VIRGINIA COMMONWEALTH UNIVERSITY

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2004

Purification and Characterization of a Methyl-DNA Binding Protein Complex from Primary Erythroid Cells

Evan Paul Kransdorf
Virginia Commonwealth University

Follow this and additional works at: http://scholarscompass.vcu.edu/etd

 Part of the Medicine and Health Sciences Commons

© The Author

Downloaded from
http://scholarscompass.vcu.edu/etd/1026

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

PURIFICATION AND CHARACTERIZATION OF A METHYL-DNA BINDING PROTEIN COMPLEX FROM PRIMARY ERYTHROID CELLS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Evan Paul Kransdorf
B.S., University of Virginia, Charlottesville, Virginia, 1999

Director: Gordon D. Ginder
Professor
Departments of Internal Medicine, Human Genetics, and Microbiology and Immunology

Virginia Commonwealth University
Richmond, Virginia
December, 2004

Acknowledgements

First and foremost I would like to thank my advisor, Dr. Gordon Ginder. He is a man of great intelligence, warmth and humor and it has been a great honor to work with him. I am greatly indebted to Shou Zhen Wang and Sheng Zu Zhu for their advice, support and friendship. Shou Zhen was especially helpful to me as she cheerfully assisted me in performing the large-scale chicken erythrocyte nuclear extract preparations.

My fellow graduate students Jeremy Rupon and Quintesia Grant were excellent companions and I thank them for their friendship and support. Tanya Sargent, David Barrett and Jing Wang answered my many questions during the early stages of my project.

Several people outside of Ginder lab were also very helpful to me during the course of my project. I would like to sincerely thank Dr. Dan Conrad who was very giving of his time and effort in performing cell sorting. Dr. Shirley Taylor offered me advice on strategies for protein and antibody production for chicken MBD2.

Dedication

This work is dedicated to my loving family: Kaity, Mark Jay, Judy and Pam.

Table of Contents

List of Tables

List of Figures

Abbreviations

ATP	adenosine triphosphate
BAF60b	BRG1-associated factor 60 kD – b isoform
BBSRC	Biotechnology and Biological Sciences Research Council
BLAST	Basic Local Alignment Search Tool
bp	base pairs
°C	degrees celcius
cDNA	DNA complementary to mRNA
cHS4	chicken beta-globin locus hypersensitive site 4 core region
CEF	chicken embryonic fibroblast
ChIP	chromatin immunoprecipitation
CIP	calf intestinal phosphatase
cMBD2	chicken MBD2
CMV	cytomegalovirus
contig	"contiguous" DNA sequence produced by alignment of overlapping subfragments
CpG	cytosine-guanine dinucleotide
^{m}CpG	5-methylcytosine–guanine dinucleotide
CV	column volume
dATP	2'-deoxyadenosite 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
DNMT1	DNA methyltransferase 1
DNMT3	DNA methyltransferase 3
DTT	dithiothreitol
ECGA	Ensembl Chicken Genome Annotation
EDTA	ethylenediaminetetraacetic acid
eIF	eukaryotic translation initiation factor
EMSA	electrophoretic mobility shift assay
EST	expressed sequence tag
FACS	fluorescence activated cell sorting
FITC	fluorescein isothiocyanate
FMIP	Fms-interacting protein
FPLC	fine pressure liquid chromatography
FRT	FLP recombination target
GFP	green fluorescent protein

GM-CSF	granulocyte macrophage-colony stimulating factor
GR_{11}	glycine-arginine repeat domain of MBD2
GST	glutathione S-transferase
H3-K4-Me_2	dimethylation of lysine 4 of histone H3
H3-K4-Me_3	trimethylation of lysine 4 of histone H3
H3-K9-Me_2	dimethylation of lysine 9 of histone H3
Hb	hemoglobin
HDAC1	histone deacetylase 1
HDAC2	histone deacetylase 2
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HSS	DNase-I hypersensitive site
IgG	immunoglobulin G
IL-4	interleukin-4
IP	immunoprecipitation
ISWI	imitation switch
KCl	potassium chloride
kD	kilodalton
LCR	locus control region
LDS	lithium dodecyl sulfate
LiCl	lithium chloride
β-ME	beta-mercaptoethanol
M-ρ248	methylated ρ-globin exon 1/intron 1 fragment
M-CG11	methylated CG11 probe
mRNA	messanger ribonucleic acid
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
MBD1	methyl-CpG binding domain protein 1
MBD2	methyl-CpG binding domain protein 2
MBD3	methyl-CpG binding domain protein 3
MBD4	methyl-CpG binding domain protein 4
MCBP	methyl-CpG binding proteins
MeCP1	methyl-CpG binding protein complex 1
MeCP2	methyl-CpG binding protein 2
MeCPC	erythroid methyl-cytosine binding protein complex
MEL	murine erythroleukemia
MENT	mature erythrocyte nuclear termination stage-specific protein
mg	milligram
mL	milliliter
MLL	mixed lineage leukemia
MLys	*Micrococcus lysodeikticus*
MOPS	3-(N-morpholino)propanesulfonic acid
MTA1	metastatsis-associated 1
MTA2	metastatsis-associated 2
MWCO	molecular weight cut off
NaCl	sodium chloride

NuRD	Nucleosome Remodeling and Deacetylase complex
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMF	peptide mass fingerprint
PMSF	phenylmethylsulfonyl fluoride
PTR	proximal transcribed region
Q	quaternary ammonium
RbAp46	retinoblastoma-associated protein 46 kD
RbAp48	retinoblastoma-associated protein 48 kD
RBC	red blood cell
rcMBD2	recombinant GST/cMBD2 fusion protein
RMCE	recombinase-mediated cassette exchange
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RPM	revolutions per minute
RT-PCR	real time-polymerase chain reaction
SA1	stromal antigen 1 protein
SA2	stromal antigen 2 protein
SAP18	sin3a-associated protein 18 kD
SAP30	sin3a-associated protein 30 kD
SCC1	sister-chromatid cohesion 1
SCD	sickle cell disease
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
siRNA	small interfering ribonucleic acid
SMC1	structural maintenance of chromosomes 1 protein
SMC3	structural maintenance of chromosomes 3 protein
SNF2h	sucrose nonfermenting-2 homolog
SP	sulfopropyl
SWI/SNF	switch/sucrose non-fermenting
TBE	tris-borate-EDTA buffer
T_{h2}	T-helper lymphocyte, subset 2
tRNA	transfer ribonucleic acid
UD	University of Deleware
µg	microgram
µL	microliter

Abstract

PURIFICATION AND CHARACTERIZATION OF A METHYL-DNA BINDING PROTEIN COMPLEX FROM PRIMARY ERYTHROID CELLS

By Evan Paul Kransdorf

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2004

Director: Gordon D. Ginder, M.D., Professor, Departments of Internal Medicine, Human Genetics and Microbiology and Immunology

The chicken embryonic β-type globin gene, ρ, is silenced on day five of embryogenesis. Concomitant with this silencing is methylation of cytosine residues in the promoter and proximal transcribed region of the gene, which is first detected on day seven and is complete in adult cells. Once methylated, expression of the gene cannot be induced unless the methylation is removed by treatment of cells with 5-azacytidine. Therefore ρ-globin is a member of a small group of genes whose normal developmentally regulated expression is mediated at least in part by DNA methylation.

A methyl-DNA binding complex, termed the MeCPC (Erythroid Methyl Cytosine-binding Protein Complex), has been found to bind to the methylated, but not unmethylated, ρ-globin promoter and proximal transcribed region in nuclear extracts from definitive erythrocytes. This complex has a stronger binding affinity for its cognate

binding sequence, the methylated ρ-globin proximal transcribed region (M-ρ248), than for an artifical 5-methylcytosine-rich sequence (M-CG11).

To define the components of the MeCPC, we developed two chromatographic procedures to purify the complex from adult chicken red blood cell nuclear extracts (Purification Strategies I and II). Mass spectrometry was performed on the MeCPC obtained by Purification Strategy I and proteins were identified by a novel application of peptide mass fingerprint data fitting. Four components of the previously-purified MeCP1 transcriptional repression complex were identified in the sample: MBD2, RbAp48, HDAC2 and MTA1. Another identified protein, MENT, is a factor expressed only in chicken hematopoietic cells. These five proteins, as well as the MeCP1 component Mi2, were found to tightly coelute by Western blotting of gel-filtration fractions from Purification Strategy II. Therefore, we conclude that these five proteins are components of the MeCPC.

To confirm that MBD2 is associated with the ρ-globin gene *in vivo*, we perfomed the chromatin immunoprecipitation assay using anti-MBD2 antibodies. In adult erythrocytes, significant enrichment for MBD2 is seen at the transcriptionally inactive ρ-globin gene, but no enrichment is observed at the transcriptionally active β^A-globin gene. These experiments confirm that MBD2 binds to the methylated ρ-globin gene in adult chicken erythroid cells.

Chapter 1: Introduction

DNA Methylation

Epigenetic phenomena are biological processes that cause heritable changes in gene expression without affecting the sequence of the gene itself. Several epigenetic phenomena are known to exist within higher eukaryotes: DNA methylation, histone coding, and RNA interference. Of these phenomena, DNA methylation is by the far the most comprehensively documented, having first been shown to be associated with a transcriptionally silent, protein-coding eukaryotic gene in 1979 (McGhee and Ginder, 1979). The epigenetic phenomena interact both functionally and mechanistically to alternatively restrict or permit gene expression such that the genome sequence, which is identical in almost all cells of an organism, is transcribed in a pattern characteristic for each cell type.

Methylation of the 5-position of cytosine residues in DNA has an important role in the regulation of gene expression in higher eukaryotes. Although the total contribution of DNA methylation to the biology of the cell is not yet known, numerous examples support a role for DNA methylation in maintaining genes in a transcriptionally inactive state. Furthermore, by permanently silencing expression of particular genes in a cell-type and temporally restricted pattern, DNA methylation contributes to maintenance of cell

differentiation and tissue-restricted gene expression and as well as to oncogenesis (Hutchins et al., 2002; Futscher et al., 2002; Jones and Baylin, 2002).

Cytosine is the only base known to be methylated in eukaryotic DNA. Methylation occurs almost exclusively on cytosines in the context of CpG dinucleotides, although low frequencies of methylation of cytosines in CpA dinucleotides have been reported in stem cells (Ramsahoye et al., 2000; Dodge et al., 2002). The 5-methylcytosine is itself extremely mutagenic, with a recent survey finding that deamination of 5-methylcytosine to T at CpG sites accounts for more that 20% of all base changes that result in genetic disease (Krawczak et al., 1998). As a result of this instability, the CpG dinucleotide has been selectively eliminated from the genome of vertebrate organisms through spontaneous deamination to thymine, to a level that is 20% of what would be expected based on the base composition of DNA (Bird, 1980).

Regions of the genome that have maintained the expected density of CpG dinucleotides have been termed "CpG islands." These regions were originally defined by Gardiner-Garden and Frommer as genetic regions that range from 200 bases to several kilobases in length, with a G + C content of greater than 50%, and having a ratio of CpG to GpC greater than 0.6 (Gardiner-Garden and Frommer, 1987). It is believed that CpG islands represent regions of the genome that are normally unmethylated in the germline, and as such over evolutionary time have not been subject to elimination of CpGs. Estimates of the total number of CpG islands in the human genome and the number of CpG islands overlapping with gene promoters vary, but in an initial analysis of the human genome sequence, Venter *et al.* reported that 42% of first exons overlap with a

CpG island (Venter et al., 2001). While the CpG dinucleotides in an island remain unmethylated, most (approximately 80%) of CpGs in the genome outside an island are methylated (Bird, 1995).

Ponger and colleagues performed a detailed analysis of the characteristics of CpG islands within the human genome and found that 90% of housekeeping genes (defined as genes active in 17 of 24 tissue types) have a CpG island overlapping the promoter (Ponger et al., 2001). This association is not dependent on the CG-density of the chromosomal isochore in which the gene is embedded. In contrast, the percentage of tissue-specific genes (defined as genes active in 1 of 24 tissue types) that have a CpG island overlapping the promoter depends on the CG-density of the isochore the gene is embedded in. Only 25% of tissue-specific genes embedded in CG-poor isochores have CpG islands, whereas 64% of tissue-specific genes embedded in CG-rich isochores have CpG islands. The implications of these findings with respect to the structure and function of the human and chicken β-globin loci is discussed below.

Promoter CpG islands are of particular interest because of their role in gene regulation. An inverse correlation between methylation of a promoter CpG island and the transcriptional activity of the corresponding gene has been documented at numerous loci. The first example of this relationship was described by McGhee and Ginder for the chicken adult β-globin gene (McGhee and Ginder, 1979). Other examples include methylation of the $p15^{Ink4b}$ tumor suppressor in leukemia, the tissue specific methylation of the *serpinb5* gene, and methylation of the ρ-globin gene during ontogeny in the chicken (Cameron et al., 1999; Futscher et al., 2002; Singal et al., 1997). Numerous *in*

vivo studies have confirmed that DNA methylation of 5'-CpG islands inhibits transcription of the corresponding gene (Siegfried et al., 1999; Singal et al., 2002; Curradi et al., 2002; Stirzaker et al., 2004). This transcriptional repression is absolute, in that cells must be treated with the DNA methylation inhibitor 5-azacytidine to eliminate the methylation, before transcription of the gene can occur.

The Methyltransferase Machinery

Three functional DNA methyltransferase enzymes have been identified in mice and man. The first enzyme to be identified, DNA Methyltransferase-1 (DNMT1), has been dubbed the "maintenance" methyltransferase. DNMT1 has an N-terminal domain that targets it to the replication fork during S-phase and enzymatically displays a higher preference for methylation of hemi-methylated DNA than for unmethylated DNA (Leonhardt et al., 1992). Two other genes, DNMT3a and 3b, are the "*de novo* methyltransferases," displaying a higher enzymatic preference for unmethylated DNA than DNMT1 (Yokochi and Robertson, 2002). The actual *in vivo* situation is not so distinct, with DNMT1 cooperating with DNMT3b to silence genes in human cancer cells (Rhee et al., 2002). Gene targeting experiments in mice have shown that deletion of any of the three known DNA methyltransferase enzymes is lethal to the developing embryo (Li et al., 1992; Okano et al., 1999).

A third species of DNA methyltransferase, DNMT2, has been identified in mice and humans (Yoder and Bestor, 1998; Okano et al., 1998). The role of this enzyme has

not been clearly determined, and it has only recently has it been shown to possess methyltransferase activity (Hermann et al., 2003).

Transcriptional Repression Complexes

Extensive studies in yeast and mammalian cells have shown that RNA polymerase II (RNAPII), the major producer of mRNA in eukaryotic cells, exists in multiple complexes within the cell. Several of these complexes have been found to contain chromatin modifying enzymes, such as the histone acetyltransferases p300 and PCAF (Cho et al., 1998; Wittschieben et al., 1999). Acetylation of the N-terminal tails of histones H3 and H4 by these enzymes is thought to contribute to the initiation and elongation phases of transcription, by creating a more open chromatin configuration (Struhl, 1998).

Similarly, numerous transcriptional repression complexes have been identified that contain histone deacetylase activities. Two of these complexes, the mSin3a and NuRD complexes, have been shown biochemically to be connected to DNA methylation. The mSin3a protein was first identified as the mammalian homolog of a yeast protein, SIN3, which is required for transcriptional repression of the yeast HO gene (Sternberg et al., 1987). Immunopurification of mSin3a from HeLa cells identified several associated proteins: Histone Deacetylase-1 and -2 (HDAC1 and HDAC2), Retinoblastoma Associated Proteins of 46kD and 48kD (RbAp46 and RbAp48), as well as Sin3-Associated Proteins of 18kD and 30kD (SAP18 and SAP30) (Zhang et al., 1997).

The Nucleosome Remodeling and Deacetylase (NuRD) complex was originally purified on the basis of its histone deacetylase and nucleosome disruption activity (Tong et al., 1998; Xue et al., 1998). This complex was found to share components with and possess distinct components from the mSin3a complex. The mSin3a and NuRD complexes share a set of factors that are thought to form a "histone deacetylase core" complex: RbAp46, RbAp48, HDAC1 and HDAC2. In particular, the NuRD complex contains: a) Mi2, an ATP-dependent nucleosome remodeling factor b) Metastasis-Associated-2 (MTA2), a transcriptional repressor and c) Methyl-CpG Binding Domain Protein-3 (MBD3), a methyl-CpG binding domain protein (Wang and Zhang, 2001; Zhang et al., 1999). Figure 1 is a schematic, showing the overlapping components of the major transcriptional repression complexes.

Further functional complexity is achieved in the cell by association of these multi-protein complexes. While purifying an Imitation Switch-class (ISWI) nucleosome-remodeling protein known as SNF2h, Hakimi *et al.* found this protein associated with the NuRD chromatin-remodeling complex as well as with the cohesin complex (Hakimi et al., 2002). The cohesin complex is a set of four proteins: Sister Chromatid Cohesion-1 (SCC1), Structural Maintenance of Chromosomes-1 (SMC1), Structural Maintenance of Chromosomes-3 (SMC3), and Stromal Antigen-1 or -2 (SA1 or SA2), that is required for sister chromatid cohesion (Losada et al., 1998; Losada et al., 2000; Sonoda et al., 2001). Interestingly, the SNF2h/NuRD/cohesin complex was found to bind to a subset of human Alu elements that possess dimethylation at lysine K4 of histone H3 (H3-K4-Me_2), and bind to an enlarged subset of Alu elements after treatment of the cells with the DNA

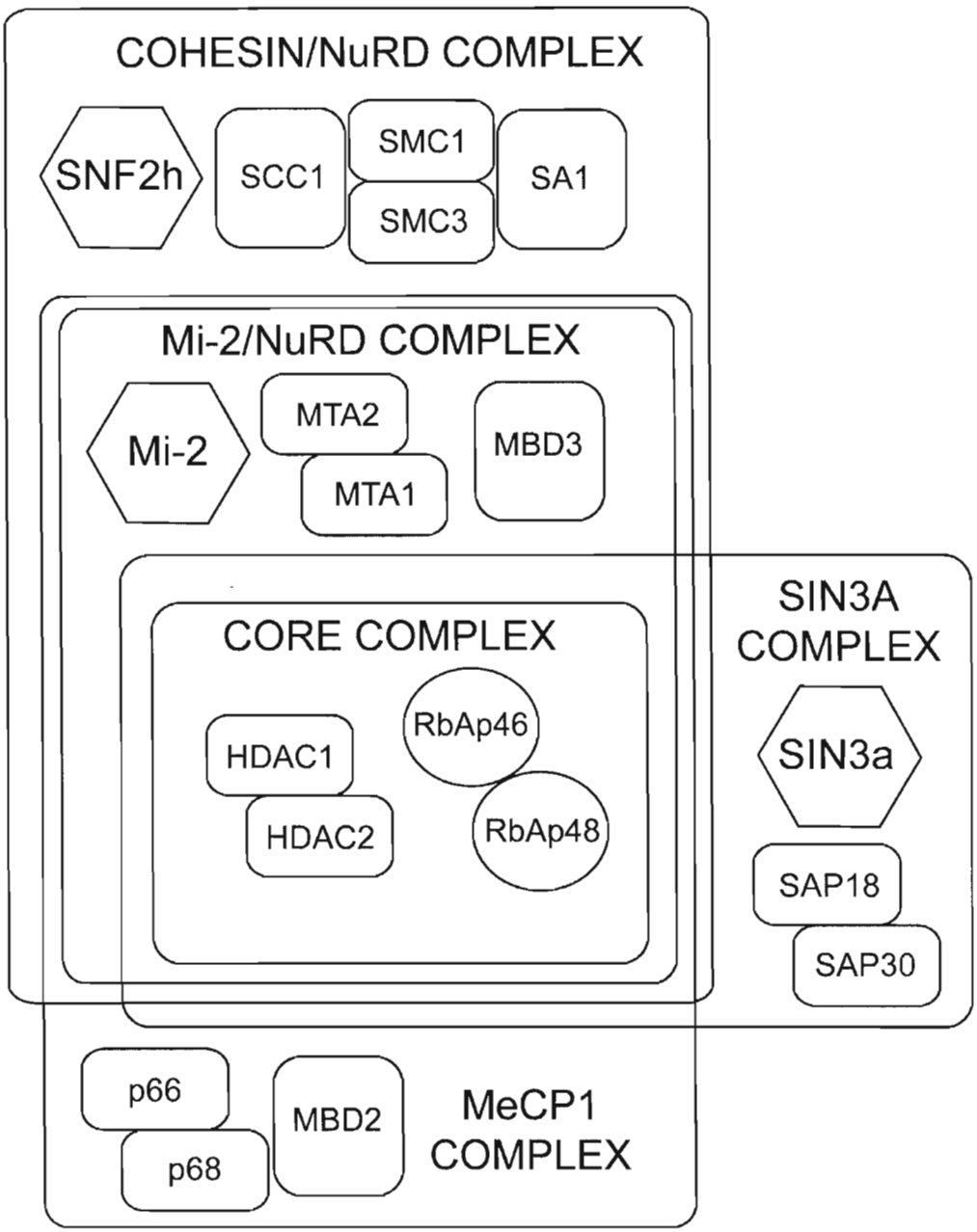


Figure 1. Schematic showing the overlapping components of the major transcriptional repression complexes. The chromatin remodeling proteins are shown as large hexagons since these proteins are thought to contribute the major enzymatic activity of each complex.

methyltransferase inhibitor 5-azacytidine. This latter finding shows that this supercomplex binds to unmethylated sites because the number of binding sites increased with genomic demethylation. This in turn implies that this complex may be involved in maintenance of transcriptional activity. Indeed, both the Sin3a and NuRD complexes have been shown to be part of a multiprotein supercomplex containing Mixed Lineage Leukemia (MLL) that binds to the actively transcribed *HoxA9* gene (Nakamura et al., 2002). These observations show that the mSin3a and NuRD complexes can exist in a transcriptional supercomplex which can mediate both transcriptional repression and activation.

The Methyl-DNA Binding Domain Proteins

Because DNA methylation has a demonstrated causative role in transcriptional repression and can be targeted to developmentally regulated genes and tumor suppressor genes, intense study has been directed towards elucidating the mechanisms of DNA methylation. The earliest theory postulated that the transcriptional repression was mediated by a direct inhibition of transcription factor binding via methylation of cytosines within the factor's binding site (Doerfler, 1983). Although several methylation-sensitive transcription factors are known, only a small number of definitive *in vivo* examples of this mechanism exist (Robertson et al., 1995b; Takizawa et al., 2001). The discovery of a family of proteins that specifically recognize methylated DNA, the methyl-CpG binding proteins (MCBP), has led to the observation of these proteins as the mediators of DNA methylation in a variety of gene systems (Hutchins et

al., 2002; Chen et al., 2003; Sarraf and Stancheva, 2004; Magdinier and Wolffe, 2001; Lin and Nelson, 2003). Thus, it is generally accepted that in most cases the 5-methylcytosine "signal" is read indirectly by MCBP, which then affect transcription.

There are six well-characterized MCBP in higher eukaryotes. Five of these proteins contain a distinct domain, known as the Methyl-CpG Binding Domain (MBD), which is necessary and sufficient for binding of the proteins to methylated-DNA (Nan et al., 1993). These proteins are MECP2, MBD1, MBD2, MBD3 and MBD4. Other MBD-containing proteins have been identified in eukaryotes but their significance and functions are not yet known (Roloff et al., 2003; Hung and Shen, 2003). The sixth protein, Kaiso, is able to bind to methylated DNA via its zinc-finger domains and lacks a true MBD (Daniel et al., 2002). Because these six proteins are all able to bind specifically to methylated DNA, we have termed this family the methyl-CpG binding proteins (MCBP). Previous to the discovery of Kaiso, the extant MCBP all contained a MBD and were thus known as the methyl-CpG binding domain proteins (MBDP). We prefer the previous since all the members of this family have been shown to have a role in methylation-mediated transcriptional repression. A summary of the properties of the MCBP can be found in Table 1.

The MeCP1 complex was the first methyl-CpG specific binding activity to be described and was identified by electrophoretic mobility shift assays (EMSA) using nuclear extracts incubated with an artificial ^{m}CpG-rich probe (M-CG11) (Meehan et al., 1989). The complete purification of MeCP1 confirmed that MBD2 was the methyl-CpG binding component of MeCP1 (Feng and Zhang, 2001). The other components of this

Table 1. Properties of the methyl-CpG binding proteins (MCBP) in mammals. Protein size is given for the murine isoform listed in NCBI RefSeq. Preference of MCBP for methylated/unmethylated DNA was taken from Fraga *et al.* {Fraga, 2003 28 /id}. ND: not determined.

Protein	Number of Known Isoforms	Protein Size (kD)	Function	Associated Proteins	Methyl/Unmethyl Preference
MeCP2	2	52	transcriptional repression	Sin3a complex	2.7
MBD1	5	70	transcriptional repression	SETDB1/CAF-1	ND[3]
MBD2	2	29, 44	transcriptional repression	Mi2/NuRD complex	69.8
MBD3	2	32	transcriptional repression	Mi2/NuRD complex	1.1
MBD4	1	63	DNA repair	MLH1	ND
Kaiso	1	74	transcriptional repression	N-CoR complex	ND

complex were the complete subunits of the Nucleosome Remodeling and Deacetylation (NuRD) complex. Using purified MeCP1, it was shown that this complex can bind to both methylated naked-DNA and methylated nucleosomes. Based on its NuRD functionalities, the complex can remodel methylated nucleosomes and deacetylate histones (Feng and Zhang, 2001). The purified MeCP1 also contained two units (one of which was post-translationally modified) of a novel zinc-finger protein, termed p66 (Feng et al., 2002). p66 was found to be a transcriptional repressor in transfection assays and to interact specifically with MBD3.

Biochemical studies have shown that the MCBP are members of distinct and mostly non-overlapping transcriptional repression complexes. MeCP2 was the first MCBP to be associated with a transcriptional repression complex. Two groups found that MeCP2 associated with the Sin3a histone deacetylase complex (Nan et al., 1998; Jones et al., 1998), although a recent study by Klose and Bird failed to find a stable association of MeCP2 with the Sin3a complex in the absence of a target gene (Klose and Bird, 2004). MBD1 has recently been identified as a critical component of an S-phase specific complex that propagates the DNA methylation signal into a dimethylation of lysine 9 of histone H3 (H3-K9-Me_2) signal during DNA replication (Sarraf and Stancheva, 2004). MBD2 is the methyl-CpG binding component of the MeCP1 transcriptional repression complex (Feng and Zhang, 2001). MBD3 is a core component of the NuRD transcriptional repression complex that can be recruited by DNA-binding proteins such as ikaros and MBD2 (Kim et al., 1999; Zhang et al., 1999). Kaiso has been found to be a component of the N-CoR transcriptional repression complex (Yoon et al.,

2003). MBD4, however, does not have a role in gene regulation. MBD4 functions as a mismatch-specific T/U DNA glycosylase, recognizing the product of spontaneous deamination of a ^{m}CpG dinucleotide (TpG) and excising it (Hendrich et al., 1999).

The elucidation of the distinct biochemistry of the MCBP has been accompanied by progress in uncovering which genes and genetic elements are regulated by each protein. Currently, the *in vivo* roles of MBD2 and MeCP2 are the most well-defined. Shortly after the discovery of MBD2, Magdinier and Wolffe showed that MBD2 was recruited to methylated tumor suppressor genes in cancer cell lines (Magdinier and Wolffe, 2001). This finding has been demonstrated by other groups (Bakker et al., 2002; Lin and Nelson, 2003; Stirzaker et al., 2004) and *in vivo*, MBD2 is essential for efficient tumorigenesis in the murine *Apc*$^{Min/+}$ (Min) model of intestinal tumorigenesis (Sansom et al., 2003). An important role for MBD2 in regulating expression of IL-4 in T_{h2} cells has also been documented (Hutchins et al., 2002).

Data on the *in vivo* role of MeCP2 has also begun to accumulate. Although MeCP2 has been shown to be recruited to methylated tumor suppressor genes in cancer cell lines (Nguyen et al., 2001; El Osta et al., 2002), the most important biological role for MeCP2 is in the brain. Mutations in the *MECP2* gene result in the developmental disorder Rett Syndrome in both humans and mice (Amir et al., 1999; Guy et al., 2001). The *MECP2* gene is located on the X-chromosome, and thus the disease is inherited in a X-linked dominant fashion (D'Esposito et al., 1996). After approximately six months of normal perinatal development, affected children undergo a progressive loss of intellectual and motor skills and develop stereotypic hand movements (Weaving et al., 2005). The

mechanistic basis of Rett Syndrome has yet to be clearly elucidated. Despite this fact, several MeCP2 target genes within the brain have recently been identified. Two papers have shown an important role for MeCP2 in regulation of the *Bdnf* gene (Chen et al., 2003; Martinowich et al., 2003). Furthermore, Horike and colleagues identified the homoeobox gene *Dlx5* as a MeCP2 target (Horike et al., 2005). It remains unclear how dysregulation of these genes leads to the phenotypic features of Rett Syndrome. A more comprehensive understanding of the cadre of genes regulated by each MCBP is necessary if manipulation of MCBP is to be used for therapeutic benefit.

The Chicken Beta-Globin Locus

Vertebrate organisms produce red blood cells in two distinct waves during their lifetime. The first wave of cells is of primitive red blood cells that are produced during the embryonic period. These primitive cells are large, nucleated cells produced in the yolk sac. They exclusively express embryonic globin genes. The second wave of cells is of definitive red cells produced from the fetal liver and later from the bone marrow that express adult globin genes. This process, by which red cells in the blood stream containing predominantly one hemoglobin are gradually replaced by red cells containing predominantly another hemoglobin, was termed "hemoglobin switching" by Nienhuis and Stamatoyannopoulos (Nienhuis and Stamatoyannopoulos, 1978).

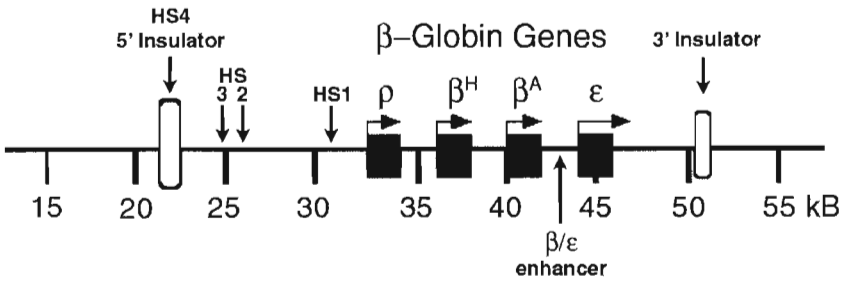
The chicken has served as an excellent model in which to study the embryonic to adult hemoglobin switch, mainly due to the relative ease with which nucleated primitive and definitive erythrocytes can be obtained from chicken embryos. Remarkably,

although 300 million years of evolution separate chickens from man (Kumar and Hedges, 1998), regulation of the β-globin gene locus is functionally identical. Transgenic mice carrying the chicken β-globin locus express the chicken globin genes in the correct developmental pattern (Mason et al., 1995). Work in our laboratory has shown that the converse situation is also true; human β-type globin constructs transfected into chicken erythrocytes are expressed at appropriate relative levels for the developmental stage of the erythrocyte (Basu et al., 2004).

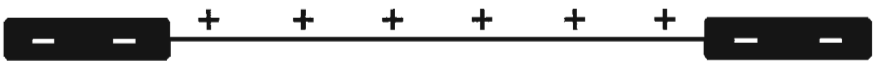
The chicken β-globin locus is composed of four globin genes and several well-characterized regulatory elements. Unlike the human and murine loci, the chicken genes are not positioned within the locus in order of their developmental expression. The two embryonic genes ρ-globin and ε-globin flank the centrally positioned fetal-like β^H-globin and the adult β^A-globin (see Figure 2, panel A). A complex regulatory region known as the locus control region (LCR) lies at the 5' region of the locus (Reitman et al., 1990). This region is composed of four elements that are more sensitive to cleavage by DNase-I than the rest of the locus and are therefore known as DNase-I hypersensitive sites (HSS) (Reitman and Felsenfeld, 1990). These elements, named HSS1 through 4, have strong transcriptional enhancing activity and can conger position-independent expression on a linked transgene. The specific functions of the HSS have been characterized and are detailed in Table 2.

The chromatin structure of the approximately 33 kb chicken β-globin locus has also been well characterized. Early studies showed that the locus is sensitive to DNase-I cleavage in erythroid cells but resistant to cleavage in non-erythroid cells, leading to the

A.Locus Map



B.DNase-I Sensitivity



C.Transcriptional Status

	ρ	β^H	β^A	ε
5 DAY RBC	+	–	–	+
14 DAY RBC	–	+	+	–
adult RBC	–	–	+	–

D.Methylation Status

	ρ	β^H	β^A	ε
5 DAY RBC	--	++	++	--
14 DAY RBC	+-	--	--	+-
adult RBC	++	--	--	++

Figure 2. Features of the chicken β-globin locus. (A) Map showing the relative locations of the globin genes and hypersensitive sites in genomic DNA. (B) The DNase-I sensitivity profile of the locus indicates that globin genes reside within a 33 kb DNase-I sensitive chromatin domain flanked by heterochromatin. (C) Transcriptional status of the four β-globin genes in erythrocytes at three different developmental stages. (D) Methylation status of the four β-globin genes in erythrocytes at three different developmental stages.

