
Table of Contents

Index of Figures and Tables.....	4
Abbreviations and nomenclature.....	6
Abstract.....	8
1. Introduction.....	10
1.1 Rett syndrome.....	10
1.2 Mental Retardation.....	11
1.2.1 Mutations in MECP2 cause Rett syndrome.....	12
1.3 Gene expression regulation.....	14
1.3.1 Transcription regulation.....	14
1.3.2 Epigenetic gene expression regulation.....	17
1.3.3 DNA methylation.....	18
1.4 MBD proteins.....	20
1.4.1 MBD1.....	21
1.4.2 MBD2.....	22
1.4.3 MBD3.....	24
1.4.4 MBD4.....	24
1.5 MECP2.....	25
1.5.1 Human MECP2: Gene structure and mutations	25
1.5.2 MECP2: Protein structure.....	27
1.5.3 MECP2: Function.....	29
1.5.4 MECP2: Expression.....	30
1.6 Techniques.....	31
1.6.1 DNA microarrays.....	31
1.6.2 Chromatin immunoprecipitation.....	32
1.7 Open questions and approaches.....	33
1.8 Goals of the thesis.....	34
2. Materials and Methods.....	35
2.1 Materials.....	35
2.1.1 Chemicals.....	35
2.1.2 Labware.....	35
2.1.3 Kits.....	35
2.1.4 Vectors.....	35
2.1.5 Bacterial strains.....	35
2.1.6 Oligonucleotides.....	36
2.1.7 Antibodies.....	36
2.1.8 Animals.....	36
2.1.9 Cell culture.....	36

2.1.10 Cell lines and primary cells.....	37
2.1.11 Bioinformatics databases and tools.....	37
2.2 Methods.....	38
2.2.1 Bioinformatics methods.....	38
2.2.1.1 Database screens.....	38
2.2.1.2 Blast searches.....	39
2.2.1.3 Sequence comparison.....	40
2.2.1.4 Perl scripts.....	40
2.2.2 cDNA microarrays.....	41
2.2.2.1 Microarray production.....	41
2.2.2.2 Microarray hybridization and washing.....	42
2.2.2.3 Image acquisition and data analysis.....	43
2.2.3 Cell culture.....	44
2.2.3.1 Coating of culture flasks.....	44
2.2.3.2 Primary fibroblasts.....	44
2.2.3.3 Lymphoblastoid cell line.....	45
2.2.3.4 Primary neurons.....	45
2.2.3.5 Glucocorticoid treatment of primary neurons.....	45
2.2.3.6 CCRF-CEM.....	46
2.2.3.7 NEURO-2A.....	46
2.2.4 Molecular biology techniques.....	46
2.2.4.1 DNA isolation.....	46
2.2.4.2 RNA isolation.....	47
2.2.4.3 Standard PCR.....	47
2.2.4.4 cDNA synthesis and subsequent gene-specific PCR (RT-PCR).....	48
2.2.4.5 Agarose gel electrophoresis.....	49
2.2.4.6 Sequencing.....	49
2.2.4.8 Chromatin immunoprecipitation.....	51
2.2.4.9 Bisulfite sequencing.....	53
3. Results.....	55
3.1 MBD protein family.....	55
3.1.1 Objective.....	55
3.1.2 Human MBD proteins.....	56
3.1.3 Domain analysis.....	59
3.1.4 MBD proteins in other species.....	61
3.1.5 Expression patterns of MBD genes.....	63
3.2 Search for MECP2 paralogues.....	64
3.2.1 Objective.....	64
3.2.2 Proteins with an overall sequence similarity to MECP2.....	65

3.3 MECP2 target genes.....	68
3.3.1 Objective.....	68
3.3.2 Genes differentially expressed in Mecp2-null mice – a microarray study.....	69
3.3.3 Localization of MECP2, FKBP5, and SGK in mouse brain.....	71
3.3.4 Establishment of chromatin immunoprecipitation (ChIP).....	73
3.3.5 MECP2 binding sites in the genomic regions of Fkbp5.....	74
3.3.6 MECP2 and the glucocorticoid receptor compete for binding to the GRE_2 locus.....	78
4. Discussion.....	80
4.1 MBD protein family.....	80
4.1.1 Expression of new MBD protein family members	80
4.1.2 The MBD domain of the new family members.....	81
4.2 Search for MECP2 paralogues.....	84
4.2.1 NEFH.....	84
4.2.2 MLLT2.....	85
4.3 MECP2 target genes.....	86
4.3.1 MECP2 regulates Fkbp5 expression.....	86
4.3.2 Northern blot and quantitative real-time PCR results support the microarray findings.....	87
4.3.3 Up-regulation of glucocorticoid–inducible Fkbp5 and Sgk in Mecp2-null animals is not due to elevation of glucocorticoid levels	89
4.3.4 Rett syndrome: a stress pathway disease - supporting evidence from the literature.....	91
4.3.5 Repression by MECP2: global and total versus local and partial.....	92
4.3.6 DNA microarray studies and chromatin immunoprecipitation.....	92
4.3.7 Future studies.....	93
4.4 Conclusions.....	94
5. Zusammenfassung.....	95
6. Annex.....	97
6.1 Sequence alignment of MECP2 and NEFH.....	97
7. References.....	98
Acknowledgments.....	110
Curriculum Vitae and Publications.....	111

Index of Figures and Tables

Tables

Abbreviations	6
Table 1. Mental retardation is the most important cost factor for the health care system.....	11
Table 2. The four RTT mouse models.....	13
Table 3. Primary antibodies used for ChIP and immunohistochemistry.....	36
Table 4. Secondary antibodies used for immunohistochemistry.....	36
Table 5. Media and other cell culture components.....	37
Table 6. Databases used in the different projects of this thesis.....	37
Table 7. Bioinformatics tools used in this thesis.....	38
Table 8. Primer pairs used for the amplification of plasmid inserts for the array production.....	41
Table 9. PCR conditions used for the amplification of cDNAs for the microarray.....	42
Table 10. Scheme depicting the amounts and concentrations of chemicals used for glucocorticoid treatment experiments.....	46
Table 11. Standard PCR temperature profile and PCR mix.....	48
Table 12. Touchdown PCR temperature profile and PCR mix.....	48
Table 13. Primers used for the amplification of gene specific PCR products from cDNA templates.....	49
Table 14. Temperature profile and PCR mix for sequencing reactions with M13 primers.....	50
Table 15. PCR mix for SR1 and SR2 reaction.....	52
Table 16. Temperature profile for SR1 and SR2 PCR reaction.....	52
Table 17. Primer used to amplify DNA fragments precipitated by ChIP.....	53
Table 18. Composition of the ChIP buffers not included in the Upstate ChIP Kit.....	53
Table 19. Primers used for bisulfite sequencing of the regions 1_4, 2_1, 2_2, and 2_4.....	54
Table 20. Protein domain and protein family databases.....	55
Table 21. Summary of the human MBD polypeptides.....	57
Table 22. Mouse homologues to human MBD proteins.....	62
Table 23. Expression patterns of the mouse/human MBD genes.....	63
Table 24. Results of the database queries.....	66
Table 25. Gene expression changes in the brain of an <i>Mecp2^{-ly}</i> animal compared to a <i>wt</i> litter mate control detected by microarray hybridizations.....	70
Table 26. The existence of amino acids important for m ⁵ CpG binding in MBD1 in other MBD protein family members.....	82
Table 27. Residues of the MBD of MECP2 and their function.....	83
Table 28. Classification of <i>Mecp2</i> -null mice according to the manifestation of phenotype.....	88

Figures

Fig. 1. Transcription initiation.....	16
Fig. 2. Representation of an eukaryotic promoter.....	17
Fig. 3. Summary of epigenetic mechanisms.....	18
Fig. 4. Cytosines can get methylated at the 5' position.....	19
Fig. 5. Structure of the MBD domain of human MECP2.....	21
Fig. 6. Distribution and frequency of mutations in <i>MECP2e2</i>	25
Fig. 7. <i>MECP2</i> genomic, transcript, and protein structure.....	28
Fig. 8. Scheme of chromatin immunoprecipitation.....	32
Fig. 9. Alignment of the human methyl-CpG-binding proteins according to Swissprot.....	57
Fig. 10. Sequence logo of the eleven human MBD sequences.....	58
Fig. 11. Unrooted dendrogram depicting methyl-CpG-binding domains of the human protein family.....	59
Fig. 12. Northern blots for <i>KIAA1461/MBD5</i> and <i>KIAA1887/MBD6</i>	64
Fig. 13. DotPlot output of a comparison of MECP2 to itself.....	65
Fig. 14. Alignment of MECP2 (486 aa) and NEFH (1020 aa) derived from the SSDB.....	66
Fig. 15. Alignment of MECP2 (486 aa) and MLLT2 (2311 aa) derived from the SSDB.....	67
Fig. 16. Expression of MECP2, FKBP5, and SGK in the brain of adult female mice.....	72
Fig. 17. DNA fragment distribution obtained by sonication of chromatin.....	74
Fig. 18. Comparison of human <i>FKBP5</i> and mouse <i>Fkbp5</i> genomic regions.....	75
Fig. 19. PCR using mouse brain cDNA with primers specific for mouse <i>Fkbp5</i>	75
Fig. 20. Chromatin immunoprecipitation shows binding of MECP2 to the genomic region of <i>Fkbp5</i>	76
Fig. 21. Schematic representation of the mouse <i>Fkbp5</i> gene.....	77
Fig. 22. Methylation status of CpGs in <i>Fkbp5</i> genomic DNA determined by bisulfite sequencing.....	77
Fig. 23. <i>Fkbp5</i> intron/exon structure and the localization of GR response elements.....	78
Fig. 24. MECP2 and GR both bind to the GRE_2 genomic region.....	79
Fig. 25. Solution structure of the MBDS of MBD1 (left) and MECP2 (right).....	81
Fig. 26. Graphical representation of MECP2 with the low complexity regions depicted in blue.....	84
Fig. 27. Structure of the NEFH protein.....	85
Fig. 28. Model of <i>Fkbp5</i> gene regulation.....	87
Fig. 29. Northern blot and real-time PCR results confirm microarray data.....	89
Fig. 30. Hormone-dependent and hormone-independent induction of <i>Fkbp5</i> and <i>Sgk</i> in mouse brain.....	90
Fig. 31. MECP2 binding to the <i>Fkbp5</i> gene is not abolished in <i>wt</i> mouse brain upon corticosterone infusion as shown by CHIP with placebo and corticosterone-treated mice.....	90

Abbreviations and nomenclature

Abbreviations

aa	Amino acid
ACTH	Adrenocorticotropic hormone
ALS	Amyotrophic lateral sclerosis
ANOVA	ANalysis Of Variance between groups
BLAST	Basic local alignment search tool
bp	Base pairs
ChIP	Chromatin immunoprecipitation
CXXC	Cysteine rich domain
DAPI	4',6-Diamidino-2-phenylindole
DMEM	Dulbecco's modified eagle medium
DNMT	DNA methyl transferase
DPBS	Dulbecco's phosphate buffered saline
EBV	Eppstein-Barr virus
FASTA	Fast-All (sequence alignment algorithm)
FBS	Fetal bovine serum
FITC	Fluorescein-Isothiocyanat
GR	Glucocorticoid receptor
GRE	Glucocorticoid responsive element
h	Hours
HDAC	Histone deacetylase
HMM	Hidden markov model
IQ	Intelligence quotient
kb	Kilo base
LC	Low complexity
lof	Loss of function
m ⁵ CpG	CpG pair symmetrically methylated at the 5' position of the cytosine
MBD1	Methyl-CpG binding domain protein 1
MBD2	Methyl-CpG binding domain protein 2
MBD3	Methyl-CpG binding domain protein 3
MBD4	Methyl-CpG binding domain protein 4
MBD5	Methyl-CpG binding domain protein 5; formerly KIAA1461 protein
MBD6	Methyl-CpG binding domain protein 5; formerly KIAA1462 protein
MECP2	Methyl-CpG binding protein 2
MECP2e1	MeCP2 isoform starting at exon 1
MECP2e2	MeCP2 isoform starting at exon 2

min	Minutes
MR	Mental retardation
MSI	Microsatellite instability
NLS	Nuclear localization signal
nt	Nucleotide
P/S	Penicillin and streptomycin
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMSF	Phenylmethanesulfonyl fluoride
Pol II	RNA polymerase II
Pomc	Pro-opiomelanocortin
PSV	Preserved speech variant (of Rett syndrome)
RT	Room temperature
RTT	Rett syndrome
SDS	Sodium dodecyl sulfate
sec	Seconds
TBP	TATA binding protein
TFIIB	Transcription factor II B
TRD	Transcription repression domain
wt	Wild-type
WWBR	WW domain binding region
XLMR	X-linked mental retardation

Nomenclature

The nomenclature proposed by the human genome organization (HUGO) was applied.

