaCl from AE100 to AE600. Both activities eluted in the AE300 elution (Figure 6.6A). The AE300 elution was dialysed into CE150 and applied to an 8 ml Sp-Sepharose column. Activity 1 was recovered by a linear gradient elution from CE150 to CE600 in fractions between 200 and 280 mM NaCl. These fractions were then dialysed into PE600 and applied to a 2 ml Phenyl-Sepharose hydrophobic interaction column and eluted with a downward gradient of ammonium sulfate from 600 mM to 0 mM. The majority of activity eluted between 276 mM and 0 mM ammonium sulfate. The active fractions were pooled and dialysed into CE150 and applied to 1 ml heparin sulfate affinity column. The majority of the MeCP2 kinase activity was found in the flow through from the Heparin column. The flow through was dialysed into AE150 and applied to a 1 ml Fractogel-TMAE column and eluted with a linear salt gradient from 150 mM to 600 mM. The MeCP2 kinase activity eluted between 200 and 325 mM NaCl. The active fractions from the Fractogel-TMAE column were pooled and directly applied to a Superose 12 size exclusion column. The fractions from the size exclusion column were used in a kinase assay (Figure 6.6B), and the resulting activity had a apparent molecular weight of ~ 300 kDa indicating that activity 1 has been specifically enriched. Aliquots from the Superose 12 fraction were separated by SDS-PAGE and analyzed by silver staining. Four major polypeptides appear to co-fractionate with the observed MeCP2 kinase activity (Figure 6.6C).

6.2.5 CDK1 corresponds to the peak of MeCP2 kinase activity

The peak fraction from the Superose 12 size exclusion step was electrophoresed on a 10 % SDS PAGE gel and visualized with Sypro-Ruby stain (Figure 6.7). Ten major polypeptides were observed in the active fraction, of which four proteins co-elute with the MeCP2 kinase activity on the Superose 12 column (Figure 6.6C). The isolated proteins were subjected to identification by mass spectrometry. The only kinase identified in the in the ten bands analyzed was human CDK1, which was identified as of two individual proteins (Figure 6.7). Because the relative molecular weight of the MeCP2 kinase activity is ~ 300 kDa by size exclusion chromatography it is possible that the activity is contained in a protein complex containing CDK1. In this purification one of the other bands that co-migrates with CDK1 by size exclusion

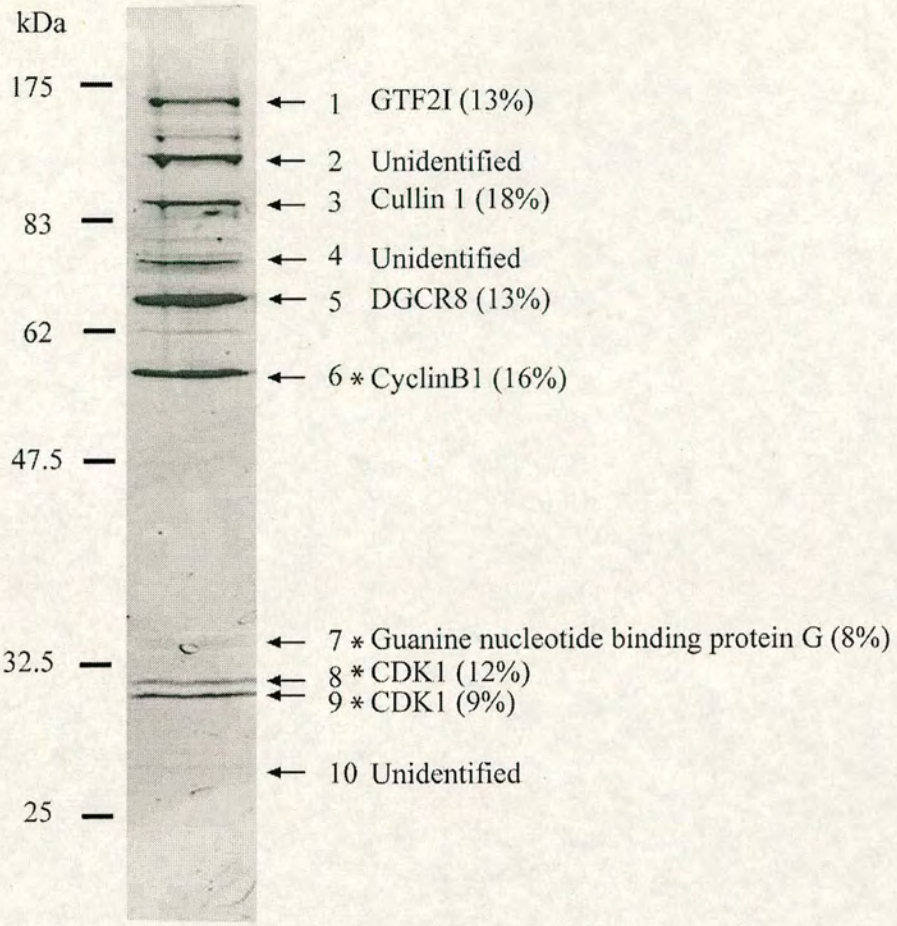


Figure 6.7- MeCP2 kinase activity 1 contains CDK1/CyclinB1

The proteins from the peak MeCP2 kinase activity fraction in the Superose 12 size column step were separated on a 10 % SDS PAGE gel and analysed by Sypro-Ruby staining. Bands were identified by mass spectrometry (right) and the percentage (%) peptide coverage of the protein indicated in brackets

chromatography is CyclinB1. Previous studies have shown that CyclinB1 is stably associated with CDK1 at most stages of the cell cycle. The fact that CDK1/CyclinB1 purify together as an MeCP2 kinase activity suggests that MeCP2 is modified by the HeLa cell equivalent of the maturation promoting factor (MPF) protein complex which consists of CDK1 and CyclinB1 (Doree and Hunt, 2002). An additional protein that co-elutes with CDK1, Guanine nucleotide binding protein G, is not likely to be a component of the MPF complex, as it did not co-migrate with CDK1/CyclinB1in a previous purification step (data not shown). Further purification of the MeCP2 kinase activity to homogeneity will clarify if there are other constituents of the MeCP2 kinase complex in addition to CDK1/CyclinB1.

6.3 Discussion

6.3.1 Phosphorylation of MeCP2 in baculovirus does not inhibit DNA binding.

In rodent neurons, activation of calcium signalling pathways causes phosphorylation of MeCP2 which appears to inhibit its DNA binding capacity (Chen et al., 2003b). In this Chapter, a phosphorylated fragment of MeCP2 consisting of amino acids 1-205 was expressed in SF9 cells. Phosphorylation did not inhibit binding to methylated DNA as tested by EMSA analysis. Since there is no insect homologue of MeCP2, phosphorylation of exogenous MeCP2 in this protein expression system is likely the result of an insect kinase with broad substrate specificity. These data indicate that residues modified by phosphorylation in MeCP2 1-205 do not interfere with DNA binding, but this does not rule out the possibility that phosphorylation of MeCP2 in C-terminus (amino acids 205-486) may contribute to inhibition of DNA binding. Assuming that phosphorylation of MeCP2 over the MBD domain will be involved in regulating DNA binding, the inhibition of DNA binding which is observed in rodent neurons is likely to be the result of a kinase with different modification target sites to those modified the baculovirus MeCP2 protein. It may be of interest to identify modified sites in baculovirus-expressed MeCP2 to eliminate these residues as potential modification sites involved in inhibition of DNA binding in rodent neurons.

6.3.2 CDK1 phosphorylates Mecp2

CDK1 is the mammalian homologue of the fission yeast cdc2 (bakers yeast cdc28) protein kinase. The major cellular function of CDK1 is to promote progression through the G2/M phases of the cell cycle (Murray, 2004). Activation of the CDK1/Cyclin B1 containing maturation promoting factor (MPF) complex allows phosphorylation of a variety of nuclear and cytoplasmic substrate molecules (Murray, 2004). The exact role of CDK1 in modifying proteins required for cell cycle progression remains incompletely defined. CDK1 is mostly cytoplasmic prior to catalytic activation at which point it is translocated to the nucleus (Kao et al., 1999; Riabowol et al., 1989). At the G2/M transition there is an influx of active CDK1 into the nucleus, and this correlates with phosphorylation of nuclear substrates. An active CDK1 complex is found in nuclear extracts of rapidly dividing cells (Kao et al., 1999), and one of the first homogenous purifications of CDK1 from mammalian cells was obtained by salt extraction of rat tissue culture cell nuclei (Chambers and Langan, 1990). Rapidly dividing HeLa cells also contain abundant nuclear MPF as other groups have observed a five fold increase in enzyme recovery by addition of a nuclear extraction step to the typical cytoplasmic preparation (Bischoff et al., 1990). In this chapter CDK1 was purified from a nuclear preparation of asynchronously growing HeLa cells, using MeCP2 as a substrate. Given that CDK1 was purified from the nucleus in a catalytically active state this likely corresponds to the active sub-population of CDK1/CyclinB1found in the nucleus prior to the breakdown of the nuclear envelope at the G2/M transition. Interestingly, the linker histone H1 is also a substrate for MPF, suggesting that phosphorylation of chromatin associated proteins may be an important function of CDK1/CyclinB1 in cycle progression. It will be important to further purify the MeCP2 kinase activity to determine whether other factors in addition to CDK1/CyclinB1 are components of the MeCP2 kinase complex.

6.3.3 The MeCP2 kinase complex, a work in progress...

How relevant phosphorylation of MeCP2 by CDK1/CyclinB1 is *in vivo* remains unknown, but the nature of this kinase suggests that MeCP2 may be modified during the cell cycle. It would be premature to speculate on the biological relevance of the CDK1 phosphorylation of MeCP2 without verifying that CDK1 does indeed modify MeCP2 *in vitro* and *in vivo*. To try and determine whether CDK1 is a *bona fide* MeCP2 kinase this initial partial purification will be followed up by two avenues of *in*

vitro and *in vivo* analysis. To test the functional importance of MeCP2 phosphorylation by CDK1, active recombinant CDK1/CyclinB1, which is commercially available, will be used to modify MeCP2 substrates. Large amounts of phosphorylated MeCP2 will be used to identify modified sites through mass spectrometry. In addition, the purified CDK1 enzyme will be used to verify modification sites using commercially available MeCP2 peptide arrays. The consequences of MeCP2 phosphorylation on DNA binding will be analyzed using purified *in vitro* phosphorylated MeCP2 and EMSA analysis. If an affect of phosphorylation on DNA binding is observed, mutagenesis will be utilized to determine which residues are important for phosphorylation-mediated inhibition of DNA binding. To address whether CDK1 is responsible for phosphorylation of MeCP2 *in vivo*, the modification state of MeCP2 will initially be analyzed during the cell cycle using synchronized HeLa cells. The aim of this experiment will be to determine whether phosphorylation of MeCP2 occurs at the same time as temporal peaks of CDK1 activity observed throughout the cell cycle. Through combined *in vitro* and *in vivo* analysis the relevance of CDK1 kinase activity in the modification of MeCP2 will be determined.

6.3.4 Are there other important MeCP2 kinases?

MeCP2 is phosphorylated by CDK1 *in vitro,* but this is unlikely to account for the modification of MeCP2 observed in terminally differentiated non-dividing neurons (Chen et al., 2003b; Martinowich et al., 2003). It is highly possible that MeCP2 is modified by several kinases, and that different activities target MeCP2 in different tissues and cell types. Recent studies have identified mutations in the STK9/*CDKL5* gene in patients with a phenotypic variant of Rett syndrome (Scala et al., 2005; Tao et al., 2004; Weaving et al., 2004). Individuals with mutations in *STK9/CDKL5* have similar phenotypic characteristics to patients with mutations in *MECP2* except that they also suffer from infantile spasms (Scala et al., 2005; Tao et al., 2004; Weaving et al., 2004). Interestingly, the predicted STK9/CDK5L protein has homology to known serine-threonine kinases (Montini et al., 1998). It is possible that STK9/CDKL5 could phosphorylate MeCP2 and regulate some aspect of normal MeCP2 function. This possibility is attractive given the overlap in symptoms between patients with mutations in *MECP2* and *STK9/CDKL5*, and the recent observation that MeCP2 is

phosphorylated in neurons. In future studies it will be important to determine if STK9/CDKL5 is catalytically active and whether MeCP2 is a substrate.

7. Chapter Seven-
Conclusions / Future Directions

7.1 Learning more about MeCP2 using biochemistry

In aim of this study was to use a biochemical approach to better understand MeCP2 function. There were three basic questions that were posed to initiate these studies; (1) What are the native biochemical properties of MeCP2? (2) What specific DNA sequences does MeCP2 bind? and (3) What are the affects of post-translational modification on MeCP2?

To investigate the native biochemical properties of MeCP2, rat brain MeCP2 was purified to near homogeneity (Section 3.3). This purification showed that mammalian MeCP2 is not a stable component of the Sin3a chromatin remodelling complex. Subsequent biochemical studies using *Xenopus laevis* oocyte extract also demonstrated that MeCP2 is not a component of the *Xenopus* Sin3a complex. Biophysical analysis of purified recombinant MeCP2 showed that MeCP2 is a highly elongated molecule. Determining why MeCP2 is elongated remains a topic for future investigation, but this feature may aid in protein-protein interaction or be required for chromatin associated functions.

In collaboration with Dr Irina Stancheva and Dr. Lars Schmiedeberg the DNA sequence binding specificity of MeCP2 was analysed both *in vitro* and *in vivo*. Differential occupancy of methyl-CpG containing DNA binding sites *in vivo* complemented *in vitro* studies that uncovered a new DNA binding requirement for efficient recognition of methyl-CpG by MeCP2 (Section 5.2.4). Using a newly developed SELEX technique, called methyl-SELEX, it was shown that MeCP2 requires an $[A/T]_{\geq 4}$ adjacent to methyl-CpG for high affinity DNA binding. The importance of this DNA binding requirement was highlighted by the observation that an MeCP2 binding site in the *Bdnf* promoter requires an $[A/T]_{\geq 4}$ for efficient binding to the methyl-CpG. Interestingly, binding sites with an $[A/T]_{\geq 4}$ result in asymmetrical association of MeCP2 with CpG dinucleotide, suggesting that high affinity binding sites may result in directional orientation of MeCP2 with respect to surrounding DNA elements. This observation suggests that directional association of MeCP2 with

DNA may impact how recruited co-factors are oriented with respect to target genes. The relevance of directional DNA binding by MeCP2 to transcriptional repression remains an interesting topic for further investigation.

The observation that MeCP2 is phosphorylated during gene activation events at the *Bdnf* promoter (Chen et al., 2003b) has sparked a floury of interest within the field regarding the affects of phosphorylation on MeCP2 repression. In order to study the affects of phosphorylation on MeCP2 it will be important to identify modified residues and potential kinases. A preliminary study in this thesis indicates that the phosphorylation that inhibits DNA binding of MeCP2 in rodent cells is likely due to modification of specific target residues, and not simply a protein charge affect, as phosphorylated MeCP2 obtained from insect cells bound methylated DNA efficiently. Using MeCP2 as a bait molecule for nuclear kinases, CDK1 was identified as a potential MeCP2 kinase. Future studies on this enzyme will help to map sites in MeCP2 that are modified by phosphoryation, and provide a basis for examining the affects of these modifications on DNA binding and MeCP2 function *in vivo*.

7.2 Determining co-factors required for MeCP2 function

Given that other MBD molecules are part of multi-protein complexes, it was unexpected to discover that MeCP2 was not stably associated with other co-repressors. Other MBD proteins appear to rely on stably associated co-factors to repress transcription and remodel chromatin (Ng et al., 1999; Sarraf and Stancheva, 2004; Zhang et al., 1999). This raises questions regarding which cellular co-repressors are important for transcriptional repression by MeCP2, and a future challenge will be to delineate which transiently interacting molecules are used by MeCP2 to carry out its normal functions within the cell. The observation that MeCP2 is a monomeric molecule does suggest that MeCP2 may use a varying repertoire of co-factors in regulating context dependent transcription. As more MeCP2 target genes are identified a unique opportunity will arise to interrogate which MeCP2 binding co-factors are used at specific promoters to repress transcription. It will be of particular interest to understand the contribution of MeCP2 to re-establishing silencing of inducible genes like *Bdnf*. For example, when the expression of an inducible gene is no longer required, is there a hierarchal recruitment of silencing factors to the promoter that culminates with the arrival of MeCP2 to lock the gene into a silenced

state, or is MeCP2 associated with the initiation of silencing through recruitment of co-factors like HDAC's and HMT's? The ease with which endogenous protein levels can be manipulated using techniques like RNA interference, it will be possible to use model gene systems to determining how DNA methylation, and specifically MeCP2, is associated with gene silencing.

7.3 The epigenome and MeCP2 targeting

One exciting aspect of this study was uncovering a new DNA binding requirement for MeCP2. The observation that MeCP2 requires $[A/T]_{\geq 4}$ in addition to methyl-CpG for efficient DNA binding will likely impact how the repressive capabilities of MeCP2 are targeted to sites of DNA methylation within the cell. Teasing out the biological affect of this binding preference is challenging as there appears to be an inherent flexibility in the distance that an $[A/T]_{\geq 4}$ run can be situated from the methyl-CpG, and therefore the occurrence of these sequences within the genome is high. To efficiently use this knowledge to predict MeCP2 target genes is even more challenging given the lack of information regarding the precise DNA methylation patterns near promoters and regulatory elements. The advancement of epigenome sequencing projects will provide valuable information regarding DNA methylation patterns in cells (Gius et al., 2004; Rakyan et al., 2004). Epigenome information will allow categorization of potential MeCP2 binding loci based on knowledge of both methylation status and the presence or absence of an $[A/T]_{\geq 4}$. A more defined set of potential promoter associated MeCP2 binding sites will facilitate targeted investigation of candidate genes to determine whether methylation and MeCP2 is important for their regulation. The advantage of this approach is the potential to exclude CpG dinucleotides that are not methylated, and concentrate on gene associated methyl-CpG sites that are good substrates for MeCP2.

7.4 Modulating epigenetic readout by post translational modification

Many transcription factors are post-translationally modified (Khidekel and Hsieh-Wilson, 2004). Common post-translational modifications include: glycosylation, methylation, phosphorylation, acetylation, ubiquitination, and sumoylation. For many nuclear factors, like P53, the nature and extent of these modifications can have profound affects on their function (Bode and Dong, 2004). Preliminary evidence

suggests that MeCP2 is phosphorylated and this affects its DNA binding capacity. Currently, efforts are being invested in identifying kinases that modify MeCP2 and determining the affect that phosphorylation has on MeCP2 function. These data will begin to elucidate the molecular switch that controls MeCP2/DNA association. The observation that MeCP2 is phosphorylated is interesting but also raises questions regarding other modifications which may exist on MeCP2, and what functional affects might these modifications have? The streamlining of mass spectrometry technology over the last decade has made identifying posttranslational modification more routine, and determining the complement of modifications on endogenous MeCP2 will be of great interest.