Table 2. Specific functions of the HSS in the chicken β-globin locus.

HSS	Known function	Distance from the ρ-globin gene* (kb)	Reference
HSS1	unknown	– 1.6	
HSS2	enhancer	– 4.4	(Abruzzo and Reitman, 1994)
HSS3	enhancer	– 6.4	(Abruzzo and Reitman, 1994)
HSS4	boundary element	– 11.0	(Chung et al., 1993)
3' β/ε	enhancer	+ 10.0	(Choi and Engel, 1986) (Reitman et al., 1990)

* Distance from the ρ-globin gene is listed as negative if the HSS is upstream of the ρ-globin gene, and positive if the HSS is downstream of the ρ-globin gene.

idea that the locus functioned as an independent "chromosomal domain" (Stalder et al., 1980). The locus is flanked on the upstream side by a gene-depleted segment of condensed chromatin and on the downstream side by a neurally-expressed odorant receptor gene. Insulator elements, capable of blocking the action of an enhancer, have been described at both the 5' (within HSS4) and 3' boundaries of the locus (Chung et al., 1997; Saitoh et al., 2000). These elements are thought to contribute to maintenance of the chicken β-globin locus as an independent chromatin domain by serving as attachment points for proteins that physically loop the locus and by recruiting histone modifying proteins that support transcriptional activity (Yusufzai et al., 2004).

Another component of chromatin structure, the histone modification state, has been mapped in detail at the chicken β-globin locus. In pre-erythroid cells that do not express any of the globin genes, very low levels of histone acetylation are seen across the locus, with the exceptions of a peak of acetylation at HSS4 and the intergenic enhancer (Litt et al., 2001b). Interestingly, the locus maintains uniformly high levels of histone acetylation in 10-day erythrocytes that express β^h- and β^A-globin (Litt et al., 2001b). Furthermore, only subtle two-fold changes in the levels of histone acetylation can be correlated with the changes in globin gene transcription that occur during development (Hebbes et al., 1992). Taken together, these data indicate that the chicken β-globin locus is maintained as an independent, open chromatin structure once transcription within the locus commences.

A frequently overlooked aspect of the chicken β-globin locus is its CG-density. The chicken β-globin locus differs significantly from the human β-globin locus in that

three of the four globin genes of the chicken locus contain CpG islands that overlap the promoter (ρ, β^A, ε), whereas none of the genes within the human locus contain CpG islands that overlap the promoter (Kransdorf and Ginder, unpublished observations) . This lack of CpG islands within the human β-globin locus has been cited as evidence that the human β-genes are not regulated by DNA methylation (Harju et al., 2002). On the contrary, abundant data in the literature support a role for DNA methylation in the regulation of human γ-globin expression (Busslinger et al., 1983; Ley et al., 1984).

One possible explanation for this disparity is that the CG-density of the chromosomal isochores in which the chicken and human β-globin loci reside is different. Ponger *et al.* found that for tissue-specific genes like the β-globin genes, the likelihood of a gene having a promoter CpG island was proportional to the CG-density of the surrounding isochore (Ponger et al., 2001). It is likely that the chicken β-globin locus resides within a CG-rich isochore, while the human β-globin locus resides within a CG-poor isochore. Further studies are required to verify this hypothesis.

Structure and Silencing of the ρ-Globin Gene

The embryonic to adult β-globin switch in the chicken serves as an interesting model for transcriptional silencing. ρ-globin is expressed abundantly in primitive erythrocytes produced in the yolk sac from embryonic days one to five, and is not expressed in the definitive erythrocytes which are produced beginning at day five from the fetal liver. The definitive cells do not express ρ-globin due to transcriptional silencing (Lois and Martinson, 1989).

Transfection experiments performed with globin gene constructs have shown that the promoters of the β^A- and ε-globin genes contain the *cis* elements required for developmentally-regulated expression of these genes (Hesse et al., 1986). Likewise, Minie *et al.* found that 500 bp of the ρ-globin promoter was sufficient for proper regulation of the gene in primitive and definitive cells (Minie et al., 1992). Dissection of the promoter identified GATA-1, Sp1, and TFIID sites that were required for expression in primitive cells (see Figure 3, panel B). The ρ-globin gene itself is composed of three exons. The first exon and first intron of the gene are CpG-rich and have been termed by our laboratory the "proximal transcribed region" (PTR).

Work in our laboratory has established that every CpG site in the promoter and PTR of the ρ-globin gene is methylated in adult erythrocytes but unmethylated in primitive erythrocytes (Ginder and McGhee, 1981; Singal et al., 1997; Singal et al., 2002). 5-methylcytosine can first be detected in the CpG-rich proximal transcribed region of the ρ-globin gene at day five and then spreads downstream to exon 3 and upstream to the promoter (Singal and vanWert, 2001). The density and location of DNA methylation at the ρ-globin gene during development is diagrammed in Figure 4. Methylation of the promoter and PTR has a 20-fold and 5-fold inhibitory effect on transcription in adult erythrocytes, respectively. These observations are functionally significant, since expression of ρ-globin in adult chickens can be induced by treating the animals with 5-azacytidine, an inhibitor of DNA methyltransferase (Ginder et al., 1984).

Our laboratory has also shown that an Erythroid Methyl-CpG Binding Protein Complex (MeCPC) forms on the methylated ρ-globin promoter and PTR *in vitro* using

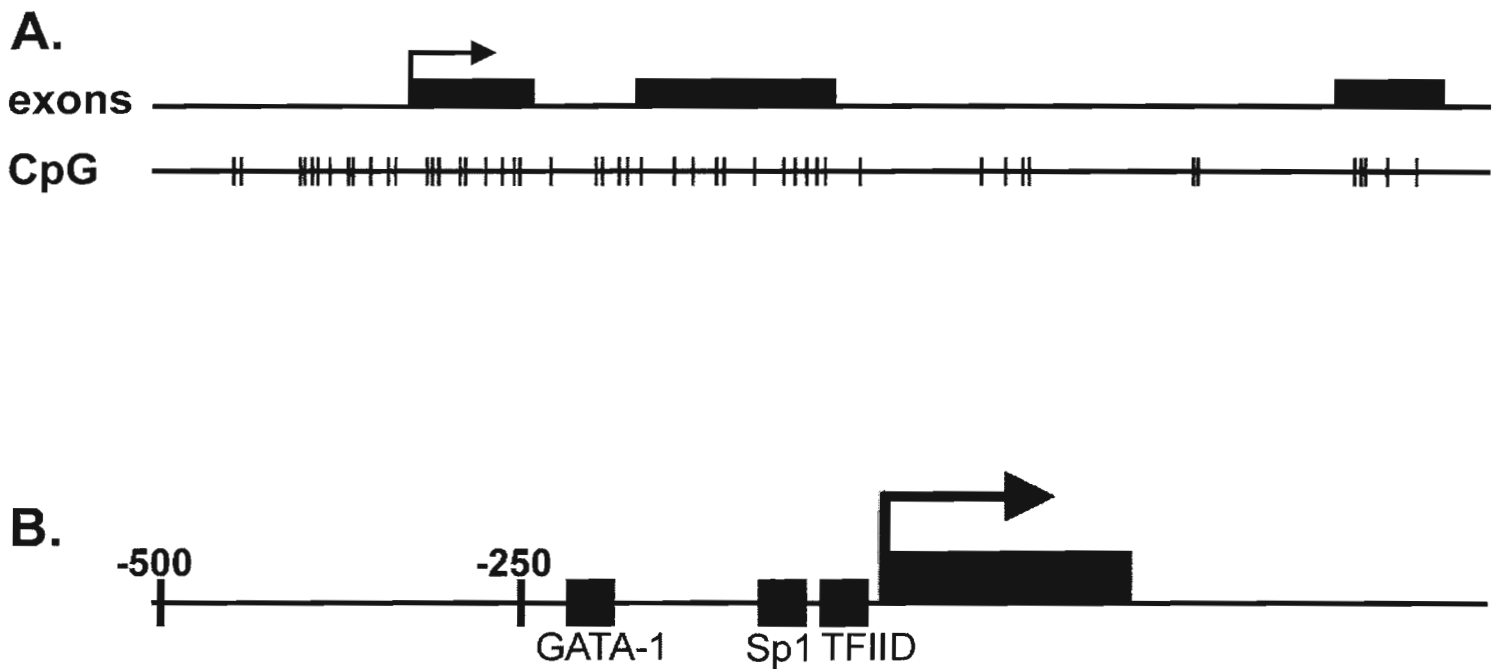


Figure 3. Schematic of the ρ-globin gene. A) The gene is composed of three exons and two introns. The 5' region of the gene (promoter and PTR) are CpG-rich, although this region is too small to fulfill the traditional definition of a CpG-island. B) The promoter of the ρ-globin gene has three elements that have been shown to be essential for expression in primitive erythroid cells: a GATA-1 site, an Sp1 site, and a TFIID site. These elements are all located within the proximal 250 bp of the promoter.

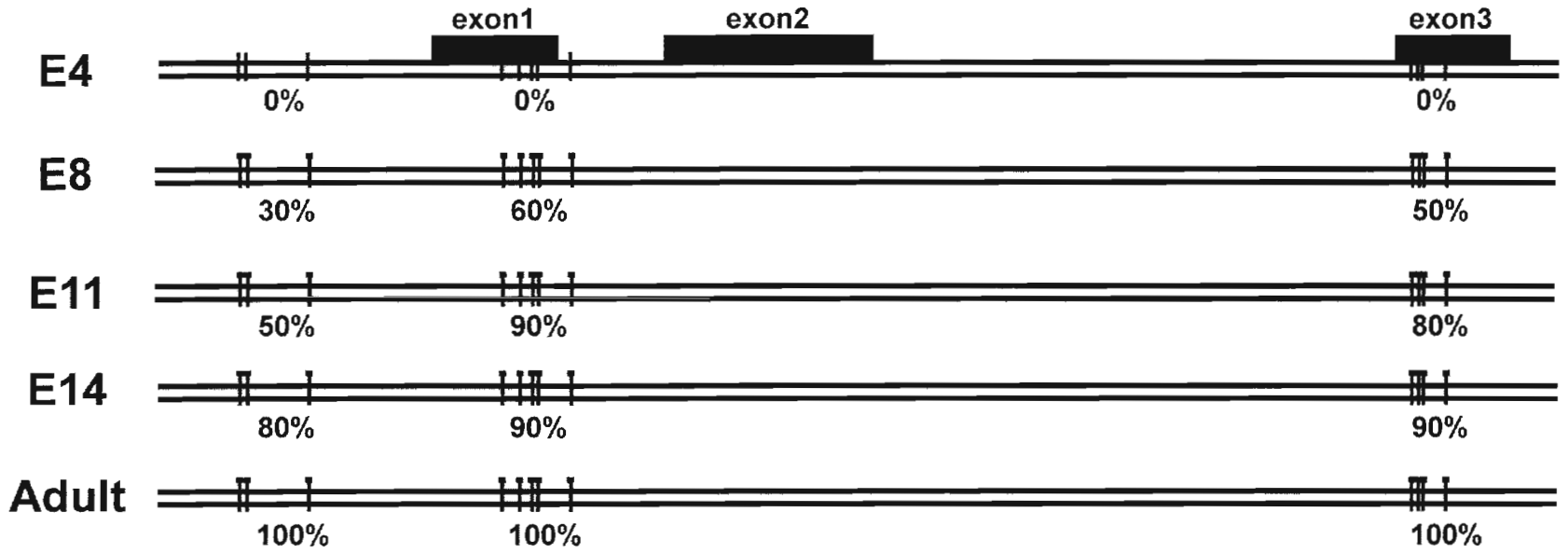


Figure 4. The density and location of DNA methylation at the ρ-globin gene during development. Methylation can first be detected in the CpG-rich proximal transcribed region of the ρ-globin gene at day five and then spreads (increases in density) downstream to exon 3 and upstream to the promoter. Every CpG in the promoter and PTR are methylated in adult erythrocytes.

chicken erythroid nuclear extracts (Singal et al., 1997; Singal et al., 2002). This complex is likely of functional significance since these sequences are methylated at every CpG *in vivo*. The DNA-binding affinity of this primary erythroid-cell derived complex differs significantly from that of MeCP1, and hence is termed the MeCPC. The MeCP1 complex exhibits a significantly greater DNA-binding affinity for its canonical binding sequence, methylated CG11 (M-CG11), than for the methylated ρ-globin proximal transcribed sequence. Surprisingly the MeCPC complex exhibits a greater DNA-binding affinity for its canonical binding sequence, the methylated ρ-globin PTR (M-ρ248), than for M-CG11 in spite of a higher ^{m}CpG content in M-CG11. Based on these data, we hypothesize that a factor or factors present in the MeCPC complex but not in the MeCP1 complex, confers sequence-context preferential binding to the MeCPC.

Significance

The β-hemoglobinopathies are a group of diseases characterized by the functional abnormality of the β-chain of hemoglobin. The β-hemoglobinopathies can be divided into two groups: those resulting from a genetic mutation in the β-globin gene and those resulting from genetic defects that affect the synthesis of the β-globin gene. Sickle cell disease (SCD) is the prototypical example of the first group, an autosomal recessive disease resulting from a mutation in codon six of the β-globin gene which replaces glutamate with valine. The second group is composed of the β-thalassemias, diseases characterized by partial or total deficiencies in β-globin production. The β-thalassemias

are very genetically heterogeneous, with over 100 distinct mutations that have been identified (Weatherall, 2004).

SCD is one of the most common monogenic diseases in the United States, affecting one in 500 African-Americans. Patients are able to achieve periods of good health, but these periods are frequently interrupted by "crises," acute periods of disease manifestation. Sadly, these patients also have a significantly decreased life expectancy (Platt et al., 1994). The most common type of crisis, the vaso-occlusive crisis, is caused by sickled blood cells occluding small vessels. The sickling of red blood cells is caused by the polymerization of the deoxygenated hemoglobin S (Hb S, which is composed of $\alpha_2\beta_2^{6\ Glu\rightarrow Val}$) into long filaments which distort the morphology of the erythrocyte. The major direction of therapy for SCD has been to inhibit hemoglobin S polymerization via induction of fetal hemoglobin (Hb F, which is composed of $\alpha_2\gamma_2$), which has been shown to directly inhibit Hb S polymerization (Nagel et al., 1979).

Several compounds have been found to have therapeutic benefit in SCD through Hb F induction. The first such agent was 5-azacytidine, which was shown to increase Hb F levels both in anemic baboons and in a single patient with severe β-thalassemia (DeSimone et al., 1982; Ley et al., 1982). Recent studies with 5-aza-deoxycytidine have also shown promising increases in Hb F levels and decreases in vaso-occlusive pathophysiology in small clinical trials involving patients with SCD (Koshy et al., 2000; Saunthararajah et al., 2003). Based on the original success of 5-azacytidine in inducing Hb F synthesis in model systems, the cytotoxic drug hydroxyurea was investigated as another potential Hb F-inducing agent. Early studies in anemic monkeys showed that

hydroxyurea was also able to induce Hb F (Letvin et al., 1984). Hydroxyurea was approved by the FDA in 1998, having been found to significantly increase Hb F levels and reduce the frequency of pain episodes, hospitalizations, and the need for blood transfusions (Charache et al., 1995). Although these pharmacological approaches are encouraging, these agents do not fully alleviate the clinical manifestations of SCD, are not able to induce a response in all treated patients and may not be safe for long-term therapy, due to their carcinogenic potential (Jackson-Grusby et al., 1997).

5-azacytidine is a well-characterized inhibitor of DNA methylation (Jones and Taylor, 1980). Thus, the Hb F induction observed in patients treated with 5-azacytidine is most likely due to increases in γ-globin transcription. In support of this, Koshy *et al.* found that treating SCD patients with 5-aza-deoxycytidine (an analog of 5-azacytidine) increased γ-globin transcription from an average of 3.19% ± 1.43% of non-α-globin before treatment to13.66% ± 4.35% of non-α-globin after treatment. Currently, the pattern of CpG methylation at the human γ-globin globin genes unknown. Recent data by Lowrey indicated that methylation of the γ^{G}-globin promoter increased from 38% of all CpG sites in the fetal liver to 73% of all sites in the adult bone marrow (Lowrey, unpublished observations). Taken together, these data indicate that DNA methylation contributes to the maintenance of transcriptional silencing of the γ-globin gene in adult human erythrocytes.

Therefore, given that DNA methylation contributes to maintenance of transcriptional silencing of the fetal β-globin genes in humans, therapeutic methodologies directed against the effector mechanisms of DNA methylation could be used to de-repress

γ-globin expression, leading to amelioration of the clinical symptoms of SCD. In this way, the potential carcinogenicity of agents that directly affect DNA methylation, such as 5-aza-deoxycytidine, could be avoided. This strategy is especially plausible in light of recent evidence that the MeCP1 methyl-DNA binding complex contains a novel zinc-finger protein (termed p66) that has a role in targeting the complex to specific loci (Feng et al., 2002). Work in our lab has shown that the MeCPC which forms on the methylated ρ-globin gene has a stronger affinity for its cognate DNA sequence than for a generic ^{m}CpG-rich sequence (Singal et al., 2002). Taken together, these findings suggest that pharmacologic agents specific for the targeting factor or factors in the γ-globin MeCPC may be able to de-repress γ-globin by preventing the transcriptional repressive actions of the MeCPC. To that end, it is necessary to precisely define the composition and function of the ρ-globin MeCPC and to extend these findings from the chicken β-globin locus to the human β-globin locus.

DNA methylation has also been shown to contribute to the transcriptional silencing of tumor suppressing genes in cancer (Jones and Baylin, 2002). As discussed above, several studies have shown that MBD2 binds to methylated tumor suppressor genes such as *p16^{Ink4a}* and *GSTP1* in cancer cell lines, and that abrogating MBD2 function can de-repress a methylated gene (Magdinier and Wolffe, 2001; Bakker et al., 2002). Despite this data, no studies have investigated the role of methyl-CpG binding protein complexes in methylation-mediated transcriptional repression of tumor suppressing genes. The *p16^{Ink4a}* gene, like the ρ-globin gene, has a CpG-rich promoter and first exon. These regions have been found to be methylated in primary tumors of the

breast and colon (Herman et al., 1995). As such, our identification of MeCPC component factors and our functional dissection of which of these factors is required for transcriptional repression of the ρ-globin gene, may be highly relevant to understanding the mechanism of transcriptional repression of the methylated $p16^{Ink4a}$ in cancer.

Chapter 2: Characterization of Chicken MBD2

Introduction

An abundance of data support the role of MCBPs as the major effectors of the transcriptional silencing mediated by DNA methylation (Wade, 2001). Originally discovered in humans, members of this family of proteins are present in most metazoans, including organisms without significant levels of DNA methylation, such as *Drosophila* (Ballestar et al., 2001).

Data from our laboratory has suggested that MBD2 is a component of the MeCPC (Singal et al., 2002). In addition to confirming this finding, we wanted to characterize the chicken homolog of MBD2 (cMBD2) and to generate antibodies against this protein. Although antibodies against mouse and human MBD2 are available commercially, high avidity antibodies are required for several of the experiments we wished to pursue, such as immunoprecipitation (IP) and chromatin immunoprecipitation (ChIP). In our hands these commercially produced anti-MBD2 antibodies cross-reacted poorly with cMBD2.

Furthermore, we wished to determine if other members of the well-characterized mammalian MCBPs are present in chicken cells. Several studies have shown that both MBD2 and MeCP2 can bind to the same methylated region of DNA *in vivo* (Nguyen et al., 2001; Magdinier and Wolffe, 2001; Koizume et al., 2002). This implies that DNA methylation can recruit more than one type of transcriptional repressor complex.

As discussed in Chapter 1, MeCP1 was the first methyl-CpG binding activity to be described. The original studies on this activity showed that it required 10-15 methyl-CpG residues per DNA fragment for binding (Meehan et al., 1989). Feng and Zhang purified the MeCP1 complex from HeLa cells and found it to be composed of all the components of the NuRD complex and MBD2 (Feng and Zhang, 2001). However, sensitive binding studies performed by Fraga *et al.* using recombinant mouse MBD2b (which is the region of mouse MBD2 homologous to cMBD2) indicated that this protein was able to bind to a DNA fragment containing a single methyl-CpG. This result agrees with the experimentally derived crystal structure of MBD1, which shows that the MBD of MBD1 contacts the 5-methylcytosine on each strand of the ^{m}CpG duplex simultaneously (Ohki et al., 2001). This implies that MBD2 may require multiple-CpGs for binding in the context of the MeCP1 complex or that binding by MBD2 is methyl-CpG density-dependent, with stronger binding observed for DNA fragments containing more methyl-CpGs. We hypothesized that the latter is true and in the process of characterizing cMBD2 wished to test this hypothesis.

Methods

Bioinformatics and Sequencing

The nucleotide sequences of the mouse MeCP2 mRNA (NCBI Genbank ID NM_010788), MBD1 mRNA (NM_013594), MBD2 mRNA (NM_010773), MBD3 mRNA (NM_013595), MBD4 mRNA (NM_010774), and Kaiso mRNA (NM_020256) were used to search the University of Delaware Chicken EST Database

(http://www.chickest.udel.edu) and the BBSRC Chicken EST Database (http://chick.umist.ac.uk) for homologous clones via the BLAST algorithm.

The clone 'pat.pk0057.c7.f' was obtained from the University of Delaware Database. This clone contains an EST insert homologous to mouse MBD2 in the pcDNA3 vector. The EST insert was sequenced using the SP6 and T7 primers present in the vector, as well as an internal primer.

Cloning of cMBD2 Into pGEX/6P

The cDNA segment encoding chicken MBD2 was PCR amplified from EST clone pat.pk0057.c7.f using the primers 5'-TCGAATTCACGGACTGCCCGGCCCTT (forward) and 5'-AGCTCGAGAGGTCACTCGTCTTGA (reverse) using Platinum Taq Hi-Fidelity Polymerase (Invitrogen). The PCR product was purified with the Rapid PCR Purification Kit (Marligen) and cloned into pGEM-T Easy (Promega) resulting in the plasmid pTA/cMBD2. The pGEX/6P-1 and pTA/cMBD2 plasmids were digested with EcoRI and XhoI and the appropriate bands were gel purified (Qiagen). The resulting cMBD2 EcoRI/XhoI fragment was ligated into the cut pGEX/6P-1 vector. The sequence of pGEX/6P/cMBD2 was confirmed by DNA sequencing.

Recombinant Protein Production

Recombinant cMBD2 was produced in Rosetta(DE3)pLacI *E. coli* (Novagen) as a GST fusion protein. These cells contain the pRARE chloramphenicol-resistance plasmid that supplies tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons to enhance translation

of eukaryotic genes in prokaryotic cells. Briefly, a 12 mL culture of LB containing 50 μg/mL ampicillin and 34 μg/mL chloramphenicol was inoculated with a single colony and grown overnight at 37°C with shaking at 225 RPM. The next morning, 10 mL of culture was inoculated into 250 mL of 2xYT medium and grown approximately 3 hours at 37°C with shaking at 225 RPM. When the OD_{600} reached 0.8, the flask was transferred to 30°C and cells were induced with 1 mM IPTG. The culture was incubated at 30°C with shaking at 225 RPM for 4 hours. The cells were pelleted by spinning at 6000g for 15 min at 4°C in a J2-21M centrifuge using the JS 14 rotor (Beckman). The supernatant was removed and the cell pellet was resuspended in 20 mL PBS and transferred to 40 mL Oak Ridge type centrifuge tubes (Nalgene). The tubes were spun at 6000g for 15 min at 4°C using the JS 13.1 rotor (Beckman). The supernatant was removed and the cell pellet was stored frozen at -80°C. The next day the cell pellet was thawed and resuspended in BugBuster Reagent (Novagen) at 10 mL/3g of cell pellet containing 10 μL / 10 mL benzonase nuclease (Novagen), 0.5 μL/10 mL recombinant lysozyme (Novagen), 10 μg/mL leupeptin, 10 μg/mL pepstatin, 10 μg/mL aprotinin, 0.2 mM PMSF, and 1 mM DTT. The lysate was incubated at room temperature for 20 minutes with inversion by rocker. After 20 minutes, 1/10 volume of 10x PBS was added to the lysate which was then incubated at 4°C for 20 minutes. The lysate was then spun 13000g for 15 min at 4°C using the JS 13.1 rotor (Beckman). The supernatant was removed and frozen at -80°C until used.

Purification of GST/cMBD2 from *E coli* Lysate: GST Chromatography

Purification was performed using a BioLogic DuoFlow system (Bio-Rad). First the *E coli* lysate was thawed, spun at 15000g for 20 min at 4°C using the JS 13.1 rotor in the J2-21M centrifuge (Beckman) and filtered using Millex 0.45 μm low protein binding filters (Millipore). DTT was added to the lysate to a final concentration of 2 mM. 60 to 75 mL of lysate was injected to a 5 mL GSTrap High Performance column (Amersham Biosciences). The flow-through was collected during injection. The column was washed with 4 CV of 1x PBS and then the bound protein eluted with 6 CV of Buffer GE (50 mM Tris-HCl pH 8.5, 100 mM NaCl, 0.1% Triton X-100, 10 mM glutathione, 1 mM DTT). The column was washed with 4 CV of 1x PBS and then the flow-through from run I was re-injected and the column was washed and bound protein eluted as detailed above. Protein containing fractions from runs I and II were pooled and concentrated via spinning in an Ultrafree-15 at 2500g for 115 min at 10°C in a 5810R centrifuge using the A-4-62 rotor (Eppendorf).

Purification of GST/cMBD2 from *E coli* Lysate: SP Chromatography

To remove excess glutathione and GST in the GST/cMBD2 sample, SP chromatography was performed using the BioLogic DuoFlow system (Bio-Rad). The concentrated GST/cMBD2 sample obtained from GST chromatography was dialyzed against buffer H100 overnight (10 mM sodium phosphate pH 6.5, 100 mM NaCl, 0.2 mM EDTA, 10% glycerol, 1 mM DTT). The approximately 5 mL sample was injected to a 1 mL HiTrap SP Fast Flow column (Amersham Biosciences). The column was washed with 4 CV of H200 and then the bound protein eluted with a 8 CV linear gradient from H200 to H600.

The column was washed with 4 CV of H800 and 4 CV of H1000 to ensure all bound protein had eluted. GST/cMBD2 containing fractions were pooled and concentrated via spinning in an Ultrafree-4 at 3800g for 60 min at 10°C in the JS 13.1.

Preparation of Antibodies Against GST/cMBD2

After production of ~13 mg recombinant GST/cMBD2, the purified protein samples were run on a preparative 12% Tris-glycine SDS-PAGE gel. The gel was stained with Bio-Safe Coomassie (Bio-Rad). The band corresponding to full-size GST/cMBD2 at 55 kD was excised from the gel and sent to Cocalico Biologicals. Cocalico Biologicals was contracted to immunize two rabbits with the recombinant protein.

Anti-cMBD2 IgG Purification

The antiserum from the two rabbits (VCU1 and VCU2) was received from Cocalico Biologicals. Purification was performed using a BioLogic DuoFlow system (Bio-Rad). The serum from animal VCU1 was thawed and 25 mL was injected to a 5 mL HiTrap rProtein A Fast Flow column (Amersham Pharmacia). The column was washed with 3 CV of 1x PBS and then the bound protein eluted with 3 CV of Buffer G (100 mM glycine-HCl pH 2.8, 10% ethylene glycol). 50 µL of Tris-HCl pH 8.8 was added to each collection tube so that the pH of the eluate would immediately be neutralized. The column was washed with 3 CV of 1x PBS and then with 3 CV of cleaning buffer (100 mM glycine pH 3.0, 2 M urea, 1 M LiCl). The purification procedure was repeated as

above with serum from animal VCU2. Protein containing fractions from each run were pooled and dialyzed against 1x PBS overnight at 4°C.

Western Blotting

Nuclear extracts were prepared by the Dignam method (Dignam et al., 1983). Protein samples were run on 10% Tris-glycine SDS-PAGE ReadyGels (Bio-Rad) in Running buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS). The gel was equilibrated in Transfer buffer (25 mM Tris-HCl pH 8.0, 192 mM glycine, 20% methanol) and then transferred to activated PVDF overnight at 30V. The next day the membrane was washed twice for 10 minutes with 1x PBS and then blocked for 1 hour at room temperature with rocking in 10 mL MPBST (5% (w/v) non-fat milk, 1x PBS, 0.05% Tween-20). The blocking solution was removed and the primary antibody was incubated with the membrane for 2 hours at room temperature with rocking in MPBST. This solution was removed and the membrane was washed three times with 10 mL PBS-T for 5 min. The secondary antibody was then incubated with the membrane for 1 hour at room temperature with rocking in MPBST. This solution was removed and the membrane was washed three times with 10 mL PBS-T for 5 min and one time with 10 mL PBS-T plus 0.1% Triton X-100. Bands were visualized with ECL Plus (Amersham Pharmacia) and exposed to film.

Electrophoretic Mobility Shift Assay Probes

The 248 bp ρ exon 1/intron 1 DNA fragment (ρ248) was generated by digestion of

plasmid pTA/ρ248 with NheI and StuI. This fragment has a blunt end at the 3' end (from the StuI cut) and a sticky end at the 5' end (from the NheI cut). The probe fragment was isolated by gel purification (Qiagen). The 50 bp ρ exon 1/intron 1 DNA subfragments (a0CpG, a1CpG, a2CpG, a3CpG) containing 0, 1, 2 or 3 CpG were synthesized (Integrated DNA Technologies), annealed, and purified from 10% polyacrylamide gels. Regions used for probes are indicated in Figure 5.

Electrophoretic Mobility Shift Assay

DNA fragments were methylated using SssI methylase (New England Biolabs) according to the manufacturer's instructions. The DNA was phenol-chloroform extracted and precipitated using glycogen as a carrier. The M-ρ248 fragment was labeled by incubation with [α-^{32}P]dCTP (MP Biomedicals) and Klenow fragment (New England Biolabs). The a0CpG, a1CpG, a2CpG and a3CpG fragments were labeled by incubation with [γ-^{32}P]ATP and polynucleotide kinase (New England Biolabs). The assay was performed by adding 15 ng purified cMBD2 or 8 μL of DEAE-purified chicken RBC nuclear extract to a 21.5 μL binding reaction containing 20 fmol of probe, 2 μg of sonicated *M. Lys* DNA, 20 mM HEPES [pH 7.9], 3 mM $MgCl_2$, 1 mM EDTA, 10 mM β-ME, 0.1% Triton X-100, 4% glycerol. Binding reactions were incubated on ice for 20 minutes and then loaded onto a 2% agarose gel and resolved in 0.5x TBE buffer. For supershift experiments, protein A-purified anti-MBD2 antibodies were added to the binding reaction after the 20 minute incubation on ice and incubated at room temperature for 15 minutes.

A.

```
1                                                          50
CTAGCAAGC TCTGAGTGCT CCCACAGCCG CACGCCAACC CCGCTGCCACC

51                                                        100
ATGGTGCAC TGGTCCGCCG AGGAGAAGCA GCTCATCACC AGCGTCTGGAG

101                                          ▼1CpG        150
CAAAGTCAA CGTGGAGGAA TGCGGTGCCG AAGCCCTGGC CAGGTGGGTCT

151                                    1CpG▼              200
GCTCCTGGG ATCGGCTCAG CTGCACCCTG GGTACTGAAA CCACTGCAGTT

201                                                       250
TTAGGAGGC AGCGCTAACG GTGTGTGCTT GTGTCCCCCG TCTCTCCGCAG

251
G
```

B.

```
a0CpG GCCAGGTGGGTCTGCTCCTGGGATtGGCTCAGCTGCACCCTGGGTACTGA
      CGGTCCACCCAGACGAGGACCCTAaCCGAGTCGACGTGGGACCCATGACT

a1CpG GCCAGGTGGGTCTGCTCCTGGGATCGGCTCAGCTGCACCCTGGGTACTGA
      CGGTCCACCCAGACGAGGACCCTAGCCGAGTCGACGTGGGACCCATGACT

a2CpG GCCAGGTGGGTCcGCTCCTGGGATtGGCTCAGCTGCACCCcGGGTACTGA
      CGGTCCACCCAGgCGAGGACCCTAaCCGAGTCGACGTGGGgCCCATGACT

a3CpG GCCAGGTGGGTCcGCTCCTGGGATCGGCTCAGCTGCACCCcGGGTACTGA
      CGGTCCACCCAGgCGAGGACCCTAGCCGAGTCGACGTGGGgCCCATGACT
```

Figure 5. Regions of the ρ-globin gene used probes in EMSA. A) Sequence of the ρ-globin PTR (ρ248). CpGs are in bold. The subfragment used for EMSA with recombinant cMBD2 is underlined and indicated with arrows. B) 50 bp ρ248 subfragments used for EMSA. Fragment a1CpG is derived directly from ρ248 without modification. Fragments a0CpG, a2CpG, a3CpG have been mutated to contain 0 CpG, 2 CpGs or 3 CpGs, respectively. Bases that are mutated from fragment a1CpG are in lower case and underlined. CpGs are in bold.