Abstract

Mutations in the gene coding for the methyl-CpG binding protein 2 (*MECP2*) cause a severe form of mental retardation known as Rett syndrome. Almost exclusively girls are affected by this disease. The first mutations in the X-chromosomal gene *MECP2* have been described in 1999, but the molecular mechanisms underlying the disorder are still unknown. *MECP2* can act as a transcriptional repressor and only two neuronal target genes (*Bdnf* and *Dlx5*) have been identified so far. While *MECP2* is expressed ubiquitously, the phenotype of the disease is primarily neuronal. This suggests that *MECP2* has an important function in the brain, whereas, in peripheral tissues, loss of function of *MECP2* might be compensated by functionally redundant proteins.

To find proteins that could potentially mediate such a compensation, two strategies were applied. In a first project, a bioinformatics approach was used to find additional polypeptides that contain an methyl-CpG binding domain (MBD), the domain of *MECP2* which binds to methylated CpGs. Six new such proteins could be detected and were studied for their expression and domain structure.

A second project aimed at identifying proteins with an overall amino acid similarity to *MECP2*. Such proteins could point to additional, so far unknown functions of *MECP2*. Two proteins were identified by database screens and their properties are discussed in this thesis. The structure of one of them, the neurofilament NEFH, suggests that *MECP2* has an elongated shape.

To elucidate the target genes of *MECP2* in the brain, chromatin immunoprecipitation (ChIP) was established and combined with a cDNA microarray approach. An animal model for Rett syndrome was used for this analysis. The microarray experiment showed, that *Mecp2*-null animals differentially express several genes that are induced during stress response by glucocorticoids. Increased levels of mRNAs for plasma glucocorticoid-inducible kinase 1 (*Sgk*) and FK506-binding protein 51 (*Fkbp5*) were observed. Immunohistochemistry revealed, that in mouse brains *Mecp2* and *Fkbp5* as well as *Sgk* are expressed in the same cells. These results suggests a modulating function of *MECP2* in gene expression regulation rather than a total repression since the transcriptional repressor *MECP2* and its target genes are expressed in the same cells..

In *Fkbp5*, three MECP2-binding regions could be determined by CHIP. One of the regions is also a target site for the glucocorticoid receptor (GR). Therefore a model can be proposed in which MECP2 and GR compete for a binding site in *Fkbp5* and regulate its transcription. In Rett patients this regulation would be disturbed due to the loss of function of MECP2 leading to a constant overexpression of glucocorticoid pathway downstream targets and potentially to several features of the Rett syndrome phenotype.

1. Introduction

1.1 Rett syndrome

Rett syndrome (RTT) is a genetic neurodevelopmental disorder. It was originally described by Andreas Rett in 1966 (Rett, 1966) and brought to attention of the international scientific community by Bengt Hagberg and colleagues in 1983 (Hagberg *et al.*, 1983). The disease is almost exclusively found in females. With an incidence of 1 in 10,000 - 15,000 it is one of the most common forms of mental retardation in girls.

The peculiar course of the disease with an onset at the age of 6-18 months after an apparently normal period of growth and development is intriguing and so far unexplained. In a first phase Rett syndrome clinically presents with regression, loss of speech and loss of purposeful hand use. Other features include autism, ataxia, stereotypic hand movements such as hand-washing or -wringing, deceleration of head growth, and epilepsy (Hagberg *et al.*, 1983). After this first period of regression, patients enter a period of apparent stability that can last for decades. A final stage is characterized by reduced mobility, i.e. even previously mobile patients lose their ability to walk. The life expectancy is around 50 years (IRSA homepage <http://www.rettsyndrome.org/main/life-expectancy.htm>).

In 1994, Hagberg and Skjeldal defined clinical criteria to classify Rett patients. According to their categorization scheme, girls of ten years or more have to meet a minimum of three out of six primary criteria and five out of eleven supportive criteria to qualify as classical RTT (Hagberg and Skjeldal, 1994). Patients who do not manifest all the necessary features are considered to have a variant of the disease, among which the preserved speech variant (PSV) is the most common one (Zappella 1992, Zappella *et al.* 1998, De Bona *et al.* 2000).

Only few male patients have been described, and most of them are categorized as PSV patients. In general, the disease is considered as embryonically lethal in males. Genetic mosaicism is believed to be the reason for the survival of some individuals (Clayton-Smith *et al.*, 2000; Renieri *et al.*, 2003). In addition, a boy with 47,XXY karyotype has been described as well (Schwartzman *et al.*, 1998).

Besides the already mentioned characteristics, mental retardation (MR) is a prominent feature of RTT.

1.2 Mental Retardation

Mental retardation, which is defined as an intelligence quotient (IQ) < 70, significant limitations in two or more adaptive skill areas and an onset before the age of 18, presents a big medical and social problem (see Table 1) (definition by the World Health Organization, <http://www.who.int>) Severe forms of mental retardation with an IQ of 50 or below have a prevalence of about 0.3 - 0.4% in the population, while mild MR (IQ of 70-50) is as frequent as 2-3% (Birch *et al.*, 1970, Laxova *et al.*, 1977).

A male excess of 20% - 40% has been found in MR and the different number of sex chromosomes between males and females is regarded as responsible for that observation. Females can compensate a mutation in a gene on the X-chromosome by the second, wild-type (*wt*) allele, whereas in males the mutation will lead to MR.

Developmental disability	Rate ^{b)}	Direct medical costs ^{c)} (millions)	Direct non-medical costs ^{d)} (millions)	Indirect costs ^{e)} (millions)	Total costs (millions)	Average costs per person
Mental retardation	12.0 ‰	7061 \$	5249 \$	38927 \$	51237 \$	1014000 \$
Cerebral palsy	3.0 ‰	1175 \$	1054 \$	9241 \$	11470 \$	921000 \$
Hearing loss	1.2 ‰	132 \$	640 \$	1330 \$	2102 \$	417000 \$
Vision impairment	1.1 ‰	159 \$	409 \$	1915 \$	2484 \$	566000 \$

a) Present value estimates, in 2003 dollars, of lifetime costs for persons born in 2000, based on a 3% discount rate.

b) Of children aged 5-10 years, on the basis of Metropolitan Atlanta Developmental Disabilities Surveillance Program data for 1991- 1994.

c) Includes physician visits, prescription medications, hospital inpatient stays, assistive devices, therapy and rehabilitation (for persons aged < 18 years), and long-term care (for persons aged 18-76 years), adjusted for age-specific survival.

d) Includes costs of home and vehicle modifications for persons aged < 76 years and costs of special education for persons aged 3-17 years.

e) Includes productivity losses from increased morbidity (i.e. inability to work or limitation on the amount or type of work performed) and premature mortality for persons aged < 35 years with mental retardation, aged <25 years with cerebral palsy and aged < 17 years with hearing loss and vision impairment.

Table 1: Mental retardation is the most important cost factor for the health care system. Estimated prevalent and lifetime economic costs ^{a)} for mental retardation, cerebral palsy, hearing loss, and vision impairment, by cost category; United States 2003 (Adopted from Honeycutt *et al.*, 2003)

Many mental handicaps have genetic causes and about 13% of all MR cases have been estimated to be caused by X-chromosomal gene defects (reviewed by Lubs *et al.*, 1999, updated by Hamel *et al.*, 2000). X-linked mental retardation (XLMR) has a prevalence of 2.6

cases per 1000 population (Stevenson and Schwartz, 2002). In spite of recent advances in the elucidation of the molecular causes of MR (for a review see Ropers and Hamel, 2005) only for the smaller part of the known forms of MR the relevant genes have been elucidated so far. But also for the few forms of XLMR, where the underlying genetic cause has been unraveled, very little is known about the relevant pathogenetic mechanisms. This even holds true for the most frequent disorder of this group, the fragile X syndrome, where the relevant gene, FMR1, has been identified in 1991 (Verkerk *et al.*, 1991) and for which scientists only slowly start to unravel the pathomechanism (Zalfa and Bagni, 2004).

Studying the pathogenesis of mental retardation is difficult because of the large number of brain-expressed genes whose functions and interactions are still mostly unknown. In addition, many of these genes are not expressed in tissues that are readily accessible which greatly hampers the search for target genes or interacting proteins. Postmortem brain tissue can sometimes be obtained from patients with MR but these tissues usually only reflect the late stages of the disease.

Animal models therefore can be very helpful to study tissues affected by the disease in detail at various time points. However, it is sometimes difficult to analyze effects on cognitive functions in such model organisms.

1.2.1 Mutations in *MECP2* cause Rett syndrome

After narrowing the disease locus of RTT to Xq28 by exclusion mapping, a systematic gene screening approach lead to the identification of mutations in the gene *MECP2* (methyl-CpG binding protein 2) as the cause of RTT (Amir *et al.*, 1999). Many reports on heterozygous mutations in this gene associated with RTT have been published since (for a review see Percy and Lane, 2004). Fewer than 1% of all Rett syndrome cases are familial and therefore, the vast majority of *MECP2* mutations (99%) are sporadic (Neul and Zoghbi, 2004). Asymptomatic female carriers have been explained as mosaics or by preferential inactivation of the defective X chromosome (Amir *et al.*, 1999; Wan *et al.*, 1999).

Three Rett syndrome mouse models have been created so far (Chen *et al.*, 2001, Guy *et al.*, 2001, Shahbazian *et al.*, 2002). In the first two cases, a part of the gene was eliminated using the cre/loxP system, thus disrupting *MECP2* in its MBD domain (Chen *et al.*, 2001; Guy *et al.*, 2001) (for details on this protein domain see chapter 1.4). In the third mouse model an N-terminal truncated form of *MECP2* is produced instead of the *wt* protein. These mice have a

less severe phenotype which resembles the human phenotype more closely in that the mice show stereotypic forelimb movements (Shahbazian *et al.*, 2002).

Publication	Genetic modification	Phenotypic features
Chen <i>et al.</i> , 2001	CNS-specific deletion of exon 4 by cre/loxP system with a nestin promoter	<p><i>Mecp2</i>-null mice (normal until ~5 weeks): nervousness body trembling pila erection hard breathing</p> <p><i>at later stages:</i> overweight physical deterioration hypoactive death at ~10 weeks reduced brain size and weight</p> <p><i>Mecp2</i>^{+/-} females (normal for ~4 months): weight gain reduced activity ataxic gait</p>
Guy <i>et al.</i> , 2001	Excision of exons 3 and 4 by cre/loxP system	<p><i>Mecp2</i>-null mice (normal until ~3-8 weeks): stiff, uncoordinated gate hind limb claspings irregular breathing uneven wearing of teeth misalignment of jaws rapid weight loss and death at ~54 days reduced brain size and weight males had internal testis</p> <p><i>Mecp2</i>^{+/-} females: inertia and hindlimb claspings after 3 months</p>
Shahbazian <i>et al.</i> , 2002	Premature stop codon at aa 308 leaves MBD and TRD intact	<p><i>Mecp2</i>^{308/y} mice (normal until ~6 weeks): first symptom is a subtle tremor when suspended by tail tremor worsens with age stereotypic forelimb motions and claspings when hung by tail progressive motor dysfunction decreased activity kyphosis in 40% fur oily and disheveled spontaneous myoclonic jerks and seizures normal brain size and weight</p> <p><i>Mecp2</i>^{+/-} females: milder and variable phenotype</p>
Collins <i>et al.</i> , 2004	Slight over-expression of human <i>MECP2</i> from a PAC clone	<p>normal until ~10-12 weeks forepaw claspings when hung by tail aggressiveness hypoactivity seizures spasticity kyphosis premature death</p>

Table 2. The four RTT mouse models. The table summarizes the phenotypic features of the four mouse models established for RTT as described in the original publications. This list is not comprehensive, since only the features described in the publications were listed and the mice were subjected to different analyses in the studies. In contrast to the first three mouse models, *Mecp2* is overexpressed in the one described by Collins *et al.*

The phenotypes of the models are summarized in Table 2.