7.5 A search for the cause of Rett syndrome

The majority of the work in this thesis dealt with very basic aspects of MeCP2 function, and did not directly address any of the functions of MeCP2 as they relate to the Rett syndrome phenotype. By studying the basic functions of MeCP2 the goal is to understand how the molecule functions normally and use this information to determine how its malfunction leads to disease. Using the newly identified MeCP2 DNA binding requirements it will be exciting to try and identify potential MeCP2 target genes with the aid of bioinformatics and epigenetic genome information. It will be particularly interesting to examine whether Rett syndrome mutations outside of known functional domains affect the biophysical properties of MeCP2, and what the relationship is between specific MeCP2 co-factors and the Rett phenotype. By extending our basic knowledge about MeCP2 through general scientific investigation, leads will inevitably arise to expedite the possibility of developing therapies to treat this debilitating disease.

Appendix 1

UNSELECTED METHYL-SELEX FRAGMENTS

23B CTCACCGATGCCTGCGTC**CCGG**GGTTGTATCCCTATTGCG
79B GAACCGTCAAGTGCGCGA**CCGG**GGAAGCCACGGCTATAAC
25A GGCCTGGCTCTCCTCCCC**CCGG**GCACTCCTACGATTTGTC
24A CGCGGGTCCAGGTCCCCA**CCGG**GCCAACGCAAAACTCGGG
22A GGGCTAGGATCTACGCGA**CCCG**GGTCTCAAAAACGCCGTG
1A ACCCTNGATTTGCGCACT**CCGG**GTGNCCCATTTTTCCGCG
34A CGGCAGATCAAGTGAACG**CCGG**GATGGGGATATATATCGC
61B CGGCAACGCCGTCCTTCCT**CCGG**GCATCATATCCCACGAGG
10A CTCCTCTGACCCTTGCGC**CCGG**GGAGGACCTTTACCCCGA
20A GCGGGGACGTGTTGGGCG**CCGG**CAAATGCTAGCCTTAAG
47B CAGGGGGAAGGTGTGGGG**CCGG**GAGGCGCAAATCAATCGC
71B GCAGGGTGCAAGTTGCTG**CCGG**CCCGATAACGACTCGACC
27A CGCTAGGCTGGGCCGACG**CCGG**TTTCAAAACGTCAAAGTG
5A CGCTTTGCTTCCCATCGA**CCGG**CTGCGCTAAACTTGACC
37B CCAGCGGAACGCGTGGCG**CCGG**AAAAAGAGTGACATTACG
39A CGCTAAGA**CCGG**GGGGTC**CCGG**GAGCATAACTTCCCGC
14B GTCCAGGTCGGGTGGGGCA**CCGG**AAATAGAATTCCATGGCC
11A CCGATAGGGAGCTCGGGG**CCGG**GCCTTTTCATACCCTCCC
17A CGATGGGGTGGTCGAGC**CCGG**ACGATAACGCGCAGTGGG
8B CGTGCTGATAAGGTCACGG**CCCG**GGATATTGAGGAGAACCATCGA
6B G**CCGG**GCCGTTTGGTCCC**CCGG**GGATATGAAACCTCGGAC
70B GGGCGGGAGGGCGGGCAGG**CCGG**TGAATAGCCACCTGTAGC
15B GGGAAGGGCGGGGCATAG**CCGG**GCAATACAGAGCGTTGCA
44B AATCGAGCCAGCGGGCTC**CCGG**AAATAACGCGTCGGACGA
26A GGTCGGGTCTGGGCGGTA**CCGG**TATTGCCAGGTACGGTGG
8A GGGCGGATAGGGGAGCAC**CCGG**TATAGATGGCGCCGACGG
54B GTCACGTCGGCCCATTGC**CCGG**ATATCTAGGGCCCCAGGG
63B GTGTCCGCGGCCTTACGC**CCGG**GTTATCGTTCCCC
78B GTCACGTCGGCCCATTGC**CCGG**ATATCTAGGGCCCCAGGG
36A CCTATCTGAA**CCGG**CCCC**CCGG**GTGTAACCCCCTCGCT
3A GGCACATGGGTGTATCAT**CCGG**G**CCGG**GACCTACCTCCCC
6A TGCGCGGGATGGATGCAC**CCGG**CTGGAGGGGATACGTAAT
16A CCCAAAGTCGCTAACCAG**CCGG**GGTCCCGCCTCCTCCCAG
32A CCGAGAGTTGCAACCCCT**CCGG**CGGTCAACCCGCCGCCCA
28A GGCAGGGGGCG**CCGG**GA**CCGG**GAGCGCAACCAGGGCCAG
31A GCATCCTGAAGAGTTCCC**CCCG**GGACCCCCCCTGCCTGGG
29A GTCCAAGACCATTGCCGC**CCGG**GGCCCATGCACCAGCGAC
4A TTTCGTGAGACCCAACCC**CCGG**TAGGATGGCCCCATCTCC
12A CCCCATGTTACTGTCGAC**CCGG**GGGCGCCGAGCACGTGCC
35A GACAGAGCTTCAGCTCTC**CCGG**TGGCCCAACTCCGCGCC
3A GGGACATGGGTGTATCAT**CCGG**G**CCGG**GACCTACCTCCCC
36A CCTATCTGAA**CCGG**CCCC**CCGG**GTGTAACCCCCTCGCT

40A TCAACGTTGCTCGACCCA**CCGG**GTCAACTTTGGTACTCCC
37A GGCTGTGAACCCACCGAG**CCGG**AGGCTATCGCCCCGCGGC
38A CCCATGGCTATCCACCGC**CCGG**CAAGCAGGAGACGTGGCC
23A TCCTGCGGGAGGGTTCAT**CCGG**CGCACACGACTAAGGCCG
13A GCCCTCTCTACCGATTGT**CCGG**GCCGAGCCTGCACCAC
9B GGGGTGGTACAGGGAGGC**CCGG**GAGAGGCGCGGCTCTGCG
5B GCAGAGGCGGTGGGGGC**CCGG**GACCAGTAACACTCTCGA
16B GGTGTGCGGACTAACCT**CCGG**GACGTTGTAGTCCGAGGC
32B ACGGGCGCTCCGTGGCT**CCGG**GTGCTTGGGATCTCTTGC
53B GGAACGTGTGTGCATCC**CCGG**TGGTTCCGCGTCCCAGGG
48B GAGGGGATCATCGGGTAG**CCGG**CCGTGCAATGTCTGG
17B GCGGATAGGAGTTGGAGG**CCGG**GTGAGCTCTAGCATAGCC
51B CAGAGGGGCGTGACAAAG**CCGG**ATGAGAGCCACGGTGCC
13B TGCGCGTCGGACTTCAAC**CCGG**TCAACCGTCCGCTGGGTG
52B CGCGGGGAGGGACGCCCG**CCGG**GAGGTGCCTAACAAAGTT
45B TAGGGGGAGCGGGGCTCT**CCGG**GAGGGGGCTGGCGCTACG
62B GACGGGTGCGACAAGGCGTAG**CCGG**CCCTCCCCCCCCCTCCCC
46B GGGGGGTGCTAGGGTGGC**CCGG**AGTGCTCCAGGCTGCACT
29B CACAGCGTCGGTCGTCGT**CCGG**ACCCCACAGGACCATAGA
43B AGACGAGTGGGACGGGGC**CCGG**GAGGGCGCCG**CCGG**CGGC
20B GGAGGGGATAGGGGACAC**CCGG**TGTGGACATGACGGAAGG
39B GAGCGGGTTGGAGTCA**CCGG**GGATAGCCGCGCTTGACT
4B GCGCGAGA**CCGG**TATGTA**CCGG**GGACTTCCCGCCCAGCCT
80B TCGGGTGGCTGCGTACA**CCGG**CCGTGAGGAAGCTCTGCC
27B TGCCCGTTGGGTAGCTAG**CCGG**GGCCACCAGGCCTTCGGG
31B AGGGGGATGAAGGGTCTT**CCGG**CAGGCGGGAATGATTCGC
22B GCAGGGGCGGGGGTGCCC**CCGG**GCCAGAGGGCTGAATGAG
64B ACGCGGTGTTAGGGTCCC**CCGG**TGCGCCAATGGCTGCTGA
60B TACGGGGTGTAGAGTACG**CCGG**GAGCCGCCAGTCGCTCTG
35B AGGCCGTGGATGTAAGTC**CCGG**CAAACCAGACGCTGTCTG
72B GG**CCGG**CCCGTA**CCGG**AGTGTTAGTGTCTAGCGA
58B GGCGCTCGGGGGGTTGGC**CCGG**ATCCCTGGCACCTCTAGT
18B TGCGGTTCATAGGCCCG**CCGG**GTCGGAAGGGTCATTGCC
21B CGAAGGTAGGGCGTTCGC**CCGG**GTCCAGAGGGGCATGCGG
34B GGAAGGGAGTTTC**CCGG**GAAGAAACTTTCGGGGCT
33B TGAGGGGTTGGGGGGAGA**CCGG**G**CCGG**AGAAGCCCTTGTG
24B GAGGGGGGCAGTATGGTC**CCGG**ATGGGTACGTCAAAGAAG
1B ATGGAAGGGACGGGGCCG**CCGG**GAGGGCATGGCCTGACGC
75B CAAGCGGCGGCGTGGCAT**CCGG**TCGGGATGTATCGGTAAG
57B GAGAGGGCGGTGTGAGAT**CCGG**GTCCCCGACCTCCTTGCG
19B GAAGAGCGGAGGGGTTAG**CCGG**GAT**CCGG**AGCTCTAGCAA
2B GGCGCTCGGGGGGTTGGC**CCGG**ATCCCTGGCACCTCTAGT
68B CATCGTGGGGCGGAGGCA**CCGG**CAGCAACGTCTCGGTAAC
10B TGAGTTACGGGCCGACCC**CCGG**GGGGCGCAGTTCTGGCGT

SELECTED METHYL-SELEX FRAGMENTS

67 GGCCAAGCTCTGATCAGC**CCGG**GCTCATTCCCGCATTTCG
65 CCGAAGCCCCGCTGACCC**CCGG**ATTCCACGCCCGATTAT
93 CACATAGACGCCTCCTCT**CCGG**GATCCCAAATTCTGTGG
110 CAGTTACAAAGGTTCCG**CCCG**GGGACATTATACCGTGGG
84 GGCACCCAAGTCTTCCTT**CCGG**GGGCCAaCCATTTTGCAG
73 CAGATAGAGGGACAACCTC**CCGG**CGGATACAAATCTTCTCC
109 CACAGAGCCACGTCGCTC**CCGG**AGCTTCCTATTTCAGTCG
75 GCGAGGGGGGTACTCTC**CCGG**GAGCCATGTATAAGGTGG
88 CCACGTGATCTTGAGGAC**CCGG**CTACAGCTAATTTCTGG
92 CGGTGGAGAGGGGGTAAT**CCGG**GATCAGTATTTATTTGGG
85 CGCCCGCAGTATGCAACC**CCGG**AATGTGAATAACTGTGTG
71 GGCAAGGAGGCTGGTGAC**CCGG**CCTTACTCATTTCTGCG
11 CACAAGGAGGGATAGTAG**CCGG**ACAGAAGAAATCTCTGGG
52 CNCAGGCTTAGCGACAAC**CCGG**CCATAGGATTAGCTTTGG
29 CCCAAGAGGTGCGC**CCGG**CAAGTTCATTTCGGTTTG
37 CGGAAGCGGGGGGTTCCC**CCGG**ACTATCAATAATCGTGGC

36 CACAACCAATCTCCACCC**CCGG**CCTCCGCTTTAAACTCGG
54 CCCAACGGACAGCC**CCGG**C**CCGG**AGGACCCTATTTCTCGCC
44 CCAAGGGATGGCGCTCCC**CCGG**ACCGAGCATATAGTG
60 GCGCTTGACTACCCAGAC**CCGG**GGCGATCAATTACCGTTG
14 GGGGCAGTTGCGGTCCCC**CCGG**CGGAAGCATATTTGGGCC
31 GGGACTGGCGACTACCTG**CCGG**CTTCAACATTTTCGCGTG
47 CAACGGCGGGGGAGTT**CCGG**GCGTAAGATTTCTGGTGCG
106 GGCGGGGGCGGGTGTGTC**CCGG**GTGCGGAAATTCAATGGG
33 CGGCGGGGAGGATGACCT**CCGG**GTGAAATAAAAGTGTGCG
118 CGGGTGAGAATTGCAAGG**CCGG**AATTAGCGGTTTATG *
77 CCGATGGAATAGTCGATT**CCGG**AAATTTCCCCTGTCCCGC
98 CGAGAGCTGTCCCCTAG**CCGG**AATAAAGAATGCTGTGTG
80 CAACCGCGGAACTAGCAC**CCGG**AGATTAGCGAGACGTAGG
61 GGTGGGGAATGGGGCTCC**CCGG**CTTTAAGGTATCTGTGGG
105 CGCACCGTGAGCCGACAC**CCGG**GTTTTACACCAACTGCGG
101 GACACGGAAATCCCTCC**CCGG**GGGTAATCCCTCTGTACG
120 CCATAGGGAACTAGACTA**CCGG**CTTTTAACATTTATAG
99 CAACAGCTCGTTCGGACC**CCGG**GTCTAATAATTCTTTGGC
108 CCGAAGGTGGTTC**CCGGCCGG**ACTTAACGGACCTCTGTC
115 CGCATTAACTACCACCC**CCGG**CCTTTACTCACTCGGCC
62 AAATGCTGGCTTACTT**CCGG**GCAATTCCAAGCACAGTG
63 CGCAGCACGGTGGGGGG**CCGG**AGTTAAGGACTCGTTGTC
103 GGCAGGCGGCCAAAGCTC**CCGG**AGTTAATTCTCCCGCCCG
100 GGGGTGTGGGGCGC**CCGGCCGG**CTTTAAGCAAAGTGTGTC
86 GGCCGCTGATGTGGATC**CCGG**AAGAAATAAACTCCTATGC
90 GGGAAGGTGTGGGTTCCC**CCGG**GTTTAGCATATCGTTTTG
87 C**CCGG**GGCGTAAGGGGAG**CCGG**ATTATGCGCAACCGTGTC
91 CACAGGGACGTCAGTCAG**CCGG**CCTATTTCTTCTTT**CCGG**
81 GAATCGCCGACTGATGGC**CCGG**ACGTAATAAATCGGTGGG
95 GGCCAGATGGTGATGCCC**CCGG**ACTATATGACCCTCGTGC
89 CACGGGGGAGATAGAGTC**CCGG**ACATAAGGAAGCGGTTTG

```
66          CAAGCATACATGCCCGGATTATCGCGGTcTTTACG
72      CATACAGATCTCGGCCACCCGGGTTTAAGTATCACCTGTG
82      GGGGCTGTCCAGCTGGGGCCGGCTTTAATCACTCCAGGTG
70      CTAAATGTCTGACAATGGCCGGGGAATTGTGCTATTCGG
69      CCAACAGCGACCAGCGCGCCGGGATTAGCCCCTCTCTGCC
68      CAACAGATGAGGCTAACTCCGGGAATTATCGATCTGCGG
116       CCGAAGTTTGTACTCCCCGGCTATTCGAATACTTTGGC
74      CGCGCACGCATGTCCCCCCCGGCCTTTATGGAGGTTTAGG
107      GGGAATGGGAGGCCTCCCCGGAGTTTATAACGCCGCGTG
104     GCTACCAAGCGGCCACCTCCGGTAATTACAAAACGGCCGG
3       GGGGGGTGAGGCGGTTGTCCGGAATATCAAAAGCCGGGGG
18      CCCATGGCTCCGCTTTCTCCGGTGATTATGCTCNCTCCGG
21      CCAACCGTCGCACCCCTCCCGGCTTTTACAATCCCTTGGC
34        CCCAGGGCTGTGATGCCCGGAATTAAGCCGACGTTCTG    *
49       CCAACGCACGGGGGGTCCCGGAAATAAGGGGGCTGTCCG
6       CACAGGGACGTCAGTCAGCCGGCCTATTTCTTCTTTCCGG
58      GGCTAGGGGCAGAGAGGGCCGGAAATAGTCAAGCAGTTGG
56      GGAGAGGGGGAAGTGTAGCCGGACATAACGAGGCAGTTGC
4       CCAACGGAGTGCTTACGCCCGGAAAATGTAACTCGATTCG
19         CGAATGTCTTAAAGCCGGGTTATTTGGCGGCCTGGG
41      CCCAAAAAAGGTTAACTCCCGGGGGATTAGGGTCTGTCCG
39      CACAGGGACGTCAGTCAGCCGGCTTATTTCTTCTTTCCGG
1       CTGAAGAGTCTGGAGACCCCGGGATTAAACTCGCGCCCTG
15      CCAGGCGTTCGACCTGCACCGGGATAAGGCTAATTCTCCGG
7       CAACCGCGGAACTAGCACCCGGAGATTAGCGAGACGTAGG

16      CCCCCCCACTTAGACGCCCCGGCTAAATCATCTTGGT
42       ACAGGGACGTCAGTCAGCCGGCTTATTTCTTCTTTCCGG
48       CCCAAGGTTGGTCGACCCCGGAGATAACGAAGCTCGGTG
55       CACACTAAAATTCCGGCCCGGAATTAAAGAGTCCGTGTG    *
51      CGCANGGTGTTATCGAGGCCGGAGATAACGAGTCTGCCGG
26       GGCCCGACGAGGTAGCCCCGGAATTAGCGACTCTCTATG    *
8       CGAAGGTTGAGGCGCCACCCGGAATTAAGGCTGAGTG      *
9         CCAGCGCATAACGCCCCCGGAATTTAGCAAACATCCGG    *
23       CACCCGGACAAGGAGCCCCGGCCTTTAGGACACCATTTG
13      GGGGGGGTGTGTTGAGCCCCGGAATATCTCTCTCCGTTGC
53      CTAGAGTCGAAGATACGCCCGGAGCTAAGCGTTCCATCTG
20      GCCAGTGAGATGGAACGGCCGGGGTTAGTGGATCCGAGG
28       CGCTAGAACAGTTACGCCCGGGGGGTCTCAATCGGGAGG
46      CCACGAGGCCTGTGGACACCGGAACTAAGATAGCGGGTTG
83      CGCACTCTGGAAGATACGCCGGGAGTAAGCGAGCTGTAGG
111     CCAAAGAGGTGCTTGCGTCCGGCAACTACAAAGCCGTGTA
```

* BDNF -148 site CGGAATT

References

Aapola, U., Liiv, I., and Peterson, P. (2002). Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. Nucleic Acids Res *30*, 3602-3608.

Ahringer, J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends Genet *16*, 351-356.