Results

In an effort to understand the epigenetic mechanisms operating at the chicken β-globin locus, we sought to identify homologs of mammalian MCBP present in chicken cells. As discussed in Chapter 1, six murine proteins are known to have specific methyl-CpG binding activity: MeCP2, MBD1, MBD2, MBD3, MBD4, and Kaiso. Weitzel *et al.* previously identified a chicken homolog of the MeCP2 protein (Weitzel et al., 1997), although the structure of this protein differed significantly from that of the mammalian protein. Zhu *et al.* also reported the identification of a putative chicken homolog of MBD4 (Zhu et al., 2000).

At the time of initiation of these studies, the genome sequence of the chicken had not yet been completed, but two expressed sequence tag (EST) databases were available. We searched these two databases, the University of Delaware (UD) and the BBSRC Chicken EST Databases with the cDNA sequence of mouse MeCP2, MBD1, MBD2, MBD3, MBD4, and Kaiso. The BBSRC EST Database was an especially useful resource because it is composed of 339,314 ESTs which have been assembled *in silico* into cDNA contigs (Boardman et al., 2002). Once the Ensembl chicken genome annotation (ECGA) became available from the Ensembl project in July 2004 we also searched this database with the mouse cDNA sequences to confirm the identifications made from the EST databases (Birney et al., 2004).

No clones matching the previously identified chicken MeCP2 protein were identified in the two EST databases or the ECGA. Although the structure of chicken MeCP2 differs significantly from that of mammalian MeCP2, there is a 125 amino acid

region of MeCP2 that is 96% homologous to that of the mammalian protein and thus likely represents its avian homolog. No clones homologous to mouse MBD1 were identified in the two EST databases or the ECGA. Despite this absence, the possibility of an MBD1 homolog in the chicken cannot be excluded because 1) the chicken EST databases are likely incomplete 2) the chicken genome sequence is only 90% complete, with under representation of high-GC content sequences. The proximal exonic regions of the human *MECP2*, *MBD2*, *MBD3* genes are CG rich, with the 5' portion of the *MBD2* gene being composed of 80% GC.

Several clones encoding a homolog of mammalian MBD2 were identified in both EST databases but not in the ECGA. The nucleotide sequences of the clones from the UD and BBSRC EST databases matched exactly. The cDNA encoded a 28-kD open reading frame that has 83% amino acid similarity to a 241 amino acid region of human and mouse MBD2 (see Figure 6). The region of homology extends over the methyl-CpG binding domain (MBD) of the mammalian MBD2 proteins (Nan et al., 1993) and key amino acids known to be involved in the recognition of the methyl-CpG (Fraga et al., 2003; Saito and Ishikawa, 2002) are conserved in cMBD2, indicating that cMBD2 is likely to have methyl-DNA binding activity *in vitro* and *in vivo*.

Clones encoding a homolog of mammalian MBD3 were also identified in both EST databases. The nucleotide sequences of the clones from the UD and BBSRC EST databases matched exactly. The cDNA encoded a 26-kD open reading frame that has

```
        1                                                   50
hMBD2   MRAHPGGGRC CPEQEEGESA AGGSGAGGDS AIEQGGQGSA LAPSPVSGVR
mMBD2   MRAHPGGGRC CPEQEEGESA AGGSGAGGDS AIEQGGQGSA LAPSPVSGVR
cMBD2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        51                                                  100
hMBD2   REGARGGGRG RGRWKQAGRG GGVCGRGRGR GRGRGRGRGR GRGRGRPPSG
mMBD2   REGARGGGRG RGRWKQAARG GGVCGRGRGR GRGRGRGRGR GRGRGRPQSG
cMBD2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        101                                                 150
hMBD2   GSGLGGDGGG CGGG...GSG GGGAPRREPV PFPSGSAGPG PRGPRATESG
mMBD2   GSGLGGDGGG GAGGCGVGSG GGVAPRRDPV PFPSGSSGPG PRGPRATESG
cMBD2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~MDKQ

        151                                                 200
hMBD2   KRMDCPALPP GWKKEEVIRK SGLSAGKSDV YYFSPSGKKF RSKPQLARYL
mMBD2   KRMDCPALPP GWKKEEVIRK SGLSAGKSDV YYFSPSGKKF RSKPQLARYL
cMBD2   GRTDCPALPP GWKKEEVIRK SGLSAGKSDV YYFSPSGKKF RSKPQLARYL

        201                                                 250
hMBD2   GNTVDLSSFD FRTGKMMPSK LQKNKQRLRN DPLNQNKGKP DLNTTLPIRQ
mMBD2   GNAVDLSSFD FRTGKMMPSK LQKNKQRLRN DPLNQNKGKP DLNTTLPIRQ
cMBD2   GNAVDLSCFD FRTGKMMPSK LQKNKQRLRN ESLHPNKGKP DLNTALPIRQ

        251                                                 300
hMBD2   TASIFKQPVT KVTNHPSNKV KSDPQRMNEQ PRQLFWEKRL QGLSASDVTE
mMBD2   TASIFKQPVT KFTNHPSNKV KSDPQRMNEQ PRQLFWEKRL QGLSASDVTE
cMBD2   TASIFKQPVT KVTNHPDNKV RSDPQRLADQ PRQLFWEKRL RGLSASDVGQ

        301                                                 350
hMBD2   QIIKTMELPK GLQGVGPGSN DETLLSAVAS ALHTSSAPIT GQVSAAVEKN
mMBD2   QIIKTMELPK GLQGVGPGSN DETLLSAVAS ALHTSSAPIT GQVSAAVEKN
cMBD2   EILRAMELPR GLQALGPVPD DVTLLSAVAS ALHVGSVPVT GQLSSAAEKN

        351                                                 400
hMBD2   PAVWLNTSQP LCKAFIVTDE DIRKQEERVQ QVRKKLEEAL MADILSRAAD
mMBD2   PAVWLNTSQP LCKAFIVTDE DIRKQEERVQ QVRKKLEEAL MADILSRAAD
cMBD2   PAVWLNSSQP LCRAFVVTDD DIRKQEERVR RVRKKLEEAL LAGDPAGSRG

        401
hMBD2   TEEMDIEMDS GDEA
mMBD2   TEEVDIDMDS GDEA
cMBD2   QDE
```

Figure 6. Comparison of the amino acid sequences of human MBD2, mouse MBD2, and chicken MBD2. Chicken MBD2 is 83% identical to human MBD2 and mouse MBD2 over an overlapping 241 amino acid region (underlined above) that corresponds to the MBD2b isoform found in mammalian cells. Translation initiation codons are in bold. Amino acids conserved between chicken MBD2 and mouse or human MBD2 are shaded grey.

90% amino acid similarity to mouse and human MBD3. A gene for MBD3 was also identified in the ECGA that matched the sequence of the EST clones. The high sequence conservation of MBD3 between chickens and mammals is not surprising, given its essential role in development (Hendrich et al., 2001).

Clones encoding homologs of mammalian MBD4 and Kaiso were also identified. Given that MBD4 functions mainly in DNA repair (Hendrich et al., 1999) and Kaiso has mainly been observed to be involved in epithelial cell regulation (Kondapalli et al., 2004; Kelly et al., 2004), we did not expect a role for these genes in methylation-mediated transcriptional silencing of the ρ-globin gene. Furthermore, as discussed in Chapter 3, we did not identify either MBD4 or Kaiso as components of the MeCPC. A summary of the properties of the identified chicken MCBP homologs can be found as Table 3.

We produced recombinant cMBD2 (rcMBD2) in prokaryotic cells as a GST fusion protein to examine its binding specificity. rcMBD2 was incubated with various methylated and unmethylated DNA probes and subjected to analysis by EMSA. The purified GST/cMBD2 was able to bind to the canonical MeCPC substrate M-ρ248 (Figure 7), as well as to 50 bp subfragments of ρ248 containing 1 methyl-CpG, 2 methyl-CpGs or 3 methyl-CpGs. No binding was detected to a 50 bp ρ248 subfragment containing no methyl-CpGs. These data confirm that cMBD2 is able to bind specifically to methylated DNA fragments containing as few as one methyl-CpG. Furthermore, as can be seen from Figure 7, the amount of probe shifted increases with the number of methyl-CpGs on the fragment. Thus, binding by MBD2 is methyl-CpG density-dependent, with stronger binding observed for DNA fragments containing more

Table 3. Properties of the identified chicken MCBP homologs. Protein size for the mammalian proteins is given for the murine isoform listed in NCBI RefSeq. Percent similarity was computed using the LALIGN program on the mouse protein and the cooresponding chicken open reading frame. Similarity was calculated over the homologous region between the proteins.

Protein	Size in Mouse (kD)	UDel Clone ID	BBSRC Clone Contig ID	Ensembl Gene ID	ORF Size (kD)	% Similarity
cMeCP2	52	not present	not present	not present		
cMBD1	70	not present	not present	not present		
cMBD2	29, 44	pat.pk0057.c7.f	337626.2 (14 clones)	not present	28	83
cMBD3	32	pgm2n.pk008.h4	340483.1 (4 clones)	ENSGALG 00000001101	26	90
cMBD4	63	pat.pk0066.c4.f	332486.4 (27 clones)	not present	45	63
cKaiso	74	not present	335692.1 (6 clones)	ENSGALG 00000008611	56	60

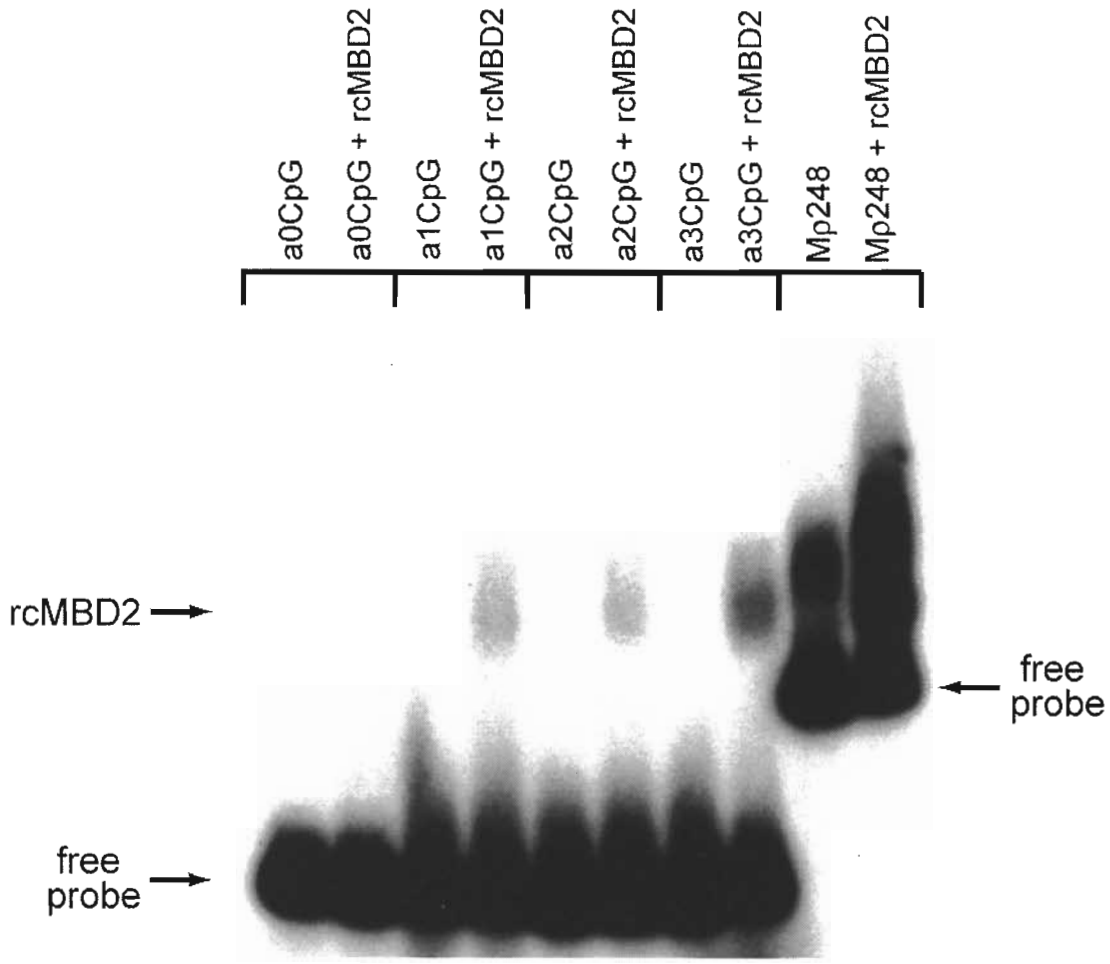


Figure 7. cMBD2 is a *bona fide* MCBP *in vitro.* rcMBD2/GST was incubated with 50 bp ρ248 subfragments containing 0 ^{m}CpG (a0CpG), 1 ^{m}CpG (a1CpG), 2 ^{m}CpG (a2CpG), 3 ^{m}CpG (a3CpG) and 14 ^{m}CpG (M-ρ248) and subjected to EMSA. The amount of probe shifted increases linearly with the number of methyl-CpGs on the probe fragment. The bands generated by the shifted probe (marked "rcMBD2") and free DNA probe (marked "free probe") are indicated with arrows.

methyl-CpGs. We interpret this to mean that MBD2 binds stoichiometrically to methylated DNA, with approximately one MBD2 molecule binding to each methyl-CpG.

To produce antibodies against cMBD2, two rabbits were immunized with rcMBD2. Total IgG was purified from the antiserum obtained from each animal (named VCU1 and VCU2) by protein A chromatography. Western blotting performed with the VCU1 and VCU2 antibodies shows that both antibodies recognize the recombinant cMBD2/GST fusion protein used for immunization and both antibodies recognize a protein of approximately 28 kD in chicken embryonic fibroblast (CEF), DT40, and adult red blood cell nuclear extracts (see Figure 8). Since the clones from the UD and BBSRC EST databases encoded a 28-kD open reading frame similar to that of human and mouse MBD2 and antibodies generated against this polypeptide recognize an endogenous protein of the same size, we conclude that this protein is endogenous cMBD2. Western blots using VCU2 were repeated in the presence of excess rcMBD2 and the band at 28 kD was competed away, but the other bands, such as those at 35 and 60 kD, were still present. Thus competition experiments confirm that the band at 28 kD corresponds to endogenous cMBD2.

Using the anti-MBD2 antibodies, we wished to confirm that MBD2 was a component of the MeCPC complex we observed by EMSA. Experiments were peformed by adding the VCU1 or VCU2 antibodies to the EMSA binding reactions. Although VCU1 antibodies were not able to affect the migration of the complex, the addition of VCU2 antibodies to the EMSA binding reactions gave a definite and reproducible

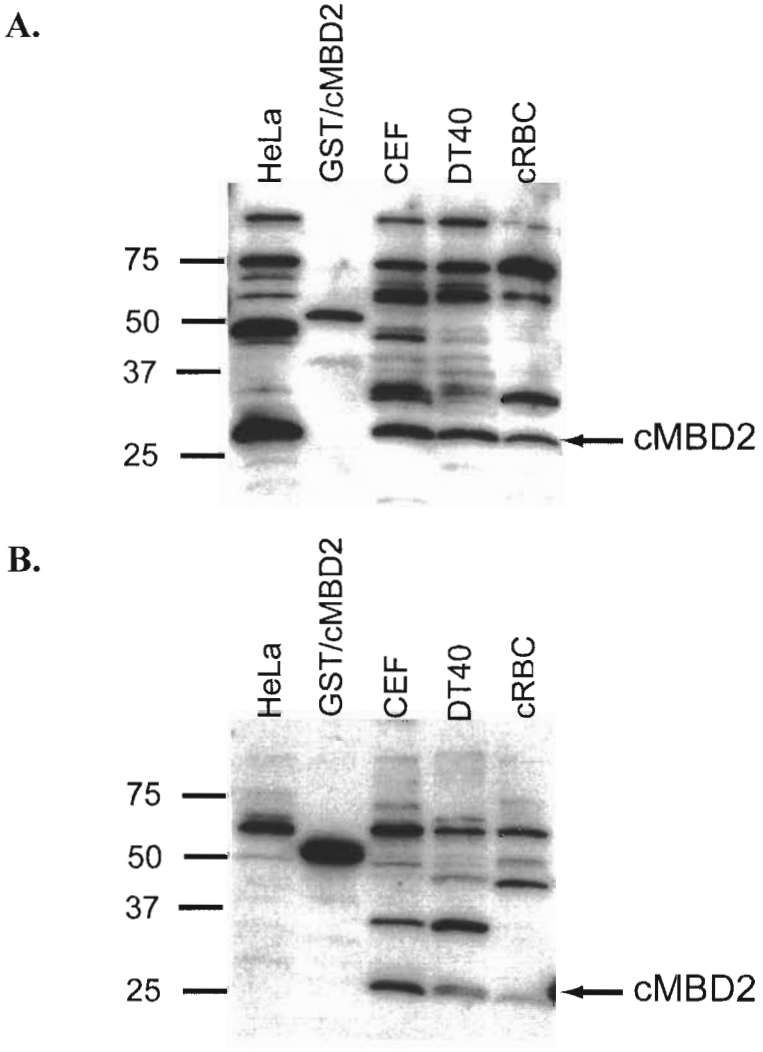


Figure 8. Anti-cMBD2 antibodies recognize a protein at 28 kD in multiple chicken cell types. A) Western blot with anti-cMBD2 antibody VCU1 on HeLa nuclear extract, recombinant GST/cMBD2, chicken embryonic fibroblast (CEF) nuclear extract, chicken pre-B cell (DT40) nuclear extract, and chicken red blood cell nuclear extract. B) Western blot with anti-cMBD2 antibody VCU2 with samples as above. Only the band at 28 kD was abolished by incubation of the antibody with an excess of recombinant GST/cMBD2.

retardation ("supershift") of the MeCPC complex's mobility (see Figure 9). These experiments confirm that MBD2 is a component of MeCPC complex.

Discussion

We have identified the chicken homolog of mammalian MBD2 as a 28 kD protein expressed in fibroblasts as well as in hematopoietic cells. Interestingly, cMBD2 corresponds closely to the 'b' isoform of MBD2 found in mammalian cells (Hendrich and Bird, 1998). This form is differentiated from the 'a' isoform by the lack of 152 amino acids of N-terminal polypeptide and is thought to arise from translational initiation at an internal methionine (both forms contain the same C-terminal 261 amino acids). Currently, no studies have carefully examined the relative amounts, biochemistry or function of the MBD2a and b isoforms *in vivo*. The MBD is wholly contained within the shared C-terminal region and thus it is not surprising that it is this region that is conserved between chicken and mammalian MBD2 proteins. The absence of this region is not unique to the chicken, as MBD2 in other vertebrates (*X. laveis*, *D. rerio*) are also missing this N-terminal extension (Kransdorf and Ginder, unpublished observations).

Whatever function the N-terminal extension present in mammalian MBD2 has, it is either lacking in chicken or is provided for by other proteins and mechanisms. The N-terminal region of mammalian MBD2 contains a glycine-arginine repeat domain (GR_{11}). Two observations have been made regarding this domain. First, Jeffery and Nakielny have found that this region contains a RNA binding domain that functions independently of the MBD (Jeffery and Nakielny, 2004). Second, Fujita *et al.* have observed that the

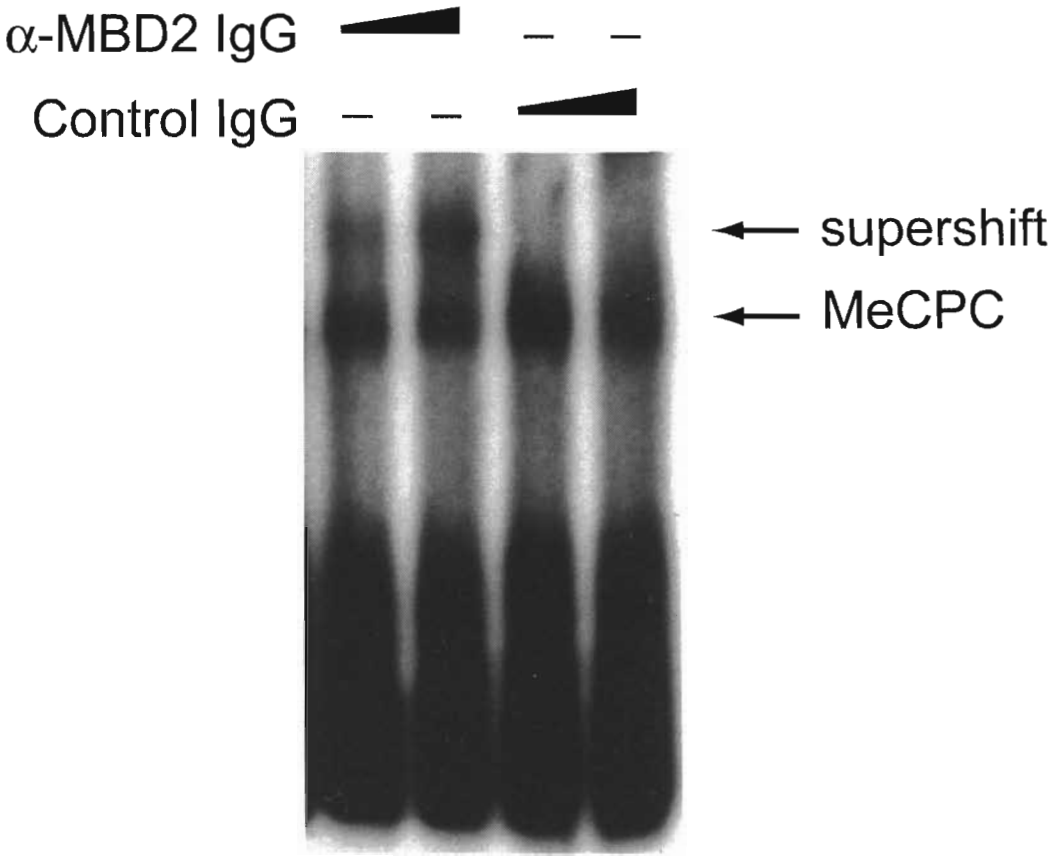


Figure 9. Anti-cMBD2 antibodies supershift the MeCPC. Inclusion of VCU2 anti-MBD2 IgG, but not control IgG, retarded the mobility of the MeCPC complex ('supershift') during EMSA.

GR_{11} domain interacts with RNA helicase A, and that this interaction can activate transcription of unmethylated, cAMP-responsive genes (Fujita et al., 2003). These data indicate that the RNA binding domain of MBD2 is involved in transcriptional activation.

Using rcMBD2, we have also investigated the methyl-CpG requirements for complex formation. We observed that rcMBD2 is able to bind to 50 bp fragments of the ρ-globin gene containing as few as one methyl-CpG. This is in agreement with the results of Fraga *et al.*, but contradicts the requirement for 10-15 methyl-CpGs found by Meehan *et al.* The disparity is explained by the methyl-CpG density-dependence of MBD2 binding, such that at low concentrations of the MeCP1 complex, the amount of probe shifted by MeCP1 on a probe containing a single methyl-CpG would be below the level of detection.

If MBD2 only requires one methyl-CpG for binding, why are entire CpG islands that contain multiple methyl-CpGs methylated in the cell? For example, the ρ-globin PTR region contains 14 methyl-CpGs. Two rationales could explain this observation. First, if MBD2 binding is methyl-CpG density-dependent, then an increased number of methyl-CpGs in a genetic region will increase the number of MBD2-recruited transcriptional repression complexes in the region. This in turn would ensure transcriptional repression of the methylated region. Second, methylation at multiple CpGs is most likely required to maintain the integrity of the DNA methylation pattern during DNA replication. That is, maintenance methylation that occurs during DNA replication is slightly error-prone and methylation at regions with low methylation density can be lost (Lorincz et al., 2002). Methylation extending over numerous CpGs

requires that the overall methylation pattern will be maintained even if methylation at a small number of sites is lost.

Chapter 3: Purification and component identification of the Erythroid Methyl-Cytosine Binding Protein Complex (MeCPC)

Introduction

Transcription is a highly complex process necessitating numerous enzymatic activities to be present at a single genetic site. To meet this need, cells possess large protein complexes composed of numerous subunits that function coordinately. For example, RNA polymerase II (RNAPII) itself is a complex of 12 proteins (Young, 1991). *In vitro*, transcription requires two additional large protein complexes: the general transcription factors and the mediator complex (Hampsey, 1998). The general transcription factors TFIIB, TFIID, TFIIE, TFIIF and TFIIH, each of which is a distinct complex of proteins, mediate assembly of the transcription preinitiation complex. The mediator complex is necessary for response to transcriptional activators and is composed of 25 subunits in yeast (Guglielmi et al., 2004).

Protein purification by liquid chromatography has served as the major route to understanding the subunit composition of transcriptional protein complexes. Protein purification is necessary for several reasons. First, most transcriptional activities are present in low concentrations in crude cellular preparations and therefore purification is necessary to increase the concentration of the activity to sufficient levels for functional analysis and subunit identification. Second, purification is necessary to remove proteins

not involved in the activity under study, so that the complexity of the final purified sample is reduced. This facilitates the identification and characterization of proteins in the final purified sample.

As discussed in Chapter 1, the MeCP1 complex was the first methyl-CpG binding activity to be described (Meehan et al., 1989). This activity was identified using an artificial ^{m}CpG-rich (24 ^{m}CpG / 135 bp of DNA) probe termed M-CG11. The activity was found to be present in nuclear extracts from several cell types of cells, including HeLa cells. Subsequently, Feng and Zhang purified MeCP1 by liquid chromatography and found the complex to be composed of MBD2 and all the subunits of the NuRD complex (Feng and Zhang, 2001).

Work in our laboratory has shown that a similar complex, the MeCPC, forms on the methylated ρ-globin gene *in vitro* (Singal et al., 2002). This complex exhibits electrophoretic properties similar to those of MeCP1: an inability to enter a polyacrylamide gel and a similar migration distance in agarose gels during EMSA. As shown in Chapter 2 and previously by our laboratory, supershift data support the presence of MBD2 in the MeCPC (Singal et al., 2002). Despite these similarities, the MeCP1 complex displays a marked preference for its canonical binding sequence M-CG11, than for M-ρ248. Similarly, the MeCPC complex exhibits a greater DNA-binding affinity for its canonical binding sequence M-ρ248, than for M-CG11, in spite of a lower ^{m}CpG content in M-ρ248 (14 ^{m}CpG / 248 bp of DNA). Based on these data, we hypothesize that a factor or factors present in the MeCPC complex but not in the MeCP1 complex, confers sequence-context preferential binding to the MeCPC.

Chicken erythrocytes are an excellent system in which to study transcriptional protein complexes. Unlike the erythrocytes of mammals, chicken erythrocytes do not undergo enucleation during maturation. Thus, transcriptional protein complexes remain in the nucleus of these cells despite the fact that the mature erythrocytes are largely transcriptionally inactive (Berkowitz and Doty, 1975). Given this fact, using chicken erythrocytes as a source of nuclear protein offers several advantages over cultured cell lines. First, very large quantities of chicken blood can be obtained as a by-product of chicken agriculture. Second, vast quantities of chicken erythrocytes can be obtained from chicken blood. Typically 0.5 L of blood yields 60 mL of packed erythrocytes, which when used as raw material for the Dignam nuclear extract procedure generates approximately 1.6g of nuclear extract (Kransdorf and Ginder, unpublished observations). Even after deduction of the high quantities of hemoglobin in the extract (approximately 95% by mass), every 0.5 L of chicken blood yields around 80 mg of non-hemoglobin nuclear protein. Third, this system is significantly more cost effective than using cultured cell lines as the raw material for the Dignam nuclear extract procedure (the approximate cost for 80 mg of HeLa nuclear extract is $1600, compared to $100 for chicken erythrocytes). Most importantly, this system allows us to address numerous interesting and medically relevant biological questions on the mechanisms of globin regulation in erythroid cells.

Methods

Note that for the buffer compositions given below, the buffer is denoted by a letter followed by a number which indicates its salt content in millimoles (*e.g.* Z100 denotes buffer Z containing 100 mM KCl).

Electrophoretic Mobility Shift Assay

The 248 bp ρ exon 1/intron 1 DNA fragment (ρ248) was generated by digestion of control DNA with NheI and StuI. The 135 bp CG11 DNA fragment (CG11) was generated by digestion of pCG11 with EcoRI and BamHI. The probe fragments were isolated by gel purification (Qiagen). The isolated fragments were then methylated using SssI methylase (New England Biolabs). The DNA was phenol-chloroform extracted and precipitated using glycogen as a carrier. The probe was labeled by incubation with [α-^{32}P]dCTP or [α-^{32}P]dATP and Klenow fragment. The assay was performed by adding 2-10 μL of column fractions to a 21.5 μL binding reaction which contained 10 or 20 fmol of probe, 2 μg of sonicated *M. Lys* DNA, 20 mM HEPES [pH 7.9], 3 mM $MgCl_2$, 1 mM EDTA, 10 mM β-ME, 0.1% Triton X-100 and 4% glycerol. Binding reactions were incubated on ice for 20 minutes, loaded onto a 2% agarose gel and resolved in 0.5x TBE buffer.

Isolation of Chicken Erythrocytes

500 mL whole chicken blood with sodium citrate as an anti-coagulant was obtained (Pel-Freez). 32 mL of blood was layered on top of 12 mL of Histopaque 1083 (Sigma) in a 50

mL conical (Greiner). 15 conicals were prepared in this way from each 500 mL of blood. The conicals were spun at 3100g for 5 minutes at 4°C in a RT7 Plus centrifuge using the RTH-250 rotor (Sorvall). The supernatant was then aspirated with a vacuum pump, leaving a pellet of 8 mL of erythrocytes. The erythrocytes were resuspended in 1x cold PBS and aliquoted to 12 centrifuge tubes (Nalgene). The tubes were spun at 3000g for 5 minutes at 4°C in a J2-21M centrifuge using the JS 13.1 rotor (Beckman). The volume of this packed erythrocyte pellet was 5 mL per tube, or 60 mL per 500 mL of blood.

Preparation of Nuclear Extracts

Nuclear extracts were prepared using the method of Dignam *et al.* with modifications (Dignam et al., 1983). Briefly, 5 mL of packed erythrocytes were resuspended with 25 mL of Buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl) and incubated for 10 minutes on ice. The tubes were then spun at 4000 RPM for 10 minutes at 4°C in the JS 13.1. The supernatant was removed with a vacuum pump and the pellet was resuspended with 10 mL of Buffer A. 3 tubes of pellet were transferred to a 40 mL dounce (Wheaton) and the cells were fully lysed with 15 strokes of a 'B' type pestle. The cells were then aliquoted to 6 centrifuge tubes and spun at 10000g for 10 minutes at 4°C in the JS 13.1. The supernatant was again removed with a vacuum pump and the pellet was resuspended in 1 volume (~ 14 mL) of Buffer C (20 mM HEPES pH 7.9, 25% glycerol, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA) and incubated for 35 minutes on ice with gentle vortexing every 5 minutes. The nuclei were spun out at 10000g for 10 minutes at 4°C in the JS 13.1. The nuclear extract was removed and spun at 15000g for

20 minutes at 4°C to pellet any cellular debris. Nuclear extract was then frozen at –80°C until use.

MeCPC Purification Strategy I: Ion-Exchange Chromatography

All Fine Performance Liquid Chromatography (FPLC) was performed using a BioLogic DuoFlow system (Bio-Rad). 120 mL of nuclear extract was desalted into buffer Z100 (20 mM Tris-HCl pH 7.9 @ 4°C, 100 mM KCl, 0.2 mM EDTA, 10% glycerol, 0.5 mM DTT, 0.2 mM PMSF) with a HiPrep Desalting 16/20 column (Amersham Biosciences) and then loaded in tandem onto a 20 mL DEAE Sepharose Fast Flow column (Amsersham Biosciences). The column was washed with 2.5 column volumes (CV) of Z100 and then bound protein was eluted with a 2.5 CV linear gradient from Z100 to Z500. The active fractions were determined by EMSA. The DEAE column step was repeated resulting in two sets of DEAE fractions, which were then pooled (total volume = 43.5 mL). The pooled DEAE fractions were desalted into Z100 with the HiPrep Desalting 16/20 and then one-half of the total desalted protein was injected onto a UnoQ6 column (Bio-Rad). The column was not washed and then bound protein was eluted with an 8 CV linear gradient from Z100 to Z500. The active fractions were determined by EMSA. The UnoQ column step was repeated with the remaining protein, resulting in two sets of UnoQ fractions, which were then pooled (total volume = 22 mL).

MeCPC Purification Strategy I: Gel Filtration Chromatography

The pooled UnoQ fractions were aliquoted to two Ultrafree-15 centrifugal concentrator devices with a 5K molecular-weight cut-off (MWCO, Millipore). The protein was concentrated by spinning the devices in the JA-17 rotor at 3700 RPM for 120 minutes at 10°C (final sample volume = 1 mL). 240 µL of the final concentrated sample was injected onto a Superose 6 10/30 column (Amersham Biosciences) washed with Buffer S500 (20 mM HEPES pH 7.9, 500 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA, 5% glycerol, 0.2 mM PMSF, 0.5 mM DTT). Proteins were eluted isocratically, with a flow rate of S500 at 0.2 mL / min. This procedure was repeated an additional two times to accommodate the rest of the concentrated Q pooled sample. Fractions 8 through 23 were aliquoted in pairs (*e.g.* fraction 8 and 9) to Ultrafree-4 centrifugal concentrator devices with a 5K MWCO (Millipore). The protein was concentrated by spinning the devices in the JA-13.1 rotor at 4500 RPM for 20 minutes at 4°C. The active fractions were determined by EMSA and pooled (total volume = 0.6 mL).