In 2004, a mouse model, in which the human MECP2 was mildly overexpressed, was described (Collins *et al.*, 2004). Interestingly, these animals presented with forepaw clasping when hung by the tail, aggressiveness, kyphosis, and hypoactivity characterized by a freezing-like behaviour. Furthermore, with age they developed seizures and spasticity. Apart from the aggressiveness these are all features found in RTT.

These models mimic different degrees of severity of the human phenotype with large deletions of the protein leading to more serious forms of the disease. Luikenhuis and colleagues showed that a Tau-MECP2 transgene expressed exclusively in neurons could rescue the phenotype of MECP2 mutant animals (Luikenhuis *et al.*, 2004). This experiment substantiates that the phenotypes of these models are indeed due to loss of function of MECP2, particularly in neurons. The mouse models therefore seem to be well suited to study the molecular mechanisms underlying the disease.

1.3 Gene expression regulation

Before being implicated in RTT, MECP2 had been described as a transcriptional repressor (Nan *et al.*, 1997). Transcription comprises six distinct steps: (1) pre-initiation complex assembly, (2) promoter opening, (3) transcription initiation, (4) promoter escape, (5) transcription elongation, and (6) transcription termination.

1.3.1 Transcription regulation

Regulation of the above mentioned steps occurs via cis-acting elements and trans-acting factors. Cis-acting elements are DNA sequences in the vicinity of the structural portion of a gene that are required for proper gene expression (e.g. promoters, enhancers, silencers). Trans-acting factors are proteins that bind to cis-acting elements to control gene expression.

There are three types of trans-acting factors: subunits of the RNA polymerases, general transcription factors, and specific transcription factors.

In eukaryotes, subunits of the RNA polymerases bind to the promoter (see Fig. 1 and Fig. 2) of a gene with the help of general transcription factors (in contrast to prokaryotes, where the

polymerase itself recognizes the promoter). General transcription factors (also called basal transcription factors) are required for the initiation of RNA synthesis at all promoters (see Fig.1). Specific transcription factors on the other hand bind to DNA and interact with other trans-acting factors. Many transcription factors bind preferentially to a specific DNA motif.

In addition, co-repressors and co-activators can have an influence on transcription in that they interact with transcription factors (Fig. 1). Co-repressors and co-activators can be: (1) small molecules that change the properties of a transcription factor, (2) proteins that interact with transcription factors and have an influence on the function of the factor, or (3) protein complexes that associate with the transcription factor and mediate its function (e.g. co-repressor complexes that contain histone deacetylases).

Transcription regulation of a gene can be controlled at various levels:

- By direct influence of specific transcription factors as activators and repressors on the assembly of the initiation complex. Activator or repressor proteins directly bind to DNA (often to specific DNA motifs) or interact with DNA-binding molecules. Their activity is often modulated by phosphorylation.
- By changes in DNA sequence
- By changes in DNA structure / conformation (ATP-dependent chromatin remodeling)
- By changes in DNA methylation
- By chromatin protein alterations (histone modification, histone exchange, non-histone proteins, ATP-dependent chromatin remodeling)
- By non-coding RNA molecules

These mechanisms are not mutually exclusive but interact to form a complex network regulating transcription temporally and spatially.

In eukaryotes, mRNA coding genes are transcribed by RNA polymerase II (Pol II). This enzyme is recruited to the transcription start site in the promoter of a gene by the TATA-binding protein (TBP) and the general transcription factor II B (TFIIB). Together with general transcription factors TFIIA, TFIIIE, TFIIF, and TFIIH the pre-initiation complex (PIC) is formed and transcription can take place (Fig. 1).

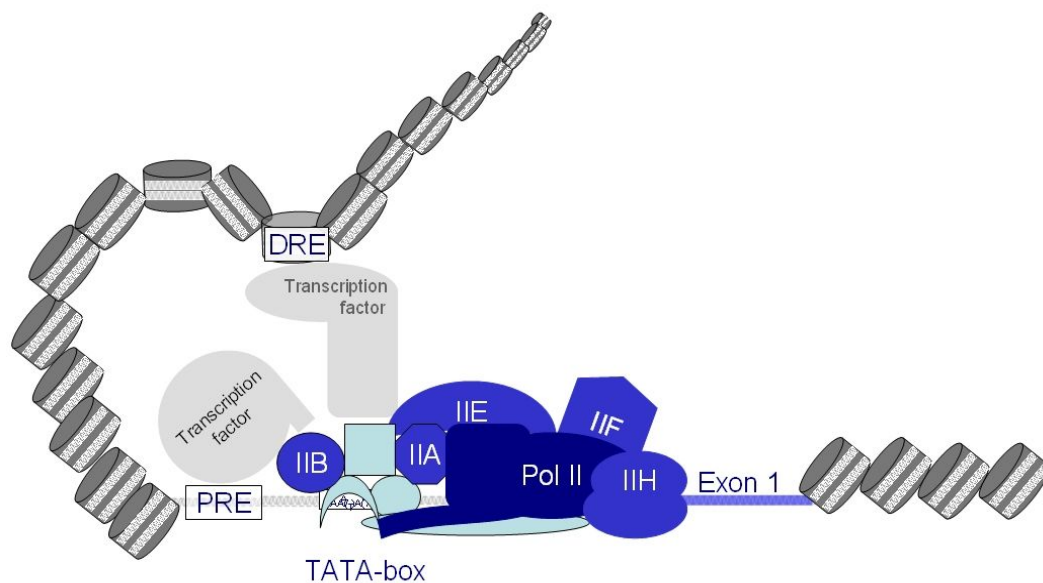


Fig. 1. Transcription initiation. This scheme shows the factors build up a pre-initiation complex at the promoter of a gene. The polymerase is shown in dark blue (Pol II), the general transcription factors (IIA, IIB, IIE, IIF, IIH) are depicted in blue. The TATA-binding protein and associated factors have a turquoise color. Proximal response elements (PRE) and distal response elements (DRE) can be bound by transcription factors (in light grey) that can act as repressors or activators of transcription.

The region around the transcription start site is called promoter and varies from gene to gene. It can be divided into three subregions: core, proximal and distal promoter. The core promoter consists of up to 40 bases upstream of the transcription start site and usually contains the TATA-box (with a consensus sequence of $TAA^{(A/T)}A^{(A/T)}$) somewhere between positions -20 and -30 (Fukue *et al.*, 2004). The proximal promoter spans the region up to about -200 bases, usually with a CAAT-box which is located at position -70 to -80 with respect to the transcription start site. Finally, the distal promoter encompasses regulatory sequences up to 2000 bp or even further away from the transcription start. This region contains response elements, i.e. DNA sequences specifically recognized by transcription factors such as hormone receptors (see Fig. 1 and Fig. 2).

Other typical sequence features of promoters are DNA stretches of high CpG content (CpG islands). They are found at about half of all tissue-specific promoters (Suzuki *et al.*, 2001) and mostly at promoter regions of RNA polymerase II-transcribed genes (Antequera and Bird, 1993; Macleod *et al.*, 1998). CpG islands are defined as regions of 500 bp or more with a

CpG content of at least 55% and a ratio of observed CpGs versus expected CpGs ≥ 0.65 (Takai and Jones, 2002). CpG islands therefore usually cover the core and proximal promoter but can extend into the distal promoter as well (Fig. 2).

In contrast to CpGs elsewhere in the genome, CpG islands are generally hypomethylated. CpG island methylation is often linked to transcription repression in that hypermethylation of a promoter-associated CpG island leads to silencing of gene expression (Cedar *et al.*, 1983, Herman *et al.*, 1994, 1996). However, the opposite effect of CpG island methylation has been observed as well (see section 1.3.3).

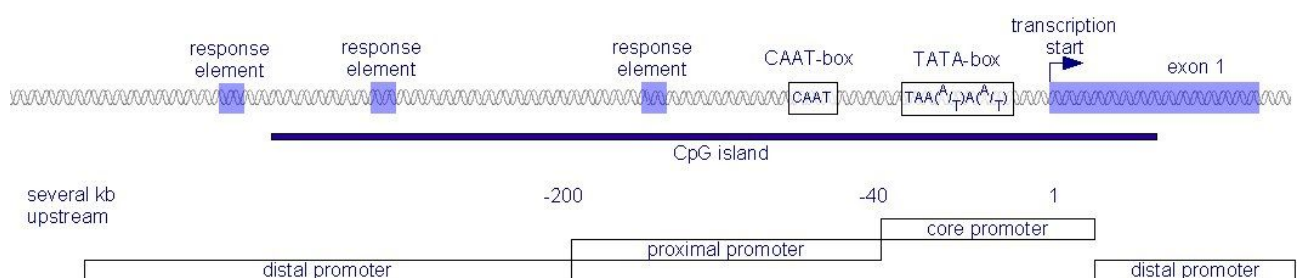


Fig. 2. Representation of an eukaryotic promoter. The relative position of transcription start, TATA-box (-20 to -30), CAAT-box (-70 to -80), and the CpG island can be seen. The existence and position of response elements varies from promoter to promoter.

1.3.2 Epigenetic gene expression regulation

In eukaryotes, transcription regulation has an additional level of complexity due to the organization of DNA in chromatin. Epigenetics has originally been defined by Conrad Hal Waddington in 1942, and meant to describe the process by which genotype gives rise to phenotype, i.e. through causal interactions among genes (gene networks) and their products (Waddington, 1942). Waddington originally defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being” (Waddington, 1942).

Due to the increasing complexity of the field and the elucidation of underlying mechanisms, that were not known to Waddington, the definition of epigenetics has changed over time. The discovery of DNA and its organization as chromatin has led to new insights in the correlation of genotype and phenotype and hence to a new definition of epigenetics as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Russo *et al.*, 1996).

Recently, another modification of the definition of epigenetics has been proposed, motivated by the fact that the mechanisms underlying changes in the chromatin state are also relevant for cells that are not dividing anymore (e.g. differentiating stem cells):

Epigenetics is the study of changes in gene transcription that cannot be explained by changes in DNA sequence (Roloff and Nuber, 2005).

Mechanisms underlying epigenetic gene expression regulation are summarized in Fig. 3.

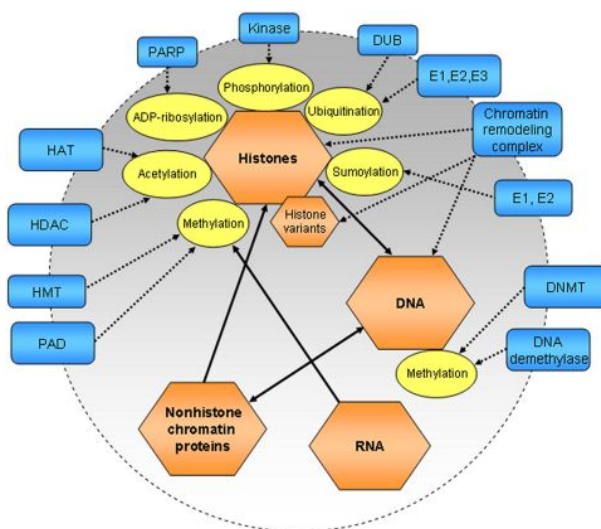


Fig. 3. Summary of epigenetic mechanisms. The basic components of epigenetic gene regulation are depicted in orange, possible modifications to DNA or proteins are shown in yellow, and enzyme classes mediating the modifications are colored in blue. (Adopted from Roloff and Nuber, 2005)

1.3.3 DNA methylation

One major mechanism of epigenetic gene expression regulation is the methylation of DNA. In mammals, this primarily occurs at the 5' position of the cytidine ring structure of CpGs (Fig. 4). Ramsahoye and colleagues (2000), showed that cytosines followed by a base other than G (especially CpA) can also be methylated, albeit to a lesser extent.

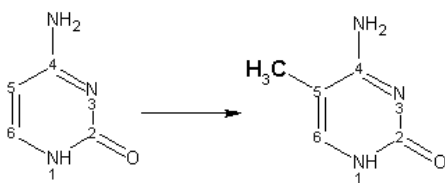


Fig. 4. Cytosines can get methylated at the 5' position .