Alland, L., Muhle, R., Hou, H., Jr., Potes, J., Chin, L., Schreiber-Agus, N., and DePinho, R. A. (1997). Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. Nature *387*, 49-55.

Amir, R., Dahle, E. J., Toriolo, D., and Zoghbi, H. Y. (2000). Candidate gene analysis in Rett syndrome and the identification of 21 SNPs in Xq. Am J Med Genet *90*, 69-71.

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet *23*, 185-188.

Antequera, F., Boyes, J., and Bird, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell *62*, 503-514.

Antequera, F., Macleod, D., and Bird, A. P. (1989). Specific protection of methylated CpGs in mammalian nuclei. Cell *58*, 509-517.

Aravind, L., and Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res *26*, 4413-4421.

Arney, K. L. (2003). H19 and Igf2--enhancing the confusion? Trends Genet *19*, 17-23.

Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J., and Matzke, M. (2002). RNA-directed DNA methylation in Arabidopsis. Proc Natl Acad Sci U S A *99 Suppl 4*, 16499-16506.

Ayer, D. E., Lawrence, Q. A., and Eisenman, R. N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. Cell *80*, 767-776.

Ayton, P. M., Chen, E. H., and Cleary, M. L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol Cell Biol *24*, 10470-10478.

Bachman, K. E., Park, B. H., Rhee, I., Rajagopalan, H., Herman, J. G., Baylin, S. B., Kinzler, K. W., and Vogelstein, B. (2003). Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell *3*, 89-95.

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., and Wyllie, A. (1999). Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. Oncogene *18*, 8044-8047.

Ballestar, E., Paz, M. F., Valle, L., Wei, S., Fraga, M. F., Espada, J., Cigudosa, J. C., Huang, T. H., and Esteller, M. (2003). Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. Embo J *22*, 6335-6345.

Ballestar, E., Pile, L. A., Wassarman, D. A., Wolffe, A. P., and Wade, P. A. (2001). A Drosophila MBD family member is a transcriptional corepressor associated with specific genes. Eur J Biochem *268*, 5397-5406.

Ballestar, E., Ropero, S., Alaminos, M., Armstrong, J., Setien, F., Agrelo, R., Fraga, M. F., Herranz, M., Avila, S., Pineda, M., *et al.* (2005). The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. Hum Genet *116*, 91-104.

Balmer, D., Arredondo, J., Samaco, R. C., and LaSalle, J. M. (2002). MECP2 mutations in Rett syndrome adversely affect lymphocyte growth, but do not affect imprinted gene expression in blood or brain. Hum Genet *110*, 545-552.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di-and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barbot, W., Dupressoir, A., Lazar, V., and Heidmann, T. (2002). Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. Nucleic Acids Res *30*, 2365-2373.

Barlow, A. L., van Drunen, C. M., Johnson, C. A., Tweedie, S., Bird, A., and Turner, B. M. (2001). dSIR2 and dHDAC6: two novel, inhibitor-resistant deacetylases in Drosophila melanogaster. Exp Cell Res *265*, 90-103.

Barlow, D. J., and Thornton, J. M. (1988). Helix geometry in proteins. J Mol Biol *201*, 601-619.

Bartolomei, M. S., Webber, A. L., Brunkow, M. E., and Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev *7*, 1663-1673.

Baylin, S. B., Hoppener, J. W., de Bustros, A., Steenbergh, P. H., Lips, C. J., and Nelkin, B. D. (1986). DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res *46*, 2917-2922.

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A *101*, 12130-12135.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc Natl Acad Sci U S A *96*, 3969-3974.

Berg, A., Meza, T. J., Mahic, M., Thorstensen, T., Kristiansen, K., and Aalen, R. B. (2003). Ten members of the Arabidopsis gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. Nucleic Acids Res *31*, 5291-5304.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., *et al.*

(2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol *203*, 971-983.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet *9*, 2395-2402.

Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M. A., Jun, Z., Walter, J., and Jaenisch, R. (2002). Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Mol Cell Biol *22*, 2124-2135.

Bird, A. (1999). DNA methylation de novo. Science *286*, 2287-2288.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Bird, A. P., and Wolffe, A. P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Bischoff, J. R., Friedman, P. N., Marshak, D. R., Prives, C., and Beach, D. (1990). Human p53 is phosphorylated by p60-cdc2 and cyclin B-cdc2. Proc Natl Acad Sci U S A *87*, 4766-4770.

Bode, A. M., and Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. Nat Rev Cancer *4*, 793-805.

Bourc'his, D., and Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Bowen, N. J., Fujita, N., Kajita, M., and Wade, P. A. (2004). Mi-2/NuRD: multiple complexes for many purposes. Biochim Biophys Acta *1677*, 52-57.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123-1134.

Brackertz, M., Boeke, J., Zhang, R., and Renkawitz, R. (2002). Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. J Biol Chem *277*, 40958-40966.

Brannan, C. I., and Bartolomei, M. S. (1999). Mechanisms of genomic imprinting. Curr Opin Genet Dev *9*, 164-170.

Briggs, S. D., Xiao, T., Sun, Z. W., Caldwell, J. A., Shabanowitz, J., Hunt, D. F., Allis, C. D., and Strahl, B. D. (2002). Gene silencing: trans-histone regulatory pathway in chromatin. Nature *418*, 498.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Brunner, E., Weitzel, J., Heitmann, B., Maurer, T., Stratling, W. H., and Kalbitzer, H. R. (2000). Sequence-specific 1H, 13C, and 15N assignments of the MAR-binding domain of chicken MeCP2/ARBP. J Biomol NMR *17*, 175-176.

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J Biol Chem *278*, 823-828.

Buhrmester, H., von Kries, J. P., and Stratling, W. H. (1995). Nuclear matrix protein ARBP recognizes a novel DNA sequence motif with high affinity. Biochemistry *34*, 4108-4117.

Bull, J. H., and Wootton, J. C. (1984). Heavily methylated amplified DNA in transformants of Neurospora crassa. Nature *310*, 701-704.

Buschdorf, J. P., and Stratling, W. H. (2004). A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. J Mol Med *82*, 135-143.

Buschhausen, G., Graessmann, M., and Graessmann, A. (1985). Inhibition of herpes simplex thymidine kinase gene expression by DNA methylation is an indirect effect. Nucleic Acids Res *13*, 5503-5513.

Buschhausen, G., Wittig, B., Graessmann, M., and Graessmann, A. (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. Proc Natl Acad Sci U S A *84*, 1177-1181.

Camporeale, G., Shubert, E. E., Sarath, G., Cerny, R., and Zempleni, J. (2004). K8 and K12 are biotinylated in human histone H4. Eur J Biochem *271*, 2257-2263.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell *15*, 57-67.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M. F., Huang, M. S., Matzke, M., and Jacobsen, S. E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol *13*, 2212-2217.

Cao, X., and Jacobsen, S. E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol *12*, 1138-1144.

Chambers, T. C., and Langan, T. A. (1990). Purification and characterization of growth-associated H1 histone kinase from Novikoff hepatoma cells. J Biol Chem *265*, 16940-16947.

Chandler, S. P., Guschin, D., Landsberger, N., and Wolffe, A. P. (1999). The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. Biochemistry *38*, 7008-7018.

Chedin, F., Lieber, M. R., and Hsieh, C. L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A *99*, 16916-16921.

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

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., and Li, E. (2003a). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol *23*, 5594-5605.

Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003b). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science *302*, 885-889.

Cheutin, T., McNairn, A. J., Jenuwein, T., Gilbert, D. M., Singh, P. B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science *299*, 721-725.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Colantuoni, C., Jeon, O. H., Hyder, K., Chenchik, A., Khimani, A. H., Narayanan, V., Hoffman, E. P., Kaufmann, W. E., Naidu, S., and Pevsner, J. (2001). Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. Neurobiol Dis *8*, 847-865.

Cooper, D. N., and Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. Hum Genet *78*, 151-155.

Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat Struct Mol Biol *11*, 1037-1043.

Cross, S. H., Meehan, R. R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet *16*, 256-259.

Crouch, T. H., Holroyde, M. J., Collins, J. H., Solaro, R. J., and Potter, J. D. (1981). Interaction of calmodulin with skeletal muscle myosin light chain kinase. Biochemistry *20*, 6318-6325.

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

de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M. A. (2005). Do protein motifs read the histone code? Bioessays *27*, 164-175.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., *et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell *7*, 663-676.

DeChiara, T. M., Efstratiadis, A., and Robertson, E. J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature *345*, 78-80.

DeChiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell *64*, 849-859.

Delaval, K., and Feil, R. (2004). Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev *14*, 188-195.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev *15*, 2940-2944.

Deplus, R., Brenner, C., Burgers, W. A., Putmans, P., Kouzarides, T., de Launoit, Y., and Fuks, F. (2002). Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucleic Acids Res *30*, 3831-3838.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dong, A., Yoder, J. A., Zhang, X., Zhou, L., Bestor, T. H., and Cheng, X. (2001). Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res *29*, 439-448.

Doree, M., and Hunt, T. (2002). From Cdc2 to Cdk1: when did the cell cycle kinase join its cyclin partner? J Cell Sci *115*, 2461-2464.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Duncan, B. K., and Miller, J. H. (1980). Mutagenic deamination of cytosine residues in DNA. Nature *287*, 560-561.

Eads, C. A., Nickel, A. E., and Laird, P. W. (2002). Complete genetic suppression of polyp formation and reduction of CpG-island hypermethylation in Apc(Min/+) Dnmt1-hypomorphic Mice. Cancer Res *62*, 1296-1299.

El-Osta, A., Kantharidis, P., Zalcberg, J. R., and Wolffe, A. P. (2002). Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. Mol Cell Biol *22*, 1844-1857.

Espada, J., Ballestar, E., Fraga, M. F., Villar-Garea, A., Juarranz, A., Stockert, J. C., Robertson, K. D., Fuks, F., and Esteller, M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J Biol Chem *279*, 37175-37184.

Esteller, M. (2000). Epigenetic lesions causing genetic lesions in human cancer: promoter hypermethylation of DNA repair genes. Eur J Cancer *36*, 2294-2300.

Esteller, M. (2005). DNA methylation and cancer therapy: new developments and expectations. Curr Opin Oncol *17*, 55-60.

Fang, J., Chen, T., Chadwick, B., Li, E., and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J Biol Chem *279*, 52812-52815.

Feinberg, A. P., and Tycko, B. (2004). The history of cancer epigenetics. Nat Rev Cancer *4*, 143-153.

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

Feng, Q., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2002). Identification and functional characterization of the p66/p68 components of the MeCP1 complex. Mol Cell Biol *22*, 536-546.

Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev *15*, 827-832.

Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M., and Surani, M. A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature *362*, 751-755.

Fischle, W., Wang, Y., and Allis, C. D. (2003). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475-479.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat Biotechnol *21*, 89-92.

Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A. M., Esteller, M., and Feil, R. (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. Embo J *21*, 6560-6570.

Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A., and Esteller, M. (2003). The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res *31*, 1765-1774.

Free, A., Wakefield, R. I., Smith, B. O., Dryden, D. T., Barlow, P. N., and Bird, A. P. (2001). DNA recognition by the methyl-CpG binding domain of MeCP2. J Biol Chem *276*, 3353-3360.

Friedl, E. M., Lane, W. S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2003). The C-terminal domain phosphatase and transcription elongation activities of FCP1 are regulated by phosphorylation. Proc Natl Acad Sci U S A *100*, 2328-2333.

Fujita, H., Fujii, R., Aratani, S., Amano, T., Fukamizu, A., and Nakajima, T. (2003a). Antithetic effects of MBD2a on gene regulation. Mol Cell Biol *23*, 2645-2657.

Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., and Nakao, M. (1999). Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol *19*, 6415-6426.

Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H., and Nakao, M. (2003b). MCAF mediates MBD1-dependent transcriptional repression. Mol Cell Biol *23*, 2834-2843.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003c). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J Biol Chem *278*, 24132-24138.

Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet *24*, 88-91.

Fuks, F., Hurd, P. J., Deplus, R., and Kouzarides, T. (2003a). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res *31*, 2305-2312.

Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003b). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem *278*, 4035-4040.

Gadgil, H., and Jarrett, H. W. (1999). Heparin elution of transcription factors from DNA-Sepharose columns. J Chromatogr A *848*, 131-138.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.

Gaudet, F., Rideout, W. M., 3rd, Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol Cell Biol *24*, 1640-1648.

Gebauer, F., and Hentze, M. W. (2004). Molecular mechanisms of translational control. Nat Rev Mol Cell Biol *5*, 827-835.

Geiman, T. M., Sankpal, U. T., Robertson, A. K., Zhao, Y., Zhao, Y., and Robertson, K. D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. Biochem Biophys Res Commun *318*, 544-555.

Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A., and Hansen, J. C. (2003). Chromatin compaction by human MeCP2: Assembly of novel secondary chromatin structures in the absence of DNA methylation. J Biol Chem.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Dong, X., Parthun, M., and Jacob, S. T. (2002). Inhibitors of Histone Deacetylase and DNA Methyltransferase Synergistically Activate the Methylated Metallothionein I Promoter by Activating the Transcription Factor MTF-1 and Forming an Open Chromatin Structure. Mol Cell Biol *22*, 8302-8319.

Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Garrick, D., Ayyub, H., and Higgs, D. R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet *24*, 368-371.

Gius, D., Cui, H., Bradbury, C. M., Cook, J., Smart, D. K., Zhao, S., Young, L., Brandenburg, S. A., Hu, Y., Bisht, K. S., *et al.* (2004). Distinct effects on gene expression of chemical and genetic manipulation of the cancer epigenome revealed by a multimodality approach. Cancer Cell *6*, 361-371.

Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of Drosophila melanogaster contains 5-methylcytosine. Embo J *19*, 6918-6923.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem.

Gregory, R. I., Randall, T. E., Johnson, C. A., Khosla, S., Hatada, I., O'Neill, L. P., Turner, B. M., and Feil, R. (2001). DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes Snrpn and U2af1-rs1. Mol Cell Biol *21*, 5426-5436.

Guschin, D., Wade, P. A., Kikyo, N., and Wolffe, A. P. (2000). ATP-Dependent histone octamer mobilization and histone deacetylation mediated by the Mi-2 chromatin remodeling complex. Biochemistry *39*, 5238-5245.

Guy, J., Hendrich, B., Holmes, M., Martin, J. E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet *27*, 322-326.

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., and Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A *96*, 14412-14417.

Harikrishnan, K. N., Chow, M. Z., Baker, E. K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J. M., Jones, P. L., Sif, S., and El-Osta, A. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet *37*, 254-264.

Harrer, M., Luhrs, H., Bustin, M., Scheer, U., and Hock, R. (2004). Dynamic interaction of HMGA1a proteins with chromatin. J Cell Sci *117*, 3459-3471.

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., and Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983-1993.

Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. Gene *189*, 227-234.

He, Y. Y., Stockley, P. G., and Gold, L. (1996). In vitro evolution of the DNA binding sites of Escherichia coli methionine repressor, MetJ. J Mol Biol *255*, 55-66.

Heitmann, B., Maurer, T., Weitzel, J. M., Stratling, W. H., Kalbitzer, H. R., and Brunner, E. (2003). Solution structure of the matrix attachment region-binding domain of chicken MeCP2. Eur J Biochem *270*, 3263-3270.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev *15*, 710-723.

Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

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

Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., Markowitz, S., Willson, J. K., Hamilton, S. R., Kinzler, K. W., *et al.* (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A *95*, 6870-6875.

Hermann, A., Schmitt, S., and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. J Biol Chem *278*, 31717-31721.

Herr, A. J., Jensen, M. B., Dalmay, T., and Baulcombe, D. C. (2005). RNA Polymerase IV Directs Silencing of Endogenous DNA. Science.

Holmes, K. C. (1997). The swinging lever-arm hypothesis of muscle contraction. Curr Biol *7*, R112-118.

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P., and Bradbury, E. M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. J Biol Chem *268*, 305-314.

Hong, R., Macfarlan, T., Kutney, S. N., Seo, S. B., Mukai, Y., Yelin, F., Pasternack, G. R., and Chakravarti, D. (2004). The identification of phosphorylation sites of pp32 and biochemical purification of a cellular pp32-kinase. Biochemistry *43*, 10157-10165.

Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet *37*, 31-40.

Hörz, W., and Altenburger, W. (1981). Nucleotide sequence of mouse satellite DNA. Nucleic Acids Research *9*, 683-697.

Hotchkiss, R. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem *175*, 315-332.

Howell, C. Y., Bestor, T. H., Ding, F., Latham, K. E., Mertineit, C., Trasler, J. M., and Chaillet, J. R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829-838.

Hung, M. S., Karthikeyan, N., Huang, B., Koo, H. C., Kiger, J., and Shen, C. J. (1999). Drosophila proteins related to vertebrate DNA (5-cytosine) methyltransferases. Proc Natl Acad Sci U S A *96*, 11940-11945.

Hutchins, A. S., Mullen, A. C., Lee, H. W., Sykes, K. J., High, F. A., Hendrich, B. D., Bird, A. P., and Reiner, S. L. (2002). Gene silencing quantitatively controls the function of a developmental trans-activator. Mol Cell *10*, 81-91.

Huynh, K. D., and Lee, J. T. (2001). Imprinted X inactivation in eutherians: a model of gametic execution and zygotic relaxation. Curr Opin Cell Biol *13*, 690-697.

Imai, S., Armstrong, C. M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature *403*, 795-800.

Jackson, J. P., Lindroth, A. M., Cao, X., and Jacobsen, S. E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jacobs, S. A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. Embo J *20*, 5232-5241.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jeddeloh, J. A., Stokes, T. L., and Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet *22*, 94-97.

Jeffery, L., and Nakielny, S. (2004). Components of the DNA methylation system of chromatin control are RNA-binding proteins. J Biol Chem *279*, 49479-49487.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jepsen, K., and Rosenfeld, M. G. (2002). Biological roles and mechanistic actions of co-repressor complexes. J Cell Sci *115*, 689-698.

Johnson, T. B., and Coghill, R. D. (1925). Research on pyrimidines. C111. The discovery of 5-methylcytosine in tuberculinic acid, the nucleic acid of tubercule bacillus. J Amer Chem Soc

47, 2838-2845.

Jones, L., Ratcliff, F., and Baulcombe, D. C. (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Curr Biol *11*, 747-757.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-191.

Jorgensen, H. F., Ben-Porath, I., and Bird, A. P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol *24*, 3387-3395.

Kaludov, N. K., and Wolffe, A. P. (2000). MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. Nucleic Acids Res *28*, 1921-1928.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489-501.

Kantor, B., Kaufman, Y., Makedonski, K., Razin, A., and Shemer, R. (2004). Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. Hum Mol Genet *13*, 2767-2779.

Kao, G. D., McKenna, W. G., and Muschel, R. J. (1999). p34(Cdc2) kinase activity is excluded from the nucleus during the radiation-induced G(2) arrest in HeLa cells. J Biol Chem *274*, 34779-34784.