MeCPC Purification Strategy I: Heparin Chromatography

The pooled Superose 6 fractions were dialyzed against a 400x volume of buffer H150 (10 mM Sodium Phosphate pH 6.8, 150 mM NaCl, 0.2 mM EDTA, 10% glycerol, 1 mM DTT). The sample was injected onto a 1mL HiTrap Heparin Sepharose High Performance column (Amersham Biosciences). The column was washed with 4 CV of H400 and then bound protein was eluted with an 8 CV linear gradient from H400 to H900. The active fractions were determined by EMSA and pooled (total volume = 2.4

mL, total protein = 360 μg). 30 μg of pooled, heparin-purified sample was dialyzed against PTE buffer (20 mM Tris-HCl pH 7.8, 0.5 mM EDTA, 1 mM DTT) overnight.

MeCPC Purification Strategy II: Ion-Exchange Chromatography

The second strategy for purification of the MeCPC was similar to the first, except that the DEAE column step was repeated four times resulting in four sets of DEAE fractions (DEAE Runs I – IV). MeCPC containing fractions from DEAE Run I and II were pooled and desalted into H100 with the HiPrep Desalting 16/20 and then injected to a 20 mL HiPrep SP Sepharose Fast Flow column (Amersham Biosciences). The column was washed for 0.25 CV with H100 and then bound protein was eluted with a 4 CV linear gradient from H100 to H600. The column was washed with 8 CV of H100 and then fractions from Run III and IV were pooled and the above procedure repeated. The active fractions from each SP run were determined by EMSA. The SP fractions were pooled and then desalted into Z100 with the HiPrep Desalting 16/20 and then one-half of the total desalted protein was injected onto a UnoQ6 column (Bio-Rad). The column was not washed and then bound protein was eluted with an 8 CV linear gradient from Z100 to Z500. The active fractions were determined by EMSA. The Q column step was repeated with the remaining protein, resulting in two sets of Q fractions, which were then pooled. Gel filtration was then performed on the pooled Q fractions as described for purification Strategy I above.

Protein Electrophoresis and Staining

To remove any contaminating substances such as glycerol and salt prior to protein electrophoresis, 30 μg of heparin-purified sample was precipitated with the ReadyPrep 2-D Cleanup Kit (Bio-Rad). The precipitated protein was resuspended in 40 μL of LDS-PAGE sample buffer (Invitrogen). The sample was heated at 70°C for 10 minutes and then loaded to two wells of a NuPAGE Novex 10% Bis-Tris gel (Invitrogen). The gel was run using MOPS running buffer at 180V for one hour. The gel was then removed and fixed in 100 mL of 10% methanol, 7% acetic acid for 30 minutes. The gel was stained in 100 mL SYPRO Ruby (Molecular Probes) protein stain overnight at room temperature with gentle agitation. The next day the stain was removed and the gel was washed twice with 100 mL of 10% methanol, 7% acetic acid for 30 minutes.

Identification of Proteins by Mass Spectrometry

The entire gel was submitted to the Michael Hooker Proteomics Core Facility at the University of North Carolina Chapel Hill. Proteins were digested in-gel with trypsin. Gel bands up to 90 kD were excised by robot and labeled A1 through A12, and B1 through B9. Peptide mass fingerprint data was acquired on the spotted digests using a model 4700 Proteomics Discovery System (Applied Biosystems).

Mass Fingerprint Data Analysis

Overall there were 18 bands that were submitted for identification. Peptide mass fingerprint (PMF) data was first submitted to the MS-Fit algorithm that is part of Protein Prospector (http://prospector.ucsf.edu/). Protein Prospector identified four bands based

on the high homology of these chicken proteins with the corresponding human proteins. For the unidentified proteins, the PMF data set for each band was stripped of peptides derived from human keratin and bovine trypsin as identified with FindPept (http://au.expasy.org/tools/findpept.html) using a 10 ppm tolerance (Gattiker et al., 2002). The PMF data set for each band was then separated into 2 to 3 lists of peptides (depending on the length) and submitted to ChickPep (http://chick.umist.ac.uk/). ChickPep searches were performed using the following parameters: monoisotopic mass, 0 charge, 0-1 maximum missed cleavages, 80 ppm error tolerance and variable methionine oxidation. The resulting list of cDNA contigs from each ChickPep query was examined to find those contigs whose putative protein had a molecular weight similar to that (within 5 kD) of the band the PMF data set was derived from. The resulting "candidate proteins" were used as queries to the Ensembl Chicken Genome Annotation (ECGA, http://www.ensembl.org) to find the chicken gene for each candidate protein. The "reference protein" obtained from the ECGA for each candidate protein was digested with trypsin *in silico* and compared to the experimentally-derived PMF data set for each band using FindPept. FindPept comparisons were performed using the following parameters: monoisotopic mass, 80 ppm error tolerance, methionine oxidation and trypsin digestion. The candidate protein that matched the experimentally-derived PMF data set most closely (*i.e.* the largest number of matching peptides) was considered the match. Only peptides resulting from 0, 1, or 2 missed trypsin cleavages were considered.

Western Blotting

Protein samples were run on Criterion 10% Tris-glycine SDS-PAGE gels (Bio-Rad) in running buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS). The gel was equilibrated in transfer buffer (25 mM Tris-HCl pH 8.0, 192 mM glycine, 20% methanol) and then transferred to activated PVDF overnight at 30V. The next day the membrane was washed twice for 10 minutes with 1x PBS and then blocked for two hours at room temperature with rocking in 10 mL MPBST (5% (w/v) non-fat milk, 1x PBS, 0.05% Tween-20). The blocking solution was removed and the primary antibody was incubated with the membrane for 2 hours at room temperature with rocking in MPBST. This solution was removed and the membrane was washed 3 times with 10 mL PBS-T for 5 minutes. The secondary antibody was incubated with the membrane for 1 hour at room temperature with rocking in MPBST. This solution was removed and the membrane was washed three times with 10 mL PBS-T for 5 minutes and 1 time with 10 mL PBS-T plus 0.1% Triton X-100. Bands were visualized with ECL Plus (Amersham Pharmacia) and exposed to film.

Results

To understand the binding specificity and functional capabilities of the MeCPC, we first sought to determine what the protein components of this complex are. The strategy we chose to accomplish this was to first purify the MeCPC by chromatography and then identify the proteins in the sample using MALDI-TOF mass spectrometry. The initial step in this strategy was to identify chromatographic columns that allowed us to elute acceptable amounts of MeCPC activity from the column. We found that the DEAE

Sepharose Fast Flow, SP Sepharose Fast Flow, UnoQ, Heparin Sepharose High Performance and Superose 6 matrices yielded purification of the MeCPC with acceptable amounts of loss (less than 50% of input). In contrast, the Phenyl Sepharose High Performance, Phenyl Superose, and MonoS matrices yielded very low amounts of MeCPC and thus were not valuable chromatography steps. It is noteworthy that the Phenyl Sepharose and MonoS matrices were employed successfully in the purification of the MeCP1 complex (Feng and Zhang, 2001). This disparity may be another indication that the MeCPC and MeCP1 are distinct protein complexes.

Based on the high fold-purification achievable with affinity chromatography (Gadgil et al., 2001), we also tried developing two separate affinity chromatography procedures to purify the MeCPC. The first procedure utilized biotinylated M-ρ248 as the affinity target, bound to a strepavidin-agarose matrix. After scaling-up the procedure so that the matrix could be placed in a 1 mL column, we found that the MeCPC failed to bind to the matrix at even low salt concentrations (data not shown). The second procedure utilized biotinylated M-ρ248 as the affinity target, but used magnetic strepavidin beads as the matrix. We were not able to elute the MeCPC off the beads once it had bound (data not shown). Therefore we were not able to use either of these affinity chromatography procedures to purify the MeCPC.

Using the chromatography columns we previously identified as able to successfully elute acceptable amounts of MeCPC activity, we developed two multi-step

chromatographic procedures to purify the MeCPC from adult chicken red blood cell nuclear extracts. The two procedures used are diagrammed in Figure 10. EMSA was performed on the eluate from each column to determine which fractions contained the MeCPC activity. These fractions were pooled and then injected to the next column step. The total fold-purification for each procedure is difficult to estimate, because there is a non-linear relationship between the activity concentration and the amount of probe shifted in the EMSA. Nevertheless, the total fold-purification for Purification Strategy I is estimated to be approximately 100-fold as determined by EMSA.

The first step of each procedure was a tandem injection of chicken adult red blood cell nuclear extract to a HiPrep Desalting column and then to a 20 mL DEAE Sepharose Fast Flow column. The DEAE Sepharose column was an excellent capture step, because as a weak anion-exchanger, the DEAE binds to the nuclear protein within the extract, while the large amount of hemoglobin within the extract flows through the column. In purification Strategy I, the MeCPC containing fractions from two DEAE runs were pooled and injected to a 6 mL UnoQ column, a strong anion-exchanger. The MeCPC containing fractions from two UnoQ runs were pooled and injected to a 24 mL Superose 6 gel-filtration column. The MeCPC eluted from the Superose 6 column between the void volume (2000 kD) and 667 kD. The MeCPC containing fractions were pooled and injected to a 1 mL Heparin Sepharose High Performance column.

The EMSA on the eluted fractions from the Heparin Sepharose column of purification Strategy I is shown in Figure 11. The MeCPC elutes in fractions 32 through 36 and displays strong methyl-CpG binding activity after four column chromatography

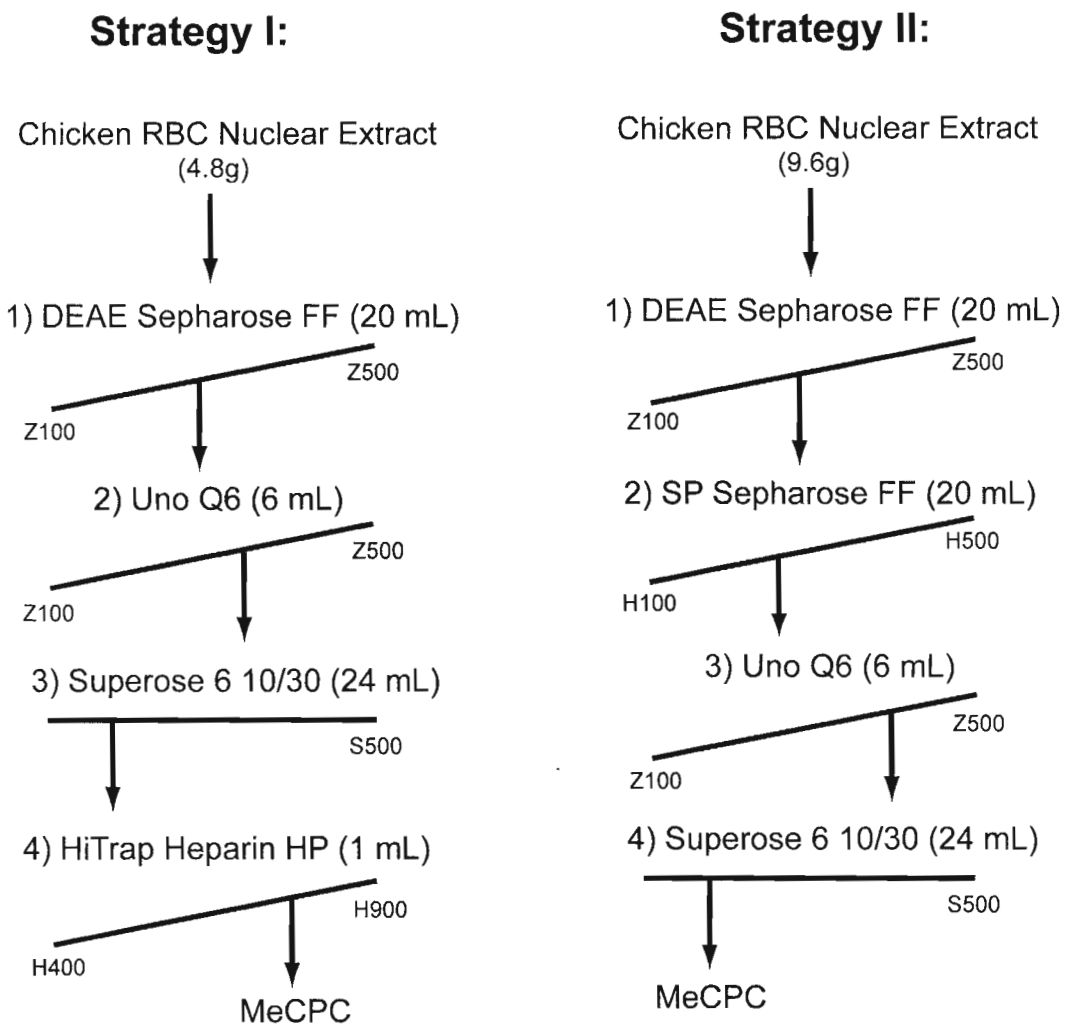


Figure 10. Chromatographic strategies used to purify the MeCPC from primary erythroid cells. Purified material resulting from Strategy I was analyzed by mass spectrometry and purified material resulting from Strategy II was analyzed by Western blotting with previously identified MeCPC components.

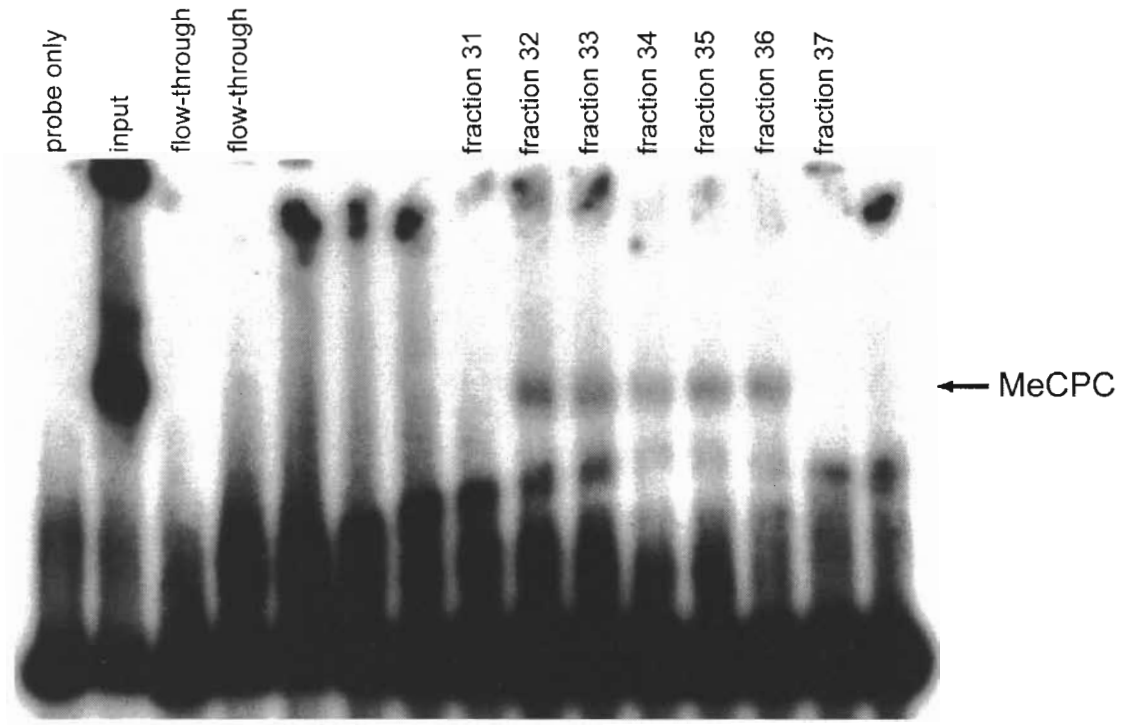


Figure 11. EMSA on the eluted fractions from the final column of MeCPC purification Strategy I (Heparin Sepharose High Performance). M-ρ248 was used as the probe. The MeCPC complex elutes in fractions 32 through 36. After four column chromatography steps the complex remains intact.

steps. To identify the proteins present in this sample, the fractions were pooled and separated by electrophoresis on a 10% SDS-PAGE gel. The gel was stained with the Sypro Ruby protein gel stain. The final protein gel showed approximately 40 protein bands of varying staining intensity, with half the bands under 85 kD. The gel was sent to the Michael Hooker Proteomics Core Facility at the University of North Carolina Chapel Hill for identification of the 18 discrete bands under 85 kD via matrix assisted laser desorption ionization-time of flight-time of flight (MALDI-TOF-TOF) mass spectrometry. Identification of the bands in the sample over 85 kD was not possible under these conditions, because electrophoresis on a 10% gel does not separate proteins over 100 kD sufficiently to permit excision of each band separately. A picture of the stained SDS-PAGE gel submitted for analysis by mass spectrometry can be found as Figure 12.

The Proteomics Core Facility at the University of North Carolina Chapel Hill was able to identify four proteins in the sample by querying Protein Prospector with the peptide mass fingerprint (PMF) data obtained by MALDI-TOF mass spectrometry. Protein Prospector is a set of bioinformatics algorithms designed primarily for the identification of proteins by mass spectrometry (Clauser et al., 1999). Protein Prospector searches standard protein sequence databases, such as Swiss-Prot (Bairoch and Boeckmann, 1991) and OWL (Bleasby et al., 1994), that contain human and mouse proteins almost exclusively. Only a small percentage of proteins within these protein databases is from the chicken. For example, there are over 250,000 human and mouse proteins in Swiss-Prot, but only 4000 chicken proteins. Thus the chicken proteins

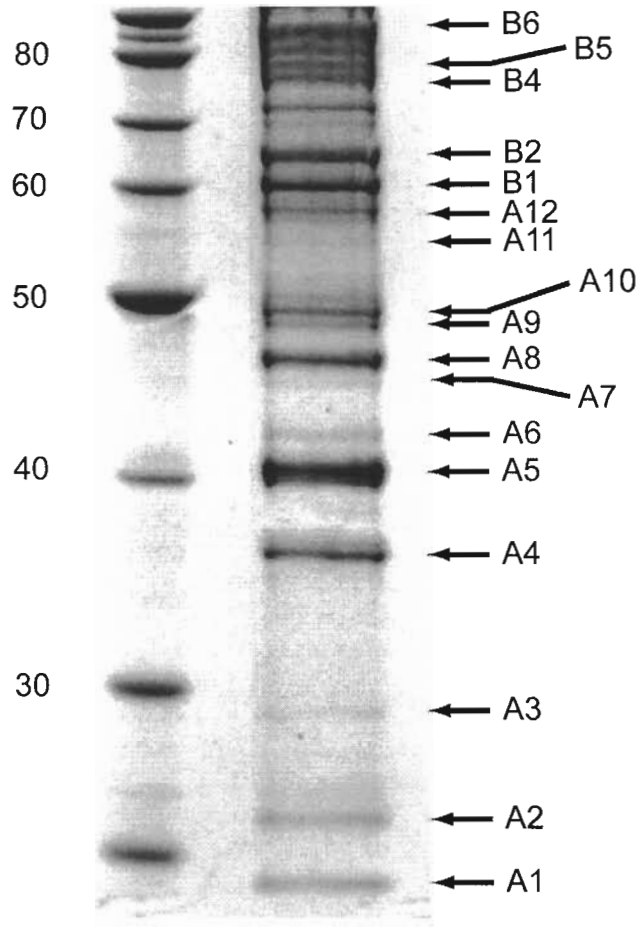


Figure 12. Sypro Ruby stained protein gel on 15 μg of purified MeCPC. The gel was submitted to the Proteomics Core Facility at the University of North Carolina Chapel Hill for analysis by mass spectrometry. Approximately 40 bands are present on the gel, with 18 bands below 85 kD. Since the bands above 85 kD were not resolved, mass spectrometry was performed on the bands below 85 kD only. Only the bands analyzed by mass spectrometry are shown.

identified from searches with Protein Prospector were identified because they are highly homologous to the corresponding human and mouse proteins within the Swiss-Prot and OWL databases. The average amino acid similarity between the human and chicken proteins was 95%, suggesting that it is possible to identify chicken proteins from their human homologs as long as the amino acid similarity between the homologs is quite high. The other 14 bands were not identified from the Protein Prospector searches because the matching chicken protein sequence was not present in the standard protein sequence databases.

There are currently two resources for chicken proteomic information: the BBSRC Chicken EST Databse (Boardman et al., 2002) and the ECGA (Birney et al., 2004; Curwen et al., 2004). At this time there is only one algorithm that allows the user to query a database of chicken proteins with PMF data in order to match that query to a chicken protein. This is the ChickPep algorithm implemented by Ian Overton and Simon Hubbard at the University of Manchester. ChickPep takes as input a set of PMF data and produces as output a list of cDNA contigs whose predicted open reading frames (ORF) match the query data set. In addition, the BBSRC Database entry for each cDNA contig contains the BLAST data for the predicted ORF of the cDNA, allowing comparison of the predicted ORF to its likely homolog. For example, cDNA contig 345208.3 from the BBSRC Database encodes an ORF of 307 amino acids that when queried against the SwissProt protein sequence database by BLAST, matches mouse DNMT3a. In this way, the "putative protein" encoded by each cDNA contig in the database is known.

Each data set from the 14 unidentified protein bands was analyzed by the method outlined in Figure 13. First, the data set was stripped of peptides derived from human keratin and bovine trypsin using the program FindPept (Gattiker et al., 2002). Next the data set was split into two or three lists of approximately 40 peptides (this is the maximum number of peptides that can be entered into ChickPep at one time). Each list was entered into ChickPep, and the resulting list of cDNA contigs from each ChickPep query was examined to find those contigs whose putative protein had a molecular weight similar to that (within 5 kD) of the band from which the PMF data set was derived. The resulting "candidate proteins" were used as queries to the ECGA to find the chicken gene for each candidate protein. This last step was necessary because many of the ORFs predicted for cDNA contigs in the BBSRC Chicken EST Database are incomplete or contain errors derived from DNA sequencing. Each gene in the ECGA is linked to its predicted transcripts and protein products; consequently, if the gene is present in the ECGA database, then the complete "reference protein" encoded by the gene is present.

The candidate protein or proteins suggested from each ChickPep search were then evaluated. The reference protein obtained from the ECGA for each candidate protein was digested with trypsin *in silico* and compared to the experimentally-derived PMF data set for each band. The candidate protein that matched the experimentally-derived PMF data set most closely (*i.e.* the largest number of matching peptides) was considered to be the match. An arbitrary minimum of five matching peptides was required to be considered as a positive match. Table 4 contains the final protein match for each band in the gel which

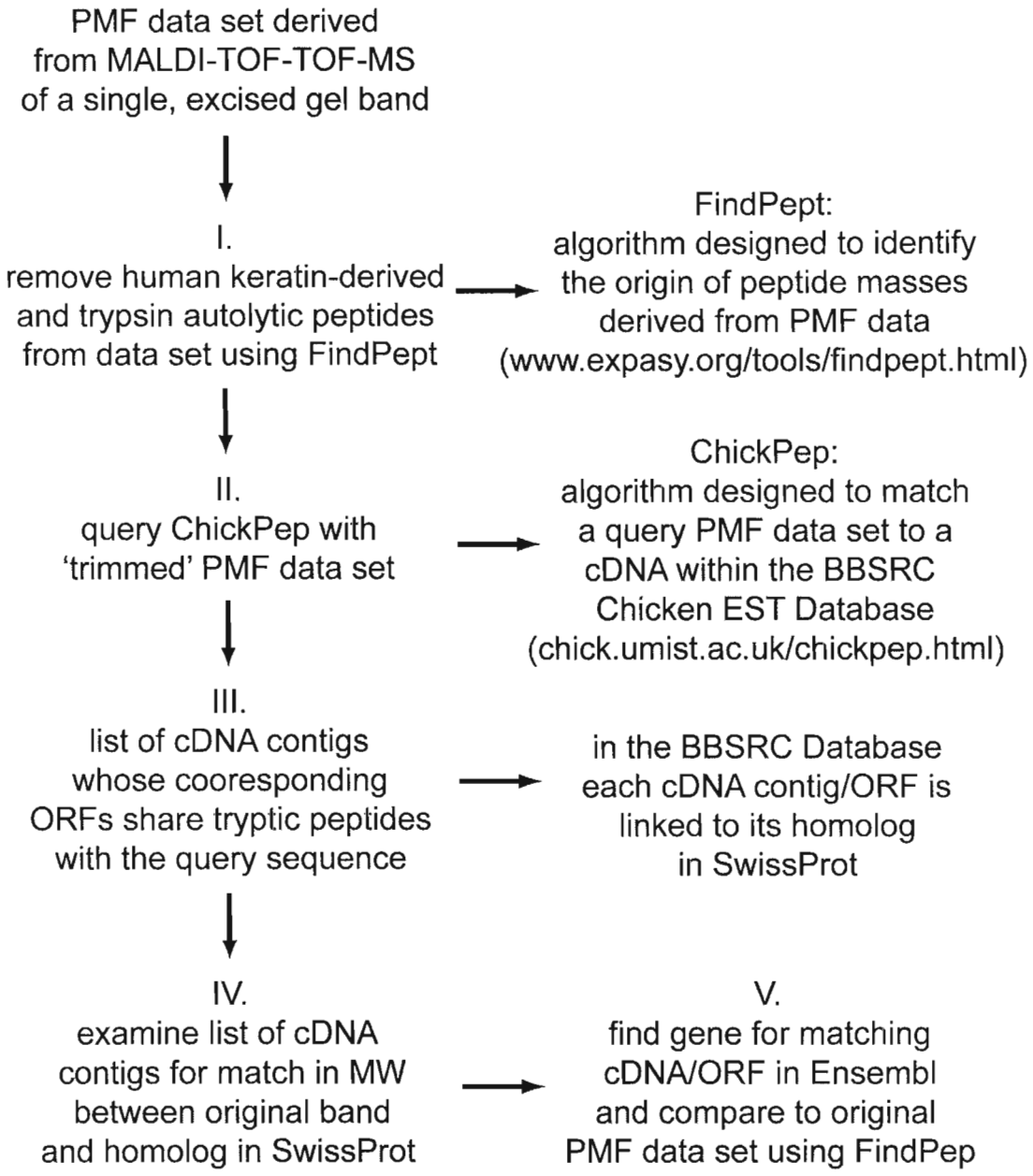


Figure 13. Method for peptide mass fingerprint data analysis using ChickPep.

was submitted to the Proteomics Core Facility at the University of North Carolina Chapel Hill.

Overall, 16 of the 18 submitted proteins were matched to a chicken protein. Only bands A1 and A2 were not successfully matched. 14 of the 16 identified proteins were matched to a reference protein present in the ECGA. Bands A3 and A11 were matched to predicted ORFs present in the BBSRC Database, because the genes for these proteins were not present in the ECGA. The stained SDS-PAGE gel submitted for MS analysis with the identified proteins labeled can be found in Figure 14.

Four components of the MeCP1 transcriptional repression complex were identified in the purified MeCPC sample: MBD2, HDAC2, RbAp48 and MTA1 (Feng and Zhang, 2001). In light of our previous data indicating that MBD2 was a component of the MeCPC (Figure 9 in Chapter 2) and published data which showed the association of these proteins *in vivo*, these four proteins were considered putative components of the MeCPC. Another identified protein, MENT, was also considered to be a putative MeCPC component, due to published data showing it is highly expressed in chicken erythrocytes and lymphoid cells, is capable of binding to DNA, and can establish a repressive chromatin configuration (Grigoryev and Woodcock, 1993; Grigoryev and Woodcock, 1998; Istomina et al., 2003). Furthermore, the cohesin complex subunit SCC1 (as well as a SCC1 cleavage product) was identified, indicating that the SNF2h/NuRD/cohesin complex may be present in these fractions (Hakimi et al., 2002).

Five components of the eukaryotic translation Initiation Factor 3 complex (eIF-3) were also identified: eIF-3d, eIF-3e, eIF-3h, eIF-3i, and an eIF-3e interacting protein

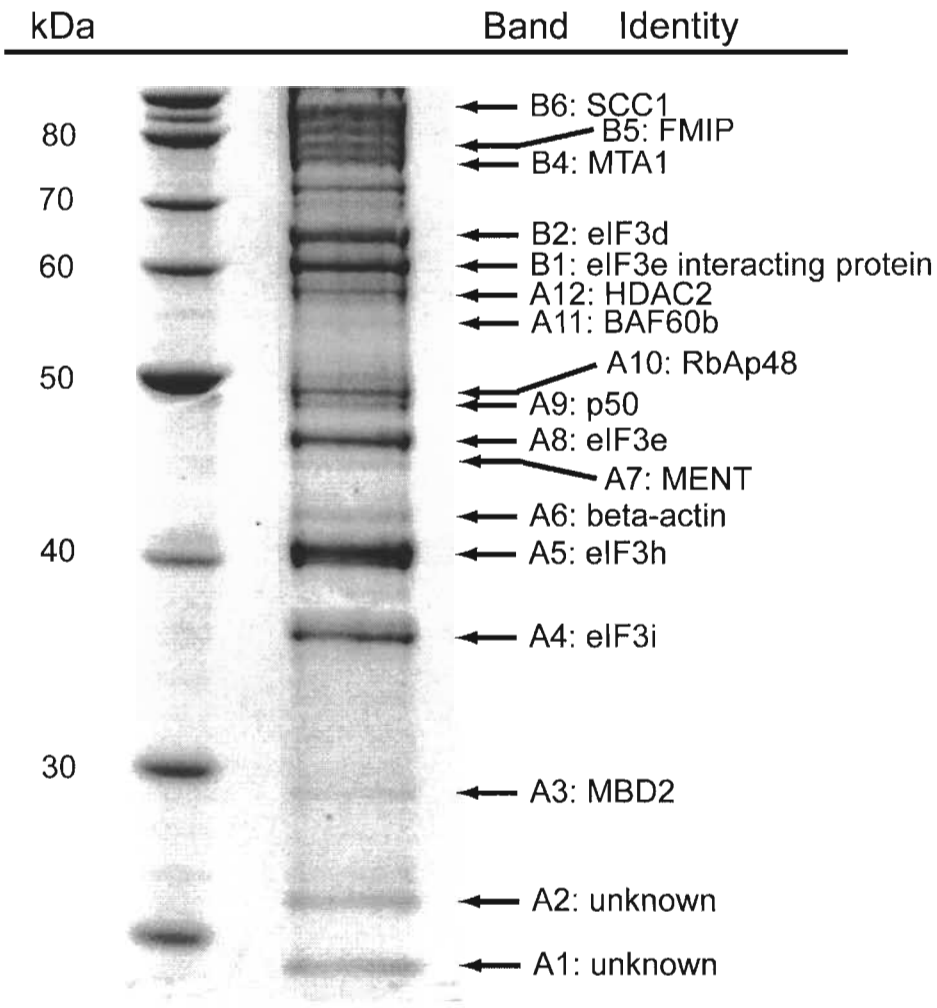


Figure 14. Sypro Ruby stained protein gel on 15 μg of purified MeCPC showing the identity of the bands in the sample.

Table 4. List of proteins identified by mass spectrometry from the purified MeCPC sample. Proteins were identified by the method outlined in Figure 13. The final step of the method was to compare tryptic peptides from *in silico* digestion of candidate proteins from each ChickPep search to the the experimentally-derived PMF data set for each band. The candidate protein that matched the experimentally-derived PMF data set most closely (*i.e.* the largest number of matching peptides) was considered the match. An arbitrary minimum of five matching peptides was required to be considered as a positive match and only peptides matching with 0, 1 or 2 missed trypsin cleavages were counted.

Band	Apparent M_r	Protein Identification	Abbreviation	Database / ID	Theoretical M_r	Matching Peptides
A1	20	NO MATCH				
A2	22	NO MATCH				
A3	28	METHYL-CpG BINDING DOMAIN PROTEIN 2	MBD2	BBSRC 337626.1	28	5
A4	36	EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 2	eIF-3i	Ensembl ENS 5271	37	9
A5	40	EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 3	eIF-3h	Ensembl ENS 25924	40	12
A6	42	BETA-ACTIN		Ensembl ENS 02101	42	6

Table 4 continued.

Band	Apparent M_r	Protein Identification	Abbreviation	Gene ID	Theoretical M_r	Matching Peptides
A7	45	MYELOID AND ERYTHROID NUCLEAR TERMINATION	MENT	Ensembl ENS 20960	47	9
A8	46	EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 6	eIF-3e	Ensembl ENS 25882	52	20
A9	50	UNCHARACTERIZED CONSERVED PROTEIN	p50	Ensembl ENS 19305	56	6
A10	50	CHROMATIN ASSEMBLY FACTOR 1 SUBUNIT C	RbAp48	Ensembl ENS 5565	51	10
A11	56	BRAHMA RELATED GENE ASSOCIATED FACTOR 60b	BAF60b	BBSRC 045096.1	57	18
A12	58	HISTONE DEACETYLASE 2	HDAC2	Ensembl ENS 24133	55	19

Table 4 continued.