The DNA methylation of many gene regulatory regions inversely correlates with gene expression. Here, DNA methylation does not seem to be essential for the initiation but rather for the maintenance of gene silencing (Bird, 2002). Although true for many genes, DNA methylation is not generally linked to gene silencing. For example, the unmethylated H19 imprinting control region leads to repression of the maternal *IGF2* allele and its methylation leads to the expression of the paternal *IGF2* allele (Holmgren *et al.*, 2001).

DNA methylation is established and maintained by DNA methyltransferases (DNMT) while demethylation takes place by passive mechanisms (i.e. absence of maintenance DNMT activity) or enzymatic activity. Different reports on active demethylation in mammals have been published. DNA glycosylase activities were described (Zhu *et al.*, 2000, Vairapandi 2004) and a direct removal of the methyl group from the cytosine has been proposed but also disputed (Ramchandani *et al.*, 1999, Wade *et al.*, 1999); finally, proteins of the Aid/Apobec1 family can convert cytosine to uracil by deamination, thus leading to excision repair and replacement of the uracil by an unmethylated cytosine (Morgan *et al.*, 2004).

DNA methylation is carried out by the DNA methyl transferase protein family that, in mammals, can be split into three distinct classes: DNMT1, DNMT2 and DNMT3. DNMT3a and DNMT3b have originally been described as *de novo* methyltransferases (Okano *et al.*, 1998, Okano *et al.*, 1999) while DNMT1 maintains methylation. However, recent findings suggest a direct interaction between DNMT1 and DNMT3a/3b (Hattori *et al.*, 2004, Kim *et al.*, 2002; Rhee *et al.*, 2002) and hence involvement of all proteins in both mechanisms. DNMT2 was found as a homologue to the *Schizosaccharomyces pombe* DNA methyltransferase (Okano *et al.*, 1998) and weak methyltransferase activity of this protein has been shown in *Drosophila*, mouse, and man (Tang *et al.*, 2003, Hermann *et al.*, 2003).

Finally, the number of methylated cytosines in CpA and CpT motifs has been shown to be decreased in differentiated tissue as compared to ES cells (Ramsahoye *et al.*, 2000). This raises the question whether these methylated motifs, in addition to CpG methylation, play a role in gene expression and chromatin state regulation and what proteins mediate these

methylation signals.

DNA methylation is associated with many genetic diseases, for example: Prader-Willi syndrome (Prader *et al.*, 1956), Angelman syndrome (Kishino *et al.*, 1997), and ICF (Immunodeficiency - Centromeric instability - Facial anomalies) (Xu *et al.*, 1999). In addition, developmental abnormalities and certain tumors can originate from abnormal DNA methylation (Yoder *et al.*, 1996). Rett syndrome (Amir *et al.*, 1999) is a genetic disease indirectly associated with DNA methylation since the product of the disease causing gene binds to m⁵CpG.

There is a global and a local aspect of transcription regulation by DNA methylation. While changes in activity of DNMTs can have an effect on gene expression in the whole genome, single genes have to be regulated specifically by methylation of CpGs at prominent positions (e.g. in the promoter or in other regulatory sequences of these genes). How the site-specific regulation works in detail is poorly understood.

DNA methylation at CpGs has been shown to often correlate with histone deacetylation. This link can be explained by the interaction of methyl-CpG binding proteins (see chapter 1.4.) with complexes containing histone deacetylases (HDAC) (Nan *et al.*, 1998, Ng *et al.*, 2000; Zhang *et al.*, 1999).

1.4 MBD proteins

The effect of DNA methylation is mediated by proteins that recognize m⁵CpGs, namely the members of the methyl-CpG binding domain (MBD) protein family and Kaiso (Prokhortchouk *et al.*, 2001; Hendrich and Tweedie, 2003; Roloff *et al.*, 2003). These proteins recognize m⁵CpGs and can interact with different co-repressor complexes and hence act as translators of DNA methylation patterns for transcription regulation.

MBD proteins contain a methyl-CpG binding domain which consists of ~70 residues in an α/β -sandwich fold built of three to four β -twisted sheets and a helix with a characteristic hairpin loop in the opposite layer (Ohki *et al.*, 1999, Wakefield *et al.*, 1999) (see Fig. 5). The prototype of this domain has been shown to allow specific binding to symmetrically methylated CpGs (Nan *et al.*, 1993).

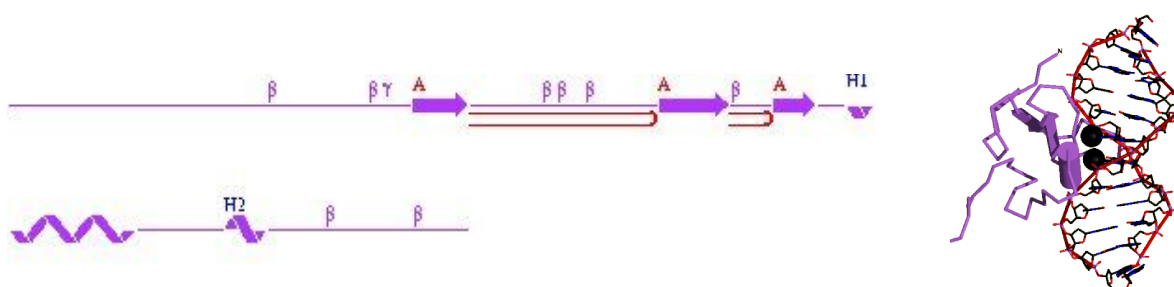


Fig. 5. Structure of the MBD domain of human MECP2. The MBD domain consists of three β -twisted sheets (indicated by arrows) and one main α -helix (indicated by H1 and H2) (protein chain from PDB protein data bank). On the right, the 3D structure of a methyl-CpG binding domain interacting with DNA is shown. The two methyl groups are depicted in black, the DNA is represented on the right, the MBD is on the left. (Adopted from PDB)

Members of this protein family are found in vertebrates as well as invertebrate animals. At the beginning of this thesis in 2001, five vertebrate MBD proteins were known: MBD1, MBD2, MBD3, MBD4, and MECP2 (for a review see: Ballestar and Wolffe, 2001, Hendrich and Tweedy, 2003). Except for MBD4, all of them are associated with histone deacetylases (HDAC), and a transcriptional repression mechanism mediated by the recruitment of HDACs has been shown for MECP2, MBD1, and MBD2 (Nan *et al.*, 1993, Ng *et al.*, 2000, Prokhortchouk *et al.*, 2001).

1.4.1 MBD1

MBD1 binds symmetrically methylated CpGs, preferably if they are separated by few bases (Ballestar and Wolffe, 2001). Five isoforms of MBD1 are generated by alternative splicing, resulting in proteins that contain one MBD domain, two to three cysteine-rich (CXXC) domains, and different C-termini (Expasy database: MBD1_human). All isoforms repress transcription from methylated promoters *in vitro* (Fujita *et al.*, 1999). In addition, isoforms with three CXXC domains also repress unmethylated promoter activity (Fujita *et al.*, 1999). Suv39h1, a histone methyltransferase, enhances MBD1-mediated transcriptional repression by binding the MBD, not the C-terminal transcriptional repression domain of MBD1. MBD1 indirectly binds to histone deacetylases through Suv39h1, causing methylation and deacetylation of histones that are associated with gene inactivation (Fujita *et al.*, 2003a).

Recent data show that MBD1 forms a stable complex with histone H3-K9 methylase SETDB1 and binds to CHAF1B (p60 subunit chromatin assembly factor I (CAF-I) required for the assembly of histone octamers onto newly-replicated DNA) (Sarraf and Stancheva, 2004). A molecular link between p59 OASL and MBD1 has been established in the context of an interferon-stimulated cell (Andersen *et al.*, 2004). p59 OASL is a protein that binds to RNase L leading to repression of gene expression by general RNA degradation.

Besides its repressive function, MBD1 interacts with the transcription activator MCAF via the transcriptional repression domain of MBD1 (Fujita N *et al.*, 2003b).

Apart from gene expression regulation, MBD1 also regulates cell cycle G1-S transition and apoptosis via the p53/p21 (Waf1) pathway (Kano *et al.*, 2004).

In addition, MBD1 interacts with MPG (a methylpurine-DNA glycosylase which excises damaged bases from substrate DNA) via its transcription repression domain (TRD). Upon DNA damage by methylmethanesulfonate or a similar alkylating agent, MBD1 binding to DNA is lost, MPG binds to the damaged sites, blocks gene expression and reverses the damage (Watanabe *et al.*, 2003).

A knockout mouse model for *Mbd1* (*Mbd1*^{-/-}) shows more or less normal development but increased genome instability, decreased adult neurogenesis, and impaired spatial learning (Zhao *et al.*, 2003).

1.4.2 MBD2

The MBD2 protein contains an MBD domain and a coiled coil region. It binds to m⁵CpGs in highly methylated regions (Ballestar and Wolffe, 2001), but in complex with MBD3, a binding to hemi-methylated DNA is also possible (Tatematsu *et al.*, 2000).

There are three transcript variants: MBD2a is the full transcript, MBD2b is lacking the first 149 aa and the third variant has a different C-terminus (Expasy database: MBD2_human).

MBD2 has been shown to recruit the co-repressor complex Mi-2/NuRD to methylated DNA *in vitro* (Zhang *et al.*, 1999), thereby mediating chromatin condensation via histone deacetylation and gene expression repression. Interestingly, the Mi-2/NuRD complex contains MBD3 (see chapter 1.4.3). Two ubiquitously expressed and highly related p66 proteins seem to be part of this repression complex. In addition, both interact with MBD2, albeit with different binding domains (Brackertz *et al.*, 2002). MBD2 also binds to the Sin3A repressor via the MBD domain (Boeke *et al.*, 2000).

Bakker and colleagues showed that MBD2 represses transcription from hypermethylated pi-class glutathione S-transferase gene promoters in hepatocellular carcinoma cells (Bakker *et al.*, 2002).

In contrast to the examples of unspecific repression shown above, sequence specificity of MBD2 mediated repression can be achieved by the interaction with MIZF. This zinc finger protein binds to CGGACGTT motifs if methylated and, by interplay with MBD2, leads to repression of the genes in that region (Sekimata and Homma, 2004). Furthermore, MBD2 has a role in the methylation-mediated inhibition of ribosomal RNA gene expression (Ghoshal *et al.*, 2004). MBDin – a GTPase shown to bind to MBD2a at the extreme C-terminus - relieves MBD2 repression and reactivates transcription from methylated promoters (Lembo *et al.*, 2003).

On the other hand, MBD2a and RNA helicase A cooperatively enhance CREB-dependent gene expression (Fujita H *et al.*, 2003), thus activating instead of repressing gene expression at a methylated promoter element.

The MBD2 protein has also been proposed to act as a DNA demethylase *in vitro* and *in vivo* (Bhattacharya *et al.*, 1999; Cervoni and Szyf 2001, Detich *et al.*, 2002) leading to activation of CpG sites within the promoter region of reporter genes. This however has been questioned (Ng *et al.*, 1999, Wade *et al.*, 1999, Boeke *et al.*, 2000).

MBD2 is also interesting in the context of cancer. A study by Patra *et al.* shows that *MBD1* is expressed in different prostate cancer tumor cell lines, but *MBD2* and *MECP2* are not (Patra *et al.*, 2003).

MBD2 gene expression may be a significant factor in tumorigenesis, in that high levels of *MBD2* expression correlate with reduced risk of cancer (Zhu *et al.*, 2004). This is in contrast to the observation, that antisense oligoDNA for MBD2 suppresses tumor growth in nude mice (Campbell *et al.*, 2004).

MBD2 not only suppresses transcription of Pol II-transcribed genes but also of Pol I-transcribed rRNA genes.

A knockout mouse model for *Mbd2* (*Mbd2*^{-/-}) has been created by Bird and colleagues (Hendrich *et al.*, 2001). These mice are viable and fertile but the nurturing behavior of *Mbd2*^{-/-} mothers is disturbed. Furthermore, the *Mbd2*^{-/-} mice show normal DNA methylation patterns, suggesting that Mbd2 does not act as a DNA demethylase or at least is not the only DNA demethylase.

Mbd2^{-/-}/*Mecp2*^{-y} double knockout mice showed the same phenotype as *Mecp2*^{-y} mice suggesting no direct genetic interaction between *Mecp2* and *Mbd2* (Guy *et al.*, 2001). Reduced survival of mice with mutations of *Mbd3* in an *Mbd2*^{-/-} background, however, suggested a genetic interaction between these two proteins (Hendrich *et al.*, 2001).