Kass, S. U., Landsberger, N., and Wolffe, A. P. (1997). DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol *7*, 157-165.

Katan-Khaykovich, Y., and Struhl, K. (2002). Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. Genes Dev *16*, 743-752.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211-217.

Kelly, K. F., Spring, C. M., Otchere, A. A., and Daniel, J. M. (2004). NLS-dependent nuclear localization of p120ctn is necessary to relieve Kaiso-mediated transcriptional repression. J Cell Sci *117*, 2675-2686.

Khidekel, N., and Hsieh-Wilson, L. C. (2004). A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. Org Biomol Chem *2*, 1-7.

Kim, S. W., Park, J. I., Spring, C. M., Sater, A. K., Ji, H., Otchere, A. A., Daniel, J. M., and McCrea, P. D. (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. Nat Cell Biol *6*, 1212-1220.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem *278*, 4806-4812.

Klekamp, M. S., and Weil, P. A. (1987). Properties of yeast class III gene transcription factor TFIIIB. Implications regarding mechanism of action. J Biol Chem *262*, 7878-7883.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357-369.

Klose, R., and Bird, A. (2003). Molecular biology. MeCP2 repression goes nonglobal. Science *302*, 793-795.

Klug, S. J., and Famulok, M. (1994). All you wanted to know about SELEX. Mol Biol Rep *20*, 97-107.

Koch, C., and Stratling, W. H. (2004). DNA binding of methyl-CpG-binding protein MeCP2 in human MCF7 cells. Biochemistry *43*, 5011-5021.

Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. J Biol Chem *276*, 34115-34121.

Kondo, Y., Shen, L., Yan, P. S., Huang, T. H., and Issa, J. P. (2004). Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. Proc Natl Acad Sci U S A *101*, 7398-7403.

Kophengnavong, T., Carroll, A. S., and Blackwell, T. K. (1999). The SKN-1 amino-terminal arm is a DNA specificity segment. Mol Cell Biol *19*, 3039-3050.

Kouzminova, E., and Selker, E. U. (2001). dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. Embo J *20*, 4309-4323.

Kriaucionis, S., and Bird, A. (2003). DNA methylation and Rett syndrome. Hum Mol Genet *12 Spec No 2*, R221-227.

Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res *32*, 1818-1823.

Krithivas, A., Fujimuro, M., Weidner, M., Young, D. B., and Hayward, S. D. (2002). Protein interactions targeting the latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus to cell chromosomes. J Virol *76*, 11596-11604.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Dragich, J., Schanen, C., and Tamura, M. (2001). Functional analyses of MeCP2 mutations associated with Rett syndrome using transient expression systems. Brain Dev *23 Suppl 1*, S165-173.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C., and Tamura, M. (2003). Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. J Med Genet *40*, 487-493.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083-5090.

Kuo, M. H., and Allis, C. D. (1998). Roles of histone acetyltransferases and deacetylases in gene regulation. Bioessays *20*, 615-626.

Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). Mol Cell Biol *22*, 835-848.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E., and Eisenman, R. N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349-356.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis *35*, 88-93.

Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, T. I., Causton, H. C., Holstege, F. C., Shen, W. C., Hannett, N., Jennings, E. G., Winston, F., Green, M. R., and Young, R. A. (2000). Redundant roles for the TFIID and SAGA complexes in global transcription. Nature *405*, 701-704.

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol *13*, 1192-1200.

Leone, G., Teofili, L., Voso, M. T., and Lubbert, M. (2002). DNA methylation and demethylating drugs in myelodysplastic syndromes and secondary leukemias. Haematologica *87*, 1324-1341.

Lever, M. A., Th'ng, J. P., Sun, X., and Hendzel, M. J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. Nature *408*, 873-876.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905-914.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., *et al.* (2004). Dual histone H3

methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J *23*, 4286-4296.

Lorincz, M. C., Schubeler, D., and Groudine, M. (2001). Methylation-mediated proviral silencing is associated with MeCP2 recruitment and localized histone H3 deacetylation. Mol Cell Biol *21*, 7913-7922.

Luckow, V. A., Lee, S. C., Barry, G. F., and Olins, P. O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. J Virol *67*, 4566-4579.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. Science *298*, 1747-1752.

Lusser, A., and Kadonaga, J. T. (2003). Chromatin remodeling by ATP-dependent molecular machines. Bioessays *25*, 1192-1200.

Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Ma, H., Baumann, C. T., Li, H., Strahl, B. D., Rice, R., Jelinek, M. A., Aswad, D. W., Allis, C. D., Hager, G. L., and Stallcup, M. R. (2001). Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. Curr Biol *11*, 1981-1985.

Macleod, D., Charlton, J., Mullins, J., and Bird, A. P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev *8*, 2282-2292.

Magdinier, F., and Wolffe, A. P. (2001). Selective association of the methyl-CpG binding protein MBD2 with the silent p14/p16 locus in human neoplasia. Proc Natl Acad Sci U S A *98*, 4990-4995.

Makedonski, K., Abuhatzira, L., Kaufman, Y., Razin, A., and Shemer, R. (2005). MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affect UBE3A expression. Hum Mol Genet.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. Embo J *21*, 6842-6852.

Manuelidis, L. (1981). Consensus sequence of mouse satellite DNA indicates it is derived from tandem 116 basepair repeats. FEBS Letters *129*, 25-28.

Marhold, J., Brehm, A., and Kramer, K. (2004a). The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2. BMC Mol Biol *5*, 20.

Marhold, J., Kramer, K., Kremmer, E., and Lyko, F. (2004b). The Drosophila MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA. Development *131*, 6033-6039.

Marmorstein, R. (2001). Structure and function of histone acetyltransferases. Cell Mol Life Sci *58*, 693-703.

Martens, J. H., O'Sullivan R, J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J *24*, 800-812.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y. E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890-893.

Masure, H. R., Alexander, K. A., Wakim, B. T., and Storm, D. R. (1986). Physicochemical and hydrodynamic characterization of P-57, a neurospecific calmodulin binding protein. Biochemistry *25*, 7553-7560.

Matzke, M. A., and Birchler, J. A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet *6*, 24-35.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res *20*, 5085-5092.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499-507.

Meggio, F., and Pinna, L. A. (2003). One-thousand-and-one substrates of protein kinase CK2? Faseb J *17*, 349-368.

Migeon, B. R. (2003). Is Tsix repression of Xist specific to mouse? Nat Genet *33*, 337; author reply 337-338.

Migeon, B. R., Lee, C. H., Chowdhury, A. K., and Carpenter, H. (2002). Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. Am J Hum Genet *71*, 286-293.

Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science *297*, 403-405.

Min, J., Zhang, Y., and Xu, R. M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev *17*, 1823-1828.

Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J., Jones, J. R., Scherer, S. W., Schanen, N. C., Friez, M. J., *et al.* (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat Genet *36*, 339-341.

Montini, E., Andolfi, G., Caruso, A., Buchner, G., Walpole, S. M., Mariani, M., Consalez, G., Trump, D., Ballabio, A., and Franco, B. (1998). Identification and characterization of a novel serine-threonine kinase gene from the Xp22 region. Genomics *51*, 427-433.

Moreno, J. M., Sanchez-Montero, J. M., Ballesteros, A., and Sinisterra, J. V. (1991). Hydrolysis of nucleic acids in single-cell protein concentrates using immobilized benzonase. Appl Biochem Biotechnol *31*, 43-51.

Morris, K. V., Chan, S. W., Jacobsen, S. E., and Looney, D. J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289-1292.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J. E., Ott, M., Verdin, E., and Zhou, M. M. (2002). Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. Mol Cell *9*, 575-586.

Murray, A. W. (2004). Recycling the cell cycle: cyclins revisited. Cell *116*, 221-234.

Nan, X., Campoy, F. J., and Bird, A. (1997a). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nan, X., Meehan, R. R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res *21*, 4886-4892.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nan, X., Tate, P., Li, E., and Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol *16*, 414-421.

Ng, H. H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol *20*, 1394-1406.

Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet *23*, 58-61.

Nguyen, C. T., Gonzales, F. A., and Jones, P. A. (2001). Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. Nucleic Acids Res *29*, 4598-4606.

Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103-107.

Nishioka, K., Rice, J. C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., *et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Mol Cell *9*, 1201-1213.

O'Farrell, F., Loog, M., Janson, I. M., and Ek, P. (1999). Kinetic study of the inhibition of CK2 by heparin fragments of different length. Biochim Biophys Acta *1433*, 68-75.

O'Neill, L. P., Randall, T. E., Lavender, J., Spotswood, H. T., Lee, J. T., and Turner, B. M. (2003). X-linked genes in female embryonic stem cells carry an epigenetic mark prior to the onset of X inactivation. Hum Mol Genet *12*, 1783-1790.

Ogbourne, S., and Antalis, T. M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J *331 (Pt 1)*, 1-14.

Ogryzko, V. V. (2001). Mammalian histone acetyltransferases and their complexes. Cell Mol Life Sci *58*, 683-692.

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell *105*, 487-497.

Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. Embo J *18*, 6653-6661.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okano, M., Xie, S., and Li, E. (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.

Okano, M., Xie, S., and Li, E. (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

Paik, W. K., and Kim, S. (1973). Enzymatic demethylation of calf thymus histones. Biochem Biophys Res Commun *51*, 781-788.

Palmer, B. R., and Marinus, M. G. (1994). The dam and dcm strains of Escherichia coli--a review. Gene *143*, 1-12.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Park, C. W., Chen, Z., Kren, B. T., and Steer, C. J. (2004). Double-stranded siRNA targeted to the huntingtin gene does not induce DNA methylation. Biochem Biophys Res Commun *323*, 275-280.

Paz, M. F., Fraga, M. F., Avila, S., Guo, M., Pollan, M., Herman, J. G., and Esteller, M. (2003). A systematic profile of DNA methylation in human cancer cell lines. Cancer Res *63*, 1114-1121.

Peters, A. H., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell *12*, 1577-1589.

Peterson, C. L., and Laniel, M. A. (2004). Histones and histone modifications. Curr Biol *14*, R546-551.

Pinarbasi, E., Elliott, J., and Hornby, D. P. (1996). Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. J Mol Biol *257*, 804-813.

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd, and Grant, P. A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434-438.

Proffitt, J. H., Davie, J. R., Swinton, D., and Hattman, S. (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol *4*, 985-988.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev *15*, 1613-1618.

Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., *et al.* (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol *2*, e405.

Rangwala, S. H., and Richards, E. J. (2004). The value-added genome: building and maintaining genomic cytosine methylation landscapes. Curr Opin Genet Dev *14*, 686-691.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet *2*, 21-32.

Rhee, I., Bachman, K. E., Park, B. H., Jair, K. W., Yen, R. W., Schuebel, K. E., Cui, H., Feinberg, A. P., Lengauer, C., Kinzler, K. W., *et al.* (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552-556.

Riabowol, K., Draetta, G., Brizuela, L., Vandre, D., and Beach, D. (1989). The cdc2 kinase is a nuclear protein that is essential for mitosis in mammalian cells. Cell *57*, 393-401.

Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., *et al.* (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat Genet *23*, 266-268.

Rietveld, L. E., Caldenhoven, E., and Stunnenberg, H. G. (2002). In vivo repression of an erythroid-specific gene by distinct corepressor complexes. Embo J *21*, 1389-1397.

Roh, T. Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev *19*, 542-552.

Roh, T. Y., Ngau, W. C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nat Biotechnol *22*, 1013-1016.

Rountree, M. R., Bachman, K. E., and Baylin, S. B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet *25*, 269-277.

Russell, P. J., Welsch, J. A., Rachlin, E. M., and McCloskey, J. A. (1987). Different levels of DNA methylation in yeast and mycelial forms of Candida albicans. J Bacteriol *169*, 4393-4395.

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E., and Meehan, R. R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. Development *131*, 6185-6194.

Ryan, J., Llinas, A. J., White, D. A., Turner, B. M., and Sommerville, J. (1999). Maternal histone deacetylase is accumulated in the nuclei of Xenopus oocytes as protein complexes with potential enzyme activity. J Cell Sci *112 (Pt 14)*, 2441-2452.

Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol *225*, 294-303.

Sado, T., Okano, M., Li, E., and Sasaki, H. (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development *131*, 975-982.

Samaco, R. C., Hogart, A., and LaSalle, J. M. (2005). Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. Hum Mol Genet *14*, 483-492.

Sansom, O. J., Berger, J., Bishop, S. M., Hendrich, B., Bird, A., and Clarke, A. R. (2003). Deficiency of Mbd2 suppresses intestinal tumorigenesis. Nat Genet *34*, 145-147.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Santos-Rosa, H., Schneider, R., Bernstein, B. E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2003). Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. Mol Cell *12*, 1325-1332.

Sarg, B., Helliger, W., Talasz, H., Koutzamani, E., and Lindner, H. H. (2004). Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. J Biol Chem *279*, 53458-53464.

Sarraf, S. A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell *15*, 595-605.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet *34*, 65-69.

Scala, E., Ariani, F., Mari, F., Caselli, R., Pescucci, C., Longo, I., Meloni, I., Giachino, D., Bruttini, M., Hayek, G., *et al.* (2005). CDKL5/STK9 is mutated in Rett syndrome variant with infantile spasms. J Med Genet *42*, 103-107.

Scarano, M. I., Strazzullo, M., Matarazzo, M. R., and D'Esposito, M. (2005). DNA methylation 40 years later: Its role in human health and disease. J Cell Physiol.

Schmiedeberg, L., Weisshart, K., Diekmann, S., Meyer Zu Hoerste, G., and Hemmerich, P. (2004). High- and low-mobility populations of HP1 in heterochromatin of mammalian cells. Mol Biol Cell *15*, 2819-2833.

Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol *6*, 73-77.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev *18*, 1251-1262.

Schramke, V., and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science *301*, 1069-1074.

Schreiber-Agus, N., Chin, L., Chen, K., Torres, R., Rao, G., Guida, P., Skoultchi, A. I., and DePinho, R. A. (1995). An amino-terminal domain of Mxi1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor SIN3. Cell *80*, 777-786.

Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., *et al.* (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev *18*, 1263-1271.

Selker, E. U., Freitag, M., Kothe, G. O., Margolin, B. S., Rountree, M. R., Allis, C. D., and Tamaru, H. (2002). Induction and maintenance of nonsymmetrical DNA methylation in Neurospora. Proc Natl Acad Sci U S A *99 Suppl 4*, 16485-16490.

Selker, E. U., Tountas, N. A., Cross, S. H., Margolin, B. S., Murphy, J. G., Bird, A. P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893-897.

Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron *35*, 243-254.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Siegel, L. M., and Monty, K. J. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochim Biophys Acta *112*, 346-362.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Silverstein, R. A., and Ekwall, K. (2005). Sin3: a flexible regulator of global gene expression and genome stability. Curr Genet *47*, 1-17.

Simmen, M. W., Leitgeb, S., Charlton, J., Jones, S. J., Harris, B. R., Clark, V. H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science *283*, 1164-1167.

Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A *93*, 1443-1448.

Sontheimer, E. J. (2005). Assembly and function of RNA silencing complexes. Nat Rev Mol Cell Biol *6*, 127-138.

Soppe, W. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I., and Fransz, P. F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo J *21*, 6549-6559.

Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H., and Meehan, R. R. (2003). A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in Xenopus embryos. Mol Cell *12*, 425-435.

Stancheva, I., and Meehan, R. R. (2000). Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. Genes Dev *14*, 313-327.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol *5*, 699-711.

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

Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol *22*, 1298-1306.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem *279*, 27816-27823.

Sun, Z. W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Surani, M. A., Barton, S. C., and Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548-550.

Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., and Oikawa, T. (2003). Direct association between PU.1 and MeCP2 that recruits mSin3A-HDAC complex for PU.1-mediated transcriptional repression. Oncogene *22*, 8688-8698.

Svoboda, P., Stein, P., Filipowicz, W., and Schultz, R. M. (2004). Lack of homologous sequence-specific DNA methylation in response to stable dsRNA expression in mouse oocytes. Nucleic Acids Res *32*, 3601-3606.

Tamaru, H., and Selker, E. U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature *414*, 277-283.

Tamaru, H., Zhang, X., McMillen, D., Singh, P. B., Nakayama, J., Grewal, S. I., Allis, C. D., Cheng, X., and Selker, E. U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nat Genet *34*, 75-79.

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O. G., Shiekhattar, R., Picketts, D. J., and Yang, X. (2004). A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. J Biol Chem *279*, 20369-20377.

Tang, L. Y., Reddy, M. N., Rasheva, V., Lee, T. L., Lin, M. J., Hung, M. S., and Shen, C. K. (2003). The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. J Biol Chem *278*, 33613-33616.

Tao, J., Van Esch, H., Hagedorn-Greiwe, M., Hoffmann, K., Moser, B., Raynaud, M., Sperner, J., Fryns, J. P., Schwinger, E., Gecz, J., *et al.* (2004). Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. Am J Hum Genet *75*, 1149-1154.

Tariq, M., and Paszkowski, J. (2004). DNA and histone methylation in plants. Trends Genet *20*, 244-251.

Tariq, M., Saze, H., Probst, A. V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A *100*, 8823-8827.

Tate, P., Skarnes, W., and Bird, A. (1996). The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat Genet *12*, 205-208.

Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Thiel, G., Lietz, M., and Hohl, M. (2004). How mammalian transcriptional repressors work. Eur J Biochem *271*, 2855-2862.

Toyota, M., and Issa, J. P. (1999). CpG island methylator phenotypes in aging and cancer. Semin Cancer Biol *9*, 349-357.

Traynor, J., Agarwal, P., Lazzeroni, L., and Francke, U. (2002). Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. BMC Med Genet *3*, 12.

Tudor, M., Akbarian, S., Chen, R. Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc Natl Acad Sci U S A *99*, 15536-15541.

Turner, B. M. (2002). Cellular memory and the histone code. Cell *111*, 285-291.

Turner, B. M. (2005). Reading signals on the nucleosome with a new nomenclature for modified histones. Nat Struct Mol Biol *12*, 110-112.

Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol *17*, 1469-1475.

Tweedie, S., Ng, H. H., Barlow, A. L., Turner, B. M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in Drosophila melanogaster? Nat Genet *23*, 389-390.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem *280*, 64-72.

von Kries, J. P., Buhrmester, H., and Stratling, W. H. (1991a). A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. Cell *64*, 123-135.

von Kries, J. P., Buhrmester, H., and Stratling, W. H. (1991b). A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. Cell *64*, 123-135.

von Kries, J. P., Rosorius, O., Buhrmester, H., and Stratling, W. H. (1994). Biochemical properties of attachment region binding protein ARBP. FEBS Lett *342*, 185-188.

Vongs, A., Kakutani, T., Martienssen, R. A., and Richards, E. J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926-1928.

Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A., and Skalnik, D. G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet *23*, 62-66.

Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P., and Barlow, P. N. (1999). The solution structure of the domain from MeCP2 that binds to methylated DNA. J Mol Biol *291*, 1055-1065.

Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M., and Nakao, M. (2003). Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. Proc Natl Acad Sci U S A *100*, 12859-12864.

Waterborg, J. H. (1993). Dynamic methylation of alfalfa histone H3. J Biol Chem *268*, 4918-4921.

Watt, F., and Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev *2*, 1136-1143.

Weaving, L. S., Christodoulou, J., Williamson, S. L., Friend, K. L., McKenzie, O. L., Archer, H., Evans, J., Clarke, A., Pelka, G. J., Tam, P. P., *et al.* (2004). Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet *75*, 1079-1093.

Weitzel, J. M., Buhrmester, H., and Stratling, W. H. (1997a). Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Mol Cell Biol *17*, 5656-5666.

Weitzel, J. M., Buhrmester, H., and Stratling, W. H. (1997b). Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Molec & Cell Biol *17*, 5656-5666.

White, I. R., Pickford, R., Wood, J., Skehel, J. M., Gangadharan, B., and Cutler, P. (2004). A statistical comparison of silver and SYPRO Ruby staining for proteomic analysis. Electrophoresis *25*, 3048-3054.

Wilkinson, C. R., Bartlett, R., Nurse, P., and Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res *23*, 203-210.

Wilson, G. G., and Murray, N. E. (1991). Restriction and modification systems. Annu Rev Genet *25*, 585-627.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A. M., Lipkin, M., and Edelmann, W. (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci U S A *99*, 14937-14942.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Xiao, B., Wilson, J. R., and Gamblin, S. J. (2003a). SET domains and histone methylation. Curr Opin Struct Biol *13*, 699-705.

Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., and Strahl, B. D. (2003b). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev *17*, 654-663.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T. L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. Proc Natl Acad Sci U S A *100*, 10635-10640.

Yamada, T., Koyama, T., Ohwada, S., Tago, K., Sakamoto, I., Yoshimura, S., Hamada, K., Takeyoshi, I., and Morishita, Y. (2002). Frameshift mutations in the MBD4/MED1 gene in primary gastric cancer with high-frequency microsatellite instability. Cancer Lett *181*, 115-120.

Yamamoto, K., and Sonoda, M. (2003). Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. Biochem Biophys Res Commun *301*, 287-292.

Yoder, J. A., and Bestor, T. H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet *7*, 279-284.

Yoder, J. A., Soman, N. S., Verdine, G. L., and Bestor, T. H. (1997a). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol *270*, 385-395.

Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997b). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Yoon, H. G., Chan, D. W., Reynolds, A. B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell *12*, 723-734.

Yu, F., Thiesen, J., and Stratling, W. H. (2000). Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. Nucleic Acids Res *28*, 2201-2206.

Yu, F., Zingler, N., Schumann, G., and Stratling, W. H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res *29*, 4493-4501.

Yusufzai, T. M., and Wolffe, A. P. (2000). Functional consequences of Rett syndrome mutations on human MeCP2. Nucleic Acids Res *28*, 4172-4179.

Zegerman, P., Canas, B., Pappin, D., and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem *277*, 11621-11624.

Zemel, S., Bartolomei, M. S., and Tilghman, S. M. (1992). Physical linkage of two mammalian imprinted genes, H19 and insulin-like growth factor 2. Nat Genet *2*, 61-65.

Zeng, L., and Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Lett *513*, 124-128.

Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell *12*, 177-185.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell *89*, 357-364.

Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zhao, X., Ueba, T., Christie, B. R., Barkho, B., McConnell, M. J., Nakashima, K., Lein, E. S., Eadie, B. D., Willhoite, A. R., Muotri, A. R., *et al.* (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc Natl Acad Sci U S A *100*, 6777-6782.

Zheng, N., Fraenkel, E., Pabo, C. O., and Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev *13*, 666-674.

Publications

chemical synthesis (*3*). This technology was recently commercialized by Thomas Swan & Co., Ltd., in a chemical plant designed for multipurpose synthesis. Together with ionic liquids (*4–6*), these alternative solvent strategies (sometimes referred to as alternative reaction media or green solvents) provide a range of options to industrialists looking to minimize the environmental impact of their chemical processes.

What are the advantages of using a room-temperature ionic liquid in an industrially relevant catalytic process? As noted above, ionic liquids have no detectable vapor pressure, and therefore contribute no VOCs to the atmosphere. But this is not the only reason for using ionic liquids. Another is that at least a million binary ionic liquids, and 10^{18} ternary ionic liquids, are potentially possible (*7*). (For comparison, about 600 molecular solvents are in use today.)

This diversity enables the solvent to be designed and tuned (*2*) to optimize yield, selectivity, substrate solubility, product separation, and even enantioselectivity. Ionic liquids can be highly conducting (*8*), dissolve enzymes (*9*), form versatile biphasic systems for separations (*10*), can form both polymers and gels for device applications (*8*), are media for a wide range of organic and inorganic reactions (*4–6*), and are the basis for at least one industrial process, called the BASIL process (see the figure) (*11*).

The BASIL process was developed and is operated by BASF. At the meeting, Matthias Maase (BASF) revealed that use of the BASIL process increases the productivity of their alkoxyphenylphosphine formation process by a factor of 80,000 compared with the conventional process. Other companies are also pursuing the use of ionic liquids. Bernd Weyershausen (Degussa) presented an ionic liquid–based process for the synthesis of organosilicon compounds. Use of an ionic liquid solvent enabled the catalyst to be easily recycled and reused without further treatment after separation from the product at the end of the reaction. Christian Mehnert (ExxonMobil) described biphasic hydroformylation with rhodium catalysts in ionic liquids.

Because research into ionic liquids is at an early stage, many of their properties remain to be elucidated. Nonetheless, ionic liquids have already provided access to new chemical processes. Recent papers describe their potential application as embalming fluids (*12*), in ion drives for space travel (*13*), for desulfurization of fuels (*14*), and as lubricants (*15*).

Ionic liquids have already found many laboratory applications in synthesis, catalysis, batteries, and fuel cells (*4–6*, *8*, *16*), and numerous new combinations of ionic liquid solvent properties are available or predicted. The next decade should see ionic liquids being used in many applications where conventional organic solvents are used today. Furthermore, ionic liquids will enable new applications that are not possible with conventional solvents. In the future, solvents will be designed to control chemistry, rather than the chemistry being dictated by the more limited range of molecular solvents currently used.

References and Notes

1. ACS Fall Meeting, 7 to 11 September 2003, New York. The Ionic Liquids symposium was sponsored by the ACS Division of Industrial and Engineering Chemistry, the Green Chemistry and Engineering Subdivision, the Separation Science and Technology Subdivision, and the Green Chemistry Institute.
2. M. Freemantle, *Chem. Eng. News* **76**, 32 (30 March 1998).
3. M. Poliakoff, J. M. Fitzpatrick, T. R. Farren, P. T. Anastas, *Science* **297**, 807 (2002).
4. R. D. Rogers, K. R. Seddon, Eds., *Ionic Liquids as Green Solvents: Progress and Prospects* (ACS Symp. Ser. 856, American Chemical Society, Washington, DC, 2003).
5. R. D. Rogers, K. R. Seddon, Eds., *Ionic Liquids: Industrial Applications for Green Chemistry* (ACS Symp. Ser. 818, American Chemical Society, Washington, DC, 2002).
6. P. Wasserscheid, T. Welton, Eds., *Ionic Liquids in Synthesis* (Wiley-VCH, Weinheim, Germany, 2003).
7. K. R. Seddon, in *The International George Papatheodorou Symposium: Proceedings*, S. Boghosian *et al.*, Eds. (Institute of Chemical Engineering and High Temperature Chemical Processes, Patras, Greece, 1999), pp. 131–135.
8. H. Ohno, Ed., *Ionic Liquids: The Front and Future of Material Developments* (CMC, Tokyo, 2003).
9. R. A. Sheldon, R. M. Lau, M. J. Sorgedrager, F. van Rantwijk, K. R. Seddon, *Green Chem.* **4**, 147 (2002).
10. K. E. Gutowski *et al.*, *J. Am. Chem. Soc.* **125**, 6632 (2003).
11. K. R. Seddon, *Nature Mater.* **2**, 363 (2003).
12. P. Majewski, A. Pernak, M. Grzymislawski, K. Iwanik, J. Pernak, *Acta Histochem.* **105**, 135 (2003).
13. M. Gamero-Castano, V. Hruby, *J. Propulsion Power* **17**, 977 (2001).
14. A. Bosmann *et al.*, *Chem. Commun.*, 2494 (2001).
15. W. M. Liu, C. F. Ye, Q. Y. Gong, H. Z. Wang, P. Wang, *Tribol. Lett.* **13**, 81 (2002).
16. R. D. Rogers, K. R. Seddon, S. Volkov, Eds., *Green Industrial Applications of Ionic Liquids* (Kluwer, Dordrecht, Netherlands, 2002), vol. 92.

17. R.D.R. acknowledges financial support from the U.S. Environmental Protection Agency; NSF; Air Force Office of Scientific Research; U.S. Department of Energy, Environmental Management Science Program and Office of Basic Energy Sciences, Office of Energy Research; National Renewable Energy Laboratory; and the PG Research Foundation. QUILL's industrial sponsors include Avecia, bp, Chevron, C-Tri, Cytec, Eastman, ICI, Merck, Novartis, SASOL, Shell, and UOP.

MOLECULAR BIOLOGY

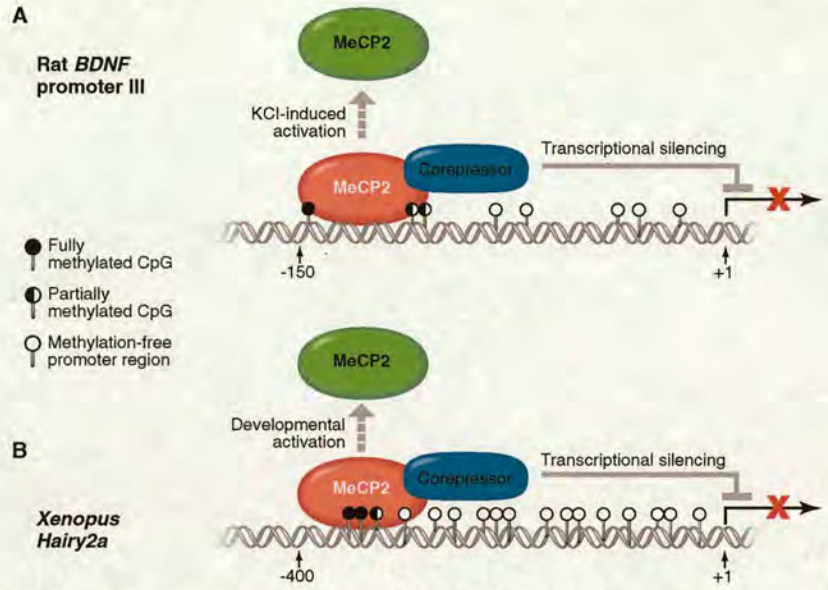
MeCP2 Repression Goes Nonglobal

Robert Klose and Adrian Bird

After replication, mammalian DNA becomes marked by the addition of methyl groups to certain cytosine bases, almost exclusively those in the sequence 5′CpG. Just how the resulting pattern of methylated and nonmethylated cytosines is converted into biological outcomes is now starting to become clear. Methyl-CpG has emerged as a gene silencing signal that usually ensures the long-term shutdown of gene expression. Likely mediators of this effect are the methyl-CpG binding domain (MBD) proteins that recruit transcriptional silencing machinery to the DNA. One of these proteins, MeCP2, is of particular interest because about 80% of patients with a profound neurological condition called Rett syndrome carry a mutation in their *MECP2* gene (*1*). Mice lacking the *Mecp2* gene exhibit several features of Rett syndrome. Furthermore, targeted deletion of *Mecp2* in mouse brain causes a Rett-like phenotype that is virtually indistinguishable from the phenotype of mice in which every tissue lacks MeCP2. Global microarray analyses designed to search for target genes that are derepressed in the brains of MeCP2-deficient mice, and thus might be causally implicated in Rett syndrome, have not yielded any clear candidates (*2*). The existence of a genuine target gene is now highlighted on pages 885 (*3*) and 889 (*4*) of this issue. Chen *et al.* (*3*) and Martinowich *et al.* (*4*) describe their discovery that normal MeCP2 regulates expression of the gene encoding brain-derived neurotrophic factor (BDNF), a secreted protein that is essential for neural plasticity, learning, and memory. Their new findings, together with recent work in the amphibian *Xenopus laevis* (*5*), confirm that MeCP2 is a methyl-CpG–dependent transcriptional repressor and reveal its unexpected role in the induction of gene expression in the nervous system.

Belonging to a set of proteins synthesized in response to neuronal activity, BDNF is thought to be essential for converting transient stimuli into long-term changes in brain activity (*6*). Understanding how BDNF is regulated in neurons is important if we are to comprehend brain development, learning, and memory. There are four BDNF promoters, one of which (promoter III) responds

The authors are at the Wellcome Trust Centre for Cell Biology, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JR, UK. E-mail: a.bird@ed.ac.uk



How MeCP2 lets go. Despite the existence of methylated CpGs throughout the vertebrate genome, loss of MeCP2 (a methyl-CpG binding domain protein) does not cause global deregulation of gene expression. MeCP2 binds to and represses the promoters of two specific genes: promoter III of the rat *BDNF* gene (**A**) and the *Hairy2a* gene promoter of *Xenopus* (**B**). In both cases, MeCP2 binds to the upstream fully methylated (black dot) or partially methylated (black and white dot) CpGs, but not to the downstream methylation-free (white dot) promoter region. MeCP2 recruits corepressor complexes, maintaining and solidifying a state of gene repression (red cross). Induction of gene expression leads to complete loss of MeCP2 from the upstream promoter region (dashed arrow), enabling transcription to proceed.

to artificial stimulation of cultured rodent primary cortical neurons treated with potassium chloride (KCl). By immunoprecipitating chromatin fragments containing MeCP2, Chen *et al.* and Martinowich *et al.* discovered that MeCP2 is bound to methylated CpG sites near promoter III of *BDNF* in resting neurons. But when the neurons were exposed to KCl—which causes membrane depolarization, calcium influx, and BDNF activation—MeCP2 dissociated from the *BDNF* gene promoter. Martinowich *et al.* further demonstrated that Sin3a—a transcriptional corepressor that forms complexes with histone deacetylases and associates with MeCP2 (*7*)—is also displaced after KCl treatment. Accordingly, loss of MeCP2 is accompanied by changes in histone modification (*8*), resulting in a transcriptionally repressive chromatin state being replaced by a permissive one.

How is MeCP2 displaced from promoter III when the *BDNF* gene is activated? One possibility is that the CpG methylation holding MeCP2 at the promoter is abruptly lost. In line with this hypothesis, Martinowich *et al.* found somewhat reduced CpG methylation in the relevant region of the activated promoter. An alternative mechanism for MeCP2 loss could involve modification of MeCP2 during the induction of gene expression such that it loses its affinity for the methylated promoter site. In support of this possibility, Chen *et al.* observed a time-dependent increase in phosphorylation of MeCP2 when neurons were stimulated with KCl. Moreover, Southwestern blot analysis indicated that the phosphorylated form of MeCP2 had a lower affinity for methylated DNA, although its presence at an unrelated methylated promoter—that of the imprinted *H19* gene—was not diminished by KCl treatment. Conceivably, displacement of MeCP2 is caused by a combination of cytosine demethylation and MeCP2 phosphorylation.

If MeCP2 is required for repression of promoter III before and after induction of *BDNF* gene expression, its absence should disrupt this process. Chen *et al.* tested this in vivo by examining inducible *BDNF* gene expression in cultured neurons from MeCP2-deficient mice. They found that *BDNF* transcription doubled in resting cells, whereas under stimulatory conditions there was no difference in *BDNF* transcription between wild-type and MeCP2-deficient neurons. While supporting the idea that MeCP2 helps to repress basal levels of BDNF expression, the effects of deleting the *Mecp2* gene appear to be subtle; uninduced *BDNF* gene expression increased from ~1% to ~2% of the induced level. It is perhaps not surprising that microarrays failed to detect such an effect.

BDNF is an important player in neuronal development, raising the possibility that misregulation of its gene because of the absence of MeCP2 may contribute to the symptoms of Rett syndrome. Other inducible promoters in neurons may also rely on repression by MeCP2. Recently, another bona fide MeCP2 target gene was discovered, this time in *Xenopus*. Stancheva and colleagues (*5*) found that reducing production of MeCP2 during early *Xenopus* embryogenesis by injecting antisense oligonucleotides caused gross defects in frog neurogenesis. By screening candidate genes known to be involved in Notch-Delta signaling cascades, they found that a transcriptional repressor protein, Hairy2a, is abnormally up-regulated in the absence of MeCP2. Hairy2a represses expression of proneuronal genes in non-neuronal cells surrounding developing neurons. Neurogenesis defects and gene expression changes caused by depletion of MeCP2 could be completely rescued by reexpressing wild-type human MeCP2, but not mutant forms of MeCP2 from Rett syndrome patients. The behavior of MeCP2 at the Hairy2a promoter has striking parallels with the BDNF story. Once again, MeCP2 is localized to the methylated upstream flank of the promoter (see the figure) and vacates the promoter upon activation of the gene (although CpG methylation in this case apparently remains unchanged). A deficiency of *Xenopus* MeCP2 leads to inappropriate activation of the *Hairy2a* gene, with severe consequences for the embryonic nervous system.

DNA methylation is often thought of as a "global" parameter in gene regulation. It is pervasive, being distributed throughout the genome. It is also an essential process—mice lacking the DNA methyltransferase enzyme, DNMT1, die during embryogenesis (*9*). Furthermore, mouse cells lacking DNMT1 display genomic instability (*10*), implying that methylation has wide-ranging effects on global genome function and integrity. Whether individual MBD proteins qualify as global gene regulators through their interpretation of methylation marks is less certain (*11*). Mice lacking MBD2, for example, are viable and fertile, but show deregulated expression of the interleukin-4 gene in T helper cells, indicating a gene-specific requirement for MBD2 (*12*). The papers discussed here establish for the first time that MeCP2 itself has specific gene targets and that it may work in concert with other factors as part of a multiprotein promoter complex. More unexpected still is its dynamic association with DNA, perhaps regulated by phosphorylation. This scenario contrasts with the conventional view that DNA methylation and MBD proteins create an almost immovable repressive environment. The new findings reveal unexpected plasticity in the biological interpretation of methylated DNA.

CREDIT: KATHARINE SUTLIFF/*SCIENCE*

Questions inevitably remain. Is the dynamic behavior of MeCP2 associated with the *BDNF* or *Hairy2a* gene promoters the exception or the rule? What other genes are induced when MeCP2 becomes phosphorylated? Which are the genes whose misregulation causes Rett syndrome? The explosion of knowledge about DNA methylation and the brain is at last making these questions experimentally accessible.