Band	Apparent M_r	Protein Identification	Abbreviation	Gene ID	Theoretical M_r	Matching Peptides
B1	62	EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 6 INTERACTING PROTEIN		Ensembl ENS 20064	66	23
B2	66	EUKARYOTIC TRANSLATION INITIATION FACTOR 3, SUBUNIT 7	eIF-3d	Ensembl ENS 20439	60	16
B3	72	SISTER CHROMATID COHESION 1 (CLEAVAGE PRODUCT)		Ensembl ENS 25921	73	6
B4	75	METASTASIS ASSOCIATED 1	MTA1	Ensembl ENS 4424	79	14
B5	78	FMS INTERACTING PROTEIN	FMIP	Ensembl ENS 13104	79	20
B6	85	SISTER CHROMATID COHESION 1	SCC1	Ensembl ENS 25921	73	31

(Hershey et al., 1996; Morris-Desbois et al., 2001). Because the eIF-3 complex has never been shown to bind to DNA or have a role in transcriptional regulation, it is likely that this complex is copurifying with the putative MeCPC components. Other proteins in the sample that are most likely copurifying with the putative MeCPC components are actin, p50, BAF60b and FMIP. Actin has been shown to have a role in RNAPII pre-initiation complex formation (Hofmann et al., 2004), but has not been shown to have a role in transcriptional repression. The protein p50 is an uncharacterized conserved protein that possesses a tetratricopeptide repeat domain but contains no significant sequence homology to any characterized proteins present within the standard protein sequence databases. BAF60b is a component of a SWI/SNF chromatin remodeling complex (Wang et al., 1996), but since no other components of this complex were identified in the MeCPC sample, its presence is unlikely to be of functional significance. FMIP is a protein that binds to the cytoplasmic domain of the receptor for the cytokine GM-CSF and appears to have a role in macrophage/granulocyte lineage choice (Tamura et al., 1999; Mancini et al., 2004).

We repeated the MeCPC purification and used Superose 6 gel-filtration as the final step in the procedure (see purification Strategy II, Figure 10). By monitoring the fractions from the Superose 6 column by both EMSA and Western blotting with antibodies against components of the MeCPC, we could determine the elution profile of the entire MeCPC as well as proteins we considered putative components of the MeCPC. Purification Strategy II contained two additional changes from purification Strategy I: a SP Sepharose Fast Flow column step was inserted after the DEAE Sepharose Fast Flow

column and the Heparin Sepharose column step was eliminated. The SP Sepharose Fast Flow column is a cation-exchange column and served as a replacement for the Heparin Sepharose column, since a significant amount of MeCPC activity eluted from this column during washing at moderate salt concentrations (0.4 M).

EMSA analysis of the final fractions from purification Strategy II shows that the MeCPC elutes in Superose 6 fractions 8 through 13, with maximal activity eluting in fractions 10 and 11 (see Figure 15, panel A). A slower-migrating "supercomplex" elutes in fractions 10 through 13 and a faster-migrating subcomplex elutes in fractions 18 and 19. The possible origin and significance of these complexes is discussed below.

Western blotting was performed on the Superose 6 fractions with antibodies against several of the components of the MeCP1 and cohesin complexes. Western blots with antibodies against the MeCP1 components MBD2, HDAC2, MTA1 and Mi2 show a strikingly similar profile, with elution of these factors in fractions 8 through 13 (see Figure 15, panel B). This elution profile matches the elution profile of MeCPC activity. Western blotting with antibodies against MENT showed that this protein coeluted with the MeCP1 factors, suggesting that this abundantly expressed protein was not a contaminant in the purified MeCPC fractions. Western blots with antibodies against the cohesin components SCC1 and SMC1, as well as the cohesin-associated protein SNF2h, show that these three factors also coelute. However, the elution profile of these factors is slightly different from that of the MeCP1 factors. SNF2h and the cohesin components elute over a slightly larger set of fractions (fractions 8 through 15) and the maximal amount of these factors is in fractions 12 and 13.

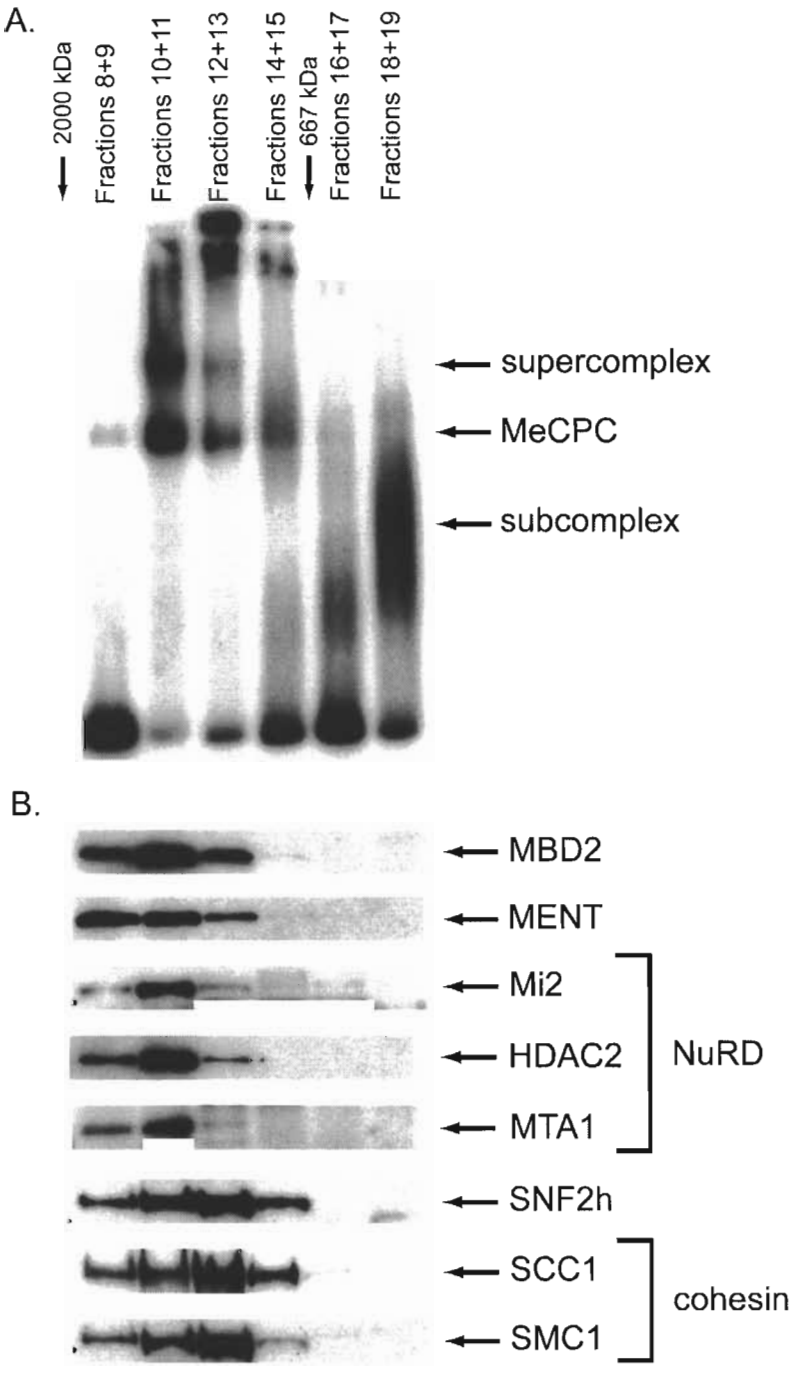

A.
2000 kDa
Fractions 8+9
Fractions 10+11
Fractions 12+13
Fractions 14+15
667 kDa
Fractions 16+17
Fractions 18+19
supercomplex
MeCPC
subcomplex
B.
MBD2
MENT
Mi2
HDAC2
MTA1
NuRD
SNF2h
SCC1
SMC1
cohesin

Figure 15. EMSA and Western blot analysis of Superose 6 fractions from MeCPC purification Strategy II. A) EMSA of Superose 6 fractions from MeCPC purification Strategy II using 20 fmol M-ρ248 as a probe. B) Western blotting for putative MeCPC and MeCPC-associated factors in Superose 6 fractions from MeCPC purification Strategy II. MBD2, MENT, HDAC2, MTA1 and Mi2 coelute with the MeCPC activty in fractions 8 through 13. SNF2h, SCC1, and SMC1 elute over these fractions as well, but their maximal factor content is slightly shifted. Maximum elution of these factors occurs in fractions 12/13 as compared to fractions 10/11 for the MeCP1-type factors.

We conclude that MBD2, RbAp48, HDAC2, MTA1, and Mi2 are components of the MeCPC complex *in vitro*. MENT coelutes with the MeCPC, but subsequent work by our laboratory has shown that MENT is not expressed in cell-types with high levels of MeCPC activity (*e.g.* 6C2 and DT40 cells), raising the possibility that it is not a component of the MeCPC (data not shown). SNF2h and the cohesin components SCC1 and SMC1 also coelute with the MeCPC, indicating that these proteins may be associated with MeCPC. However, because the elution profile of these factors differs slightly from that of the definitive MeCPC factors (MBD2 *et al.*), we conclude that not all of the SNF2h/cohesin complex may be associated with the MeCPC. Subsequent studies done in our laboratory using SCC1 knock-out DT40 cells (Sonoda et al., 2001) have shown that this protein is not required for formation of the MeCPC (data not shown). Thus, it is unlikely that SCC1 and the other SNF2h/cohesin components of the MeCPC *per se*.

Additional EMSA studies were performed with the Superose 6 fractions from purification Strategy II to determine the origin of the MeCPC supercomplex and subcomplex (see Figure 16, panel A). EMSA was performed under identical conditions using the methylated ρ-globin PTR (M-ρ248), the unmethylated ρ-globin PTR (ρ248) and an artificial ^{m}CpG-rich fragment (M-CG11) as probes. Neither the MeCPC complex nor the supercomplex formed on the ρ248 probe by EMSA (Figure 16, panel B), confirming that the MeCPC complex and supercomplex have no affinity for non-methylated DNA. The MeCPC subcomplex did form on the ρ248 probe, indicating that this complex can bind to both methylated and unmethylated DNA. The MeCPC complex and subcomplex

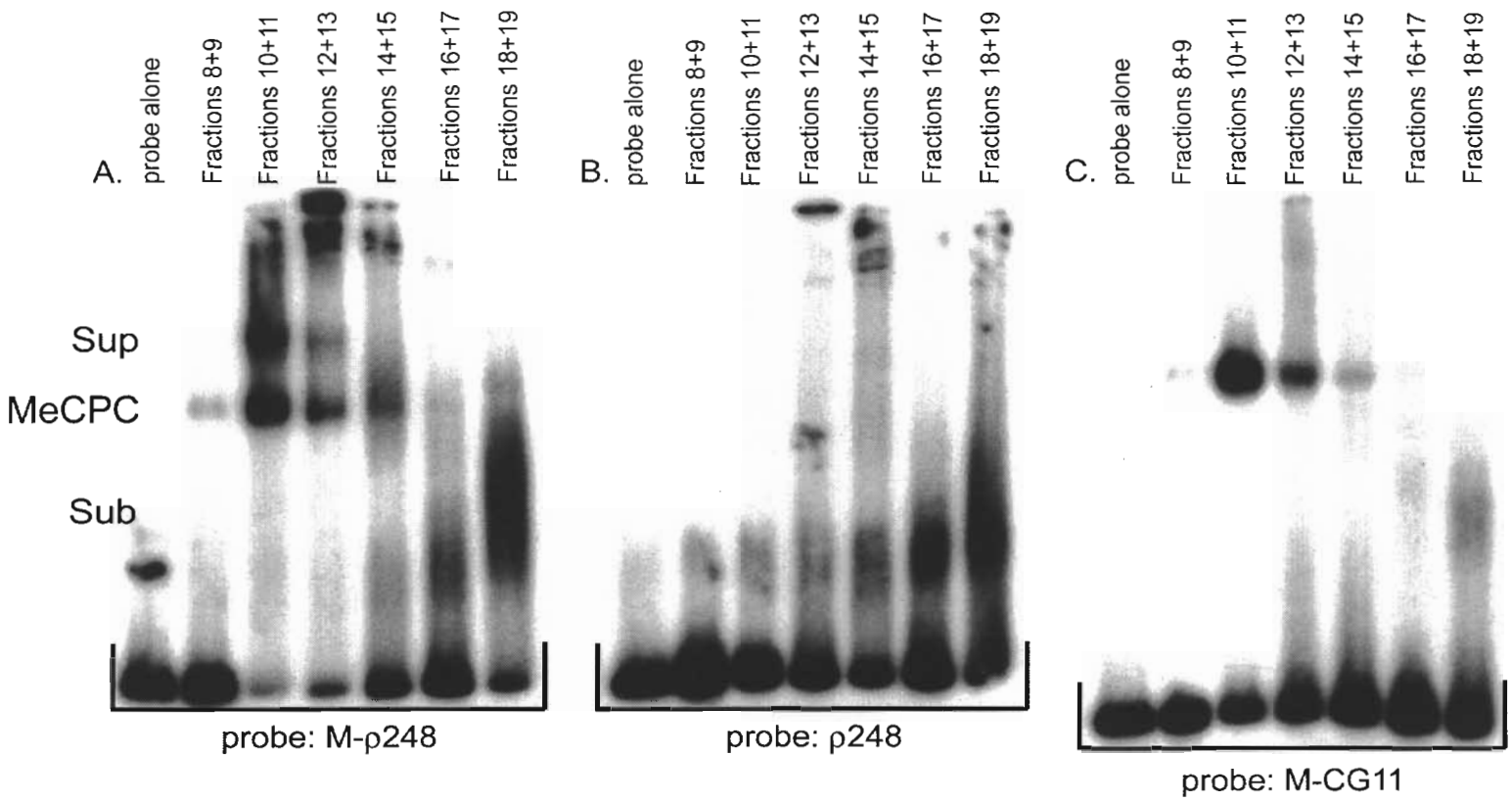


Figure 16. Extended EMSA analysis of Superose 6 fractions from MeCPC purification Strategy II. A) EMSA using Superose 6 fractions 8 through 19 with the methylated ρ-globin PTR (M-ρ248) as a probe. The band generated by MeCPC is marked "M," the band generated by the MeCPC subcomplex is marked "Sub," and the band generated by the MeCPC supercomplex is marked "Sup." B) EMSA using Superose 6 fractions 8 through 19 with the unmethylated ρ-globin PTR (ρ248) as a probe. C) EMSA using Superose 6 fractions 8 through 19 with the methylated CG11 sequence (M-CG11) as a probe.

but not the supercomplex formed on the M-CG11 probe (Figure 16, panel C). Thus the MeCPC supercomplex is specific to the M-ρ248 probe.

Western blotting on the Superose 6 fractions from purification Strategy II revealed that fractions 16 through 19, but not fractions 8 through 13, contain lower molecular weight isoforms of the putative MeCPC factor Mi2 and the putative MeCPC-associated factor SCC1 (see Figure 17, panel B). Therefore our hypothesis is that the MeCPC subcomplex results from proteolytic degradation or loss of critical post-translational modifications of certain MeCPC components, thereby yielding a complex with non-specific DNA binding affinity. Notably, a negligible amount of MBD2 is associated with the subcomplex (Figure 17, panel B). Given that the cohesin complex has non-specific DNA binding activity and has been shown to interact with the NuRD complex *in vivo* (Kagansky et al., 2004; Hakimi et al., 2002), we hypothesize that the MeCPC supercomplex results from the recruitment of the NuRD and cohesin complexes to the M-ρ248 probe by MBD2 under conditions of high DNA and protein concentration during EMSA. Cohesin has been shown to bind preferentially to DNA fragments approximately 300 bp long (Kagansky et al., 2004), thus providing a possible explanation of why this complex did not form on the M-CG11 probe which is only 135 bp long. The hypotheses presented as to the origin of the MeCPC supercomplex and subcomplex are reasonable given the data available, but as yet have not been rigorously tested.

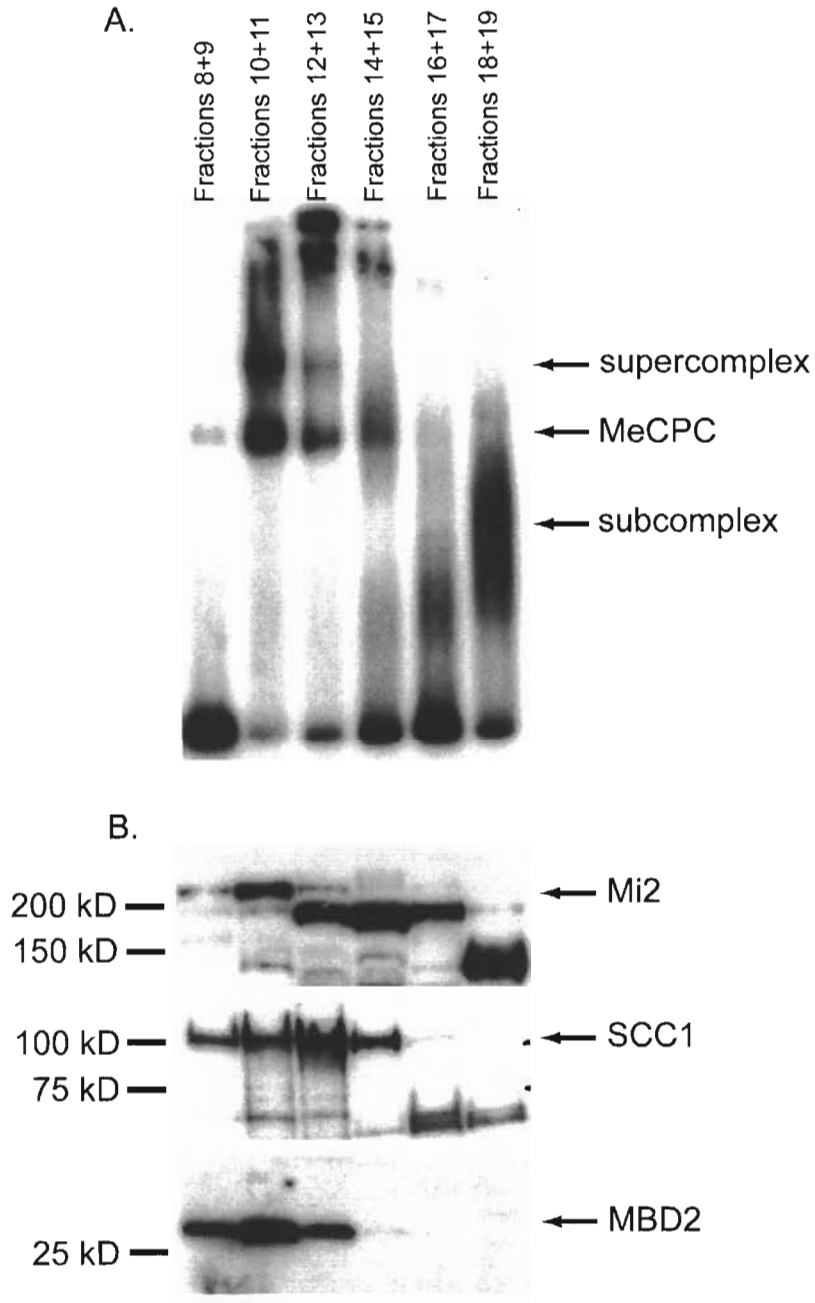

A.
Fractions 8+9
Fractions 10+11
Fractions 12+13
Fractions 14+15
Fractions 16+17
Fractions 18+19
supercomplex
MeCPC
subcomplex
B.
200 kD
150 kD
100 kD
75 kD
25 kD
Mi2
SCC1
MBD2

Figure 17. Smaller molecular-weight forms of Mi2 and SCC1 coelute with the MeCPC subcomplex. A) EMSA of Superose 6 fractions from MeCPC purification Strategy II using 20 fmol M-ρ248 as a probe. B) Western blotting with anti-Mi2 and anti-SCC1 antibodies shows that smaller molecular-weight forms of these proteins coelute with the MeCPC subcomplex activity. Western blotting with anti-MBD2 antibodies shows that MBD2 is not present in these fractions, thereby explaining the DNA methylation-independent binding of this subcomplex.

Discussion

We wished to determine what the protein components of the MeCPC complex are. The MeCPC complex was purified from adult chicken red blood cell nuclear extract using two procedures (Strategies I and II). The proteins present in the purified sample obtained through purification Strategy I were identified by MALDI-TOF mass spectrometry. Four components of the MeCP1 transcriptional repression complex were present in the sample: MBD2, RbAp48, HDAC2 and MTA1. The proteins present in the purified fractions obtained through purification Strategy II were monitored by Western blotting. These studies confirm that MBD2, RbAp48, HDAC2, MTA1 and Mi2 are components of the MeCPC complex.

Another potentially important protein identified by mass spectrometry in the purified sample obtained through purification Strategy I was MENT. MENT is a member of the serpin family of serine protease inhibitors and can inhibit several types of proteases *in vitro* (Grigoryev et al., 1999; Irving et al., 2002). Interestingly, studies have shown that MENT itself is sufficient to induce higher-order chromatin folding and nucleosome compaction (Grigoryev et al., 1992; Grigoryev and Woodcock, 1998). Several pieces of evidence support the hypothesis that MENT is the tissue-restricted MeCPC component that confers sequence-context preferential binding to the complex. First, MENT is abundantly expressed in terminally differentiated chicken hematopoietic cells such as erythrocytes, granulocytes and lymphocytes (Grigoryev and Woodcock, 1993; Grigoryev and Woodcock, 1998). Second, MENT possesses DNA-binding activity (Grigoryev and Woodcock, 1998). Third, as discussed above, MENT copurifies from the Superose 6

column with the other MeCPC components MBD2, RbAp48, HDAC2, MTA1 and Mi2. On the other hand, MENT is not expressed in cell-types with high levels of MeCPC activity, raising the possibility that it is not an intrinsic component of the MeCPC (data not shown). Additional experiments are necessary to show that MENT binds to the methylated ρ-globin gene *in vivo* and that MENT is functionally required for the transcriptional silencing of ρ-globin.

SCC1, another putative component of the MeCPC, was identified by MALDI-TOF mass spectrometry. SCC1 has been shown previously to be a component of a supercomplex containing SNF2h, the NuRD complex and cohesin (Hakimi et al., 2002). Data from purification Strategy II indicates that SNF2h and cohesin copurify with the MeCPC components, but the specificity of this interaction needs verification. As discussed above, we hypothesize that the MeCPC supercomplex is composed of the NuRD and cohesin complexes as recruited to the Mρ-248 probe by MBD2.

Although significant enrichment for the MeCPC activity was achieved through the two purification strategies detailed above, a significant number of contaminating proteins are still present in the purified MeCPC samples. For example, out of the 18 proteins sent for analysis by mass spectrometry, only one-third are likely or confirmed components of the MeCPC. Therefore, if analysis of the subunit composition of the MeCPC is to be repeated, an additional purification step would be helpful in reducing the number of contaminating proteins present in the sample. Because each chromatographic step results in a significant loss of material, an immunoaffinity procedure using antibodies against one of the known components of the MeCPC would be optimal for

further purification. On the contrary, additional purification steps could also result in the loss of MeCPC components or MeCPC associated factors due to differential binding or elution during chromatography.

Using the available chicken proteomic resources, we successfully identified 16 of the 18 chicken proteins present in the purified MeCPC sample under 85 kD. However, the method used to match the PMF data from each analyzed band to its corresponding chicken protein has several weaknesses. First, using this method, we were not able to identify two of the bands in the sample. These bands, A1 and A2, were from low molecular weight proteins (~20 kD) and therefore yield a small number of tryptic peptides. It is likely that this contributes to a low signal-to-noise ratio that makes finding the identity of the bands difficult. Second, this identification method relies on conservation of protein size across species. Although there was substantial conservation of both protein sequence and size in the chicken and mammalian proteins identified in the MeCPC sample, several proteins were not conserved. For example, although HDAC2 was approximately the same size in chicken and human cells, the size of MBD2 differs substantially. Furthermore, if a band in the sample was a proteolytic product that differed significantly in size from the parent protein, it also would not be identified by this method. Third, this method is quite time consuming and not amenable to automation. If it were necessary to routinely identify chicken proteins from PMF data, this method would be highly inefficient.

Despite these weaknesses, this method is a significant advancement in chicken proteomic methods. At the time of initiation of these studies, no algorithms or methods

were available to identify chicken proteins from PMF data. Nevertheless, it is imperative that continued progress in chicken proteomics is made. The chicken is an important model organism because of its role in agriculture, embryology and gene regulation; hence a more direct method of analysis of chicken PMF data is necessary.

Since the human and mouse protein sequence databases have accumulated data over many years, it will be necessary to take a more expedited approach to build a comprehensive protein sequence database for the chicken. The most direct approach would be to further develop the Ensembl Chicken Genome Annotation. Ensembl is an automated gene-prediction system that takes as input genomic sequence data, cDNA data and available proteomic data (Birney et al., 2004; Curwen et al., 2004). Each Ensembl gene is linked directly to its gene products. As such, the ECGA contains the most complete set of proteins encoded by the chicken genome. A comprehensive chicken proteomic resource could be established if the Protein Prospector algorithm (Clauser et al., 1999), which matches mass spectrometry derived PMF data to protein sequence data, was connected to the database of chicken proteins within the ECGA. For proper integration of the search algorithm with the ECGA, the Sanger Institute (which maintains the ECGA) would need to implement this method.

Chapter 4: Functional Analysis of the MeCPC

Introduction

Several different experimental systems have been used to study *cis* and *trans* elements that regulate globin gene transcription at the chicken β-globin locus. These systems include animal models such as adult chickens (Ginder et al., 1984) and transgenic mice carrying the chicken β-globin locus (Mason et al., 1995), as well as cellular models such as primary erythrocytes (Minie et al., 1992) and erythroid cell lines (Wandersee et al., 1996; Mutskov et al., 2002). Erythroid cell lines are obviously unsuitable for physiologic or developmental studies, but provide a useful model in which to rapidly decipher the role of particular *cis* sequences and *trans* factors in globin regulation.

The pattern of globin transcription in each erythroid cell line reflects the transcription factor environment of the cell-type from which the line was derived. An example is the commonly used erythroid cell line K562, which was derived from the pleural effusion of a patient with chronic myeloid leukemia (Lozzio and Lozzio, 1975), and after erythroid induction expresses human γ- and ε-globins but not β-globin (Dean et al., 1983). In contrast, murine erythroleukemia (MEL) cell lines, derived from animals infected with Friend Erythroleukemia Virus, display an "adult" phenotype because they express β^{maj}- and β^{min}-globins after erythroid induction (Nudel et al., 1977). The chicken

erythroid cell line 6C2 is arrested at the CFU-E stage of erythroid differentiation (Prioleau et al., 1999) and displays an erythroblastic phenotype, characterized by very low levels of β^A-globin and no ρ-globin expression upon induction (Wang and Ginder, unpublished observations). Non-erythroid cell lines, such as the chicken lymphoid cell line DT40, do not express the β-globin genes (Litt et al., 2001b).

In order to understand the function of the MeCPC in maintaining the chicken ρ-globin gene in a transcriptionally inactive state, we wish to knock-out components of the complex in an erythroid cell model. We have pursued a stably transfected cell model by introducing a chicken β-globin mini-locus into the MEL cell line by stable transfection. Because stably transfected cells can require extended periods of time for selection and expression of the introduced constructs can vary between lines due to position effect variegation (Robertson et al., 1995a), we have utilized the Flp and recombinase-mediated cassette exchange (RMCE) systems to create these cell lines. The Flp system as conceived by O'Gorman (O'Gorman et al., 1991) and the RMCE system as conceived by Feng *et al.* (Feng et al., 1999) uses a DNA recombinase to integrate transfected DNA at a specific chromosomal site, thus allowing the creation of isogenic cell lines containing different constructs at the same location.

The chromatin immunoprecipitation (ChIP) assay is a complementary technique that can be used to verify the association of a DNA-binding protein with a specific locus *in vivo* (Hebbes et al., 1988; Hebbes et al., 1992). In the first step of the ChIP assay, cells are treated with formaldehyde to crosslink the protein and DNA within chromatin. Next, the cells are lysed and the released chromatin is sonicated, so that the average size of the

chromatin fragments is approximately 300-600 bp. An antibody against a specific DNA-binding protein is then used to immunoprecipitate the chromatin fragments. PCR is performed to determine if a specific DNA sequence is present within the immunoprecipitated material. By computing the ratio of the PCR signal for the immunoprecipitated material to the PCR signal for the input chromatin, the fold-enrichment for the specific factor at the locus is determined. We have performed ChIP using the antibodies we generated against cMBD2 to determine if this protein is present at the ρ-globin gene *in vivo*.

Methods

Construction of plasmid pTA/tHS4

The chicken HS4 core insulator element (cHS4) (Chung et al., 1997) was amplified from chicken genomic DNA by PCR using the primers 5'-TAAAGCTTTTTCCCCGTATCC (forward) and 5'-GAGAGCTCACGGGGACAGCCC (reverse) with the Advantage GC 2 PCR Kit (Clontech). The PCR product was gel purified and cloned into pGEM-T Easy (Promega), resulting in the plasmid pTA/cHS4. The cHS4 element is naturally flanked by 5'-SacI and 3'-HindIII sites. To create a tandem cHS4 construct, two copies of cHS4 were cloned into the MCS of pGEM-3Z. First, pTA/cHS4 was cut with HindIII, the cut end was blunted by incubation with Klenow DNA polymerase (New England Biolabs), and the fragment was excised by digestion with SacI (giving the fragment: SacI-cHS4-blunt). pTA/cHS4 was also cut with SacI, the cut end was blunted with Klenow DNA polymerase, and the fragment was excised by digestion with HindIII (giving the

fragment: blunt-cHS4-HindIII). Second, pGEM-3Z was cut with SacI and SmaI and the SacI-cHS4-blunt fragment was ligated, resulting in the plasmid pGEM-3Z/cHS4. Third, pGEM-3Z/cHS4 was cut with BamHI, the cut end was blunted by incubation with Klenow DNA polymerase, and then the vector was cut again by digestion with HindIII. The blunt-cHS4-HindIII fragment was then ligated into the cut pGEM-3Z/cHS4 vector, resulting in the plasmid pGEM-3Z/tHS4. The sequence was confirmed by DNA sequencing.

Construction of plasmids pTA/NotI-tHS4 and pTA/Asp718-tHS4

To clone the tHS4 element into the pcDNA5/FRT/delCMV vector, the tHS4 element had to be modified by the addition of compatible restriction sites at the 5' and 3' ends of the tHS4 element. NotI sites were added to the tHS4 by performing PCR with the primers 5'-TATGCGGCCGCAGCTCACGGGGACAGCC (forward) and 5'-TCGGCGGCCGCAGCTTTTTCCCCGTATCC (reverse) and pGEM-3Z/tHS4 as the template using the Pfx Platinum polymerase (Invitrogen). The 500 bp PCR product was gel purified and A-overhangs were added by incubation with Taq polymerase (Invitrogen) for 15 minutes in the presence of 10mM dATP. This product was then cloned into pGEM-T Easy (Promega), resulting in the plasmid pTA/NotI-tHS4. Asp718 sites were added to the tHS4 by the procedure outlined above, except the primers used for PCR were 5'-TATGGTACCAGCTCACGGGGACAGCC (forward) and 5'-TCTGGTACCAGCTTTTTCCCCGTATCC (reverse). The resulting plasmid is referred

to as pTA/Asp718-tHS4. The sequence of both plasmids was confirmed by DNA sequencing.

Construction of plasmids pcDNA5/IEρ and pcDNA5/Eρ

The pcDNA5/IEρ and pcDNA5/Eρ plasmids are based on the pcDNA5/FRT vector to which the chicken HS2 and HS3 enhancer elements (IEρ and Eρ) and flanking tHS4 insulator elements (IEρ only) were added. The first step was deletion of the CMV promoter from pcDNA5/FRT (Invitrogen) by digestion of the vector with MluI and NheI, followed by blunting of the cut ends with Klenow DNA polymerase and religation with T4 DNA ligase (New England Biolabs). This plasmid is referred to as pcDNA5/FRT/delCMV. A 4 kb HS2/HS3 fragment was cut from control DNA (Singal et al., 2002) by digestion with BamHI and then gel purified. The pcDNA5/FRT/delCMV vector was cut with BamHI and dephosphorylated by treatment with CIP (New England Biolabs). The cut plasmid was then gel purified. The HS2/HS3 BamHI fragment was ligated into the BamHI cut pcDNA5/FRT/delCMV vector using T4 DNA ligase. The resulting plasmid is referred to as pcDNA5/Eρ. The tHS4 element flanked by restriction sites was cut from pTA/NotI-tHS4 and pTA/Asp718-tHS4 by digestion with NotI and Asp718, respectively. The NotI-tHS4 and Asp718-tHS4 fragments were gel purified. pcDNA5/Eρ was cut with NotI, dephosphorylated by treatment with CIP and gel purified. The NotI-tHS4 fragment was ligated into pcDNA5/Eρ using T4 DNA ligase. Colonies were screened by PCR to find a colony with the NotI-tHS4 fragment in a 5' to 3' orientation. This plasmid was then cut with Asp718, dephosphorylated by treatment with

CIP and gel purified. The Asp718-tHS4 fragment was ligated into the cut plasmid using T4 DNA ligase. Colonies were again screened by PCR to find a colony with the Asp718-tHS4 fragment in a 5' to 3' orientation. The resulting plasmid is termed pcDNA5/IEρ.