1.4.3 MBD3

MBD3 only interacts non-specifically with DNA (Fraga *et al.*, 2003) and unlike the other family members, MBD3 is not capable of binding methylated DNA. However, it is involved in repression of transcription as a component of a co-repressor complex (Zhang *et al.*, 1999). MBD3 (Expasy database: MBD3_human) is a subunit of NuRD, a multisubunit complex containing nucleosome remodeling and histone deacetylase activities. MBD3 mediates the association of metastasis-associated protein 2 (MTA2) with the core histone deacetylase complex.

Mbd3 knockout mice (*Mbd3*^{-/-}) die during early embryogenesis (Hendrich *et al.*, 2001).

The *Drosophila melanogaster* homolog of MBD3 has been shown to mediate the interaction between the MI-2 chromatin complex and CpT/A-methylated DNA (Marhold *et al.*, 2004).

Two proteins homologous to MBD3 have been found. MBD3L1 and MBD3L2 have no MBD domain and do not bind to m⁵CpGs *in vitro* (Jiang *et al.*, 2002). MBD3L1 is a transcriptional repressor that interacts with MBD2 and the NuRD complex. *Mbd3l1* knockout mice are viable and fertile suggesting that the protein is not essential and its function probably redundant to MBD3 (Jiang *et al.*, 2004).

1.4.4 MBD4

In contrast to the other MBD family members, MBD4 (Expasy database: MBD4_human) preferably binds to m⁵CpG/GpT mismatches instead of symmetrically methylated m⁵CpGs. Such mismatches can result from spontaneous deamination of methylated cytosines. MBD4 can reverse the mismatch by excision of the thymine (Hendrich *et al.*, 1999, Petronzelli *et al.*, 2000). Furthermore, *MBD4* gene mutations are detected in tumors with primary microsatellite-instability (MSI), a form of genomic instability associated with defective DNA mismatch repair, and the *MBD4* gene meets 4 of 5 criteria for a bona fide MSI target gene.

1.5 MECP2

MECP2 is the founder of the MBD protein family, in that it was the first family member to be cloned (Lewis *et al.*, 1992).

1.5.1 Human *MECP2*: Gene structure and mutations

The MECP2 gene maps to Xq28 spanning 76kb and is composed of four exons transcribed from telomere to centromere. Its transcript of 1461 nucleotides was originally described with a coding sequence in exons 2-4. A second isoform has been isolated in 2004 with a transcript encompassing exons 1, 3, and 4 (Kriaucionis and Bird, 2004, Mnatzakanian *et al.*, 2004) (Fig. 7). Since exon 2 consists of only 124 nucleotides, the size of the coding sequence is almost the same, explaining why the second isoform was not detected earlier.

The original isoform is now called MECP2e2, while the new isoform lacking exon2 is called MECP2e1 (where e1 and e2 stand for exon1 and exon2). Until 2004, the presence of MECP2e1 was not known and previous studies did not take that difference into account. In general, only MECP2e2 was studied, or both isoforms were analyzed without being aware of that fact. If not further specified, the use of the abbreviation MECP2 therefore indicates that the study did not distinguish between MECP2e1 and MECP2e2.

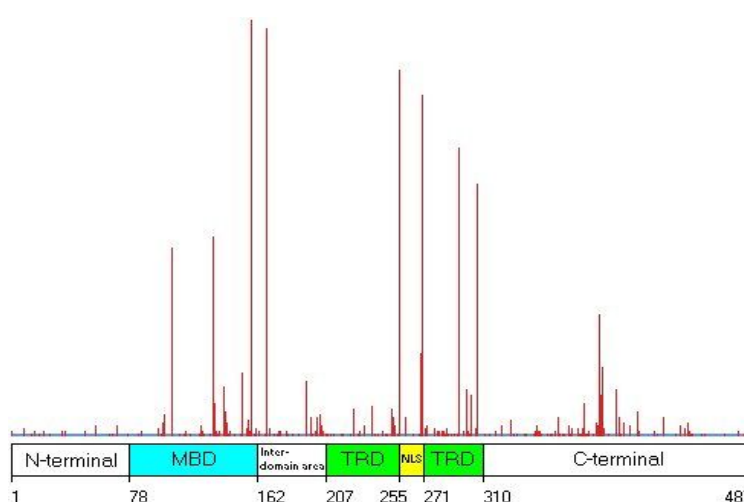


Fig. 6. Distribution and frequency of mutations in *MECP2e2*. The height of the lines indicates how many individuals carrying a mutation at this site have been found. Mutation hot spots are seen in the MBD and the TRD domain. (adopted from RettBASE).

Following the elucidation of *MECP2* as disease gene for RTT, variations in its genomic sequence have been studied intensively. RettBASE, a *MECP2* variation database features more than 290 amino acid exchanges, insertions, and deletions in a total of 1595 cases from 61 publications and from many direct entries (Fig. 6).

Early nonsense mutations (leading to a stop codon in the 5' part of the gene) likely lead to loss of the transcript via nonsense-mediated mRNA decay, whereas late nonsense mutations (leading to a stop codon in the 3' part of the gene) likely result in a truncated protein. Nonsense mutations within the methyl-CpG-binding domain have been experimentally shown to abrogate methylation-specific binding to DNA (Yusufzai and Wolffe, 2000). However, most of the truncating mutations to be found in RettBASE are distal to the MBD, and one hypothesis holds that the truncated proteins still bind methylated DNA but cannot interact with co-repressor complexes. Experimentally, nonsense mutations within the TRD were shown to result in premature truncations and the resulting proteins were unable to repress transcription (Yusufzai and Wolffe, 2000).

Investigations by Yusufzai and Wolffe (2000) revealed that missense mutations within the MDB abrogated methylation-specific binding to DNA. Different functional consequences of *MECP2* mutations are conceivable:

A complete functional loss of the protein as a transcriptional silencer might lead to transcriptional upregulation of target genes, as for instance suggested by studies on the *MDR1* gene (El-Osta and Wolffe, 2001, El-Osta *et al.*, 2002), the imprinted genes H19 (Drewell *et al.*, 2002), and U2af1-rsl (Gregory *et al.*, 2001). Altered replication timing of the inactive X chromosome in a portion of lymphocytes from RTT patients as well as a decreased compaction of heterochromatic regions on chromosome 9 have been reported together with the idea of a connection between RTT and X-chromosome replication disturbance (Vorsanova *et al.*, 1996; Vorsanova *et al.*, 1998). Moreover, indirect effects of mutations might occur. Loss of DNA binding capability could lead to an interaction with unmethylated sequences in a non-specific manner with a persistent binding to co-repressors or to no interaction with DNA at all.

As reported by Amir *et al.* (2000), the influence of the mutation type on the phenotype (13 clinical traits) was analyzed in 48 RTT patients. Correlations were only found between truncating mutations and two parameters (breathing abnormalities and low levels of

homovanillic acid in cerebrospinal fluid). Neither the overall severity score nor any of the other parameters correlated with the type of mutation. Cheadle *et al.* (2000) reported missense mutations to have significantly milder clinical consequences than truncating mutations, but these findings are not confirmed by the study of Huppke *et al.* (2000). In both studies, however, several patients with identical mutations but widely different phenotypes are described, thereby illustrating that the significant clinical variability seen in RTT cannot be ascribed to allelic differences alone.

Mutations in *MECP2* have been found in patients with widely varying phenotypes. In two reports dealing with familial *MECP2* mutations, male patients were described which present with severe non-specific X-linked mental retardation, but display few if any signs of RTT (Meloni *et al.*, 2000; Orrico *et al.*, 2000). Female patients with random X-inactivation pattern did not present with RTT syndrome, but with a mild form of mental retardation (Meloni *et al.*, 2000; Orrico *et al.*, 2000). Thus, the phenotypic variability of *MECP2* mutations can in part be explained by skewed X-inactivation and by different types and positions of mutations. In addition, a digenic model has been proposed in which RTT is considered as a disease that develops when a *de novo* mutation in *MECP2* occurs in the presence of a mutation in another gene. According to this model, *MECP2* mutations alone would produce a recessive phenotype of non-specific mental retardation, whereas a mutation in the second gene alone may produce no phenotypic effect at all (Meloni *et al.*, 2000). On the other hand, *MECP2* mutations have been found in up to 80% of sporadic and approximately 50% of familial cases, which clearly opposes the digenic model.

The 20% of missing mutations may be due to the fact that large non-coding regions of the gene as well as the promoter have generally not been screened for mutations, that the existence of microdeletions as well as large rearrangements have often been neglected, and that there might be mutations in the coding region of *MECP2e1*. One patient has been shown to have a mutation in the coding region of *MECP2e1* (Mnatzakanian *et al.*, 2004). However, another study could not find any mutations in the promoter or exon 1 of *MECP2* in 97 mutation-negative RTT patients (Evans *et al.*, 2004)

1.5.2 MECP2: Protein structure

Besides the MBD, *MECP2* also contains two AT-hooks, a TRD, two nuclear localisation sequences (NLS), a N-terminal segment with a group II WW domain binding region (WWBR)

and a C-terminal part with a histidine- and a proline-rich region. The location of the domains relative to the transcripts and the genomic arrangement can be seen in Fig.7.

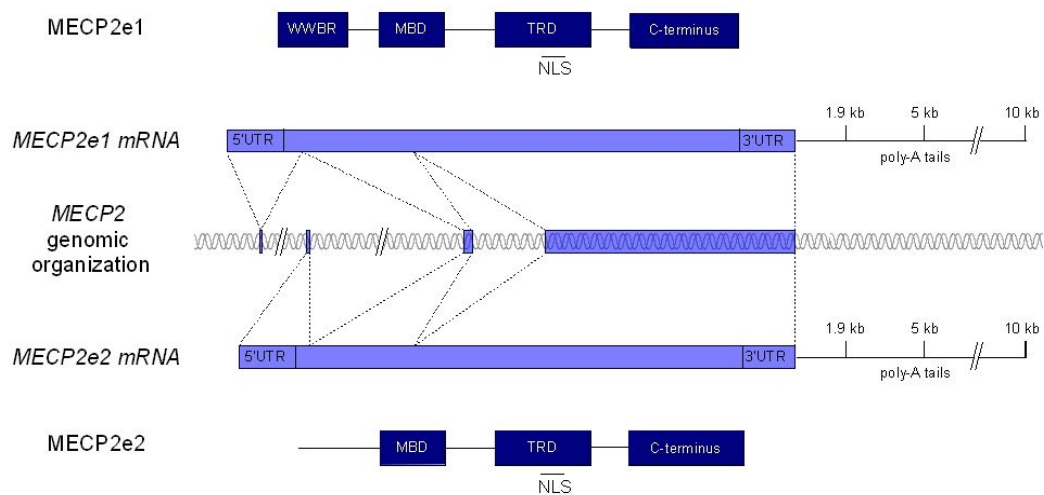


Fig. 7. MECP2 genomic, transcript, and protein structure. Human MECP2: genomic organization (in the middle), mRNAs and domain organization of MECP2e1 (top) and MECP2e2 (bottom) proteins. The three different transcripts for each splice variant are denoted by the different poly-A tail sizes that are possible.

AT-hooks are DNA binding motifs of eleven residues that allow binding to the minor groove of preferably A/T rich regions.

TRD describes a protein domain that can mediate transcription repression, usually via interaction with histone deacetylase containing protein complexes. The TRD of MECP2 can bind to TFIIB *in vitro* (Kaludov and Wolffe, 2000). Repression by MECP2 as well as by a TRD fusion product correlates with selective assembly of large nucleoprotein complexes. This suggests, that even in the presence of initiation factor components, a repressor complex can be established, preventing transcription initiation.

There are two nuclear localization signals, one lying between nucleotides 173 and 193 of MECP2e2 and one located within the TRD region.

The C-terminal segment was shown to facilitate the binding to the nucleosome core (Chandler *et al.*, 1999). WW domains, present in the C-terminal portion, are characterized by two tryptophane (W) residues and bind to proline-rich regions. According to the recognition sequence these domains are classified into four groups. Group II domains specifically bind PPLP (pro-pro-lys-pro) motifs. The domain is found at aa residues 384-387 in human

MECP2e2.

Histidine and proline-rich stretches as found in the C-terminal region are conserved in certain neural-specific transcription factors (Vacca *et al.*, 2001).