References

1. R. E. Amir *et al.*, *Nat. Genet.* **23**, 185 (1999).
2. M. Tudor *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15536 (2002).
3. W. G. Chen *et al.*, *Science* **302**, 885 (2003).
4. K. Martinowich *et al.*, *Science* **302**, 889 (2003).
5. I. Stancheva *et al.*, *Mol. Cell* **12**, 425 (2003).
6. A. E. West *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11024 (2001).
7. X. Nan *et al.*, *Nature* **393**, 386 (1998).
8. T. Jenuwein, C. D. Allis, *Science* **293**, 1074 (2001).
9. E. Li *et al.*, *Cell* **69**, 915 (1992).
10. A. Eden *et al.*, *Science* **300**, 455 (2003).
11. S. Kriaucionis, A. Bird, *Hum. Mol. Genet.* **12**, R221 (2003).
12. A. S. Hutchins *et al.*, *Mol. Cell* **10**, 81 (2002).

PHYSICS

Searching for Gravity's Hidden Strength

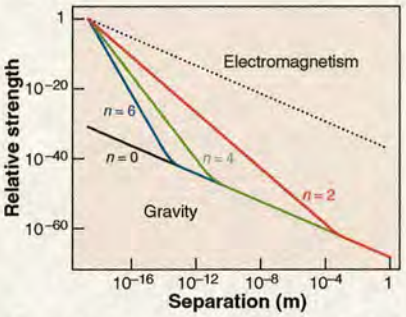
Jonathan L. Feng

Of the four known fundamental forces—gravity, electromagnetism, and the weak and strong forces—gravity is by far the weakest. The reasons for this weakness have long remained enigmatic. Recent proposals suggest, however, that the weakness of gravity may be evidence for extra spatial dimensions. Experiments ranging from tabletop tests of Newtonian gravity to searches for microscopic black holes in kilometer-scale detectors are now putting these ideas to the test.

The importance of gravity in everyday life results not from its strength but from its universality: Objects cannot be gravitationally neutral, and all bodies with mass attract. Yet as an interaction between elementary particles, gravity is extremely weak. For example, the gravitational attraction between two protons is 35 orders of magnitude weaker than their electromagnetic repulsion. This holds for protons separated by any distance r, because both gravitational and electromagnetic forces are proportional to $1/r^2$.

The observed weakness of gravity may, however, not be an intrinsic property of gravity, but may instead be an effect of extra spatial dimensions. This possibility is based on a simple consideration. Suppose that our three-dimensional (3D) world is merely a subspace of a higher-dimensional space, and that gravity propagates freely in all dimensions, but that all other forces are confined to our three dimensions. In contrast to the familiar three dimensions, the extra dimensions are curled up in small circles of circumference L. Hence, moving a distance L in the direction of any of the extra dimensions brings one back to one's starting place.

Now suppose that at some separation distance $r < L$, gravity is strong, that is,



Gravity in extra dimensions. The strength of gravity for various numbers of large extra dimensions n is compared to the strength of electromagnetism (dotted). Without extra dimensions, gravity is weak relative to the electromagnetic force for all separation distances. With extra dimensions, the gravitational force rises steeply for small separations and may become comparable to electromagnetism at short distances.

comparable to electromagnetism. As r increases, the electromagnetic force drops as $1/r^2$. However, the gravitational field spreads out in all available spatial dimensions, and the gravitational force therefore decreases much more rapidly as $1/r^{2+n}$, where n is the number of extra dimensions. This rapid drop continues until $r > L$, at which point the extra dimensions become less and less important and gravity recovers its $1/r^2$ behavior (see the figure).

If this picture is correct, then gravity is not intrinsically weak: It is as strong as electromagnetism at small length scales. It appears weak at the relatively large distances of common experience only because its effects are diluted by propagation in extra dimensions. The distance at which the gravitational and electromagnetic forces might have equal strength is unknown, but a particularly interesting possibility is that it is 10^{-19} m, the distance at which the electromagnetic and weak forces are known to unify to form the electroweak force (*1*).

A priori, the size of the extra dimensions L and their number n are independent parameters. However, to achieve equality of gravitation and electromagnetic forces at 10^{-19} m, they become constrained by the relation

$$L \approx 10^{(32/n)-19}\ \text{m} \qquad (1)$$

For large n, the strength of gravity grows very rapidly at microscopic length scales. Gravity may then deviate from its $1/r^2$ behavior only at very small distances and still be comparable to electromagnetism at 10^{-19} m.

This scenario, called "large extra dimensions" because the length L of Eq. 1 is large relative to typical length scales in particle physics, raises many more questions than it answers. When first proposed, perhaps its most surprising aspect was that such a bold modification of Newtonian gravity was not immediately excluded by data. Now, however, a wide variety of experiments are reaching the sensitivity required to test these speculative ideas. In combination, they probe all possible values for the number of extra dimensions, placing the entire scenario on the threshold of detailed investigation.

The possibility of one large extra dimension is untenable. It requires the extra dimension to be of size $L \approx 10^{13}$ m, a length scale where the $1/r^2$ gravitational force law is clearly still valid. For two extra dimensions, each extra dimension would have $L \approx 1$ mm. Sensitive tests of gravity are notoriously difficult at such length scales. Nonetheless, recent tabletop experiments with torsion pendulums have excluded significant deviations from the $1/r^2$ force law at length scales as small as 0.1 mm (*2*).

Astrophysical observations provide less direct but more stringent constraints on low numbers of extra dimensions (*3*, *4*). For two extra dimensions, for example, the gravitational force would be enhanced at large enough length scales that supernovae should release much of their energy as gravitational energy—in conflict with observations. These constraints, which were noted immediately after the proposal of large extra dimensions, exclude scenarios with few extra dimensions.

The challenge, then, has been to explore large numbers of extra dimensions, such as the six or seven favored by string theory. In such cases, tabletop and astrophysical constraints are ineffective, because the predicted deviations from Newtonian gravity oc-

The author is in the Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA. E-mail: jlf@uci.edu

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 45, Issue of November 5, pp. 46490–46496, 2004
Printed in U.S.A.

MeCP2 Behaves as an Elongated Monomer That Does Not Stably Associate with the Sin3a Chromatin Remodeling Complex*

Received for publication, July 22, 2004, and in revised form, August 17, 2004
Published, JBC Papers in Press, August 18, 2004, DOI 10.1074/jbc.M408284200

Robert J. Klose and Adrian P. Bird‡

From The Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, United Kingdom

MeCP2 is a transcription factor that recognizes and binds symmetrically methylated CpG dinucleotides to repress transcription. MeCP2 can associate with the Sin3a/histone deacetylase corepressor complex and mediate repression in a histone deacetylase-dependent manner. In extracts from rodent tissues, cultured cells, and *Xenopus laevis* oocytes, we find that only a small amount of mammalian MeCP2 interacts with Sin3a and that this interaction is not stable. Purification of rat brain MeCP2 (53 kDa) indicates no associated proteins despite an apparent molecular mass by size exclusion chromatography of 400–500 kDa. Biophysical analysis demonstrated that the large apparent size was not because of homo-multimerization, as MeCP2 consistently behaves as a monomeric protein that has an elongated shape. Our findings indicate the MeCP2 is not an obligate component of the Sin3a corepressor complex and may therefore engage a more diverse range of cofactors for repressive function.

Cytosine methylation is an important epigenetic mark on vertebrate genomes (1). In mammals, methylation occurs mostly in the context of the CpG dinucleotide and can account for about 70–80% of genomic CpGs (2). CpG methylation acts as an additional means of controlling genome function, beyond DNA base pair sequence. Because DNA methylation is copied faithfully to the newly replicating DNA strands during cell division, these marks can be maintained across development and act as a form of epigenetic memory. There are two general mechanisms by which CpG methylation is believed to function. First, modification of cytosines in the recognition sequence of DNA-binding proteins can inhibit their binding to cognate sequences and thus deny access to regulatory regions. Second, proteins have been identified that specifically bind the methyl-CpG dinucleotide via a methyl-CpG binding domain (MBD)[1] (3) or, in the case of Kaiso (4), by a zinc finger domain. These proteins can interact with methylated CpGs and affect nearby genes by repressing transcription and modulating chromatin structure (5).

There are five mammalian members of the MBD family: MeCP2 and MBD1–4. With the exception of MBD3, all MBD family members bind methylated CpG dinucleotide specifically. MBD1 (6–8), MBD2 (9–12), MBD3, and MeCP2 (13) are all transcriptional repressors. Elucidation of the relationship between MBD proteins and their partner corepressors is a prerequisite for understanding how DNA methylation represses transcription and modulates chromatin structure. Several methyl-CpG-binding proteins have been shown to associate with histone deacetylases (HDACs) (10, 14, 15) or histone methyltransferases (16, 17). MBD2 and MBD3 are stable components of the NuRD chromatin remodeling complex (10, 12), and Kaiso can be purified in a complex containing NCoR (18). In this study, we have asked whether MeCP2 is also part of a stable repression complex.

MeCP2 dysfunction is the sole identified determinant of Rett syndrome, the most common inherited form of mental retardation in females (19). The majority of Rett syndrome point mutations in the *MECP2* gene cluster in the MBD and the transcriptional repression domain (20), suggesting that methyl-CpG binding and transcriptional repression are important functional determinants of MeCP2 *in vivo*. MeCP2 has been shown to interact with the Sin3a-HDAC chromatin remodeling complex (14). In *Xenopus laevis* it was reported that MeCP2 partially cofractionated with the Sin3a complex and proposed that xMeCP2 occurs in a stable complex with xSin3a (21, 22). The link between MeCP2 and Sin3a has been strengthened recently with the identification of MeCP2 target genes. Both MeCP2 and Sin3a bind the promoter region of the brain-derived neurotrophic factor (*Bdnf*) gene (23–25) and modulate its expression. A similar situation occurs at the *X. laevis xHairy2a* gene, where MeCP2 and Sin3a bind upstream of the promoter region and repress transcription (26).[2]

In addition to Sin3a, several other factors have been reported to bind mammalian MeCP2, including DNMT1, CoREST, Suv39H1, and c-SKI (14, 27–29), although the contribution of these factors to MeCP2-mediated repression is not known. Native MeCP2 has not been purified previously from mammalian sources, leaving open the possibility that it may exist in a novel multiprotein complex. Here we investigate the association of MeCP2 with the Sin3a complex from both mammalian sources and *X. laevis*, and we purify native MeCP2 from rat brain. We conclude that MeCP2 does not stably associate with the Sin3a complex. Moreover, we find no evidence that MeCP2 forms a stable association either with itself or with other proteins in nuclear extracts. Hydrodynamic analysis of MeCP2 shows that it behaves as an elongated monomeric molecule. These findings raise the possibility that DNA-bound MeCP2 interacts differently with partner proteins compared with its unbound form. In addition, the results suggest that MeCP2 might interact with a range of cofactors in addition to Sin3a.

* This work was supported by a grant from The Wellcome Trust (to A. P. B.) and by a Wellcome Trust Prize Studentship (to R. J. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: The Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Bldg., Mayfield Rd., Edinburgh EH9 3JR, Scotland, UK. Tel.: 44-131-650-8695; Fax: 44-131-650-5379; E-mail: A.Bird@ed.ac.uk.

[1] The abbreviations used are: MBD, methyl-CpG binding domain; HDAC(s), histone deacetylase(s); DTT, dithiothreitol; Ni-NTA, nickel-nitrilotriacetic acid; BSA, bovine serum albumin; EGS, ethylene glycol-bis succinimidylsuccinate; ADH, alcohol dehydrogenase; AE, anion exchange; CE, cation exchange; r, recombinant.

[2] I. Stancheva, personal communication.

EXPERIMENTAL PROCEDURES

Cell Culture—NG-108 cells were a gift of Rod Bremner (University of Toronto) and were maintained in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% bovine calf serum, nonessential amino acids, sodium pyruvate, and antibiotics (Invitrogen).

Chromatography Solutions—Chromatography buffers were filtered through a 0.2-μm filter before application to fast protein liquid chromatography or disposable columns. Anion exchange buffer 20 mM Tris-HCl (pH 7.9), 0.2 mM EDTA, 1 mM DTT, 10% glycerol, supplemented with 100 mM NaCl (AE100), or 1000 mM NaCl (AE1000). Cation exchange buffer was 20 mM Hepes (pH 7.6), 0.2 mM EDTA, 1 mM DTT, 10% glycerol, supplemented with 100 mM NaCl (CE100), or 150 mM NaCl (CE150), or 200 mM NaCl (CE200), or 1000 mM NaCl (CE1000). Ni-Nta affinity buffers (N) were 50 mM NaH_2PO_4, 300 mM NaCl, 10% glycerol (pH 8.0), supplemented with 20 mM imidazole (N20), 100 mM imidazole (N100), or 250 mM imidazole (N250). Gel filtration buffers were made in 20 mM Hepes-KOH (pH 7.9), 3 mM $MgCl_2$, 10% glycerol, supplemented with 150 mM KCl (GF150) or 500 mM KCl (GF500).

Isolation of Rat Brain Nuclei and Nuclear Protein Extraction—Rat brains (~450 brains obtained from Pel-Freez Biologicals) were ground to a fine powder in liquid nitrogen with a mortar and pestle. The brain powder was diluted 5 volumes to 1 in ice-cold buffer A containing 10 mM Hepes (pH 7.5), 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 M sucrose, 10% glycerol, and complete protease inhibitors (Roche Applied Science) followed by homogenization in a 60-ml Dounce (Braun) on a Potter S (Braun) motorized homogenizer (five strokes at 1100 rpm). The homogenate was layered onto a 10-ml cushion of buffer A and centrifuged in pre-chilled SW28 rotor at 24,000 rpm in an Beckman XL100 ultracentrifuge for 40 min at 3 °C. Recovered nuclei were resuspended in 5 volumes of buffer B containing 10 mM Hepes (pH 7.9), 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, complete protease inhibitors (Roche Applied Science) and incubated on ice 10 min. The nuclei were pelleted at 250 × *g* and resuspended in 1 volume of buffer C containing 5 mM Hepes (pH 7.9), 26% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, and complete protease inhibitors (Roche Applied Science) supplemented 400 mM NaCl. The extraction was allowed to proceed for 1 h on ice, and then the nuclei were pelleted at 13,000 rpm for 20 min at 4 °C. The supernatant was taken as the nuclear extract and dialyzed to the indicated salt concentration.

Immunoprecipitation—MeCP2 or Gal4 antibodies were incubated with 500 μg of rat brain nuclear extract at 4 °C for 4 h. Protein A-Sepharose beads (Amersham Biosciences) were added to the reaction and incubated for 1 h at 4 °C. Beads were washed four times with 20 mM Hepes (pH 7.9), 0.1 M NaCl, 10% glycerol, 0.2 mM EDTA, 0.01% Triton X-100. Bound proteins were eluted in Laemmli buffer and run on an 8% SDS-polyacrylamide gel. MeCP2 (14), Sin3a (Santa Cruz Biotechnology SC994 K20), and topoisomerase I (SC10783 H300) were identified by Western blotting.

Large Scale Purification of Native MeCP2—All chromatography and dialyses were performed at 4 °C, and MeCP2 (14), Sin3a (Santa Cruz Biotechnology SC994 K20), and HDAC 1/2 (Santa Cruz Biotechnology SC7879 H51/SC7899 H54) containing fractions were identified by Western blotting. Rat brain nuclear extract (26 ml of 400 mM NaCl extraction) was dialyzed against CE150, and the insoluble material was pelleted at 13,000× *g* for 20 min at 4 °C. Soluble material (140 mg) was loaded onto an 8-ml SP-Sepharose (Amersham Biosciences) column and eluted with a 160-ml linear gradient of CE100 to CE1000 collecting 4-ml fractions. MeCP2-containing fractions were combined (8.8 mg) and dialyzed against AE100. Dialyzed protein was loaded onto a 1-ml MonoQ (Amersham Biosciences) column and eluted with a 40-ml linear gradient of AE100 to AE1000. MeCP2-containing fractions from the flow-through and wash were combined (4 mg) and dialyzed against CE100. The protein was loaded onto a 1-ml heparin (Amersham Biosciences) affinity column and eluted with a 40-ml linear gradient of CE100 to CE1000. MeCP2-containing fractions (0.408 mg) were combined and dialyzed into N20. Ni-NTA resin (0.65 ml; Qiagen) pre-equilibrated with N20 was added to the dialyzed protein and mixed at 4 °C for 1 h. The Ni-NTA was applied to a 10-ml disposable column (Bio-Rad). The column was washed thoroughly with N20 and then eluted in batch with N100 and N250. MeCP2 eluted at the N100 step, and these fractions were pooled (0.148 mg). An aliquot of the N100 elution was applied to a Superose 12 (Amersham Biosciences) column and eluted from the column in GF500 collecting 0.5-ml fractions. Proteins from the gel filtration column were Western-blotted to identify MeCP2-containing fractions, and subsequently trichloroacetic acid-precipitated, subjected to SDS-PAGE, and stained with Sypro-Ruby stain (Bio-Rad). Bands were excised and analyzed by mass spectrometry.

Mass Spectrometry—To identify purified proteins, excised bands were in-gel trypsinized, and peptides were eluted from the gel slice. Mass spectrometry analysis was carried out on an Applied Biosystems Voyager DE-STR matrix-assisted laser desorption ionization time-of-flight instrument using α-cyano-4-hydroxycinnamic acid matrix. Spectra were analyzed in MS-Fit and then submitted to Protein Prospector (prospector.ucsf.edu/) for peptide matching.

Xenopus Oocyte Extract and xMeCP2 Chromatography—Oocyte extract was prepared from one female as described previously (21). The extract (90 mg) was then loaded onto a 10-ml Bio-Rex 70 column exactly as described previously (21) in Bio-Rex 70 exchange buffer (BE100) containing 100 mM NaCl, 20 mM Hepes (pH 7.5), 10 mM β-glycerophosphate, 1.5 mM $MgCl_2$, 1 mM EGTA, 0.5 mM DTT, 10% glycerol, complete protease inhibitors (Roche Applied Science) and washed with three column volumes of BE100. The column was batch-eluted with Bio-Rex 70 exchange buffer containing 500 mM NaCl (BE500), and fractions were analyzed by Western blotting with three independent xMeCP2 antibodies and an xSin3a antibody (a gift of P. L. Jones). The BE500 elution (250 μl) was separated on a Superose 6 (Amersham Biosciences) gel filtration column in GF150 collecting 0.5-ml fractions. Proteins were trichloroacetic acid-precipitated and run on an 8% SDS-polyacrylamide gel and Western blotted for xMeCP2 and xSin3a.