Construction of plasmids pTA/ρ2.1 and pTA/ρ4.5

The primer sequences used to create the new ρ-globin genomic constructs ρ2.1 and ρ4.5 by PCR overlap extension (Ho et al., 1989) can be found in Table 5. The primers (Integrated DNA Technologies) were purified from 10% polyacrylamide gels. Primers AB, CD, EF, GH, and IJ were used in separate PCR reactions using Pfx Platinum polymerase (Invitrogen) and pUC8/ρ4.6 as a template to create the individual segments of the final constructs. All PCR reactions were purified using a PCR Purification Kit (Marligen). PCR products AB and CD were mixed and used as a template for PCR (as above) using primers AD. PCR products GH and IJ were mixed and used as a template for PCR (as above) using primers GJ. PCR products AD and EF were mixed and used as a template for PCR (as above) using primers AF. PCR products AF and GJ were mixed and used as template for PCR (as above) using primers AJ (which result in the construct ρ4.5) and primers IK (which result in the construct ρ2.1). Fragments ρ4.5 and ρ2.1 were gel purified and A-overhangs were added by incubation with Taq polymerase (Invitrogen) for 15 minutes in the presence of 10mM dATP. The fragments were then cloned into pGEM-T Easy (Promega), resulting in the plasmids pTA/ρ4.5 and pTA/ρ2.1. The sequence of both plasmids was confirmed by DNA sequencing.

Table 5. List of primers used to construct novel ρ-globin genomic constructs. The primers introduce mutations into the constructs that allow for excision, methylation and subsequent religation of the ρ-globin promoter and PTR.

Name	Abbreviation	Sequence (5' to 3')
rho_4.5_F	A	CGG GAC GGG GAC CTG ACG TTT CTC TCC TCT
rho_mut_5'StuI_R	B	CTG GGA AGT AAA AGG GCC CTG CAT AAG GAC AGC AG
rho_mut_5'StuI_F	C	CTG CTG TCC TTA TGC AGG GCC CTT TTA CTT CCC AG
rho_prom_mut_R	D	CCT TGC ACA GAG CAC CCC TTT CGA AGG CAC CAA ACC C
rho_prom_mut_F	E	GGG TTT GGT GCC TTC GAA AGG GGT GCT CTG TGC AAG G
mut_StuI_R	F	GGT AGA CGA TCA GAG GCC TGC GGA GAG ACG G
mut_StuI_F	G	CCG TCT CTC CGC AGG CCT CTG ATC GTC TAC C
rho_mut_3'StuI_R	H	GTG CTG GAT GAG AGA CGT CTG CCC CCA CAT G
rho_mut_3'StuI_F	I	CAT GTG GGG GCA GAC GTC TCT CAT CCA GCA C
rho_4.5_R	J	CAG AGG TGG TGA CGT CCT GAA CAG GTT GCC CAA
rho_2.1_F	K	GCA GGA GCC TTT GGT ACG GCT CTG AAC CAG GA

Assembly of plasmids IEρ2.1 and IEρ4.5

Plasmids pcDNA5/IEρ and pcDNA5/Eρ were cut with EcoRV, dephosphorylated by treatment with CIP and gel purified. Plasmids pTA/ρ4.5 and pTA/ρ2.1 were cut with NotI and the excised ρ-globin genomic fragments were gel purified. The ends of the cut fragments were blunted by incubation with Klenow DNA polymerase. The fragments were then ligated into pcDNA5/IEρ and pcDNA5/Eρ by overnight incubation with T4 DNA ligase. The resulting plasmids are termed IEρ4.5, Eρ4.5, IEρ2.1 and Eρ2.1. Colonies were screened by restriction digestion to find plasmids with the ρ-globin genomic fragment in the 5' to 3' orientation.

Preparation of methylated and mock-methylated IEρ4.5 constructs

120 μg of plasmid pTA/ρ2.1 was digested with BstBI and StuI to excise the 500 bp ρ-promoter/ρ-gene region (ρ-235/ρ-248). The resulting fragment was gel purified and methylated with *SssI* methylase (New England Biolabs). Methylated or unmethylated ρ-235/ρ-248 was phenol-chloroform extracted and precipitated with glycogen as a carrier. Plasmid IEρ4.5 was digested with BstBI and StuI and gel purified. The cut plasmid was dephosphorylated by treatment with CIP and phenol-chloroform extracted and precipitated as above. A 3:1 molar ratio of methylated-ρ-235/ρ-248 or unmethylated-ρ-235/ρ-248 was ligated into 1500 ng cut, dephosphorylated IEρ4.5 in a 20 μL reaction containing 1 mM ATP and 3000 units T4 DNA ligase overnight at 16°C.

Transfection of MEL, 6C2 and DT40 cells

Cell transfections were performed using the Nucleofector system (Amaxa). MEL cells were transfected using solution V and program T16. 6C2 cells were transfected using solution T and program G16. DT40 cells were transfected using solution T and program B09. Each plasmid transfection reaction contained 2 μg of the GFP-expressing plasmid pMaxGFP (Amaxa) and 2 x 10^6 cells. Cells were assayed for fluorescence 48 hours after transfection. Each siRNA transfection reaction contained 1 μM of a FITC-labeled siRNA against SMAD2 (Dharmacon) and 1 x 10^6 cells. Cells were assayed for fluorescence 2 hours after transfection. FACS analysis to determine transfection efficiency was performed by Dr. Dan Conrad using a MoFlo Cell Sorter (DakoCytomation).

Anti-cMBD2 IgG purification by Protein A chromatography

The antiserum from the two rabbits (VCU1 and VCU2) was received from Cocalico Biologicals. Purification was performed using a BioLogic DuoFlow system (Bio-Rad). The serum from animal VCU1 was thawed and 25 mL was injected to a 5 mL HiTrap rProtein A Fast Flow column (Amersham Pharmacia). The column was washed with 3 CV of 1x PBS and then the bound protein eluted with 3 CV of Buffer G (100 mM glycine-HCl pH 2.8, 10% ethylene glycol). 50 μL of Tris-HCl pH 8.8 was added to each collection tube so that the pH of the eluate would immediately be neutralized. The column was washed with 3 CV of 1x PBS and then with 3 CV of cleaning buffer (100 mM glycine pH 3.0, 2 M urea, 1 M LiCl). The purification procedure was repeated as

above with serum from animal VCU2. Protein-containing fractions from each run were pooled and dialyzed against 1x PBS overnight at 4°C.

Anti-cMBD2 antibody purification by affinity chromatography

Purified GST/cMBD2 was bound to a 1 mL HiTrap NHS-activated High Performance column (Amersham Biosciences) by the manufacturer's protocol. The column was washed three times with Wash buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3) alternating with three washes with Wash buffer B (0.1 M acetate, 0.5 M NaCl, pH 4). It was then washed with 5 CV of Buffer G and 15 CV of PBS to equilibrate the column for loading. 6 mL of Protein A purified VCU1 IgG was run through the column three times. The column was then washed with 5 CV 1x PBS. The bound antibodies were eluted with 10 CV buffer G. 80 μL of Tris-HCl pH 8.8 was added to each collection tube so that the pH of the eluate would immediately be neutralized. The column was then washed with 5 CV of Buffer G and 15 CV of 1x PBS. The above procedure was then repeated with Protein A purified VCU2 IgG. The antibody-containing fractions were aliquoted to Ultrafree-4 centrifugal concentrator devices with a 5K MWCO (Millipore). The protein was concentrated by spinning the devices in the 5810R centrifuge using the A-4-62 rotor (Eppendorf) at 3500 RPM for 20 min at 10°C (final sample volume = 0.5 mL).

Chromatin immunoprecipitation assay – cell preparation, crosslinking and lysis

Red blood cells (RBCs) were collected from 5-day or 18-day chicken embryos and purified by Histopaque. Cells were washed twice with 50 mL chicken culture medium

(91% Leibovitz L-15 medium, 5% fetal bovine serum, 3% chicken serum, 1% pen/strep). 4 x 10^8 RBCs were resuspended in 40 mL chicken culture medium and cross-linked by the addition of formaldehyde to a final concentration of 0.5%. After 10 minutes the cross-linking reaction was quenched by the addition of 125 mM glycine. Cells were then washed twice with 1x PBS containing protease inhibitors (1 μg/mL leupeptin, 1 μg/mL pepstatin, 1 μg/mL aprotinin, and 1 mM PMSF). The washed RBCs were pelleted by spinning at 850g for 2 minutes at 4°C and then resuspended in 1.7x the pellet volume Cell lysis solution containing protease inhibitors (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% NP-40) and incubated on ice for 10 minutes. The lysed cells were again pelleted by spinning at 850g for 2 minutes at 4°C and then resuspended in 2.6 mL SDS Lysis buffer containing protease inhibitors (50 mM Tris-HCl pH 8.1, 1% SDS, 10 mM EDTA) and incubated on ice for 10 minutes. 630 μL of lysate was aliquoted to a fresh 1.5 mL microtube and ¼ volume of glass beads were added. Each lysate tube was then sonicated with eight pulses of 30 seconds on setting 2.5, using a model W-220F sonicator (Misonix). The tubes were then spun at 10,000 RPM for 10 minutes at 4°C. The supernatant was transferred to a fresh tube.

Chromatin immunoprecipitation assay – pre-clearing and immunoprecipitation

After sonication the supernatant was diluted 10-fold with ChIP Dilution buffer (16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, 1.1% Triton X-100) and pre-cleared by adding 100 μL salmon sperm DNA/protein A-agarose (Upstate) and rotating on a wheel for 1 hour at 4°C. The tubes were spun at 2000 RPM for 2 minutes at

4°C and the pre-cleared supernatant was transferred to a fresh tube. 600 µL of the pre-cleared chromatin was saved at -20°C as the input. 10 to 20 µg of affinity-purified anti-chicken MBD2 antibodies or anti-H3-K4-Me_3 antibodies (Abcam) were added to a tube of pre-cleared chromatin and rotated on a wheel overnight at 4°C. 60 to 100 µL (depending on the amount of antibody added to the pre-cleared chromatin) of salmon sperm DNA/protein A-agarose (Upstate) was added to each tube and rotated on a wheel for 1 to 2 hours at 4°C. The agarose beads (containing bound immune complexes) were pelleted by spinning at 2000 RPM for 2 minutes at 4°C.

Chromatin immunoprecipitation assay – wash and elution

The pelleted agarose beads were washed with 1 mL LSW buffer (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) for 5 minutes at 4°C with rotation and then pelleted by spinning at 2000 RPM for 2 minutes at 4°C. The supernatant was removed and the beads were washed with 1 mL HSW buffer (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) for 5 minutes at 4°C with rotation and then pelleted by spinning as above. The supernatant was again removed and the beads were washed with 1 mL LiW buffer (10 mM Tris-HCl pH 8.1, 250 mM LiCl, 1 mM EDTA, 1% deoxycholic acid, 1% NP-40) for 5 minutes at 4°C with rotation and then pelleted by spinning as above. Finally, the supernatant was removed and the beads were washed with 1 mL TE buffer (10 mM Tris-HCl pH 6.5, 1 mM EDTA) for 5 minutes at 4°C with rotation and then pelleted by spinning as above. The TE wash was repeated and the beads were pelleted as above. Immune complexes

were then eluted from the beads by addition of 250 µL ChIP Elution buffer (0.1 M $NaHCO_3$, 1% SDS) and incubation at room temperature for 20 minutes with agitation. The beads were pelleted and the elution step was repeated.

Chromatin immunoprecipitation assay – cross-link reversal and precipitation

The eluate from the Protein A beads was pooled in a fresh microtube. The cross-links were reversed by adding 50 µL 3 M NaCl and 3 µL of 0.5 mg/mL RNase A and incubated overnight at 65°C. The tubes were then cooled to room temperature on the bench and then the chromatin was proteinase K digested by adding 10 µL 0.5 M EDTA, 20 µL 1 M Tris-HCl pH 6.5 and 5 µL of 20 mg/mL proteinase K. The tubes were incubated at 50°C for two hours. The resulting DNA was cleaned by extraction first with ½ phenol ½ chloroform, followed by only chloroform, and finally by ethanol precipitation using 40 µg glycogen as a carrier. The precipitated DNA was resuspended in TE and the ethanol precipitation was repeated. The precipitated DNA was quantified with the PicoGreen assay (Molecular Probes) according to the manufacturer's instructions.

Chromatin immunoprecipitation assay – RT-PCR and calculation of enrichment

Real-time PCR (RT-PCR) was performed on the final precipitated samples using TaqMan primers in an ABI PRISM 7900 (Applied Biosystems). A standard curve for the RT-PCR amplification was constructed using PicoGreen-quantified input DNA. This standard curve was used to calculate a correction factor, so that the fold amplification for

each sample could be normalized to the amount of DNA within that sample. The primers used for amplification of ρ-globin exon II and β^A-globin exon II are listed in Table 6. Each RT-PCR reaction was performed in triplicate. The enrichment for the immunoprecipitated protein at each locus was determined by dividing the normalized amplification from the immunoprecipitated DNA ("bound DNA"), by the amplification from the input DNA.

Results

To understand which of the biochemically identified components of the MeCPC are required for transcriptional silencing *in vivo*, we sought to knock-down components of this complex in an adult erythroid cell model. The only well-characterized adult erythroid cell line is the MEL line, which has been used in numerous studies that examined the role of particular *trans* factors in globin gene expression (Wandersee et al., 1996; Sun et al., 2004; Brand et al., 2004; Feng et al., 2001).

Novel ρ-globin genomic constructs containing mutations that allow for excision, methylation and subsequent religation of the ρ-globin promoter and proximal transcribed region (ρ-235/ρ-248) were created by PCR overlap extension (Ho et al., 1989). These constructs are referred to as IEρ4.5 and Eρ4.5 and contain two main components: a 4.5 kb ρ-globin genomic DNA fragment and the chicken β-globin LCR enhancer elements HS2 and HS3. Flanking copies of the cHS4 chromatin insulator element were introduced into construct IEρ4.5 only. The 4.5 kb genomic fragment contains the entire ρ-globin gene and approximately 2.5 kb of upstream and 1.2 kb of downstream sequence. This

Table 6. List of ρ- and β-globin primers used for real time-PCR of DNA obtained by the chromatin precipitation assay.

Name	Primer	Sequence (5' to 3')
ρ-Globin Exon II	Forward	CACCCTGGGTACTGAAACCACT
ρ-Globin Exon II	Reverse	ACGATCAGCAGCCTGCG
ρ-Globin Exon II	TaqMan	6FAM-CAGTTTTAGGAGGCAGCGCTAACGG TG-TAMRA
β^A-Globin Exon II	Forward	CGGGTGCAGTTGAAGGTGTA
β^A-Globin Exon II	Reverse	TTCCCAAAGGACGCAAAGAAC
β^A-Globin Exon II	TaqMan	6FAM-CCCATCTCTCTACAGGCTGCTGATCGT CT-TAMRA

fragment contains unique restriction sites so that the promoter and PTR regions (ρ-235/ρ-248) can be excised by restriction digestion, methylated *in vitro*, and religated into the construct. In this way, the physiologic methylation pattern can be introduced into the constructs. The HS2 and HS3 enhancer elements are necessary to ensure high-levels of expression of the unmethylated ρ-globin gene. A diagram of the IEρ4.5 and Eρ4.5 constructs can be found in Figure 18.

Although the IEρ4.5 and Eρ4.5 constructs have the same ρ-globin genomic fragment and the same enhancer elements, IEρ4.5 contains flanking copies of the cHS4 chromatin insulator element. These elements were introduced into IEρ4.5, but not Eρ4.5, to test the hypothesis that the insulator element is necessary to maintain an open chromatin structure at the methylated ρ-globin gene. In the numerous examples of methylated genes documented in the literature, almost no example exists where a methylated gene retains an open chromatin structure. Indeed, most methylated loci are associated with the loss of histone acetylation, H3-K4-Me_2 and DNase-I sensitivity (Schubeler et al., 2000; Nguyen et al., 2001; Nguyen et al., 2002; Irvine et al., 2002). However, methylation clearly occurs at the promoter and PTR of the ρ-globin gene in adult erythroid cells without the loss of these characteristics of transcriptionally active chromatin (Stalder et al., 1980; Hebbes et al., 1992; Litt et al., 2001b; Schneider et al., 2004; Wang and Ginder, unpublished observations).

In the course of assembling the IEρ4.5 and Eρ4.5 constructs, we also produced IEρ2.1 and Eρ2.1 constructs. These constructs are identical to IEρ4.5 and Eρ4.5, except that they contain a 2.1 kb ρ-globin-containing genomic fragment. Previous work in our

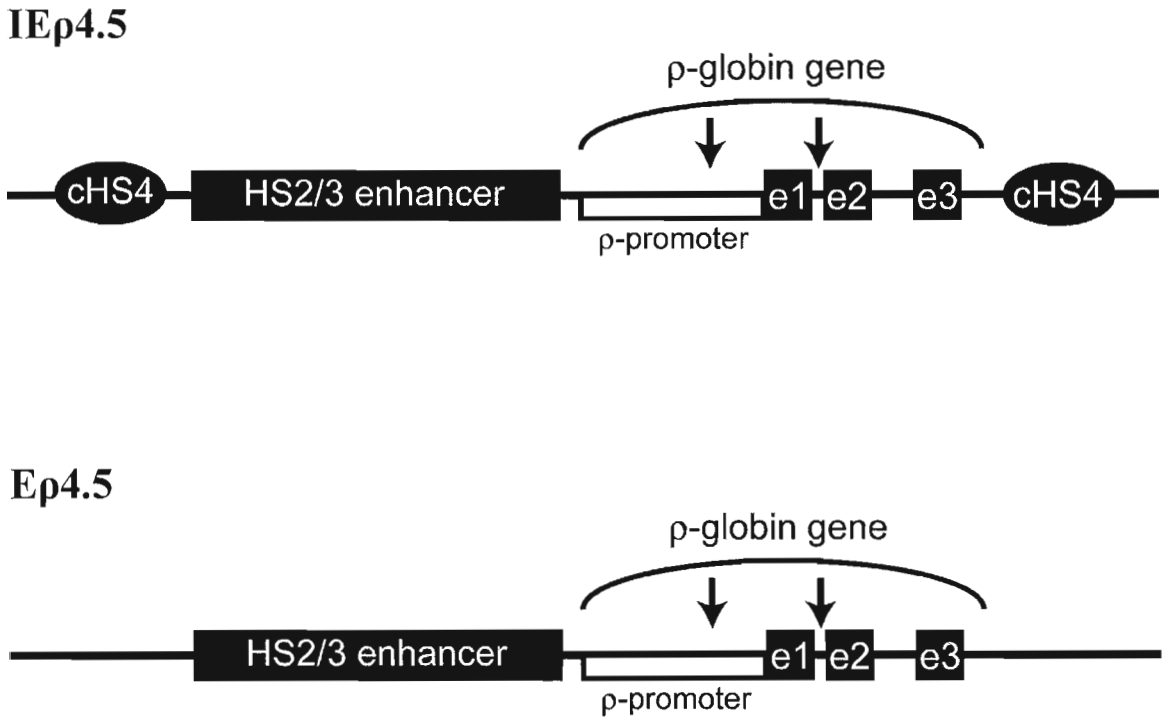


Figure 18. Diagram of the IEρ4.5 and Eρ4.5 constructs. These constructs contain contain two common components: 1) a 4.5 kb ρ-globin containing genomic fragment and 2) the LCR enhancer elements HS2 and HS3. The arrows indicate the location of restriction sites that can be used to excise the ρ-globin promoter and proximal transcribed region. Once excised, this fragment can be methylated and then religated into the construct, to create ρ-globin constructs posessing a physiologically relevant DNA methylation pattern. IEρ4.5 has the cHS4 insulator element placed at the 5' and 3' ends to test the hypothesis that these elements are necessary to maintain an open chromatin structure at the ρ-globin gene in the presence of DNA methylation.

laboratory has shown that DNA sequences 2.5 kb upstream and 2 kb downstream of the ρ-globin cap site are able to silence the effect of strong erythroid-specific enhancers on ρ-globin transcription (Wandersee et al., 1996). The IEρ2.1 and Eρ2.1 constructs were created to test the hypothesis that these upstream and downstream sequences may be necessary to maintain methylation at the ρ-globin promoter and PTR when linked to a strong erythroid-specific enhancer. No experiments were performed with these constructs, but they are available for future use.

We wished to establish stably transfected MEL cell lines containing IEρ4.5 and Eρ4.5 constructs using the Flp system. FRT^+ MEL cells were transfected with four constructs: methylated-IEρ4.5, unmethylated-IEρ4.5, methylated-Eρ4.5 and unmethylated-Eρ4.5, as well as a positive control construct (pcDNA5/FRT/CAT). For the methylated-IEρ4.5 and -Eρ4.5 constructs, the promoter and PTR region (ρ-235/ρ-248) of the ρ-globin gene were excised with restriction enzymes, methylated *in vitro*, and then religated into the construct. Cells were placed under hygromycin selection for three weeks. Unfortunately, although this procedure was repeated three times, no hygromycin-resistant cells developed. We believe that either insufficient Flp-recombinase expression is attained within the cells to exchange the transfected constructs into the FRT site or that the FRT^+ MEL cells are overly sensitive to hygromycin and are not able to survive the selection process.

After we failed to establish stably transfected MEL cell lines containing IEρ4.5 and Eρ4.5 constructs using the Flp system, we wished to adapt the constructs we had made for use with the RMCE system. This system is robust and has been used in a

number of studies (Feng et al., 2001; Eszterhas et al., 2002; Lin et al., 2003). The IEρ4.5 and Eρ4.5 constructs will be excised from the pcDNA5/FRT/delCMV vector and cloned into a vector containing flanking 1L/L1 sites. These constructs can then be methylated *in vitro* and transfected into the RL4, RL5 and RL6 MEL cells. The assembly of these constructs is underway.

Because the MeCPC was purified from chicken erythroid cells, it is possible that not all components of this complex are present in MEL cells. Indeed, the putative MeCPC component MENT, a member of the serpin family of serine protease inhibitors, is highly expressed in terminally differentiated chicken erythrocytes and lymphocytes, but has no known homolog in mouse or human (Grigoryev et al., 1999). Thus we also sought to knock-out components of the MeCPC in a chicken erythroid cell model. Unfortunately, there are no adult chicken erythroid cell lines analogous to MEL cells. The closest extant cell lines are the pre-erythroid cell line 6C2 and the lymphoid cell line DT40. EMSA performed with nuclear extracts shows that both cell types have MeCPC activity (see Figure 19). However, neither cell type expresses the embryonic or the adult β-globin genes and in fact have a "closed" chromatin configuration at the β-globin locus (Litt et al., 2001b; Prioleau et al., 2003).

The lack of ρ-globin gene expression in 6C2 and DT40 cells indicates that these cells are not ideal models in which to determine the role of particular MeCPC components in transcriptional repression of the ρ-globin gene. Basal expression of ρ-

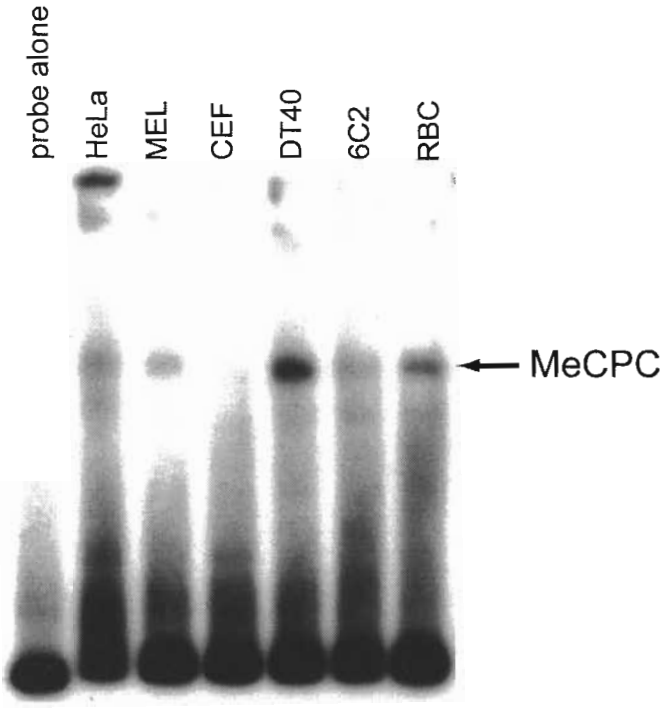


Figure 19. MeCPC activity is present in multiple cell types. EMSA was performed using 10 fmol M-ρ248 as a probe with 10 μg nuclear extract from multiple cell types. HeLa: HeLa cell nuclear extract; MEL: murine erythroleukemia cell nuclear extract; CEF: chicken embryonic fibroblast nuclear extract; DT40: chicken pre-B cell line nuclear extract; 6C2: chicken erythroblast cell line nuclear extract; RBC: primary 15-day erythrocyte nuclear extract.

globin is necessary to show that knock-down of a specific MeCPC component has an effect on ρ-globin transcription. Nevertheless, the presence of high amounts of MeCPC activity in these cells suggests that this complex is likely involved in the transcriptional repression of the other genes in these cells. Therefore, we wish to knock-down the key MeCPC components MBD2, MTA1 and Mi2 to determine which genes are de-repressed by the absence of these factors. To accomplish this, siRNAs (Hannon and Rossi, 2004) targeting these factors will be transfected into 6C2 and DT40 cells and gene expression before and after knock-down will be monitored by Affymetrix GeneChip microarray analysis. These experiments are also underway.

Initial experiments were directed at optimizing transfection of 6C2 and DT40 cells via the Nucleofector system. We transfected 6C2 and DT40 cells with a green fluorescent protein (GFP) expression plasmid as well as FITC-labeled siRNA using the Nucleofector system (Gresch et al., 2004). Expression was monitored by FACS 48 hours after nucleofection of the GFP expression plasmid and two hours after nucleofection of the FITC-labeled siRNA. The results of these experiments can be found in Table 7. In agreement with published data, we found that siRNAs enter nucleofected cells with efficiencies approaching 100%, thereby validating this approach as a method for knock-down of specific target genes.

Several lines of evidence support a role for MBD2 in the transcriptional repression of the ρ-globin gene in adult chicken red blood cells. To determine if MBD2 is associated with the methylated ρ-globin gene *in vivo*, ChIP assays were performed using affinity-purified anti-MBD2 antibodies in embryonic day five and adult

Table 7. Results of transfection optimization experiments using the Nucleofector system on 6C2, DT40 and MEL cells. A GFP-expression plasmid or FITC-labeled siRNA were transfected into the 6C2, DT40 and MEL cell lines using the Nucleofector system and the transfection solution and program indicated in the table. The percent positive cells was determined by FACS 48 hours after transfection of the GFP-expression plasmid or two hours after transfection of the FITC-labeled siRNA.

Cell Type	Solution	Program	Construct	% Positive Cells
6C2	T	G-16	GFP plasmid	22%
6C2	T	G-16	FITC siRNA	100%
DT40	T	B-09	GFP plasmid	76%
DT40	T	B-09	FITC siRNA	100%
MEL	V	T-16	GFP plasmid	10%

erythrocytes. The amount of ρ-globin or β^A-globin DNA in the immunoprecipitated DNA was then quantitated by RT-PCR. In embryonic day five erythrocytes, no enrichment for MBD2 is seen at the transcriptionally active ρ-globin gene, but there is enrichment for MBD2 at the transcriptionally inactive β^A-globin gene (see Figures 20 and 21). The β^A-globin gene is known to be methylated in primitive erythrocytes (Ginder and McGhee, 1981). Likewise, strong enrichment for a marker of transcriptionally active chromatin, H3-K4-Me_3, is seen at the ρ-globin gene but not the β^A-globin gene (Santos-Rosa et al., 2002). In adult erythrocytes the opposite pattern is observed. Significant enrichment for MBD2 is seen at the transcriptionally inactive ρ-globin gene, but no enrichment is observed at the transcriptionally active β^A-globin gene. No enrichment for H3-K4-Me_3 is seen at the ρ-globin gene but strong enrichment is seen at the β^A-globin gene. The ChIP assay therefore confirms that MBD2 binds to the transcriptionally inactive ρ-globin gene in adult chicken erythroid cells.

Discussion

We wished to determine which components of the MeCPC complex are required to maintain the methylated ρ-globin gene in a transcriptionally inactive state. Initially we pursued the development of a stably transfected cell model by introducing a chicken ρ-globin mini-locus into the MEL cell line using the Flp system. Unfortunately, using this method we were not able to obtain any cells containing the chicken ρ-globin mini-locus. We believe there are two main possible reasons that we failed to obtain these cells. First, it is likely that insufficient levels of Flp-recombinase were expressed to exchange the

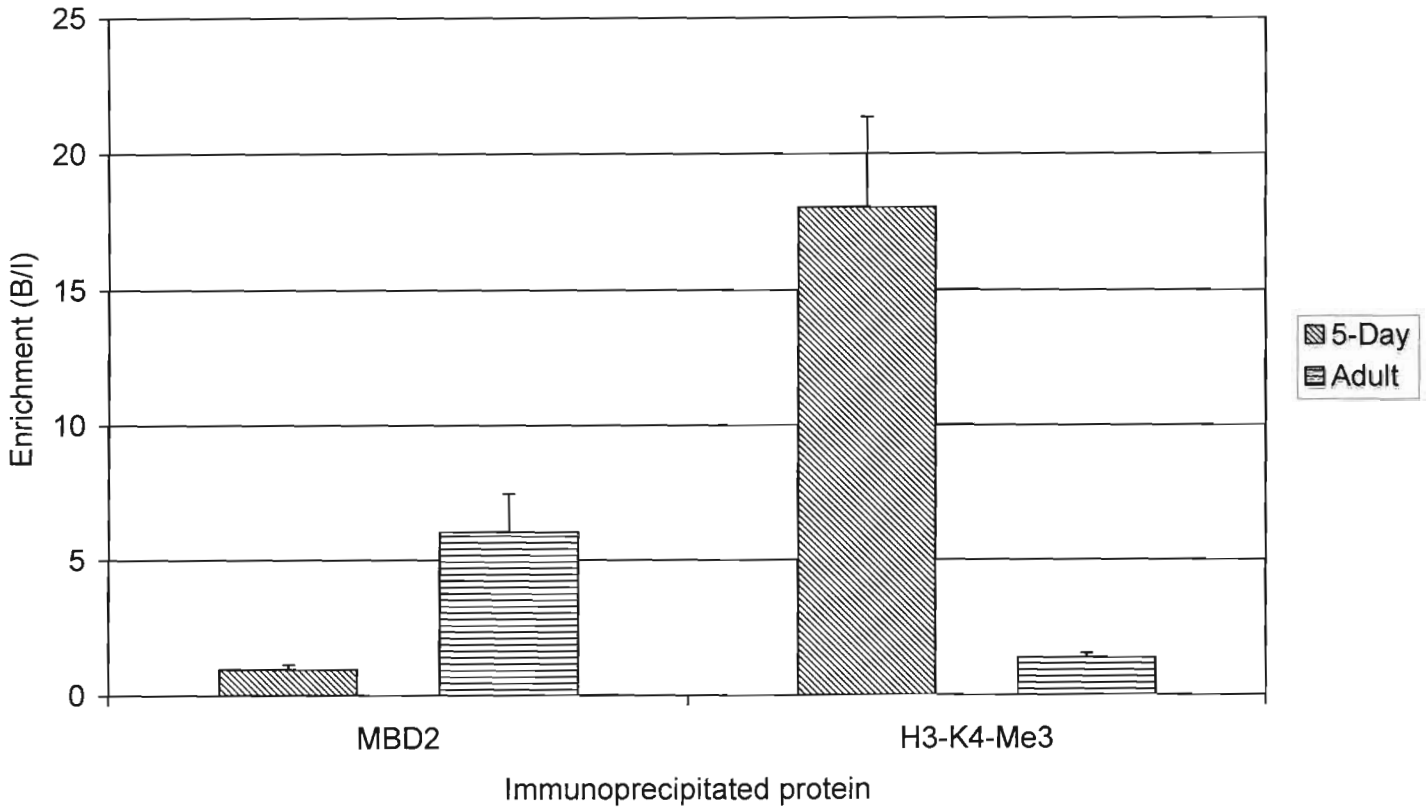


Figure 20. Enrichment for MBD2 and H3-K4-Me_3 at the ρ-globin gene in 5-day (diagonally-hatched) and adult (horizontally-hatched) erythrocytes as determined by the ChIP assay. The data show that MBD2 is depleted from the transcriptionally active ρ-globin gene in 5-day erythrocytes, but enriched at the transcriptionally inactive and methylated ρ-globin gene in adult erythrocytes. In contrast the opposite pattern is seen for the histone modification H3-K4-Me_3.