The human and mouse 3'UTR contain eight regions of closely conserved sequence similarity which has led to the suggestion that they may be involved in stabilization of the transcript as well as in post-transcriptional regulation (Coy *et al.*, 1999).

MECP2e1 furthermore features a polyalanine (poly-A) and polyglycine (poly-G) tract in its 21 aa specific N-terminal sequence. Poly-A and poly-G stretches are also found in members of the homeobox family (Mnatzakanian *et al.*, 2004; Utsch *et al.*, 2002)

1.5.3 MECP2: Function

MECP2 can function as a transcriptional repressor. The MBD of MECP2 is necessary and sufficient for DNA binding *in vitro* and allows MECP2 to preferentially recognize a single symmetrically methylated CpG in diverse sequence contexts (Nan *et al.*, 1993). MECP2 is abundantly found in the heavily methylated pericentromeric heterochromatin of mouse chromosomes (Lewis *et al.*, 1992) and a transiently expressed MECP2 fusion protein is targeted to methylated heterochromatin (Nan *et al.*, 1996). Furthermore it was found to bind to nuclear matrix attachment regions, the putative anchorage sites of chromatin loop domains (von Kries *et al.*, 1991).

Transcriptional repression has been described to be mediated by MECP2 in five different reports. The mechanism described in the first two reports rely, to a significant extent, on histone deacetylation.

1. A histone deacetylation-dependent mechanism of transcriptional repression is realized by the interaction of the TRD with the co-repressor mSin3A which in turn is part of a large co-repressor complex containing histone deacetylases HDAC1 and 2 (Nan *et al.*, 1998). The deacetylation of histones allows DNA to wind more tightly around the histone, preventing access of the transcription machinery to the promoters.

This mechanism has been questioned in 2004 by Klose and Bird. They showed that MECP2 exists as an elongated monomer and that its interaction with Sin3A is not stable. This also correlates with the observation that ARBP (the chicken MECP2) has an elongated shape (von Kries *et al.*, 1994).

2. In 2001 Kokura and colleagues showed that MECP2 interacts with Ski and N-CoR (Kokura

et al., 2001), a complex containing HDACs (for a review see Jones and Shi, 2003).

3. A histone deacetylase-independent mechanism of repression was demonstrated by transient transfection studies using a reporter plasmid containing the SV40enhancer/promoter (Yu *et al.*, 2000). In this setting, TRD-mediated transcriptional repression could not be relieved by the specific histone deacetylase inhibitor TSA. However, TRD-mediated repression using a different promoter (adenovirus major late promoter) could be relieved by TSA, indicating that the mode of action, histone deacetylase-dependent or -independent, relies on the promoter context.

4. Another way of transcription repression mediated by MECP2 is the blocking of binding sites for transcription factors. This has been shown for the E2F binding site at the Rb-1 promoter in tumor cells (di Fiore *et al.*, 1999). Also, MECP2 has been demonstrated to inhibit the assembly of the basal transcriptional machinery at methylated promoters in the absence of chromatin assembly and to associate with TFIIB *in vitro* (Kaludov and Wolffe, 2000).

5. MECP2 links DNA methylation and histone methylation by interaction with a histone methylase whose nature is yet unknown. This interaction reinforces the repressive function of DNA methylation by recruitment of proteins that can methylate histones (Fuks *et al.*, 2003a, Fuks *et al.*, 2003b).

Gene repression by MECP2 can be a dynamic process. For example, membrane depolarization of primary neurons leads to a dissociation of MECP2 from a *Bdnf* promoter and increased *Bdnf* expression. This change is accompanied by chromatin changes at this site (Chen *et al.*, 2003; Martinowich *et al.*, 2003).

MECP2 also seems to be involved in maintenance of DNA methylation. MECP2 binds to hemimethylated CpGs after DNA replication and can recruit DNMT1 which then fully methylates the CpG (Kimura and Shiota, 2003).

MECP2 therefore has a double function in repressing gene expression as well as maintaining the methylation pattern for continuous repression.

1.5.4 MECP2: Expression

The *MECP2* gene is ubiquitously expressed. Although early developmental stages show low levels of expression, *MECP2* is widely active in embryonic and adult tissues. In human tissues, originally three different transcripts (1.8 kb, ~7.5 kb, and 10 kb), resulting from differential use of polyadenylation signals, have been described in most tissues with tissue-

specific variation in expression: brain and spinal cord show higher expression of the long transcript (10 kb) whereas the smaller transcript (1.8 kb) is more abundant in other tissues, e.g. muscle and lymphoid tissues (D'Esposito *et al.*, 1996; Reichwald *et al.*, 2000). These two transcripts have similarly short half-lives, and the functional significance, if any, of their differential expression is unknown.

Due to the splice variants, a total of 6 transcripts could exist, at least in principle (2 variants with 3 different polyadenylation signals each)(Fig. 7). The *MECP2e1* variant has been shown to be ubiquitously expressed as well. In adult human brain, *MECP2e1* levels were 10 times higher than *MECPe2* while in other tissues *MECP2e2* transcripts are expressed more strongly (Mnatzakanian *et al.*, 2004).

1.6 Techniques

1.6.1 DNA microarrays

Gene expression profiling on a large scale is nowadays mostly performed using DNA microarrays (for a review on DNA microarray techniques see Stoughton, 2005).

Basically three types of DNA microarrays can be distinguished depending on the material immobilized on the array: cDNA, oligonucleotide, and genomic arrays. For this study, cDNA microarrays were used.

In such expression studies, the RNA of interest (i.e. from cell type or tissue after a certain treatment or at a certain cellular state) is labeled during reverse transcription into cDNA. The labeling is achieved by incorporating fluorescent dyes during the reverse transcription procedure. To compare two samples, the cDNAs are labeled with different dyes and then co-hybridized to a microarray containing large numbers of PCR-amplified cDNAs. Every PCR-amplified cDNA sample on the array has been spotted to a different area. After hybridization of the labeled cDNAs, a laser scanner allows to detect specific fluorescent signals for every spot on the array. The comparison of the two signals of the co-hybridized cDNA pools shows the relative presence of each cDNA sequence that corresponds to RNA levels in the target sample. The more cDNAs are spotted on the array, the more mRNA levels can be studied. For general screens such as the one planned for this thesis, the use of a comprehensive array is of advantage.

1.6.2 Chromatin immunoprecipitation

To study *in vivo* DNA-protein interactions, chromatin immunoprecipitation has become the method of choice (for a review see Das *et al.*, 2004). Basically, DNA and interacting proteins are cross-linked in cells, the chromatin is then cleaved into small pieces (e.g. by shearing) and the DNA fragments of interest (i.e. bound by the protein to be studied) are isolated by immunoprecipitation with a specific antibody. After reversal of the cross-linking, obtained DNA fragments can be analyzed in several ways such as sequence-specific PCR, sequencing, or hybridization to a genomic DNA microarray (Fig. 8).

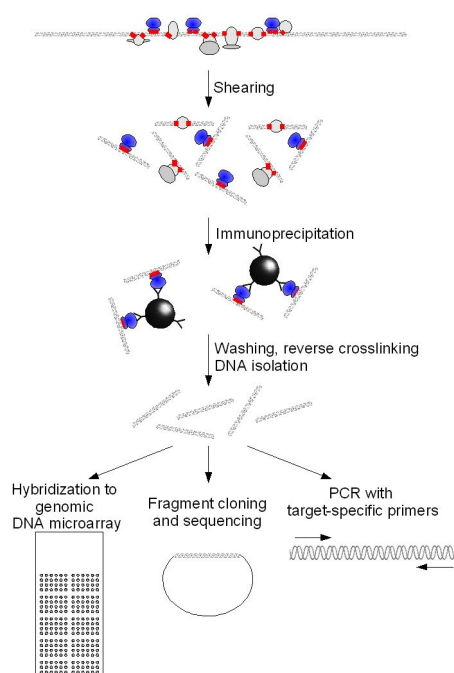


Fig. 8. Scheme of chromatin immunoprecipitation.

Protein-DNA complexes are linked (red dots) in cells by the use of formaldehyde. The chromatin is then sheared into small fragments and the protein of interest (blue) is immunoprecipitated with a specific antibody bound to protein A-agarose (black). After washing steps and unlinking, DNA fragments that were bound to the protein of interest can be isolated. To characterize these fragments basically 3 methods exist. The fragments can be labeled and hybridized to a genomic DNA microarray. They can also be cloned into a vector and sequenced. Finally, they can be amplified with primers specific for a genomic region. The last method is especially useful if potential target sites are known (e.g. from prior gene expression studies).

Many steps in this approach have to be optimized depending on the biological material used, the protein studied, and the method for the analysis of the DNA fragments obtained by ChIP. ChIP is especially suited for transcription factors with many genomic binding sites, such as MECP2, since it possibly supplies all target sequences in only one experiment.

1.7 Open questions and approaches

Even though a lot of facts have been gathered about RTT and MECP2, the molecular mechanisms underlying this disease remain unknown. To understand the pathomechanism of RTT, the transcriptional consequences of MECP2 mutations and target genes of this protein need to be identified. Gene expression profiling with DNA microarrays has become the method of choice for studying the expression of large numbers of transcripts in parallel. This technique is especially suited if the disease gene is a transcriptional regulator, such as MECP2. To confirm the binding of MECP2 to the potential target genes, ChIP was performed.

Since the beginning of this thesis, in three published studies related to RTT, the microarray technology has been applied. Post mortem RTT brains, brain tissue from a RTT mouse models and lymphoblastoid cell lines were used respectively (Colantuoni *et al.*, 2001, Tudor *et al.*, 2002, Ballestar *et al.*, 2005). None of the studies has however revealed genes regulated by MECP2 that would plausibly explain the RTT phenotype. A distinct feature of RTT is, that apart from the brain, mutations in MECP2 do not seem to have a strong effect on other organs. This could be explained by compensation of loss of function of MECP2 by proteins with similar properties. Finding such proteins could help to understand the exact role of MECP2 and why the loss of MECP2 function primarily shows an effect in the brain.

1.8 Goals of the thesis

This thesis therefore had two goals:

1. The first goal was to find proteins that could compensate for the loss of function of MECP2 in peripheral tissues.

To do so, 2 projects were designed. "**The detection of MBD protein family members formerly not recognized**" to find proteins with an methyl-CpG binding domain that might, like MECP2, act as transcriptional repressors, and "**the search for paralogues of MECP2**", that might be structurally and functionally related to MECP2.

2. The second goal was to identify "**MECP2 target genes**" in the brain, and to determine pathways involved in the pathogenesis of RTT.

These studies should give new insights into the pathogenesis of RTT and should help to explain why loss of function of the ubiquitously expressed gene *MECP2* gives rise to a syndrome that is primarily confined to the nervous system.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

Standard chemicals were ordered from Amersham (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany), Biochrom (Berlin, Germany), Calbiotech (Bad Soden, Germany), Invitrogen (Karlsruhe, Germany), Merck (Darmstadt, Germany), Roche Biochemicals (Basel, Switzerland), Roth (Karlsruhe, Germany), Sigma-Aldrich (Sigma-Aldrich Chemie, Munich, Germany) in the quality pro Analysis (p.a.). For special chemicals and solutions, the manufacturer's name is indicated in the individual methods section.

Buffers were prepared with millipore water. Aqua ad iniectabilia (Baxter, Unterschleissheim, Germany) was used to take up DNA pellets and prepare small amounts of aqueous solutions.

2.1.2 Labware

All reusable labware was autoclaved before use. Plasticware was ordered from BD Biosciences (Heidelberg, Germany) or Eppendorf (Hamburg, Germany).

2.1.3 Kits

Blood & Cell Culture DNA Maxi-kit (Qiagen, Hilden, Germany)

DNeasy tissue kit (Qiagen)

QIAquick PCR purification kit (Qiagen)

QIAGEN Plasmid Mini kit (Quiagen)

QIAquick gel extraction kit (Quiagen)

Chromatin Immunoprecipitation (ChIP) Assay Kit (Upstate/Biomol, Hamburg, Germany)

2.1.4 Vectors

pCR[®]2.1-TOPO[®] vectors from the TOPO TA cloning[®] system (Invitrogen) were used.

2.1.5 Bacterial strains

One Shot[®] TOP10F' competent cells were used for transformations.