MeCP2 Expression Plasmid—A human MeCP2 cDNA was used to PCR-amplify MeCP2 with primer pairs containing a 5′ NdeI site corresponding to the initiating ATG and 3′ EcoRI site downstream of the endogenous MeCP2 stop codon. The PCR fragment was inserted into the NdeI/EcoRI sites of the bacterial expression plasmid pET30b (Novagen) to create an untagged bacterial MeCP2 expression vector pET30bhMeCP2.

Expression and Purification of Full-length Untagged MeCP2 in Bacteria—pET30bhMeCP2 was transformed into BL21 codon plus bacteria (a gift from Robin Allshire). Bacterial cultures (usually 0.5 or 1 liter) were grown in LB at 37 °C until the culture reached an A_{600} of 0.5 absorbance units. Cultures were induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 3 h at 30 °C. Cells were pelleted and lysed in 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 0.1% Nonidet P-40, complete protease inhibitors (Roche Applied Science) by sonicating on output setting 4–5 at 30% for 3 min with a Branson 250 sonifier. Extracts were centrifuged at 4 °C for 20 min at 20,000 × *g*, and Ni-NTA beads (Qiagen) pre-equilibrated with lysis buffer were added to the supernatant. Recombinant MeCP2 was allowed to bind the beads mixing for 1 h at 4 °C and then applied to a 10-ml disposable column. The column was washed with 20 column volumes of N20 and batch-eluted with N250. MeCP2-containing fractions were identified by Coomassie Blue staining and dialyzed into CE200. The protein from the Ni-NTA elutions was loaded onto an Sp-Sepharose column and eluted with a linear gradient of CE200 to CE1000, and the MeCP2-containing fractions were combined and directly loaded onto a Sephacryl S-300 26/60 (Amersham Biosciences) column and eluted with GF500. MeCP2-containing fractions were combined and dialyzed against CE200 and then loaded onto a 1-ml MonoS column. Proteins were eluted with a linear gradient of CE200 to CE1000, and MeCP2-containing factions were combined and dialyzed into CE200 and stored at −20 °C.

Bandshift Analysis—Increasing concentrations of rMeCP2 (50, 100, 250, 500, and 750 ng) were bound to the CG11 probe (30) either methylated or unmethylated. The CG11 probe containing 27 CpGs was generated by digesting out an EcoRI/HindIII (New England Biolabs) fragment from the plasmid pCG11 (30). The 135-bp CG11 fragment was purified by gel electrophoresis and agarose gel extraction. The CpGs in the resulting fragments were methylated with SssI methyltransferase (New England Biolabs), and the methylated or unmethylated CG11 probe was end-labeled with [^{32}P]dCTP using Klenow (Roche Applied Science). The binding reactions were assembled in buffer containing 6 mM $MgCl_2$, 3% glycerol, 1 mM DTT, 150 mM KCl, 100 ng/μl poly(dA-dT) (Sigma) for 10 min at room temperature in the absence of probe, and then the probe was added and incubated a further 25 min. The reactions were run on a 1.5% agarose gel and dried onto DE-81 anion exchange paper (Whatman). The bandshifts were exposed on a phosphor screen and analyzed on a Storm 840 PhosphorImager (Amersham Biosciences). Competition with cold methylated and unmethylated probe at ~100-fold excess was used to demonstrate the specificity of the bandshift.

EGS Cross-linking—An EGS (Pierce) stock solution was made fresh to 25 mM in Me_2SO. 5 μg of rMeCP2 was incubated in CE200 (without DTT or EDTA) with increasing concentrations of EGS (0.25, 0.5, 1.0, 2.5, 5.0 mM) in a 50-μl reaction volume. BSA (5 μg, monomer) and ADH (5 μg, tetramer) were included as internal controls for cross-linking efficiency. Cross-linking reactions were carried out at room tempera-

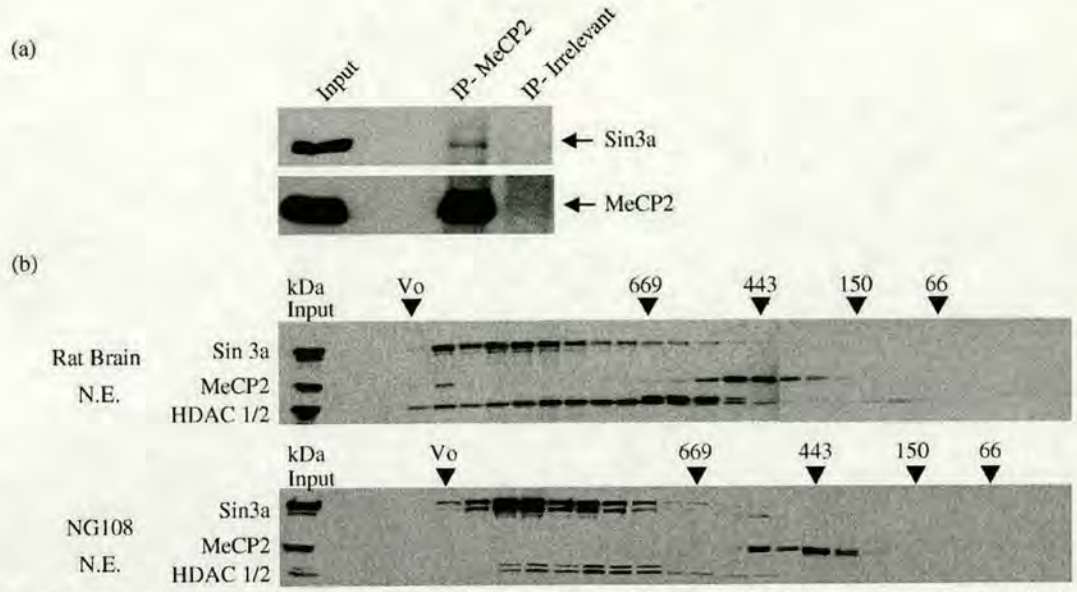


FIG. 1. *a*, Sin3a is coimmunoprecipitated from rat brain nuclear extracts by anti-MeCP2 antibodies but not an irrelevant antibody (anti-Gal4). *b*, Superose 6 size exclusion chromatography of nuclear extracts (*N.E.*) from rat brain or NG108 cells. Fractions were analyzed by Western blotting with antibodies raised against Sin3a, MeCP2, and HDAC 1/2 as indicated. The elution profile of molecular weight standards and the void volume (V_0) is indicated above each Western blot. *IP*, immunoprecipitation.

ture for 30 min. Tris-HCl (pH 8.8) (5 μl of 1.5 M) was added to quench the reaction for 15 min at room temperature. SDS-PAGE loading buffer (4×) was added, and the proteins separated on an 8% SDS-polyacrylamide gel followed by Coomassie Blue staining to visualize the proteins.

Gel Filtration and Sucrose Gradient Analysis of rMeCP2—A Superose 12 HR 10/30 gel filtration column was pre-equilibrated with gel filtration standards thyroglobin (669 kDa, R_S = 8.5), apoferritin (443 kDa, R_S = 6.1), B-amylase (200 kDa, R_S = 5.4), ADH (150 kDa), BSA (66 kDa, R_S = 3.55), and carbonic anhydrase (29 kDa, R_S = 2). Recombinant MeCP2 (125 μg) was loaded onto the column pre-equilibrated with buffer GF500. Fractions (0.5 ml) were collected, and 50 μl of each fraction was separated on an 8% SDS-polyacrylamide gel and Coomassie Blue-stained to verify the identity of protein observed from the A_{280} trace. Relative absorbance at 280 nm was plotted as the protein eluted from the Superose 12 column, and the radius was calculated by using an equation derived from the plotted standards. For sucrose gradient sedimentation ~15 μg of rMeCP2, 50 μg of apoferritin (17.7 S), 30 μg of B-amylase (8.9 S), 50 μg of ADH (7.4 S), and 50 μg of BSA (4.3 S) were loaded onto a 13-ml linear 5–20% sucrose gradient made in 0.3 M KCl, 20 mM Hepes (pH 7.9), 2 mM EDTA, 10% glycerol, 10 mM β-mercaptoethanol. The gradient was centrifuged for 19 h at 40,000 rpm in a Beckman SW40 rotor at 4 °C. Fractions (0.5 ml) were taken from the top of the gradient, trichloroacetic acid-precipitated, run on an 8% SDS-polyacrylamide gel, Western-blotted using an anti-MeCP2 antibody or on a 10% SDS-polyacrylamide gel, and Coomassie Blue-stained for the indicated standards. Densitometric analysis utilized Gene Tools Analysis Software package (SynGene), and values were adjusted for background. Intensity was plotted by fraction to determine the relative sedimentation coefficient of rMeCP2.

Molecular Weight and Frictional Coefficient Calculations—Calculations to determine molecular weight and frictional coefficient (f/f_0) were applied as described (31, 32) using Equations 1 and 2,

$$M_r = 6\pi\eta_{20,w}\cdot s_{20,w}\cdot R_S\cdot N/(1 - \rho_{20,w}\nu) \quad \text{(Eq. 1)}$$

$$f/f_0 = 6\pi\eta_{20,w}\cdot R_S/6\pi\eta_{20,w}\cdot(3\nu M_r/4\pi N)^{1/3} \quad \text{(Eq. 2)}$$

where R_S is the Stoke's radius (cm), $s_{20,w}$ is the sedimentation velocity ($S \times 10^{-13}$), $\eta_{20,w}$ is the viscosity of water at 20 °C (0.01002 g·s^{-1} cm^{-1}), N = Avogadro's number (6.022 × 10^{23}·mol^{-1}), $\rho_{20,w}$ is the density of water at 20 °C (0.9981 g·cm^3), ν is the partial specific volume (used 0.725 cm^3/g).

RESULTS

Biochemical Analysis of MeCP2 in Nuclear Extracts—Previous studies have demonstrated that *Xenopus* (21) and mammalian MeCP2 can associate with the Sin3a complex (14), but the properties of the mammalian Sin3a/MeCP2 interaction remain to be investigated. We confirmed that in rat brain nuclear extracts Sin3a is coimmunoprecipitated by antibodies against MeCP2 (Fig. 1*a*), but the amounts are small relative to input. In order to evaluate the relative amount of MeCP2 residing in a stable Sin3a complex, nuclear extracts from either rat brain or NG108 (mouse-rat neural-glial fusion) tissue culture cells were separated by size exclusion chromatography, and the resulting fractions were analyzed by Western blot with antibodies against MeCP2, Sin3a, and HDACs 1 and 2, which are components of the Sin3a complex. To our surprise most of Sin3a-containing fractions were devoid of detectable MeCP2. The majority of Sin3a eluted over the apparent molecular mass range of 500 kDa to 2 mDa, showing significant overlap with HDAC 1/2-containing fractions (Fig. 1*b*). In contrast, MeCP2 eluted with an apparent molecular mass of 400–500 kDa (Fig. 1*b*).

Biochemical Purification of MeCP2 from Rat Brain—The predicted molecular mass of MeCP2 based on its amino acid sequence is between 52.4 and 53 kDa, depending on species. This is much smaller than the apparent molecular weight observed by size exclusion chromatography, indicating that MeCP2 may exist in a multiprotein complex. To test this possibility, a large scale biochemical purification of rat brain MeCP2 was devised (Fig. 2*a*). During the purification, MeCP2, Sin3a, and HDAC 1/2 were tracked by Western blotting. Sp-Sepharose and MonoQ columns efficiently separated MeCP2 from the majority of the Sin3a and HDAC 1/2-containing fractions, confirming that the majority of MeCP2 from nuclear extract is absent from the Sin3a complex. After four purification steps, three polypeptides were detected by SDS-PAGE, of which the middle band was identified as MeCP2 by Western blot (Fig. 2*c*). A final size exclusion chromatography step was applied to fractions eluted from the Ni-NTA column. SDS-PAGE of the resulting fractions (Fig. 2*d*) demonstrated that the apparent molecular weight of MeCP2 (400–500 kDa) remained constant over the purification. The SDS-PAGE analysis also showed a 90-kDa band whose elution profile overlapped with MeCP2. This protein was identified by mass spectrometry (data not shown) as topoisomerase I. We were able to rule out

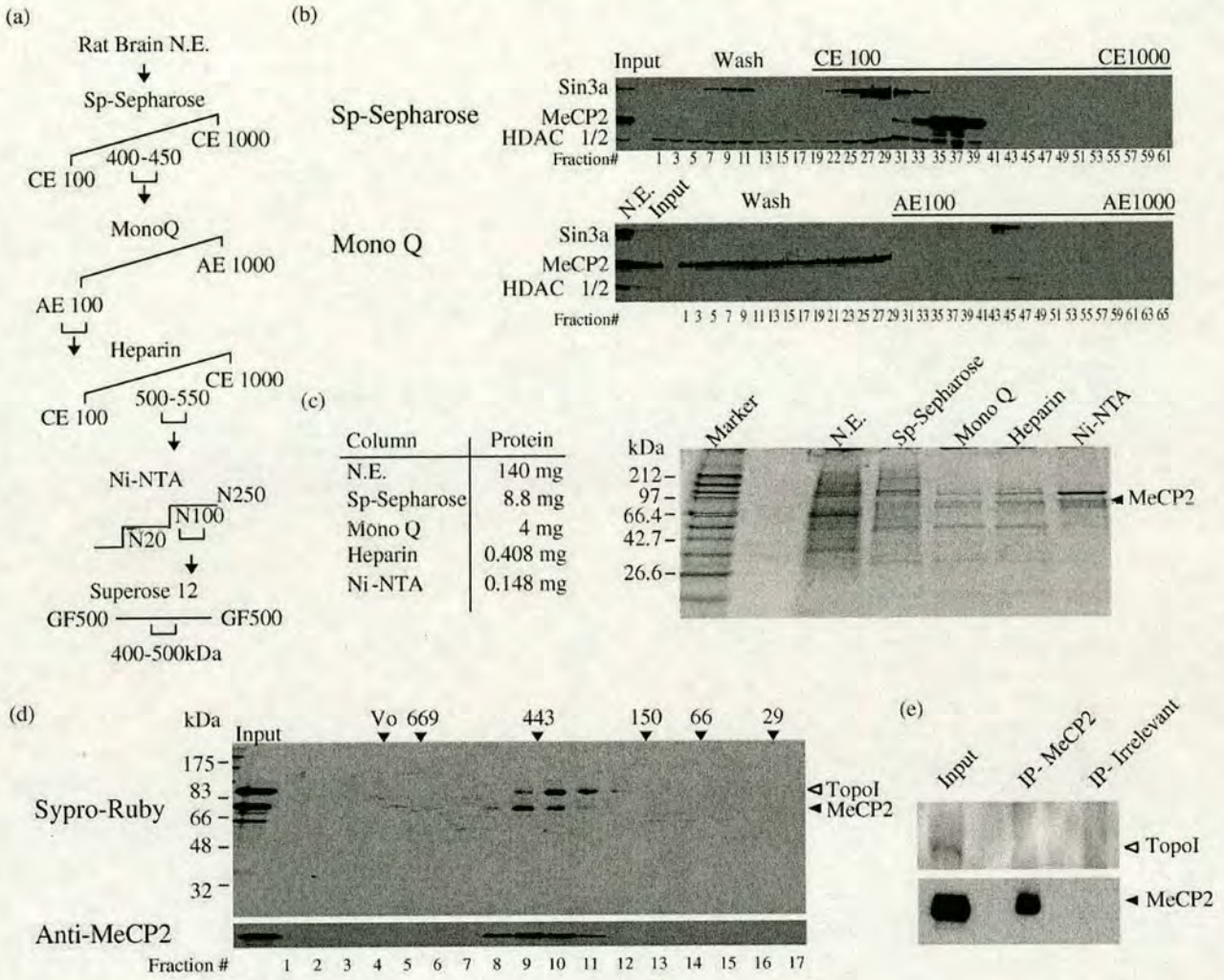


FIG. 2. *a*, MeCP2 was purified from rat brain nuclear extract (*N.E.*) by using the indicated five-step purification scheme (Sp-Sepharose, MonoQ, heparin, Ni-NTA, and Superose 12). *b*, MeCP2 is separated from Sin3a and HDAC 1/2 over the Sp-Sepharose and MonoQ purification steps as demonstrated by Western blot analysis of the collected fractions. The protein separation steps are indicated *above* each Western blot. Fraction numbers are shown *below* each blot. *c*, protein quantities (*left*) and Sypro-Ruby stained SDS-polyacrylamide gel (*right*) of proteins from each purification step are shown. The *arrowhead* indicates bands corresponding to MeCP2. *d*, Superose 12 size exclusion chromatography fractions of Ni-NTA-purified protein in *c* visualized by Sypro-Ruby stain. The *solid arrowhead* indicates the position of MeCP2, and the *open arrowhead* indicates topoisomerase I, as identified by mass spectrometry. *Below* the Sypro-Ruby stained gel are the same fractions analyzed by Western blotting with anti-MeCP2 antibodies. Elution profile of molecular weight standards (kDa) and void volume (V_0) is indicated at the *top* and the fraction number at the *bottom*. *e*, anti-MeCP2 antibodies or an irrelevant antibody (anti-Gal4) does not coimmunoprecipitate (*IP*) topoisomerase I from rat brain nuclear extract.

the possibility that MeCP2 associated with topoisomerase I as MeCP2 antibodies were unable to immunoprecipitate topoisomerase I from rat brain nuclear extract (Fig. 2*e*). Furthermore, an independent purification starting with 10 mg of rat brain extract yielded MeCP2 with the same Superose 12 size exclusion profile but in the absence of detectable topoisomerase I (data not shown). We conclude that MeCP2 in rat brain extracts, despite its large apparent molecular weight, does not exist in a complex with other proteins.

Biochemical Analysis of MeCP2 from X. laevis Oocytes—A previous report (21) suggested that *X. laevis* MeCP2 cofractionated with *X. laevis* Sin3a (xSin3a). To re-visit this observation, we made extract from *X. laevis* oocytes and separated it by Bio-Rex 70 ion exchange chromatography (Fig. 3, *a* and *b*), using the procedures described previously (21). Western blots using antibodies directed against xMeCP2 and xSin3a were used to monitor the elution of these proteins (Fig. 3*b*). Proteins from the Bio-Rex 70 step were separated by size exclusion chromatography, and fractions were analyzed by Western blot (Fig. 3*c*). The xSin3a protein elutes in a complex with an apparent mass of 500 kDa to 2 mDa, whereas xMeCP2, like mammalian MeCP2, elutes with an apparent molecular mass of 400–500 kDa. In agreement with our observations in mammalian extracts (Fig. 1 and Fig. 2) and with an independent study of *Xenopus* oocyte extracts (33), *X. laevis* MeCP2 does not coelute with fractions containing the majority of the xSin3a corepressor complex.