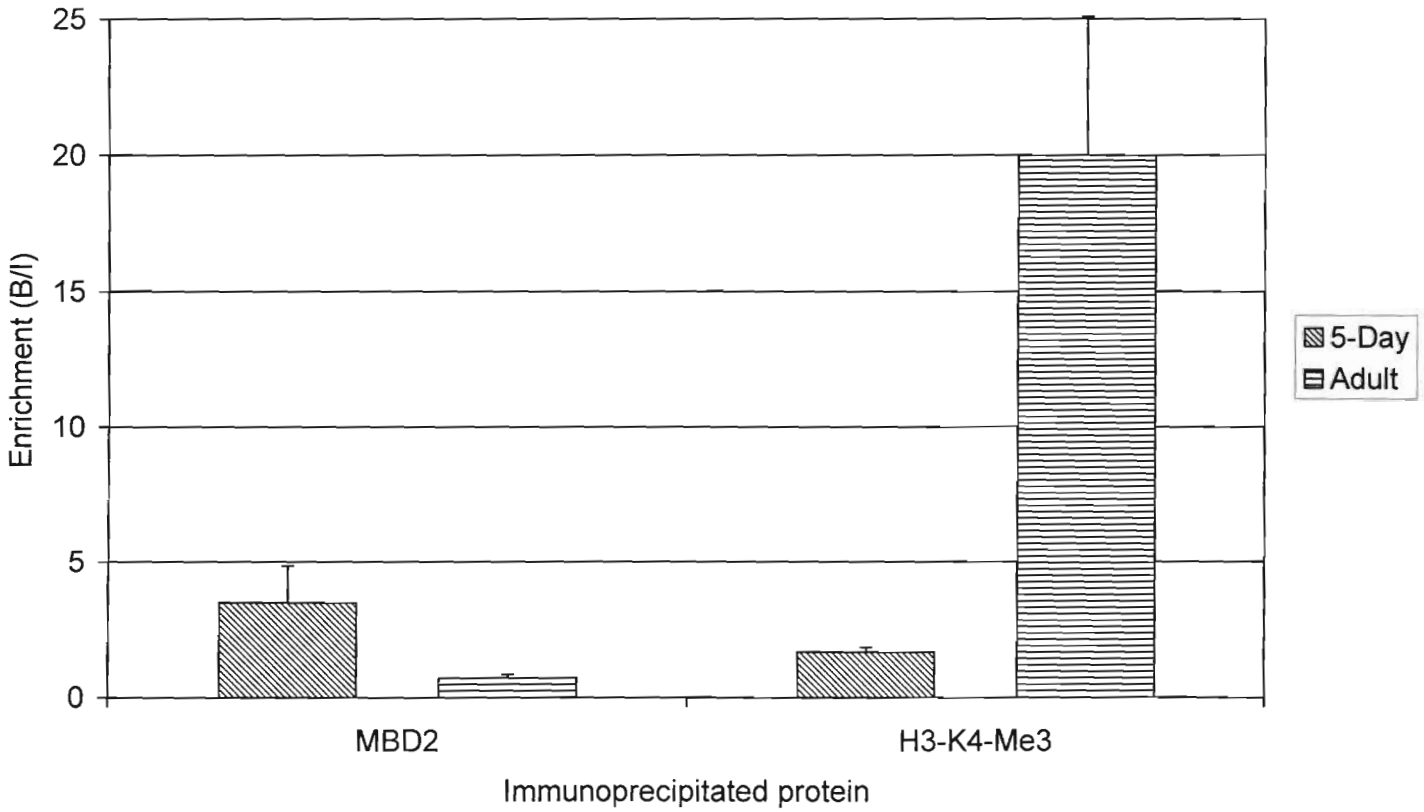


Figure 21. Enrichment for MBD2 and H3-K4-Me_3 at the β^A-globin gene in 5-day (diagonally-hatched) and adult (horizontally-hatched) erythrocytes as determined by the ChIP assay. The data show that MBD2 is enriched at the transcriptionally inactive β^A-globin gene in 5-day erythrocytes, but depleted from the transcriptionally active β^A-globin gene in adult erythrocytes. In contrast the opposite pattern is seen for the histone modification H3-K4-Me_3.

transfected constructs into the FRT site. Second, it is possible that the FRT^+ MEL cells are overly sensitive to hygromycin and are not able to survive the selection process. We are continuing to pursue the development of a stably transfected cell model by converting the constructs used with the Flp system to be compatible with the RMCE system. These experiments are ongoing.

In addition to a stably transfected cell model, it would also be beneficial to establish a transgenic mouse containing the chicken β-globin locus. Mason *et al.* established such a mouse previously and found that the chicken globin genes are expressed at the appropriate stage and location during development (Mason et al., 1995). This model would be highly useful, since mice containing deletions of the methyl-CpG binding proteins MBD2 and MeCP2 are already available. The mice containing the chicken β-globin locus could be rapidly bred to the $MBD2^{-/-}$ and $MeCP2^{-/-}$ mice to determine whether the loss of these factors has an effect on the transcriptional repression of the ρ-globin gene.

Interestingly, the epigenetic profile of the chicken β-globin locus differs from that of the mouse locus. The chromatin of the chicken locus remains highly acetylated even across transcriptionally inactive genes (Litt et al., 2001b; Litt et al., 2001a), whereas the inactive genes at the mouse locus do lose histone acetylation (Forsberg et al., 2000; Im et al., 2002). The establishment of a transgenic mouse containing the chicken β-globin locus would allow us to determine whether the differences in the epigenetic profiles of the chicken and mouse loci are due to the *cis* elements within the locus or due to *trans* factors present in the cell. Furthermore, it would be important to decipher whether the

epigenetic differences between the loci lead to differences in the ability to reactivate transcription of inactive genes *in vivo*.

Chapter 5: Summary and Future Directions

Summary

Given the need of eukaryotic cells to regulate transcription of specific genes in temporally, tissue-restricted and quantitative fashions, it is not surprising that the mechanisms of regulation are highly complex. A comprehensive understanding of a gene's regulation must account for changes in transcription based on modulation of transcription factor binding, RNA polymerase II recruitment and local chromatin structure. Additional factors, such as DNA replication (Danis et al., 2004) and RNA processing (Morillon et al., 2003) have also been found to affect transcription. Despite the difficulty in dissecting out mechanisms of transcriptional regulation of eukaryotic genes, a comprehensive model of the regulation of a single gene of clinical importance is of great value, since it suggests methods of intervention that can be used to modulate expression of that gene therapeutically.

The avian, murine and human β-globin loci have been important models for the study of transcriptional regulation. In all three species the β-globin genes are developmentally regulated and expressed at very high levels within erythrocytes, while transcriptionally silent in non-erythroid tissues. Research over the last 30 years has sought to identify the *cis* and *trans* factors responsible for these characteristics. The

overwhelming message from these studies is that numerous *cis* elements and *trans* factors are responsible for the phenotypic features of the β-globin locus.

Our laboratory has had a long standing interest in the role and mechanism of DNA methylation-mediated transcriptional silencing of the embryonic β-globin genes. Previous work has shown that a methyl-DNA binding complex termed the MeCPC forms on the methylated ρ-globin promoter and PTR *in vitro* (Singal et al., 1997; Singal et al., 2002). The MeCPC complex exhibits a greater DNA-binding affinity for its canonical binding sequence, the methylated ρ-globin PTR (M-ρ248), than for a generic ^{m}CpG-rich sequence (M-CG11). Likewise, the MeCP1 complex exhibits a significantly greater DNA-binding affinity for its canonical binding sequence, M-CG11, than for the methylated ρ-globin PTR. Based on these data, we hypothesize that a factor or factors present in the MeCPC, but not in the MeCP1 complex, confers sequence-context preferential binding to the MeCPC. To understand the binding specificity and functional capabilities of the MeCPC, we sought to determine what the protein components of this complex are.

Based on previous data from our laboratory showing that MBD2 is a component of the MeCPC (Singal et al., 2002), we characterized the chicken homolog of MBD2. Interestingly, cMBD2 lacks a 152 amino acid N-terminal extension which is present in both human and murine MBD2. The MBD is encoded within the conserved C-terminal portion of the protein and the lack of this N-terminal region is consistent with MBD2 homologs present in other non-mammalian vertebrates. cMBD2 is a *bona fide* MCBP as determined by *in vitro* binding assays, and the binding by this protein seems to be

methyl-CpG density-dependent, such that a cooperative relationship exists between the number of ^{m}CpG present in the DNA fragment and the number of MBD2 molecules binding to the fragment. Antibodies generated against cMBD2 recognize an endogenously expressed 28 kD protein in several cell types. These antibodies are also able to retard the mobility of the MeCPC ("supershift") by EMSA analysis, further showing that MBD2 is a component of the MeCPC.

The MeCPC is known to be a large protein complex based on its electrophoretic similarities to MeCP1. Thus, we wanted to determine all the components of the MeCPC complex. We purified the MeCPC complex by two methods (Strategy I and Strategy II), each composed of four sequential column chromatography steps. The MeCPC activity eluting from each column was determined by EMSA. Mass spectrometry was performed on the purified MeCPC sample (from Strategy I) resulting in a set of PMF data for each protein band in the sample. We applied a novel method of protein identification from PMF data for chicken proteins. Using this method, we identified 18 proteins present in the purified MeCPC sample. Four components of the MeCP1 transcriptional repression complex were present in the sample: MBD2, RbAp48, HDAC2 and MTA1. A tissue-restricted factor, MENT, was also identified. These five proteins, as well as the MeCP1 component Mi2, were found to coelute by Western blotting of Superose 6 gel-filtration fractions (from Strategy II). Therefore, we conclude that MBD2, RbAp48, HDAC2, MTA1, Mi2, and possibly MENT, are components of the MeCPC.

A component of a distinct protein complex that has been previously shown to associate with the NuRD complex (Hakimi et al., 2002) was also found in the MeCPC

purified sample by mass spectrometry. This protein, SCC1, was found to coelute with two other SNF2h/cohesin components, SNF2h and SMC1, by Western blotting of Superose 6 gel-filtration fractions. The elution profile of these fractions differed from that of the MeCPC fractions, with the SNF2h/cohesin factors eluting across a larger set of fractions. Additional data suggests that the MeCPC supercomplex may be composed of these SNF2h/cohesin factors in addition to the MeCPC components. Further experiments are needed to confirm this hypothesis.

Given our identification of several components of the MeCPC, we wished to determine the functional importance of these components by knocking-out these components in an adult erythroid cell model. To this end, we produced a novel ρ-globin mini-locus that could be methylated *in vitro* in the same pattern found *in vivo* and introduced these constructs into MEL cells via the Flp system. Unfortunately, using this method, we were not able to obtain any cells containing the chicken ρ-globin mini-locus. We are currently adapting our mini-locus constructs for use with the RMCE system and will attempt to introduce these constructs into MEL cells using that system. Additional ongoing functional experiments include knocking-down components of the MeCPC in the chicken cell lines 6C2 and DT40 using siRNA.

To determine if MBD2 is associated with the methylated ρ-globin gene *in vivo*, we affinity-purified anti-MBD2 antibodies and performed ChIP assays with these antibodies using 5-day and adult erythrocytes. In 5-day erythrocytes, no enrichment for MBD2 was seen at the transcriptionally active ρ-globin gene, but there was enrichment for MBD2 at the transcriptionally inactive β^A-globin gene. In adult erythrocytes the

opposite pattern was observed. Significant enrichment for MBD2 was seen at the transcriptionally inactive ρ-globin gene, but no enrichment was observed at the transcriptionally active β^A-globin gene. The ChIP assay confirmed that MBD2 binds to the methylated, transcriptionally inactive ρ-globin gene in adult chicken erythroid cells.

A question that arises from these studies is whether the MeCPC is a cell-type restricted protein complex with a DNA-binding affinity different than the MeCP1 complex, or whether the MeCPC is a gene-specific protein complex forming specifically on the ρ-globin gene. Previous data from quantitative competitive EMSA experiments showed that the MeCPC had a stronger affinity for its canonical binding sequence, M-ρ248, than for M-CG11 in spite of a higher ^{m}CpG content in M-CG11 (Singal et al., 2002). This result suggests that the MeCPC contains a factor (or factors) that binds more strongly to the DNA sequence of M-ρ248 than to the sequence of M-CG11.

On the other hand, datum presented in Chapter 4 supports the alternative hypothesis, that the MeCPC is unique to erythroid cells but not to the ρ-globin gene. As shown in Figure 19, the MeCPC activity is present at widely varying levels in nuclear extracts from multiple cell types. Generally nuclear extracts from hematopoietic cell-types (*e.g.* DT40, 6C2, RBC) have higher MeCPC activity than epithelial cell-types (*e.g.* CEF). We interpret this to indicate that a specific factor associates with MBD2 and the other NuRD components in these high-activity cell-types, mediating the strong methyl-CpG binding activity. If the MeCPC activity was mediated by a ubiquitously expressed

factor that was specific to the ρ-globin gene, we would expect that the MeCPC activity of the different cell-types would be more similar.

In reconciliation of these two experimental results, we hypothesize that the MeCPC contains an erythroid-restricted factor or factors that mediate the high MeCPC binding activity seen in nuclear extracts from hematopoietic cells. Furthermore, there is a strong binding site for this factor or factors in the ρ248 sequence, but not in the CG11 sequence. In this way, the MeCPC is a cell-type restricted protein complex since the "specificity" factor is an intrinsic part of the MeCPC and does not bind to the ρ-globin gene independent of the other, more general components of the complex (such as the NuRD components). As discussed in Chapter 3, it is possible that MENT is the erythroid-restricted factor.

Model

Based on data from this work, as well as studies performed previously in our laboratory and by other groups, we have developed a model for the transcriptional regulation of the ρ-globin gene during development. As indicated in Figure 2, ρ-globin is expressed abundantly in primitive erythrocytes produced in the embryo until day-5 of development. The definitive erythrocytes that are produced beginning at day-5 of development do not express ρ-globin due to transcriptional silencing of the gene (Lois and Martinson, 1989). Interestingly, methylation of DNA at the ρ-globin gene is first detected at the ρ-globin PTR and spreads downstream to exon 3 and upstream to the promoter (Singal and vanWert, 2001). Methylation is not complete (*i.e.* each cytosine

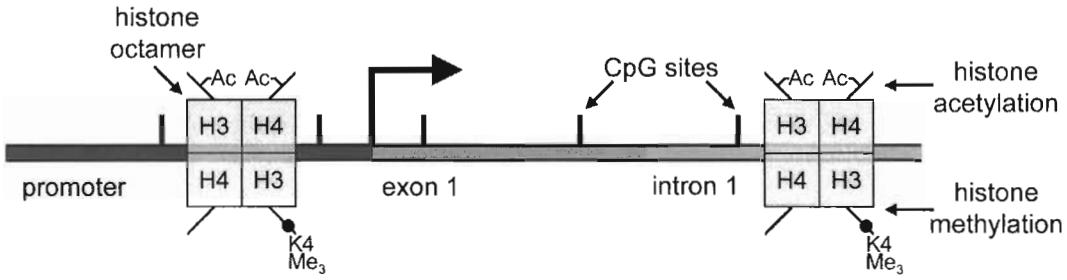
converted to 5-methylcytosine) until after day-14. Thus, it is unlikely that DNA methylation is the initiating event in the establishment of transcriptional silencing. This is consistent with data from other groups showing that transcriptional silencing precedes DNA methylation (Mutskov and Felsenfeld, 2004; Bachman et al., 2003).

We have divided the pathway of transcriptional silencing of ρ-globin into three broad steps: 1) active transcription (*e.g.* in 5-day erythrocytes) 2) early inactive transcription (8-day erythrocytes) and 3) late inactive transcription (adult erythrocytes). This model is presented diagrammatically in Figure 22. The mechanisms of β-globin gene activation have been well-studied in human and murine erythroid cell lines. Several studies have shown that RNAPII is recruited to the core HSS of the β-globin LCR prior to activation of β-globin gene transcription (Vieira et al., 2004; Johnson et al., 2003). The transcription factors GATA-1 and NF-E2 are then able to mediate transfer of RNAPII to the β-globin promoter, enabling transcription of the gene to commence (Johnson et al., 2002). Given the functional homology between the avian, murine and human β-globin LCR, the mechanisms governing transcriptional activation of the ρ-globin gene in 5-day erythrocytes are almost certainly identical.

In yeast, methylation of lysine 4 of histone H3 is performed by the enzyme Set1 (Santos-Rosa et al., 2002). Set1 is recruited to the proximal transcribed regions of yeast genes by the elongating form of RNAPII and catalyzes H3-K4-Me_3 at these target sites (Ng et al., 2003). Analysis of H3-K4-Me_3 at the chicken β-globin locus has shown that this modification is functionally conserved in vertebrates, as high-levels of enrichment for H3-K4-Me_3 were seen at the transcriptionally active ρ- and ε-globin genes in 5-day

Epigenetic silencing of the embryonic ρ-globin gene during developmental switching

A. Schematic of gene



B. Active transcription (5-day erythrocytes)

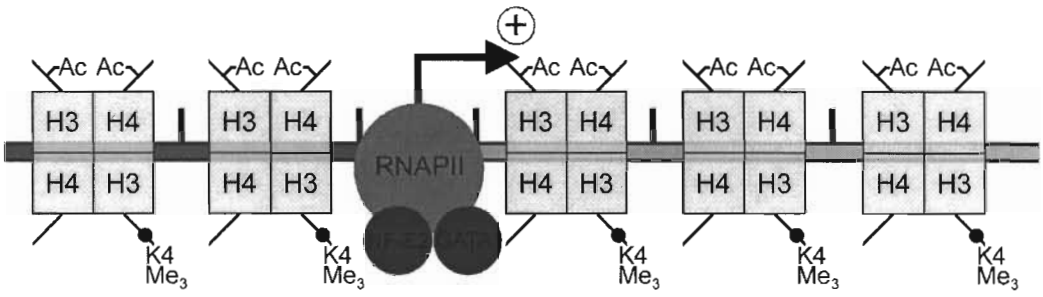
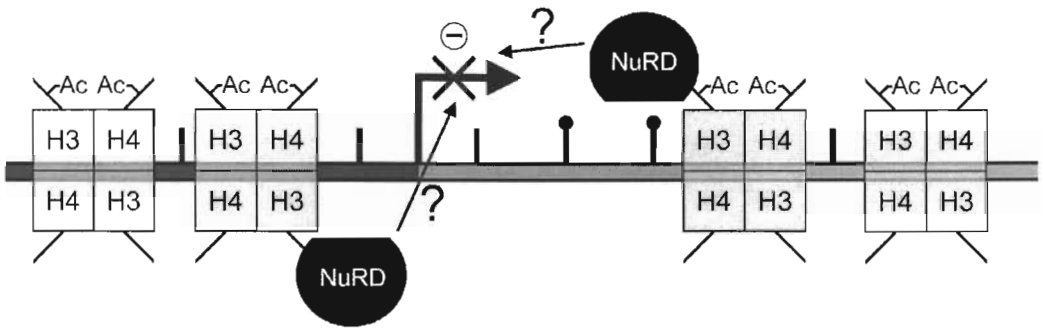


Figure 22. Model for epigenetic silencing of the embryonic ρ-globin gene during developmental switching. A) Schematic of the ρ-globin promoter and PTR, showing how histone octamers (nucleosomes) are depicted in the figure. B) In 5-day erythrocytes, ρ-globin is actively transcribed. The transcription factors GATA1 and NF-E2 recruit RNAPII to the gene, leading to transcription. RNAPII in turn recruits H3-K4-methyltransferase activity, which blocks NuRD factors from binding to the gene.

C. Early inactive transcription (8-day erythrocytes)



D. Late inactive transcription (adult erythrocytes)

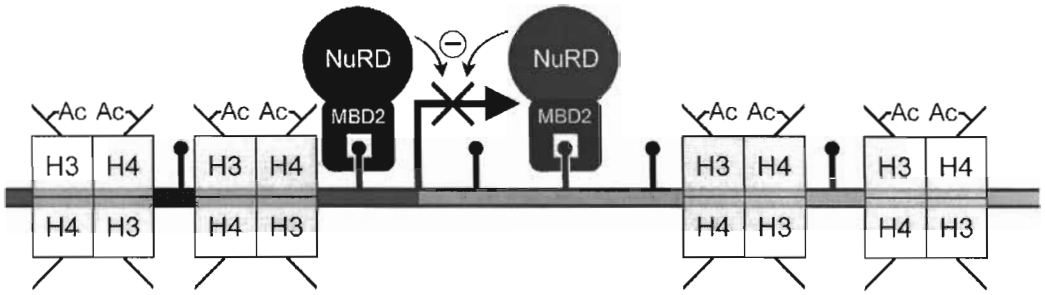


Figure 22 continued. C) In 8-day erythrocytes, ρ-globin is not actively transcribed but DNA methylation of the gene is not complete. The mechanisms of transcriptional repression at this stage are speculative. It is likely that the loss of transcription factors results in the loss of RNAPII recruitment to the promoter, which in turn results in the loss of H3-K4-Me_3 from the gene. This allows NuRD to bind to the N-terminal tails of histone H3 to initiate transcriptional repression. D) In adult erythrocytes, DNA methylation of the ρ-globin gene at the promoter and PTR is complete. DNA methylation recruits MBD2, which tethers the NuRD complex to the gene.

erythrocytes, but low levels of enrichment were seen at the transcriptionally inactive ρ- and ε-globin genes in 15-day erythrocytes (Schneider et al., 2004). Therefore, it is likely that the histone H3-K4 methyltransferase responsible for this modification is associated either physically or functionally with RNAPII in chicken erythrocytes.

Based on the copurification of SNF2h and cohesin subunits SCC1 and SMC1 with the MeCPC, we investigated the possibility that cohesin associates with the ρ-globin gene *in vivo*. Interestingly, SCC1 was found to bind to the ρ-globin gene in 5-day, but not adult, erythrocytes by ChIP assay. These data are consistent with the results of Hakimi *et al.*, who found that the SNF2h/NuRD/cohesin complex preferentially associated with unmethylated DNA sites containing H3-K4-Me. Although the association of SCC1 or the cohesin complex with active β-globin transcription has not been described previously, the significance of these findings is unclear.

Definitive erythrocytes are produced in the chicken embryo beginning on day 5 of embryogenesis. No transcription of ρ-globin in found in these cells. One component of this transcriptional switch is the significant decrease in the abundance of the transcription factors Sp1 and GATA-1 in definitive cells (Minie et al., 1992). The decrease of these factors would then be sufficient to decrease RNAPII recruitment to the ρ-globin promoter. Decreased RNAPII would in turn decrease H3-K4-Me_3 at the regions of the gene which are downstream of the promoter. Four of the MeCPC complex components have been previously identified as part of the NuRD complex: RbAp48, HDAC2, MTA1, and Mi2. Two studies have shown that H3-K4-Me_2 and H3-K4-Me_3, in the context of the histone H3 N-terminal tail, are able to block binding of NuRD to the tail *in vitro*

(Zegerman et al., 2002; Nishioka et al., 2002). Thus, in early definitive erythrocytes, decreased H3-K4-Me_3 may allow NuRD binding, which could contribute to the transcriptional repression of the ρ-globin gene. It is also possible that residual H3-K4-Me_2 and H3-K4-Me_3 at the ρ-globin gene still precludes NuRD binding. Additional experiments are necessary to test these possibilities.

In adult erythrocytes, methylation of the ρ-globin is complete such that every cytosine within the promoter and PTR has been converted to 5-methylcytosine. The extensive DNA methylation serves to recruit MBD2 to the PTR, as shown in Figure 21. MBD2 has previously been shown to tether the NuRD complex to sites of DNA methylation *in vitro* (Feng and Zhang, 2001). Therefore, the function of methylation at the ρ-globin gene is to recruit and stabilize binding of the MeCPC complex, which contains MBD2 and several NuRD components. In this way, the NuRD complex has been "locked" into place at the gene. Another possible component of the MeCPC, MENT, is a tissue-restricted factor that may serve to increase the binding affinity of the complex for the ρ-globin gene. This type of specificity would prevent the MBD2/NuRD complex from being titrated out by highly-abundant methylated DNA present at repeat sequences within the chicken genome (Sobieski and Eden, 1981).

Despite an abundance of data in the literature indicating that DNA methylation inhibits transcription through histone deacetylation and histone H3-K9 methylation, the mechanism of transcriptional repression by the MeCPC complex at the ρ-globin gene is unclear. Although HDAC2 is a component of the MeCPC, two observations strongly oppose the notion that transcriptional repression by the MeCPC is mediated through

histone deacetylation. First, studies have shown that treatment of both adult chickens and definitive erythrocytes with a histone deacetylase inhibitor, without prior demethylation, does not de-repress methylated ρ-globin gene transcription (Ginder et al., 1984; Singal et al., 2002). The lack of transcriptional activation of methylated genes with histone deacetylase inhibition alone has been seen in other systems (Cameron et al., 1999). Second, experiments done in our laboratory, as well as by others, have shown that there is no significant change in the amount of histone acetylation present at the ρ-globin gene between definitive cells and primitive cells (Wang and Ginder, unpublished; Hebbes et al., 1992; Litt et al., 2001b).

An attractive resolution to this contradiction is that HDAC2 mediates deacetylation of a non-histone protein ("factor X") required for transcription of ρ-globin. Acetylation has been shown to activate DNA binding of several transcription factors (Luo et al., 2004; Ammanamanchi et al., 2003). In this model, histone deacetylase inhibition alone does not de-repress transcription, because DNA methylation occurs across the binding site for factor X and recruits MBD2, thereby establishing a genetic competition. Such a mechanism has been shown to regulate the developmental transcription of the murine *Il4* gene (Hutchins et al., 2002). This model gains additional support from work of Dempsey *et al.*, who showed that short-chain fatty acids that activate embryonic globin transcription in adult erythroid cells (Ginder et al., 1984) can induce acetylation of Sp1 and EKLF (Dempsey et al., 2003).

Dimethylation of lysine 9 of histone H3 (H3-K9-Me_2) is another histone post-translational modification that may be involved in the transcriptional silencing of the ρ-

globin gene. Numerous studies have shown that there is reciprocal recruitment or reinforcement of DNA methylation and H3-K9-Me_2 (Fuks et al., 2003; Kondo et al., 2003; Fahrner et al., 2002; Nguyen et al., 2002). Furthermore, H3-K9-Me_2 at a gene promoter can induce transcriptional repression in the absence of DNA methylation (Snowden et al., 2002; Bachman et al., 2003). Our attempts at ChIP using anti-H3-K9-Me_2 antibodies have been unsuccessful due to the inability of the antibody to efficiently immunoprecipitate chromatin from chicken erythrocytes (Wang and Ginder, unpublished observations). Studies by Litt *et al.* did find a modest enrichment for H3-K9-Me_2 surrounding the ρ-globin gene in 10-day erythrocytes (Litt et al., 2001a).

The presence of high levels of histone acetylation and enrichment for H3-K9-Me_2 at the transcriptionally inactive ρ-globin gene may seem incongruous, since acetylation and methylation of lysine 9 of histone H3 are mutually exclusive (Wang et al., 2001). Despite this fact, forms of histone H3 containing up to three acetyl-lysines can contain a methyl group at lysine 9 of histone H3 (Zhang et al., 2004); therefore histone acetylation and H3-K9-Me_2 may be simultaneously present on nucleosomes at the ρ-globin gene. The possible presence of MENT in the MeCPC also supports a role for H3-K9-Me_2 in the regulation of ρ-globin, because MENT has been shown to specifically interact with H3-K9-Me_2, but not H3-K9-Ac, to promote the self-association of chromatin (Istomina et al., 2003). Thus, the function of MENT at the ρ-globin gene may also be to induce a localized "chromatin knot" that interferes with the recruitment of transcription factors or RNAPII.

Future directions

Above we have presented a model for the transcriptional regulation of the ρ-globin gene during development, with specific attention to the role of DNA methylation in maintaining the gene in a transcriptionally inactive state in definitive erythroid cells. Obviously, a number of experiments are needed to confirm biochemical and mechanistic aspects of the model. First, we need to confirm the interaction of the components of the MeCPC complex and supercomplex *in vitro* by immunoprecipitation from crude extracts. Second, we need to establish a stable model of ρ-globin silencing by DNA methylation to investigate the role of each MeCPC component in the maintenance of transcriptional inactivity. Third, if immunoprecipitation experiments confirm that MENT is a component of the MeCPC, recombinant MENT protein could be used to determine if this protein has a higher affinity for the ρ-globin PTR and promoter than for a generic CpG-rich sequence (CG11). ChIP experiments could then be used to verify the association of MENT with the methylated of ρ-globin gene *in vivo.*

If cohesin and SNF2h are confirmed by IP experiments to be components of the MeCPC supercomplex, it will be necessary to ascertain what their regulatory function is. As discussed above, the SCC1 component of the cohesin complex was found to bind to the ρ-globin gene in 5-day, but not adult, erythrocytes by ChIP. In contrast, SCC1 was found to bind to the β-globin gene in adult cells but not in 5-day cells. The association of cohesin with actively transcribing genes has not been shown previously in higher eukaryotes. If SNF2h and cohesin are present in the MeCPC supercomplex, along with MBD2 and NuRD components, we hypothesize that the function of this complex would

be to juxtapose transcriptional activating (SNF2h/cohesin) and repressive proteins (MBD2/NuRD) so that a switch from transcriptional activation to transcriptional repression could be rapidly achieved. Cohesin itself has been shown to be recruited to boundary regions in yeast (Laloraya et al., 2000) and components of cohesin can affect promoter-enhancer communication in *Drosophila* (Rollins et al., 1999; Rollins et al., 2004). Thus, a speculative hypothesis is that cohesin modulates RNAPII transfer from the HSS of the LCR to the ρ-globin gene.

If MENT is conclusively shown to be the tissue-restricted binding component of the MeCPC, an additional implication of this association is that MENT may function to target *de novo* DNA methyltransferase activity to the ρ-globin PTR in early definitive cells. Very few studies have addressed how *de novo* DNA methyltransferase activity is targeted to specific sites, but two studies have shown that sequence-specific DNA binding proteins can recruit DNMT3 to target regions (Di Croce et al., 2002; Fuks et al., 2001). Since DNA methylation at the ρ-globin PTR nucleates at a highly localized site on the coding strand of DNA at 5-days of development (Singal and vanWert, 2001), it is likely that a factor with a DNA-binding site in, or sequence-preference for, this region recruits DNMT3.

Clinical directions

The overall purpose of this project was to elucidate the mechanistic basis of DNA methylation-mediated transcriptional repression of the ρ-globin gene. An abundance of data indicates that the human γ-globin genes are also transcriptionally silenced by DNA

methylation. Therefore, a more comprehensive understanding of the mechanisms and factors required for methylation-mediated transcriptional repression gained from studying the ρ-globin gene could be used therapeutically to de-repress γ-globin gene expression in patients with β-hemoglobinopathies.

Therapeutic interventions aimed at de-repressing γ-globin gene expression could include administration of siRNA-expressing viral vectors (Schomber et al., 2004) targeted to molecules found to be essential for methylation-mediated repression or small-molecule inhibitors of the function of these essential molecules. Furthermore, these "de-repression therapies" could theoretically be combined with "activating therapies," such as small-molecule mimics of γ-globin activating transcriptional factors (Kwon et al., 2004), to achieve high-levels of γ-globin gene expression in anemic patients. Such therapies await continued progress in our understanding of the mechanisms of DNA methylation and globin gene regulation.

List of References

List of References

Abruzzo,L.V. and Reitman,M. (1994). Enhancer activity of upstream hypersensitive site 2 of the chicken beta-globin cluster is mediated by GATA sites. J. Biol. Chem. *269*, 32565-32571.

Amir,R.E., Van,d.V., I, 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.

Ammanamanchi,S., Freeman,J.W., and Brattain,M.G. (2003). Acetylated sp3 is a transcriptional activator. J. Biol. Chem. *278*, 35775-35780.

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.

Bairoch,A. and Boeckmann,B. (1991). The SWISS-PROT protein sequence data bank. Nucleic Acids Res. *19 Suppl*, 2247-2249.

Bakker,J., Lin,X., and Nelson,W.G. (2002). Methyl-CpG binding domain protein 2 represses transcription from hypermethylated pi-class glutathione S-transferase gene promoters in hepatocellular carcinoma cells. J. Biol. Chem. *277*, 22573-22580.

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.

Basu,P., Sargent,T.G., Redmond,L.C., Aisenberg,J.C., Kransdorf,E.P., Wang,S.Z., Ginder,G.D., and Lloyd,J.A. (2004). Evolutionary conservation of KLF transcription factors and functional conservation of human gamma-globin gene regulation in chicken. Genomics *84*, 311-319.

Berkowitz,E.M. and Doty,P. (1975). Chemical and physical properties of fractionated chromatin. Proc. Natl. Acad. Sci. U. S. A *72*, 3328-3332.

Bird,A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499-1504.

Bird,A.P. (1995). Gene number, noise reduction and biological complexity. Trends Genet. *11*, 94-100.