2.1.6 Oligonucleotides

Oligonucleotides used were designed with the Primer3 program except for primers for bisulfite sequencing, that were designed manually and ordered from MWG Biotech (Ebersberg, Germany). Stocks were stored as 100 mM solutions in H₂O at -20°C.

2.1.7 Antibodies

Antibody	Species	Company	Catalogue number
anti-acetylated-histone H3	rabbit polyclonal IgG	Upstate/Biomol	06-599
anti-FKBP51 (FKBP5)	goat polyclonal IgG	Santa Cruz Biotechnology,	sc-11518
anti-glucocorticoid receptor (Ab-2)	mouse monoclonal IgG	Oncogene (San Diego, CA, USA)	GR32L
anti-MeCP2 C-terminus	rabbit polyclonal IgG	Upstate/Biomol	07-013
anti-mSin3A	rabbit polyclonal IgG	Santa Cruz Biotechnology	sc-944
anti-SGK1	goat polyclonal IgG	Santa Cruz Biotechnology	sc-15885

Table 3. Primary antibodies used for ChIP and immunohistochemistry.

Antibody	Species	Company	Catalogue number
FITC-anti-rabbit	donkey	Dianova (Hamburg, Germany)	111-095-152
Cy3-anti-goat	donkey	Dianova	705-165-147

Table 4. Secondary antibodies used for immunohistochemistry.

2.1.8 Animals

For preparation of histological sections, C57BL/6 strains mice were used. Primary neurons, RNA and genomic DNA from brain, was prepared from in house bred CD1 strain mice.

2.1.9 Cell culture

Cell and tissue culture flasks (γ -irradiated and cell culture treated to facilitate binding of the cells to the plastic surface) were ordered from TPP Biochrom (Berlin, Germany).

DPBS s/o Ca and Mg (BioWhittaker/Biozym, Hess Oldendorf, Germany) was used to wash cells.

Media	Supplier	Other material	Supplier
RPMI1640	Invitrogen	FBS	Biochrom
DMEM	BioWhittaker	P/S (penicillin/streptomycin)	Invitrogen
DMEM:F12	BioWhittaker	L-glutamine	BioWhittaker
		Fungizone	Invitrogen
		NON-essential amino acids	Invitrogen
		Trysin	Invitrogen
		Poly-L-lysine 0.1% in H ₂ O	Sigma
		Collagen S Type I	Roche
		DPBS s/o Ca and Mg	BioWhittaker
		B27 with antioxidants	Gibco/Invitrogen

Table 5. Media and other cell culture components.

2.1.10 Cell lines and primary cells

Cell lines were ordered from DSMZ (Braunschweig, Germany) and cultured according to the manufacturer's instructions. Primary mouse cell lines were established from CD1 animals, the human lymphoblastoid cell line was established (in house) from a male control patient.

2.1.11 Bioinformatics databases and tools

Database name	Database type	Internet address
Pfam	Protein families	http://www.sanger.ac.uk/Pfam
Smart	Protein families	http://smart.embl-heidelberg.de
Prosite	Protein families and domains	http://www.expasy.org/prosite
BLink	Precalculated Blast searches	http://www.ncbi.nlm.nih.gov
SSDB	Sequence similarity database	http://ssdb.genome.jp
Celera	Sequence retrieval and analysis	http://www.celera.com
Ensembl	Sequence retrieval and analysis	http://www.ensembl.org
Unigene	Gene-oriented sequence clusters	http://www.ncbi.nlm.nih.gov/UniGene
CDD	Conserved domain database	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
Expasy	Protein sequences and analysis tools	http://www.expasy.org
PDB	Biological macromolecular 3-D structures	http://www.pdb.org
HUGE	KIAA gene database	http://www.kazusa.or.jp/huge
HUGE PPI	KIAA protein interaction database	http://www.kazusa.or.jp/huge/ppi
Pubmed	Medline biomedical articles	http://www.pubmed.org
Transfac	Transcription factor binding sites	http://www.biobase.de
RettBASE	Mutation database for MECP2	http://www.mecp2.org.uk

Table 6. Databases used in the different projects of this thesis.

Tool name	Internet address	Used in this thesis for
Ensembl Blast	http://www.ensembl.org/Multi/blastview	Genomic blasts / batch blasts
NCBI Blast	http://www.ncbi.nlm.nih.gov/BLAST	Masked protein blast
NCBI HomoloGene	http://www.ncbi.nlm.nih.gov/HomoloGene	Homologous genes
EBI Blast (TBLASTX)	http://www.ebi.ac.uk/Blast2/index.html	Translated nt query blasted against translated nt database
ClustalW	http://clustalw.genome.jp/	Sequence alignments
ClustalX Windows version	ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX	Graphical representation of alignments
Plogo	http://www.cbs.dtu.dk/~gorodkin/appl/plogo.html	Creation and display of logos
LALIGN	http://www.ch.embnet.org/software/LALIGN_form.html	Calculation of identity scores
MATCH	http://www.biobase.de	Find transcription factor binding sites
GCG package	http://www.accelrys.com/products/gcg_wisconsin_package	Sequence comparison and
Primer3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi	Primer design

Table 7. Bioinformatics tools used in this thesis.

2.2 Methods

2.2.1 Bioinformatics methods

2.2.1.1 Database screens

MBD protein family:

Using "MBD" or the MBD of human MECP2 (PDB ID: 1QK9) as query, the NCBI (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein>), Pfam (release 7), Prosite (release 17.17) and Smart (version 3.4) databases were screened for proteins with this domain. The resulting polypeptides were sorted according to their species of origin and additional identified domains within their sequence. SSDB at GenomeNet was searched with pfm:MBD as query motif.

MECP2 paralogues:

SSDB at GenomeNet was searched for paralogues with hsa:4204 (the accession number of MECP2 in GenomeNet) as query and a threshold of 200 to get only the most specific candidates. The resulting data was exported as a list and GenomeNet accession numbers were

converted into NCBI accession numbers to be comparable to the BLink output list.

For P51608 (the accession number of MECP2 in the Entrez protein database) the pre-calculated blast hits from the BLink database were sorted by taxonomy proximity and score cut-off of 150 was chosen. All entries for *homo sapiens* were exported as a list for further analysis with Perl scripts.

Mecp2 binding sites:

The genomic region of *FKBP5* and *Fkbp5* were compared in the Ensembl database using compara tool in the Contig View mode. Furthermore CpG islands, CpG content and Eponine regions were mapped to the *Fkbp5* locus.

The MATCH tool at the Transfac database to find all matches for Transfac matrices in the genomic region of *Fkbp5*. The results were compared to the findings in the Ensembl database and resulted in two regions that were considered as possible promoters and covered with primer pairs for PCR fragments of ~200 bp.

Glucocorticoid receptor response elements in the genomic region of FKBP5 were found by mapping the Transfac matrix for glucocorticoid binding sites in man, mouse and rat (Transfac accession number: M00921) to the genomic region of FKBP5. For the resulting 6 regions, primers were designed and binding of GR was confirmed by ChIP.

2.2.1.2 Blast searches

MBD protein family:

Non-redundant GenBank, high throughput genomic sequences and expressed sequence tag (EST) databases were searched using BLASTP (expect=10, word size=3, matix=BLOSUM62, gap costs=existence11, extension1) and TBLASTN (expect=10, word size=3, matix=BLOSUM62, gap costs=existence11, extension1) of the NCBI web tools.

The sequence of the human MECP2 MBD (chain A, NCBI accession number: 1QK9A) was used as a query input.

TBLASTX (matrix=default, DNA strand=both, exp.thr=default, filter=none, view filter=none, sensitivity=high) was carried out applying the web tools at the European Bioinformatics Institute (EBI).

BLASTP and BLASTN searches were carried out against the Celera database with settings corresponding to the ones used in the NCBI blast searches.

KIAA1887 was detected as an MBD protein due to a blast search with MBD of KIAA1461

(amino acids 11-81) as query for a BLASTP at NCBI using the parameters specified above.

MECP2 paralogues:

To find repetitive regions to be masked in the BLAST search, the sequence of MECP2 was compared to itself using the Compare program of the GCG Wisconsin package (Version 10.3-UNIX) (comparison window size=25, stringency=9) to calculate a matrix of points of similarity and the DotPlot program to visualize the output.

Of the MECP2 protein sequence (NCBI accession number: P51608) the region aa20 - aa90, aa160 - aa200 and aa340 - aa400 were masked in a BLASTP (standard settings as defined above) at NCBI by using the "mask lower case"-option.

2.2.1.3 Sequence comparison

MBD protein family:

Alignments and phylogenetic trees were computed using the ClustalW program at GenomeNet with standard settings and the ClustalX program (version 1.8) (Thompson *et al.*, 1997) for graphical representation. The sequence logo was constructed by means of the plogo script. File formats were converted with the GCG Wisconsin package when necessary. Sequence comparison between mouse and man was carried out using the LALIGN algorithm at EMBNET with default settings. Homologues of *KIAA1461*, *CLLL8* and *KIAA1887* were identified using the ortholog prediction in the Ensembl database.

Mecp2 target genes:

Sequence analysis was done using the reformat and pileup functions of the GCG Wisconsin package to get multiple sequence alignments that could be screened for the presence of cytosines not modified by the bisulfite treatment.

2.2.1.4 Perl scripts

Several small scripts were written using the script language Perl (version 5.8.0). File formats were converted to be suitable as input files for GCG. Accession numbers of GenomeNet derived from the SSDB searches for MECP2 paralogues were transformed into NCBI-GI accession numbers. With the help of a Perl script, sequences files formats of bisulfite treated and cloned DNA fragments were sorted and converted to be suitable as input files for the GCG PileUp program.

2.2.2 cDNA microarrays

2.2.2.1 Microarray production

Murine cDNA clones (n = 10,080; arrayTAG clone collection) were purchased from LION Bioscience (Heidelberg, Germany). These cDNA clones map to the 3' end of the respective transcript and are sequence-verified and sequence-specific (i.e. do not contain poly-A tails or large repeats). An additional 3510 murine cDNA clones, representing brain-expressed transcripts were obtained from the resource center of the German Human Genome Project (RZPD, Berlin, Germany) and 3 murine cDNA sequences were amplified with specific primers. As negative controls 17 plant cDNA sequences, amplified with specific primers, were included as duplicates resulting in 34 plant derived cDNAs on the chip. This resulted in a set of 13,627 clones that were used for the array and could be kept as plasmid inserts, stored as *E. coli* glycerol stocks.

For the array production, clone inserts were amplified by PCR (Tables 8, 9). Primers Lion 3' and Lion 5' were used for clones from LION bioscience, and primers M13for and M13rev were used for RZPD clones, as well as for the 3 murine and the plant clones. All PCR products were evaluated by agarose gel electrophoresis (see section 2.2.4.5). These steps were automated with a robot (Tecan Genesis Workstation 200, Crailsheim, Germany) in combination with a temperature controlled hotel (automatic incubator cytomat 6002, Hereaus, Hanau, Germany).

2 µl of the bacterial template clones stored in 96-well microtiter plates were mixed with 50 µl PCR master mix in a new 96-well microtiter plate.

Primer name	Primer sequence	Annealing temperature	Annealing time
Lion 3'	5'-tcgagcggccgcccgggcaggt-3'	68 °C	30 sec
Lion 5'	5'-agcgtggtcgcggccgaggt-3'	68 °C	30 sec
M13for	5'-gtaaacgacggccag-3'	54 °C	90 sec
M13rev	5'-caggaaacagctatgac-3'	54 °C	90 sec

Table 8. Primer pairs used for the amplification of plasmid inserts for the array production.

LION Bioscience clones			RZPD and plant clones			PCR master mix (total volume = 50 µl)
temperature	time	cycle numbers	temperature	time	cycle numbers	
94°C	3 min	1	94°C	5 min	1	5 µl 10 x Perkin Elmer PCR buffer
94°C	30 sec	35	94°C	45 sec	35	10 µl dNTPs (1 mM each)
68°C	30 sec		54°C	90 sec		2 U MPI Taq (made in house) (10 U/µl)
72°C	50 sec	1	72°C	2 min	1	2 µl forward primer (10 µM)
72°C	10 min		72°C	10 min		2 µl reverse primer (10 µM)
4°C	infinite	1	4°C	infinite	1	29 µl nuclease free water

Table 9. PCR conditions used for the amplification of cDNAs for the microarray.