Purification of Recombinant Untagged Human MeCP2—To examine further the discrepancy between the observed molecular weight (400–500 kDa) and the monomeric molecular weight (53 kDa) of MeCP2, we developed a four-step scheme to obtain pure recombinant MeCP2, free of protein tags, from bacterial lysates (rMeCP2) (Fig. 4, *a* and *b*). To ensure that this protein was active and suitable for further biophysical analy-

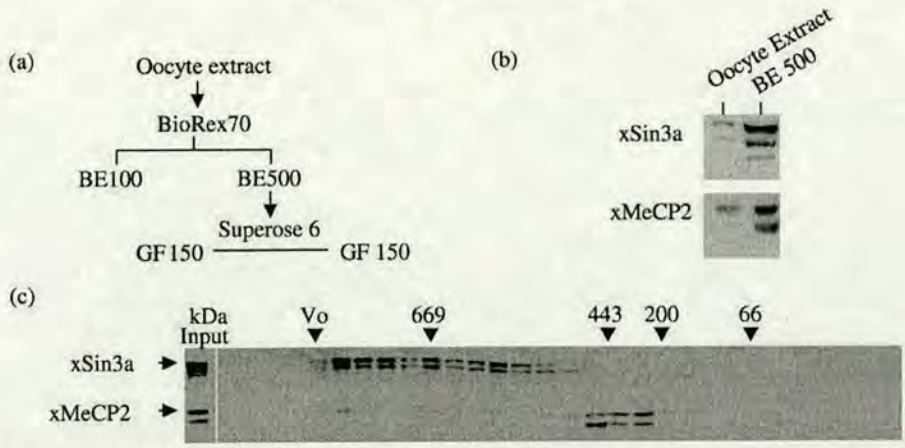


FIG. 3. *a*, *X. laevis* oocyte extract was separated by Bio-Rex 70 ion exchange chromatography followed by size exclusion chromatography on a Superose 6 column. *b*, Western blot analysis of input oocyte extract and Bio-Rex 70 exchange 500 mM NaCl (BE500) elution using xSin3a and xMeCP2 antibodies. The *lower band* in the MeCP2 Western blot is likely a degradation product. *c*, Superose 6 size exclusion chromatography of the BE500 elution Western blotted for xSin3a and xMeCP2. Elution profile of molecular weight standards (kDa) and the void volume (V_0) is indicated *above*.

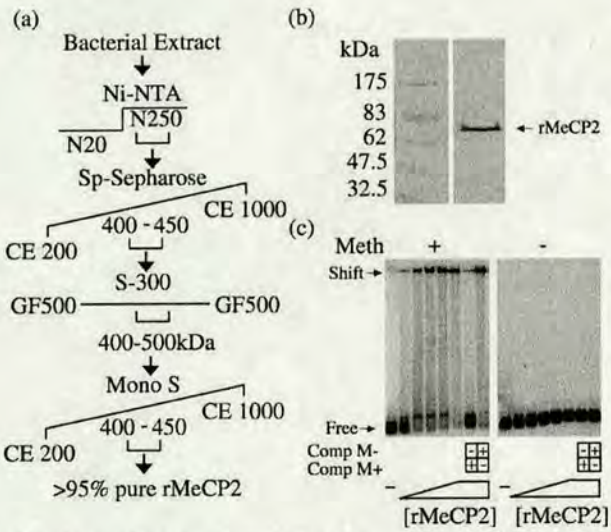


FIG. 4. *a*, a four-step chromatographic purification procedure (Ni-NTA, Sp-Sepharose, Sephacryl-S300, and Mono S) was used to purify untagged recombinant MeCP2 (rMeCP2) from bacterial extracts. *b*, Coomassie Blue stained gel showing a typical purification of rMeCP2 with protein molecular weight marker (*left*) and rMeCP2 (*right*). *c*, rMeCP2 has the capacity to specifically recognize methylated (*Meth*) DNA. Increasing amounts of rMeCP2 (50, 100, 250, 500, and 750 ng) were incubated with methylated (*left*) or unmethylated (*right*) CG11 probe. Cold competitor DNA, either methylated (*Comp M+*) or unmethylated (*Comp M−*) were added to show specificity of bandshift.

sis, we confirmed that untagged rMeCP2 specifically interacted with a methylated probe (Fig. 4*c*). This bandshift was successfully competed by cold methylated competitor but not by unmethylated competitor (Fig. 4*c*). Most importantly, rMeCP2 maintained an apparent molecular mass of 400–500 kDa by size exclusion chromatography (Fig. 5*b*), identical to that seen for native MeCP2.

Hydrodynamic Analysis of rMeCP2—Since MeCP2 does not associate stably with other proteins, we entertained the following three possible explanations for the discrepancy in observed *versus* predicted molecular weight: 1) MeCP2 self-associates to create a homomultimeric complex; 2) MeCP2 interacts with a DNA or RNA component that alters its molecular size; and 3) MeCP2 is a monomeric molecule with an abnormal shape (radius). To address whether MeCP2 might self-associate, chemical cross-linking experiments using glutaraldehyde (data not shown) or EGS were carried out. Although the control protein ADH was cross-linked as expected to give multimers, MeCP2 showed no evidence of multimerization (Fig. 5*a*). A slight reduction in the size of treated MeCP2 was observed, presumably due to compaction of the protein as a result of intramolecular cross-links. To determine whether MeCP2 is associated with a DNA or RNA component, purified rat MeCP2 was treated with Benzonase, which contains both DNase and RNase activity. Nuclease treatment had no effect on the apparent molecular weight as determined by size exclusion chromatography (data not shown).

To explore more thoroughly the biophysical properties and molecular shape of MeCP2, we carried out hydrodynamic analysis. Seigel and Monty (31) demonstrated that the mass of a protein (or protein complex) can be determined accurately by combining size exclusion chromatography and sucrose gradient sedimentation. rMeCP2 was fractionated by size exclusion chromatography, and the peak elution of protein was determined by the absorbance at 280 nm. This value was used to calculate the apparent radius of MeCP2 to ~6.15 nm (Fig 5*b*). To determine the sedimentation coefficient, rMeCP2 was separated on a 5–20% sucrose gradient and an approximate sedimentation coefficient of ~2.28 S (Fig. 5*c*) was observed. By combining the values from our hydrodynamic analysis, we calculated the derived mass of MeCP2 in solution (Table I). The derived mass (57.8 kDa) falls within ~10% of the predicted monomeric molecular weight of MeCP2 (52.4 kDa). Furthermore, the calculated frictional ratio of ~2.41 indicates that rMeCP2 is a highly elongated molecule (34).

DISCUSSION

Mammalian MeCP2 can interact with Sin3a as a recombinant protein *in vitro* or by coimmunoprecipitation from nuclear extracts (14). Also, it is reported that *Xenopus* MeCP2 partially cofractionates with xSin3a, suggesting that the proteins may exist in a stable complex (21). In this study, we have investigated the mammalian MeCP2/Sin3a interaction biochemically, and we found that MeCP2 does not exist in a stable Sin3a complex. Reinvestigation of the biochemical properties of *Xenopus* oocyte MeCP2 led us to conclude that xMeCP2 does not coelute with the majority of xSin3a either. Therefore, neither amphibian nor mammalian MeCP2 appears to form a stable complex with Sin3a in cellular extracts.

Biochemical purification of MeCP2 from rat brain demon-

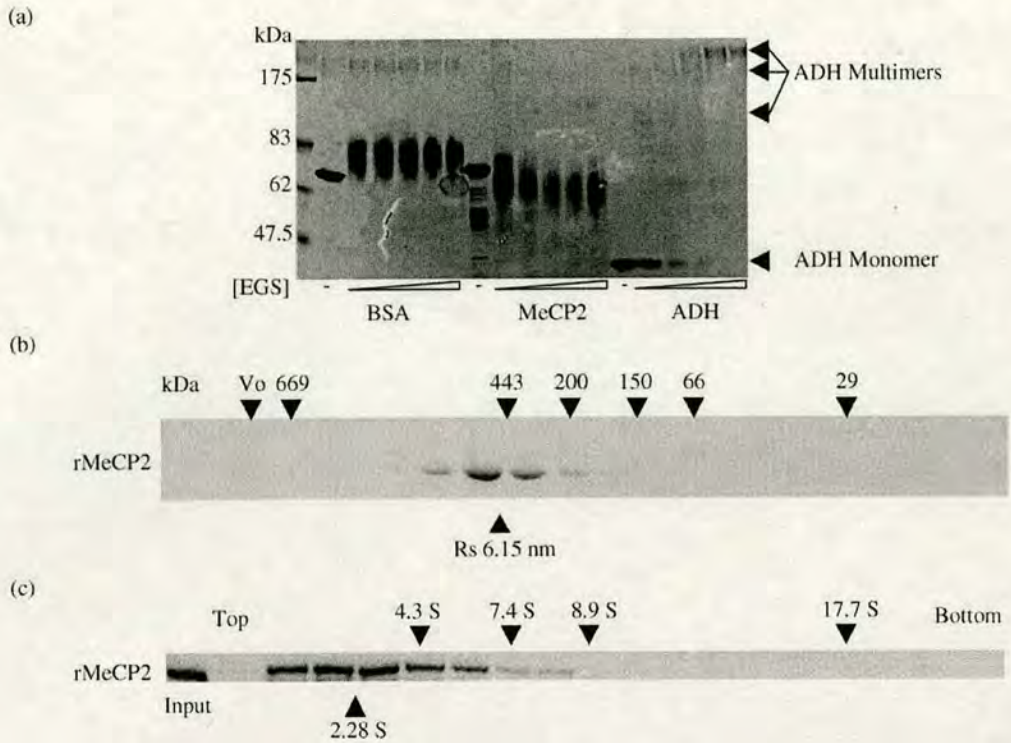


FIG. 5. *a*, BSA (monomer), rMeCP2, and ADH (tetramer) were cross-linked with increasing concentrations of EGS (0.25, 0.5, 1.0, 2.5, 5.0 mM). rMeCP2 did not cross-link, whereas the ADH monomer when cross-linked formed several multimeric species (*arrowheads right*). *b*, rMeCP2 was separated by Superose 12 size exclusion chromatography, and fractions were separated by SDS-PAGE and stained with Coomassie Blue. Elution profile of molecular weight standards (kDa) and of the void volume (V_0) is indicated *above*, and calculated radius in nm *below*. *c*, rMeCP2 was separated by 5–20% sucrose gradient sedimentation, and the fractions were separated by SDS-PAGE and Western-blotted by using anti-MeCP2 antibodies. The sedimentation profile of protein standards are indicated *above*, and the calculated sedimentation value in Svedberg units (*S*) is shown *below*.

TABLE I
Hydrodynamic Properties of rMeCP2

Theoretical mass	Sedimentation coefficient	Stokes radius	Frictional ratio	Derived mass
Da	$s_{20,w}$	*nm*	f/f_0	*Da*
52,437	~2.28	~6.15	2.41	57,880

strated that MeCP2 does not stably associate with other proteins. Nevertheless, purified rat brain MeCP2 had an apparent molecular mass by size exclusion chromatography of 400–500 kDa, which is nearly 10 times its true molecular mass (52 kDa). By generating recombinant untagged MeCP2, we were able to study in more detail the biophysical properties of MeCP2. rMeCP2 maintained the same large apparent molecular weight as the native rat and *X. laevis* proteins. This large apparent molecular weight was not because of an RNA or DNA component or self-association. Based on the experimentally determined radius and sedimentation coefficient, we were able to deduce the native molecular weight of rMeCP2 and determine the frictional ratio, which indicates molecular shape (31, 32). The derived molecular weight fits closely to the theoretical molecular weight of MeCP2, indicating that, despite its aberrant gel filtration properties, MeCP2 exists as a monomeric protein in solution. The frictional ratio of ~2.41 is consistent with an elongated molecule. Globular proteins invariably show frictional ratios of <1.5, whereas elongated proteins have frictional ratios in excess of 2.0 (34). As size exclusion chromatography separates proteins largely based on radius (*i.e.* shape), this finding can account for the large apparent molecular weight of native MeCP2. MeCP2 has 11.5% proline by amino acid composition, and the inherent rigidity of this amino acid (35) may in part give rise to the nonglobular nature of MeCP2. The biological significance of the elongated MeCP2 molecular shape is currently unclear.

Two *bona fide* target genes whose transcriptional activity responds to MeCP2 have been identified: *xHairy2a* (26) and *Bdnf* (23, 24). Sin3a is also localized to each of these promoter regions by chromatin immunoprecipitation analysis[2] (23, 24, 26). These findings provide support for the idea that MeCP2 and Sin3a interact on DNA templates. It is also noteworthy that loss of MeCP2 from the *Bdnf* promoter results in a loss of local histone methylation (23, 24), in agreement with previous evidence that MeCP2 can interact with the histone methyltransferase activity (17). To reconcile these results with the present findings, we propose that MeCP2 enters into a stable association with cofactors that modulate gene expression only when bound to DNA. Given recent evidence for a gene-specific role of MeCP2, it is possible that sustained interactions between MeCP2 and its protein partners rely on the specific DNA sequence architecture of its cognate target promoters. Thus monomeric MeCP2 may act as a multifunctional repressor that recruits Sin3a and perhaps other corepressors in a template-dependent manner. The biochemical properties of MeCP2 support this view. Efficient extraction of MeCP2 from nuclei requires more stringent conditions than does extraction of its methyl-CpG-binding counterparts MBD2 and Kaiso, which are components of large molecular weight chromatin remodeling complexes (30, 36). This biochemical distinction is compatible with the view that methyl-CpG-binding proteins belonging to free multiprotein complexes interact with DNA in a transient manner, whereas MeCP2 remains tightly associated with chromatin and recruits corepressors in its DNA-bound context.

Our findings may have relevance to Rett syndrome, which is

caused by mutations in the *MECP2* gene (19). The absence of an exclusive MeCP2-Sin3a repression complex leaves open the possibility that other protein partners, including those reported previously (14, 17, 27–29) and others as yet unidentified, play an important role in mediating the effects of MeCP2 on gene expression. Given the broad distribution of Rett mutations within the *MECP2* coding sequence, it seems likely that the surface of MeCP2 is capable of multiple intermolecular interactions that are of relevance to neuronal function. Understanding these interactions at clinically relevant MeCP2 target genes is a challenge for the future.

Acknowledgments—NG108 cells and BL21 DE3 codon plus bacteria were kindly provided by Rod Bremner and Robin Allshire. xMeCP2 and xSin3a antibodies were a gift from Peter L. Jones. We thank Skirmantas Kriaucionis, Emma Turnbull, and Helle Jørgensen for critical reading of the manuscript and the Bird laboratory for helpful discussion. We also thank Karen Wilson for technical assistance. We would especially like to thank Jim Hunt for discussion about biochemical properties of MeCP2 and the useful experiments he suggested.

Note Added in Proof—Since publication of this paper, we have become aware of an earlier study (von Kries *et al.*, 1994) reporting that the chicken protein ARBP has an elongated shape based on its Stokes radius. ARBP was subsequently found to be chicken MeCP2 (Weitzel *et al.*, 1997). The conclusion of von Kries and colleagues is confirmed by this aspect of our study and should have been cited (von Kries, J. P., Rosorius, O., Buhrmester, H., and Stratling, W. H. (1994) *FEBS Lett.* **342,** 185–188 and Weitzel, J. M., Buhrmester, H., and Stratling, W. H. (1997) *Mol. Cell. Biol.* **17,** 5656–5666).

REFERENCES

1. Li, E., Bestor, T. H., and Jaenisch, R. (1992) *Cell* **69,** 915–926
2. Bird, A. (2002) *Genes Dev.* **16,** 6–21
3. Hendrich, B., and Bird, A. (1998) *Mol. Cell. Biol.* **18,** 6538–6547
4. Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001) *Genes Dev.* **15,** 1613–1618
5. Wade, P. A. (2001) *Oncogene* **20,** 3166–3173
6. Ng, H. H., Jeppesen, P., and Bird, A. (2000) *Mol. Cell. Biol.* **20,** 1394–1406
7. Cross, S. H., Meehan, R. R., Nan, X., and Bird, A. (1997) *Nat. Genet.* **16,** 256–259
8. Jorgensen, H. F., Ben-Porath, I., and Bird, A. P. (2004) *Mol. Cell. Biol.* **24,** 3387–3395
9. Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., and Renkawitz, R. (2000) *J. Biol. Chem.* **275,** 34963–34967
10. Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) *Nat. Genet.* **23,** 58–61
11. Hutchins, A. S., Mullen, A. C., Lee, H. W., Sykes, K. J., High, F. A., Hendrich, B. D., Bird, A. P., and Reiner, S. L. (2002) *Mol. Cell* **10,** 81–91
12. Feng, Q., and Zhang, Y. (2001) *Genes Dev.* **15,** 827–832
13. Nan, X., Campoy, F. J., and Bird, A. (1997) *Cell* **88,** 471–481
14. Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) *Nature* **393,** 386–389
15. Wade, P. A., Gegonne, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999) *Nat. Genet.* **23,** 62–66
16. Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003) *J. Biol. Chem.* **278,** 24132–24138
17. Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P., and Kouzarides, T. (2003) *J. Biol. Chem.* **278,** 4035–4040
18. Yoon, H. G., Chan, D. W., Reynolds, A. B., Qin, J., and Wong, J. (2003) *Mol. Cell* **12,** 723–734
19. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999) *Nat. Genet.* **23,** 185–188
20. Kriaucionis, S., and Bird, A. (2003) *Hum. Mol. Genet.* **12,** R221–R227
21. Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998) *Nat. Genet.* **19,** 187–191
22. Jones, P. L., Wade, P. A., and Wolffe, A. P. (2001) *Methods Mol. Biol.* **181,** 297–307
23. Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003) *Science* **302,** 885–889
24. Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y. E. (2003) *Science* **302,** 890–893
25. Klose, R., and Bird, A. (2003) *Science* **302,** 793–795
26. Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H., and Meehan, R. R. (2003) *Mol. Cell* **12,** 425–435
27. Kimura, H., and Shiota, K. (2003) *J. Biol. Chem.* **278,** 4806–4812
28. Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001) *J. Biol. Chem.* **276,** 34115–34121
29. Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S. H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002) *Science* **298,** 1747–1752
30. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989) *Cell* **58,** 499–507
31. Siegel, L. M., and Monty, K. J. (1966) *Biochim. Biophys. Acta* **112,** 346–362
32. Kolb, S. J., Hudmon, A., Ginsberg, T. R., and Waxham, M. N. (1998) *J. Biol. Chem.* **273,** 31555–31564
33. Ryan, J., Llinas, A. J., White, D. A., Turner, B. M., and Sommerville, J. (1999) *J. Cell Sci.* **112,** 2441–2452
34. Sober, H. A. (1973) in *Handbook of Biochemistry* (Sober, H. A., ed) 2nd Ed., pp. c10–c12, Chemical Rubber Co., Cleveland, OH
35. George, R. A., and Heringa, J. (2002) *Protein Eng.* **15,** 871–879
36. Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992) *Nucleic Acids Res.* **20,** 5085–5092