Birney,E., Andrews,T.D., Bevan,P., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cuff,J., Curwen,V., Cutts,T., Down,T., Eyras,E., Fernandez-Suarez,X.M., Gane,P., Gibbins,B., Gilbert,J., Hammond,M., Hotz,H.R., Iyer,V., Jekosch,K., Kahari,A., Kasprzyk,A., Keefe,D., Keenan,S., Lehvaslaiho,H., McVicker,G., Melsopp,C., Meidl,P., Mongin,E., Pettett,R., Potter,S., Proctor,G., Rae,M., Searle,S., Slater,G., Smedley,D., Smith,J., Spooner,W., Stabenau,A., Stalker,J., Storey,R., Ureta-Vidal,A., Woodwark,K.C., Cameron,G., Durbin,R., Cox,A., Hubbard,T., and Clamp,M. (2004). An overview of Ensembl. Genome Res. *14*, 925-928.

Bleasby,A.J., Akrigg,D., and Attwood,T.K. (1994). OWL--a non-redundant composite protein sequence database. Nucleic Acids Res. *22*, 3574-3577.

Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R., Wilson,S.A., and Hubbard,S.J. (2002). A comprehensive collection of chicken cDNAs. Curr. Biol. *12*, 1965-1969.

Brand,M., Ranish,J.A., Kummer,N.T., Hamilton,J., Igarashi,K., Francastel,C., Chi,T.H., Crabtree,G.R., Aebersold,R., and Groudine,M. (2004). Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. Nat. Struct. Mol. Biol. *11*, 73-80.

Busslinger,M., Hurst,J., and Flavell,R.A. (1983). DNA methylation and the regulation of globin gene expression. Cell *34*, 197-206.

Cameron,E.E., Bachman,K.E., Myohanen,S., Herman,J.G., and Baylin,S.B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat. Genet. *21*, 103-107.

Charache,S., Terrin,M.L., Moore,R.D., Dover,G.J., Barton,F.B., Eckert,S.V., McMahon,R.P., and Bonds,D.R. (1995). Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. N. Engl. J. Med. *332*, 1317-1322.

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

Cho,H., Orphanides,G., Sun,X., Yang,X.J., Ogryzko,V., Lees,E., Nakatani,Y., and Reinberg,D. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. Mol. Cell Biol. *18*, 5355-5363.

Choi,O.R. and Engel,J.D. (1986). A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult beta-globin gene. Nature *323*, 731-734.

Chung,J.H., Bell,A.C., and Felsenfeld,G. (1997). Characterization of the chicken beta-globin insulator. Proc. Natl. Acad. Sci. U. S. A *94*, 575-580.

Chung,J.H., Whiteley,M., and Felsenfeld,G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505-514.

Clauser,K.R., Baker,P., and Burlingame,A.L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal. Chem. *71*, 2871-2882.

Curradi,M., Izzo,A., Badaracco,G., and Landsberger,N. (2002). Molecular mechanisms of gene silencing mediated by DNA methylation. Mol. Cell Biol. *22*, 3157-3173.

Curwen,V., Eyras,E., Andrews,T.D., Clarke,L., Mongin,E., Searle,S.M., and Clamp,M. (2004). The Ensembl automatic gene annotation system. Genome Res. *14*, 942-950.

D'Esposito,M., Quaderi,N.A., Ciccodicola,A., Bruni,P., Esposito,T., D'Urso,M., and Brown,S.D. (1996). Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. Mamm. Genome *7*, 533-535.

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.

Danis,E., Brodolin,K., Menut,S., Maiorano,D., Girard-Reydet,C., and Mechali,M. (2004). Specification of a DNA replication origin by a transcription complex. Nat. Cell Biol. *6*, 721-730.

Dean,A., Ley,T.J., Humphries,R.K., Fordis,M., and Schechter,A.N. (1983). Inducible transcription of five globin genes in K562 human leukemia cells. Proc. Natl. Acad. Sci. U. S. A *80*, 5515-5519.

Dempsey,N.J., Ojalvo,L.S., Wu,D.W., and Little,J.A. (2003). Induction of an embryonic globin gene promoter by short-chain fatty acids. Blood *102*, 4214-4222.

DeSimone,J., Heller,P., Hall,L., and Zwiers,D. (1982). 5-Azacytidine stimulates fetal hemoglobin synthesis in anemic baboons. Proc. Natl. Acad. Sci. U. S. A *79*, 4428-4431.

Di Croce,L., Raker,V.A., Corsaro,M., Fazi,F., Fanelli,M., Faretta,M., Fuks,F., Lo,C.F., Kouzarides,T., Nervi,C., Minucci,S., and Pelicci,P.G. (2002). Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science *295*, 1079-1082.

Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475-1489.

Dodge,J.E., Ramsahoye,B.H., Wo,Z.G., Okano,M., and Li,E. (2002). De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. Gene *289*, 41-48.

Doerfler,W. (1983). DNA methylation and gene activity. Annu. Rev. Biochem. *52*, 93-124.

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.

Eszterhas,S.K., Bouhassira,E.E., Martin,D.I., and Fiering,S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol. Cell Biol. *22*, 469-479.

Fahrner,J.A., Eguchi,S., Herman,J.G., and Baylin,S.B. (2002). Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. Cancer Res. *62*, 7213-7218.

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.

Feng,Y.Q., Lorincz,M.C., Fiering,S., Greally,J.M., and Bouhassira,E.E. (2001). Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. Mol. Cell Biol. *21*, 298-309.

Feng,Y.Q., Seibler,J., Alami,R., Eisen,A., Westerman,K.A., Leboulch,P., Fiering,S., and Bouhassira,E.E. (1999). Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J. Mol. Biol. *292*, 779-785.

Forsberg,E.C., Downs,K.M., Christensen,H.M., Im,H., Nuzzi,P.A., and Bresnick,E.H. (2000). Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. Proc. Natl. Acad. Sci. U. S. A *97*, 14494-14499.

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.

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

Fuks,F., Burgers,W.A., Godin,N., Kasai,M., and Kouzarides,T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J. *20*, 2536-2544.

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

Futscher,B.W., Oshiro,M.M., Wozniak,R.J., Holtan,N., Hanigan,C.L., Duan,H., and Domann,F.E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat. Genet. *31*, 175-179.

Gadgil,H., Oak,S.A., and Jarrett,H.W. (2001). Affinity purification of DNA-binding proteins. J. Biochem. Biophys. Methods *49*, 607-624.

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

Gattiker,A., Bienvenut,W.V., Bairoch,A., and Gasteiger,E. (2002). FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification. Proteomics. *2*, 1435-1444.

Ginder,G.D. and McGhee,J.D. (1981). DNA Methylation in the Chicken Adult Beta-Globin Gene: A Relationship With Gene Expression. In Organization and Expression of Globin Genes, (New York: Liss), p. 191.

Ginder,G.D., Whitters,M.J., and Pohlman,J.K. (1984). Activation of a chicken embryonic globin gene in adult erythroid cells by 5-azacytidine and sodium butyrate. Proc. Natl. Acad. Sci. U. S. A *81*, 3954-3958.

Gresch,O., Engel,F.B., Nesic,D., Tran,T.T., England,H.M., Hickman,E.S., Korner,I., Gan,L., Chen,S., Castro-Obregon,S., Hammermann,R., Wolf,J., Muller-Hartmann,H., Nix,M., Siebenkotten,G., Kraus,G., and Lun,K. (2004). New non-viral method for gene transfer into primary cells. Methods *33*, 151-163.

Grigoryev,S.A., Bednar,J., and Woodcock,C.L. (1999). MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. J. Biol. Chem. *274*, 5626-5636.

Grigoryev,S.A., Solovieva,V.O., Spirin,K.S., and Krasheninnikov,I.A. (1992). A novel nonhistone protein (MENT) promotes nuclear collapse at the terminal stage of avian erythropoiesis. Exp. Cell Res. *198*, 268-275.

Grigoryev,S.A. and Woodcock,C.L. (1993). Stage-specific expression and localization of MENT, a nuclear protein associated with chromatin condensation in terminally differentiating avian erythroid cells. Exp. Cell Res. *206*, 335-343.

Grigoryev,S.A. and Woodcock,C.L. (1998). Chromatin structure in granulocytes. A link between tight compaction and accumulation of a heterochromatin-associated protein (MENT). J. Biol. Chem. *273*, 3082-3089.

Guglielmi,B., van Berkum,N.L., Klapholz,B., Bijma,T., Boube,M., Boschiero,C., Bourbon,H.M., Holstege,F.C., and Werner,M. (2004). A high resolution protein interaction map of the yeast Mediator complex. Nucleic Acids Res. *32*, 5379-5391.

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.

Hakimi,M.A., Bochar,D.A., Schmiesing,J.A., Dong,Y., Barak,O.G., Speicher,D.W., Yokomori,K., and Shiekhattar,R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature *418*, 994-998.

Hampsey,M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol. Mol. Biol. Rev. *62*, 465-503.

Hannon,G.J. and Rossi,J.J. (2004). Unlocking the potential of the human genome with RNA interference. Nature *431*, 371-378.

Harju,S., McQueen,K.J., and Peterson,K.R. (2002). Chromatin structure and control of beta-like globin gene switching. Exp. Biol. Med. (Maywood.) *227*, 683-700.

Hebbes,T.R., Thorne,A.W., Clayton,A.L., and Crane-Robinson,C. (1992). Histone acetylation and globin gene switching. Nucleic Acids Res. *20*, 1017-1022.

Hebbes,T.R., Thorne,A.W., and Crane-Robinson,C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. *7*, 1395-1402.

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.

Herman,J.G., Merlo,A., Mao,L., Lapidus,R.G., Issa,J.P., Davidson,N.E., Sidransky,D., and Baylin,S.B. (1995). Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. *55*, 4525-4530.

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

Hershey,J.W., Asano,K., Naranda,T., Vornlocher,H.P., Hanachi,P., and Merrick,W.C. (1996). Conservation and diversity in the structure of translation initiation factor EIF3 from humans and yeast. Biochimie *78*, 903-907.

Hesse,J.E., Nickol,J.M., Lieber,M.R., and Felsenfeld,G. (1986). Regulated gene expression in transfected primary chicken erythrocytes. Proc. Natl. Acad. Sci. U. S. A *83*, 4312-4316.

Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K., and Pease,L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Hofmann,W.A., Stojiljkovic,L., Fuchsova,B., Vargas,G.M., Mavrommatis,E., Philimonenko,V., Kysela,K., Goodrich,J.A., Lessard,J.L., Hope,T.J., Hozak,P., and de Lanerolle,P. (2004). Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II. Nat. Cell Biol. *6*, 1094-1101.

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.

Hung,M.S. and Shen,C.K. (2003). Eukaryotic methyl-CpG-binding domain proteins and chromatin modification. Eukaryot. Cell *2*, 841-846.

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.

Im,H., Grass,J.A., Christensen,H.M., Perkins,A., and Bresnick,E.H. (2002). Histone deacetylase-dependent establishment and maintenance of broad low-level histone acetylation within a tissue-specific chromatin domain. Biochemistry *41*, 15152-15160.

Irvine,R.A., Lin,I.G., and Hsieh,C.L. (2002). DNA methylation has a local effect on transcription and histone acetylation. Mol. Cell Biol. *22*, 6689-6696.

Irving,J.A., Shushanov,S.S., Pike,R.N., Popova,E.Y., Bromme,D., Coetzer,T.H., Bottomley,S.P., Boulynko,I.A., Grigoryev,S.A., and Whisstock,J.C. (2002). Inhibitory activity of a heterochromatin-associated serpin (MENT) against papain-like cysteine proteinases affects chromatin structure and blocks cell proliferation. J. Biol. Chem. *277*, 13192-13201.

Istomina,N.E., Shushanov,S.S., Springhetti,E.M., Karpov,V.L., Krasheninnikov,I.A., Stevens,K., Zaret,K.S., Singh,P.B., and Grigoryev,S.A. (2003). Insulation of the chicken beta-globin chromosomal domain from a chromatin-condensing protein, MENT. Mol. Cell Biol. *23*, 6455-6468.

Jackson-Grusby,L., Laird,P.W., Magge,S.N., Moeller,B.J., and Jaenisch,R. (1997). Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. Proc. Natl. Acad. Sci. U. S. A *94*, 4681-4685.

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.

Johnson,K.D., Grass,J.A., Boyer,M.E., Kiekhaefer,C.M., Blobel,G.A., Weiss,M.J., and Bresnick,E.H. (2002). Cooperative activities of hematopoietic regulators recruit RNA polymerase II to a tissue-specific chromatin domain. Proc. Natl. Acad. Sci. U. S. A *99*, 11760-11765.

Johnson,K.D., Grass,J.A., Park,C., Im,H., Choi,K., and Bresnick,E.H. (2003). Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain. Mol. Cell Biol. *23*, 6484-6493.

Jones,P.A. and Baylin,S.B. (2002). The fundamental role of epigenetic events in cancer. Nat. Rev. Genet. *3*, 415-428.

Jones,P.A. and Taylor,S.M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell *20*, 85-93.

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.

Kagansky,A., Freeman,L., Lukyanov,D., and Strunnikov,A. (2004). Histone tail-independent chromatin binding activity of recombinant cohesin holocomplex. J. Biol. Chem. *279*, 3382-3388.

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.

Kim,J., Sif,S., Jones,B., Jackson,A., Koipally,J., Heller,E., Winandy,S., Viel,A., Sawyer,A., Ikeda,T., Kingston,R., and Georgopoulos,K. (1999). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. Immunity. *10*, 345-355.

Klose,R.J. and Bird,A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. J. Biol. Chem. *279*, 46490-46496.

Koizume,S., Tachibana,K., Sekiya,T., Hirohashi,S., and Shiraishi,M. (2002). Heterogeneity in the modification and involvement of chromatin components of the CpG island of the silenced human CDH1 gene in cancer cells. Nucleic Acids Res. *30*, 4770-4780.

Kondapalli,J., Flozak,A.S., and Albuquerque,M.L. (2004). Laminar shear stress differentially modulates gene expression of p120 catenin, Kaiso transcription factor, and vascular endothelial cadherin in human coronary artery endothelial cells. J. Biol. Chem. *279*, 11417-11424.

Kondo,Y., Shen,L., and Issa,J.P. (2003). Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer. Mol. Cell Biol. *23*, 206-215.

Koshy,M., Dorn,L., Bressler,L., Molokie,R., Lavelle,D., Talischy,N., Hoffman,R., van Overveld,W., and DeSimone,J. (2000). 2-deoxy 5-azacytidine and fetal hemoglobin induction in sickle cell anemia. Blood *96*, 2379-2384.

Krawczak,M., Ball,E.V., and Cooper,D.N. (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am. J. Hum. Genet. *63*, 474-488.

Kumar,S. and Hedges,S.B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917-920.

Kwon,Y., Arndt,H.D., Mao,Q., Choi,Y., Kawazoe,Y., Dervan,P.B., and Uesugi,M. (2004). Small molecule transcription factor mimic. J. Am. Chem. Soc. *126*, 15940-15941.

Laloraya,S., Guacci,V., and Koshland,D. (2000). Chromosomal addresses of the cohesin component Mcd1p. J. Cell Biol. *151*, 1047-1056.

Leonhardt,H., Page,A.W., Weier,H.U., and Bestor,T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.

Letvin,N.L., Linch,D.C., Beardsley,G.P., McIntyre,K.W., and Nathan,D.G. (1984). Augmentation of fetal-hemoglobin production in anemic monkeys by hydroxyurea. N. Engl. J. Med. *310*, 869-873.

Ley,T.J., Chiang,Y.L., Haidaris,D., Anagnou,N.P., Wilson,V.L., and Anderson,W.F. (1984). DNA methylation and regulation of the human beta-globin-like genes in mouse erythroleukemia cells containing human chromosome 11. Proc. Natl. Acad. Sci. U. S. A *81*, 6618-6622.

Ley,T.J., DeSimone,J., Anagnou,N.P., Keller,G.H., Humphries,R.K., Turner,P.H., Young,N.S., Keller,P., and Nienhuis,A.W. (1982). 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. N. Engl. J. Med. *307*, 1469-1475.

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

Lin,C.M., Fu,H., Martinovsky,M., Bouhassira,E., and Aladjem,M.I. (2003). Dynamic alterations of replication timing in mammalian cells. Curr. Biol. *13*, 1019-1028.

Lin,X. and Nelson,W.G. (2003). Methyl-CpG-binding domain protein-2 mediates transcriptional repression associated with hypermethylated GSTP1 CpG islands in MCF-7 breast cancer cells. Cancer Res. *63*, 498-504.

Litt,M.D., Simpson,M., Gaszner,M., Allis,C.D., and Felsenfeld,G. (2001a). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science *293*, 2453-2455.

Litt,M.D., Simpson,M., Recillas-Targa,F., Prioleau,M.N., and Felsenfeld,G. (2001b). Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. EMBO J. *20*, 2224-2235.

Lois,R. and Martinson,H.G. (1989). Chicken globin gene transcription is cell lineage specific during the time of the switch. Biochemistry *28*, 2281-2287.

Lorincz,M.C., Schubeler,D., Hutchinson,S.R., Dickerson,D.R., and Groudine,M. (2002). DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. Mol. Cell Biol. *22*, 7572-7580.

Losada,A., Hirano,M., and Hirano,T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes Dev. *12*, 1986-1997.

Losada,A., Yokochi,T., Kobayashi,R., and Hirano,T. (2000). Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. J. Cell Biol. *150*, 405-416.

Lozzio,C.B. and Lozzio,B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood *45*, 321-334.

Luo,J., Li,M., Tang,Y., Laszkowska,M., Roeder,R.G., and Gu,W. (2004). Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo. Proc. Natl. Acad. Sci. U. S. A *101*, 2259-2264.

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.

Mancini,A., Koch,A., Whetton,A.D., and Tamura,T. (2004). The M-CSF receptor substrate and interacting protein FMIP is governed in its subcellular localization by protein kinase C-mediated phosphorylation, and thereby potentiates M-CSF-mediated differentiation. Oncogene *23*, 6581-6589.

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.

Mason,M.M., Lee,E., Westphal,H., and Reitman,M. (1995). Expression of the chicken beta-globin gene cluster in mice: correct developmental expression and distributed control. Mol. Cell Biol. *15*, 407-414.

McGhee,J.D. and Ginder,G.D. (1979). Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. Nature *280*, 419-420.

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.

Minie,M.E., Kimura,T., and Felsenfeld,G. (1992). The developmental switch in embryonic rho-globin expression is correlated with erythroid lineage-specific differences in transcription factor levels. Development *115*, 1149-1164.

Morillon,A., Karabetsou,N., O'Sullivan,J., Kent,N., Proudfoot,N., and Mellor,J. (2003). Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. Cell *115*, 425-435.

Morris-Desbois,C., Rety,S., Ferro,M., Garin,J., and Jalinot,P. (2001). The human protein HSPC021 interacts with Int-6 and is associated with eukaryotic translation initiation factor 3. J. Biol. Chem. *276*, 45988-45995.

Mutskov,V. and Felsenfeld,G. (2004). Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J. *23*, 138-149.

Mutskov,V.J., Farrell,C.M., Wade,P.A., Wolffe,A.P., and Felsenfeld,G. (2002). The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev. *16*, 1540-1554.

Nagel,R.L., Bookchin,R.M., Johnson,J., Labie,D., Wajcman,H., Isaac-Sodeye,W.A., Honig,G.R., Schiliro,G., Crookston,J.H., and Matsutomo,K. (1979). Structural bases of the inhibitory effects of hemoglobin F and hemoglobin A2 on the polymerization of hemoglobin S. Proc. Natl. Acad. Sci. U. S. A *76*, 670-672.

Nakamura,T., Mori,T., Tada,S., Krajewski,W., Rozovskaia,T., Wassell,R., Dubois,G., Mazo,A., Croce,C.M., and Canaani,E. (2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol. Cell *10*, 1119-1128.

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.

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.

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.

Nguyen,C.T., Weisenberger,D.J., Velicescu,M., Gonzales,F.A., Lin,J.C., Liang,G., and Jones,P.A. (2002). Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. Cancer Res. *62*, 6456-6461.

Nienhuis,A.W. and Stamatoyannopoulos,G. (1978). Hemoglobin switching. Cell *15*, 307-315.

Nishioka,K., Chuikov,S., Sarma,K., Erdjument-Bromage,H., Allis,C.D., Tempst,P., and Reinberg,D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev. *16*, 479-489.

Nudel,U., Salmon,J.E., Terada,M., Bank,A., Rifkind,R.A., and Marks,P.A. (1977). Differential effects of chemical inducers on expression of beta globin genes in murine erythroleukemia cells. Proc. Natl. Acad. Sci. U. S. A *74*, 1100-1104.

O'Gorman,S., Fox,D.T., and Wahl,G.M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science *251*, 1351-1355.

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.

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

Platt,O.S., Brambilla,D.J., Rosse,W.F., Milner,P.F., Castro,O., Steinberg,M.H., and Klug,P.P. (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. N. Engl. J. Med. *330*, 1639-1644.

Ponger,L., Duret,L., and Mouchiroud,D. (2001). Determinants of CpG islands: expression in early embryo and isochore structure. Genome Res. *11*, 1854-1860.

Prioleau,M.N., Gendron,M.C., and Hyrien,O. (2003). Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. Mol. Cell Biol. *23*, 3536-3549.

Prioleau,M.N., Nony,P., Simpson,M., and Felsenfeld,G. (1999). An insulator element and condensed chromatin region separate the chicken beta-globin locus from an independently regulated erythroid-specific folate receptor gene. EMBO J. *18*, 4035-4048.

Ramsahoye,B.H., Biniszkiewicz,D., Lyko,F., Clark,V., Bird,A.P., and Jaenisch,R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U. S. A *97*, 5237-5242.

Reitman,M. and Felsenfeld,G. (1990). Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. Mol. Cell Biol. *10*, 2774-2786.

Reitman,M., Lee,E., Westphal,H., and Felsenfeld,G. (1990). Site-independent expression of the chicken beta A-globin gene in transgenic mice. Nature *348*, 749-752.

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., Baylin,S.B., and Vogelstein,B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552-556.

Robertson,G., Garrick,D., Wu,W., Kearns,M., Martin,D., and Whitelaw,E. (1995a). Position-dependent variegation of globin transgene expression in mice. Proc. Natl. Acad. Sci. U. S. A *92*, 5371-5375.

Robertson,K.D., Hayward,S.D., Ling,P.D., Samid,D., and Ambinder,R.F. (1995b). Transcriptional activation of the Epstein-Barr virus latency C promoter after 5-azacytidine treatment: evidence that demethylation at a single CpG site is crucial. Mol. Cell Biol. *15*, 6150-6159.

Rollins,R.A., Korom,M., Aulner,N., Martens,A., and Dorsett,D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol. Cell Biol. *24*, 3100-3111.

Rollins,R.A., Morcillo,P., and Dorsett,D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Roloff,T.C., Ropers,H.H., and Nuber,U.A. (2003). Comparative study of methyl-CpG-binding domain proteins. BMC. Genomics *4*, 1.

Saito,M. and Ishikawa,F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J. Biol. Chem. *%20;277*, 35434-35439.

Saitoh,N., Bell,A.C., Recillas-Targa,F., West,A.G., Simpson,M., Pikaart,M., and Felsenfeld,G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. EMBO J. *19*, 2315-2322.

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.

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.

Saunthararajah,Y., Hillery,C.A., Lavelle,D., Molokie,R., Dorn,L., Bressler,L., Gavazova,S., Chen,Y.H., Hoffman,R., and DeSimone,J. (2003). Effects of 5-aza-2'-deoxycytidine on fetal hemoglobin levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. Blood *102*, 3865-3870.

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.

Schomber,T., Kalberer,C.P., Wodnar-Filipowicz,A., and Skoda,R.C. (2004). Gene silencing by lentivirus-mediated delivery of siRNA in human CD34+ cells. Blood *103*, 4511-4513.

Schubeler,D., Lorincz,M.C., Cimbora,D.M., Telling,A., Feng,Y.Q., Bouhassira,E.E., and Groudine,M. (2000). Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. Mol. Cell Biol. *20*, 9103-9112.

Siegfried,Z., Eden,S., Mendelsohn,M., Feng,X., Tsuberi,B.Z., and Cedar,H. (1999). DNA methylation represses transcription in vivo. Nat. Genet. *22*, 203-206.

Singal,R., Ferris,R., Little,J.A., Wang,S.Z., and Ginder,G.D. (1997). Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells. Proc. Natl. Acad. Sci. U. S. A *94*, 13724-13729.

Singal,R. and vanWert,J.M. (2001). De novo methylation of an embryonic globin gene during normal development is strand specific and spreads from the proximal transcribed region. Blood *98*, 3441-3446.

Singal,R., Wang,S.Z., Sargent,T., Zhu,S.Z., and Ginder,G.D. (2002). Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. J. Biol. Chem. *277*, 1897-1905.

Snowden,A.W., Gregory,P.D., Case,C.C., and Pabo,C.O. (2002). Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr. Biol. *12*, 2159-2166.

Sobieski,D.A. and Eden,F.C. (1981). Clustering and methylation of repeated DNA: persistence in avian development and evolution. Nucleic Acids Res. *9*, 6001-6015.

Sonoda,E., Matsusaka,T., Morrison,C., Vagnarelli,P., Hoshi,O., Ushiki,T., Nojima,K., Fukagawa,T., Waizenegger,I.C., Peters,J.M., Earnshaw,W.C., and Takeda,S. (2001). Scc1/Rad21/Mcd1 is required for sister chromatid cohesion and kinetochore function in vertebrate cells. Dev. Cell *1*, 759-770.

Stalder,J., Larsen,A., Engel,J.D., Dolan,M., Groudine,M., and Weintraub,H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell *20*, 451-460.

Sternberg,P.W., Stern,M.J., Clark,I., and Herskowitz,I. (1987). Activation of the yeast HO gene by release from multiple negative controls. Cell *48*, 567-577.

Stirzaker,C., Song,J.Z., Davidson,B., and Clark,S.J. (2004). Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. Cancer Res. *64*, 3871-3877.

Struhl,K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. *12*, 599-606.

Sun,J., Brand,M., Zenke,Y., Tashiro,S., Groudine,M., and Igarashi,K. (2004). Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. Proc. Natl. Acad. Sci. U. S. A *101*, 1461-1466.

Takizawa,T., Nakashima,K., Namihira,M., Ochiai,W., Uemura,A., Yanagisawa,M., Fujita,N., Nakao,M., and Taga,T. (2001). DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. Dev. Cell *1*, 749-758.

Tamura,T., Mancini,A., Joos,H., Koch,A., Hakim,C., Dumanski,J., Weidner,K.M., and Niemann,H. (1999). FMIP, a novel Fms-interacting protein, affects granulocyte/macrophage differentiation. Oncogene *18*, 6488-6495.

Tong,J.K., Hassig,C.A., Schnitzler,G.R., Kingston,R.E., and Schreiber,S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature *395*, 917-921.

Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., Chen,L., Skupski,M., Subramanian,G., Thomas,P.D., Zhang,J., Gabor Miklos,G.L., Nelson,C., Broder,S., Clark,A.G., Nadeau,J., McKusick,V.A., Zinder,N., Levine,A.J., Roberts,R.J., Simon,M., Slayman,C., Hunkapiller,M., Bolanos,R., Delcher,A., Dew,I., Fasulo,D., Flanigan,M., Florea,L., Halpern,A., Hannenhalli,S., Kravitz,S., Levy,S., Mobarry,C., Reinert,K., Remington,K., Abu-Threideh,J., Beasley,E., Biddick,K., Bonazzi,V., Brandon,R., Cargill,M., Chandramouliswaran,I., Charlab,R., Chaturvedi,K., Deng,Z., Di,F., V, Dunn,P., Eilbeck,K., Evangelista,C., Gabrielian,A.E., Gan,W., Ge,W., Gong,F., Gu,Z.,

Guan,P., Heiman,T.J., Higgins,M.E., Ji,R.R., Ke,Z., Ketchum,K.A., Lai,Z., Lei,Y., Li,Z., Li,J., Liang,Y., Lin,X., Lu,F., Merkulov,G.V., Milshina,N., Moore,H.M., Naik,A.K., Narayan,V.A., Neelam,B., Nusskern,D., Rusch,D.B., Salzberg,S., Shao,W., Shue,B., Sun,J., Wang,Z., Wang,A., Wang,X., Wang,J., Wei,M., Wides,R., Xiao,C., Yan,C., Yao,A., Ye,J., Zhan,M., Zhang,W., Zhang,H., Zhao,Q., Zheng,L., Zhong,F., Zhong,W., Zhu,S., Zhao,S., Gilbert,D., Baumhueter,S., Spier,G., Carter,C., Cravchik,A., Woodage,T., Ali,F., An,H., Awe,A., Baldwin,D., Baden,H., Barnstead,M., Barrow,I., Beeson,K., Busam,D., Carver,A., Center,A., Cheng,M.L., Curry,L., Danaher,S., Davenport,L., Desilets,R., Dietz,S., Dodson,K., Doup,L., Ferriera,S., Garg,N., Gluecksmann,A., Hart,B., Haynes,J., Haynes,C., Heiner,C., Hladun,S., Hostin,D., Houck,J., Howland,T., Ibegwam,C., Johnson,J., Kalush,F., Kline,L., Koduru,S., Love,A., Mann,F., May,D., McCawley,S., McIntosh,T., McMullen,I., Moy,M., Moy,L., Murphy,B., Nelson,K., Pfannkoch,C., Pratts,E., Puri,V., Qureshi,H., Reardon,M., Rodriguez,R., Rogers,Y.H., Romblad,D., Ruhfel,B., Scott,R., Sitter,C., Smallwood,M., Stewart,E., Strong,R., Suh,E., Thomas,R., Tint,N.N., Tse,S., Vech,C., Wang,G., Wetter,J., Williams,S., Williams,M., Windsor,S., Winn-Deen,E., Wolfe,K., Zaveri,J., Zaveri,K., Abril,J.F., Guigo,R., Campbell,M.J., Sjolander,K.V., Karlak,B., Kejariwal,A., Mi,H., Lazareva,B., Hatton,T., Narechania,A., Diemer,K., Muruganujan,A., Guo,N., Sato,S., Bafna,V., Istrail,S., Lippert,R., Schwartz,R., Walenz,B., Yooseph,S., Allen,D., Basu,A., Baxendale,J., Blick,L., Caminha,M., Carnes-Stine,J., Caulk,P., Chiang,Y.H., Coyne,M., Dahlke,C., Mays,A., Dombroski,M., Donnelly,M., Ely,D., and Esparham,S. (2001). The sequence of the human genome. *291*, 1304-1351.

Vieira,K.F., Levings,P.P., Hill,M.A., Crusselle,V.J., Kang,S.H., Engel,J.D., and Bungert,J. (2004). Recruitment of transcription complexes to the beta -globin gene locus in vivo and in vitro. J. Biol. Chem.

Wade,P.A. (2001). Methyl CpG-binding proteins and transcriptional repression. Bioessays *23*, 1131-1137.

Wandersee,N.J., Ferris,R.C., and Ginder,G.D. (1996). Intronic and flanking sequences are required to silence enhancement of an embryonic beta-type globin gene. Mol. Cell Biol. *16*, 236-246.

Wang,H., Cao,R., Xia,L., Erdjument-Bromage,H., Borchers,C., Tempst,P., and Zhang,Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol. Cell *8*, 1207-1217.

Wang,H.B. and Zhang,Y. (2001). Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor. Nucleic Acids Res. *29*, 2517-2521.

Wang,W., Xue,Y., Zhou,S., Kuo,A., Cairns,B.R., and Crabtree,G.R. (1996). Diversity and specialization of mammalian SWI/SNF complexes. Genes Dev. *10*, 2117-2130.

Weatherall,D.J. (2004). Thalassaemia: the long road from bedside to genome. Nat. Rev. Genet. *5*, 625-631.

Weaving,L.S., Ellaway,C.J., Gecz,J., and Christodoulou,J. (2005). Rett syndrome: clinical review and genetic update. J. Med. Genet. *42*, 1-7.

Weitzel,J.M., Buhrmester,H., and Stratling,W.H. (1997). Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Mol. Cell Biol. *17*, 5656-5666.

Wittschieben,B.O., Otero,G., de Bizemont,T., Fellows,J., Erdjument-Bromage,H., Ohba,R., Li,Y., Allis,C.D., Tempst,P., and Svejstrup,J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. Mol. Cell *4*, 123-128.

Xue,Y., Wong,J., Moreno,G.T., Young,M.K., Cote,J., and Wang,W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851-861.

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.

Yokochi,T. and Robertson,K.D. (2002). Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. J. Biol. Chem. *277*, 11735-11745.

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.

Young,R.A. (1991). RNA polymerase II. Annu. Rev. Biochem. *60*, 689-715.

Yusufzai,T.M., Tagami,H., Nakatani,Y., and Felsenfeld,G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell *13*, 291-298.

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.

Zhang,K., Siino,J.S., Jones,P.R., Yau,P.M., and Bradbury,E.M. (2004). A mass spectrometric "Western blot" to evaluate the correlations between histone methylation and histone acetylation. Proteomics.

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.

Zhu,B., Zheng,Y., Angliker,H., Schwarz,S., Thiry,S., Siegmann,M., and Jost,J.P. (2000). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res. *28*, 4157-4165.

Vita

Evan Paul Kransdorf was born in Richmond, Virginia. He attendend the University of Virginia where he was awarded a Bachelor of Science in Chemistry with Highest Distinction. He then entered the MD/PhD program at Virginia Commonwealth Unviersity. In 2003 he married Kathryn Elizabeth Matta of East Berlin, Connecticut. They reside in Richmond with their two cats, Parker and Blakey.