Finished plates were stored in the temperature controlled hotel until use for PCR amplification in a PrimusHT multiblock thermal cycler (MWG Biotech) or alternatively were frozen at -20°C for later amplification.

The same robot could be used to mix 5 µl of PCR product from readily amplified plates with 5 µl of gel-loading buffer in a new 96-well microtiter plate. This mixture was analyzed on agarose gels using the Roboseq 4204S system (MWG Biotech) for gel loading.

45 µl (remaining after the agarose gel analysis) of each PCR product were combined with 2.5 volumes EtOH and 0.1 volumes sodium acetate 3 M pH=5. DNA was precipitated by incubation at -80°C for 30 min, and centrifugation at 20,000g for 15 min at 4°C. The pellet was resuspended in 18 µl 3 x SSC and 8 µl of each product were transferred to a well in a 384-well plate with a Multimek 96/384 robot (Beckman Coulter, Krefeld, German). The resulting 384-well plates were stored at -20°C until further use.

PCR products purified and processed in such a manner were printed on Corning GAPS II slides by using a robotic spotting device (SDDC-2 MicroArrayer, ESI, Toronto, Canada/ChipWriter Pro, Bio-Rad, Munich, Germany) with SMP3 pins from TeleChem International (Sunnyvale, CA).

2.2.2.2 Microarray hybridization and washing

Total RNA from the brain of a 74 day old *Mecp2^{-y}* animal with late symptoms (gait ataxia, hind limb claspings, breathing irregularities, uneven teeth, small stature) and a *wt* male litter mate was sent by our collaborators in Edinburgh. Labeled target cDNA was generated by direct incorporation of fluorescent nucleotide analogs in a reverse transcription reaction. Total

RNA (75 µg of each sample) was used in an oligo dT primed reaction in the presence of 100 µM Cy3- or Cy5-dUTP (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany), 200 µM dTTP and 500 µM dATP, dCTP, and dGTP. The labeled cDNA was purified using the Qiaquick PCR purification kit. Labeled cDNA targets were resuspended in hybridization solution (50% formamide, 6 x SSC, 0.5% SDS, 5 x Denhardt's solution), to which 1 µl poly-dA (10 µg/µl) and 1 µl mouse Cot-1 DNA (20 µg/µl) were added. The resulting reaction solution was concentrated to 35 µl, denatured at 95°C for 3 min and snap-cooled on ice. Hybridization took place under a coverslip at 42°C for 16 hours. The slides were washed at room temperature in 0.2 x SSC and 0.01% SDS for 5 min, followed by two washing steps in 0.2 x SSC for 5 min each. In total, two co-hybridizations (*Mecp2^{-y}* versus *wt* brain cDNAs) were performed with dyes swapped in the repeat experiment.

2.2.2.3 Image acquisition and data analysis

Fluorescence intensities of Cy3 and Cy5 were measured separately at 532 and 635 nm with a laser scanner (Affymetrix 428 Array Scanner, Affymetrix). The resulting 16-bit data files were imported into Microarray Suite image analysis software (version 2.0), which runs as an extension of IPLab Spectrum software (Scanalytics, Fairfax, VA). Raw spot intensities of Cy3 and Cy5 were locally background subtracted by the MicroArray Suite software. Empty spots and spots carrying plant gene sequences were excluded from further analysis. Each dye swap experiment was normalized by applying variance stabilization (Huber *et al.*, Bioinformatics 2002) using the vsn package of bioconductor (<http://www.bioconductor.org>). Means of normalized log-products and log-ratios of each dye swap experiment pair were used for further analysis. Normalization procedures were performed using R (<http://cran.R-project.org>). The data sets of two co-hybridization experiments with dye swap were first quality checked for pin- and PCR plate-specific effects by local regression according to each pin or PCR plate. Furthermore, background intensities across the slide were checked for homogeneity. Spots with local contaminations were excluded from further analysis. Local background-subtracted raw intensities on a logarithmic scale were normalized by Analysis of variance between groups (ANOVA) using the MATLAB code, with slight modifications. The underlying ANOVA model we applied was:

$$\log(y_{ijk}) = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (V G)_{kg} + \varepsilon_{ijk}$$

$$i=1,2; j=1,2; k=1,2; g=1, \dots, 13627;$$

(where y is the log intensity of each channel, μ is the overall mean, A is the overall array effect, D is the overall dye effect, V is the overall variety effect, G is the overall gene effect, AG describes array gene interactions and VG describes variety-gene interactions). Genes were then sorted according to their variety-gene differences ($(V G)_{k=\text{sample}, g} - (V G)_{k=\text{control}, g}$), which is an estimate of the log ratio of the sample versus control normalized intensities. Clones showing more than 2-fold changes were considered relevant. All data analyses were performed using MATLAB, Version 6.0.0.88, Release 12, MathWorks Inc., MA, USA.

2.2.3 Cell culture

2.2.3.1 Coating of culture flasks

Poly-L-lysine 0.1% in H₂O was diluted 1:10 with sterile H₂O. This solution was applied to cell culture flasks and incubated at room temperature for 5min. After three washes with sterile H₂O, flasks were left to dry under sterile conditions. Poly-L-lysine coated flasks were incubated with collagen S Type I solution diluted 1:100 in Neurobasal medium for 2 h at 37°C followed by 3 washes with DBPS, just before use.

2.2.3.2 Primary fibroblasts

Primary fibroblasts were prepared from adult CD1 mice (in house breeding facility): Dermis from dorsal skin was dissected, washed in sterile DPBS s/o Ca and Mg several times, placed in trypsin and chopped into pieces of 1-2 mm diameter. The pieces were placed in a 25 cm² cell culture flask and left to attach for three to four hours. Once the cells had attached to the flask, 5 ml DMEM:F12 medium 1:1 (BioWittacker) supplemented with 20% FBS (Biochrom, Berlin, Germany), P/S (100 U/ml, 10 µg/ml respectively) (Invitrogen), L-glutamine (2 mM) (BioWhittaker) and Fungizone (2.5 µg/ml) (Invitrogen) were added carefully. Cells were grown for 1 week to a confluency of 95% before passaging. For passaging, medium was removed, cells were washed with sterile DPBS s/o Ca and Mg twice and separated by incubation in 2 ml trypsin (Invitrogen) for 5 min at 37°C. The trypsin solution containing the cells was mixed with 13 ml fresh medium and plated in a 75 cm² flask. This step was repeated after 5 days to get a 150 cm² cell culture flask with primary fibroblasts. Once the cells had reached 95% confluence, they were harvested for DNA isolation.

2.2.3.3 Lymphoblastoid cell line

Eppstein-BarrVirus (EBV) transformed primary lymphoblastoid cells were grown as suspension cultures in RPMI1640 medium supplemented with 10% FBS, P/S (100 U/ml, 10 µg/ml), and L-glutamine (2 mM) and passaged 1:2 every 3 days..

2.2.3.4 Primary neurons

Heads of mouse embryos from a litter between embryonic days 16 to 18 were carefully prepared and placed in DPBS s/o Ca and Mg + 6 g/l glucose (Sigma). After cutting off the snout, brains were squeezed out and placed on ice in DPBS s/o Ca and Mg with 10% FBS and 6 g/l glucose until further use. Single cells were isolated mechanically by pipetting, fragmented with a cell strainer (40 µm; BD Falcon, Heidelberg, Germany), and centrifuged at 700 rpm for 1 min with slow acceleration and brake. The cell pellet was resuspended in 2ml medium (see below). The cell concentration was determined with a CASY1 Cell Counter (Schärfe System, Reutlingen, Germany). 1.5×10^7 cells were plated per 150 cm² cell culture flask coated with collagen S Type I and poly-L-lysine. The medium was changed every 3-4 days. After 7-10 days, cells were harvested for ChIP or treated with glucocorticoids for ChIP. For 500 ml medium, 485 ml Neurobasal Medium (Invitrogen), 10 ml B27 with antioxidants (Gibco; mit Antioxidantien), 5 ml P/S (100 U/ml, 10 µg/ml respectively), (Invitrogen) and 1.25 ml L-glutamin (2 mM) were used.

2.2.3.5 Glucocorticoid treatment of primary neurons

Stock solutions of dexamethasone (Sigma-Aldrich) 10 mM in EtOH and RU-486 (Mifepristone, Sigma-Aldrich) 100 mM in EtOH were prepared. Twenty hours before harvest, respective amounts of the stock solutions (Table 10) were added to the culture medium of primary neuron cultures. EtOH was added to negative control cultures to consider potential effects of EtOH on the primary neurons.

Culture	Dexamethasone (10mM in EtOH)	RU-486 (100mM in EtOH)	EtOH 100%
EtOH	-	-	2.5 µl
Ru	-	1.25 µl	1.25 µl
RU, Dex	1.25 µl	1.25 µl	-
Dex	1.25 µl	-	1.25 µl

Table 10. Scheme depicting the amounts and concentrations of chemicals used for glucocorticoid treatment experiments.

2.2.3.6 CCRF-CEM

CCRF-CEM cells were grown in RPMI1640 medium supplemented with 10% FBS, P/S (100 U/ml, 10 µg/ml), L-glutamine (2 mM) and Fungizone (2.5 µg/ml), according to manufacturer's instructions. Saturated cultures were split 1:2 every 2-4 days by detaching cells in with trypsin and reseeding in new cell culture flasks with fresh medium as described for primary fibroblasts. CCRF-CEM cells were harvested at 95% confluence for RNA isolation.

2.2.3.7 NEURO-2A

NEURO-2A cells were grown according to manufacturer's instructions in DMEM medium with 10% FBS, P/S (100 U/ml, 10 µg/ml), L-glutamine (2 mM), 1% NON-essential amino acids (NEAA) and Fungizone (2.5 µg/ml). Cells were grown to 85% confluence and split 1:4 once a week as described for primary fibroblasts. Cells were harvested at 85% for DNA isolation and ChIP.

2.2.4 Molecular biology techniques

2.2.4.1 DNA isolation

General:

Genomic DNA from CCRF-CEM cells and primary fibroblasts was isolated with the Qiagen Blood Maxi-KIT following the manufacturer's instructions.

Genomic DNA from CD1 mouse brains was isolated DNeasy tissue kit following the manufacturer's instructions.

Mecp2 binding sites:

After ChIP, DNA was isolated using the phenol/chloroform extraction followed by

precipitation of the DNA. DNA containing solutions were mixed 1 volume of phenol:chloroform 1:1, carefully inverted for 5 min, and centrifuged for 10 min at 4000rpm in a table centrifuge. The upper aqueous phase was transferred to a new tube, mixed with 1 volume of chloroform, carefully mixed for 5 min and centrifuged for 10 min at 4000rpm. The upper phase was then transferred to a new tube and combined with 2.5 volumes of ice cold EtOH, 1/10 volume sodium acetate and 1 μ l glycogen. After inversion of the tube, the solution was cooled down for 30 min at -80°C (alternatively overnight at -20°C) and then centrifuged for 45 min at 4°C. The resulting DNA pellet was washed with 70% EtOH (in water) briefly and centrifuged for 45 min at 4°C. The supernatant was discarded and the pellet dried in an concentrator 5301 speedvac (Eppendorf) at 37°C after puncturing the lid of the tube with a needle (to prevent contaminations as far as possible). The dried pellet was redissolved in 50 μ l ultra pure H₂O and subjected to SR1 and SR2 amplifications (see section 2.2.4.8.)

2.2.4.2 RNA isolation

Total RNA from CCRF-CEM cells was isolated using the Trizol reagent (Invitrogen). All plastic- and glassware used was rinsed with water containing 0.05% DEPC and autoclaved before use to inactivate RNAses. The cell pellet was dissolved carefully in 10 ml Trizol, transferred to and 30 ml tube and incubated at RT for 5 min 2 ml chloroform were added and the solution was mixed vigorously for 15 sec, incubated at RT for 2-3 min followed by a centrifugation step at 4°C and 5000rpm for 20 min. The supernatant was transferred to a new tube and well mixed with 5 ml isopropanol. After incubation at RT for 5 - 10 min the solution was centrifuged at 4°C and 10000rpm for 10 min. The supernatant was discarded and the pellet washed with 70% EtOH (prepared with RNase-free water) followed by a centrifugation at 4°C at 10000rpm for 5 min. The resulting pellet was dried and dissolved in 500 μ l RNase-free water. RNA concentration was measured using a Ultraspec 3100pro UV/VIS photometer (Amersham) and RNA quality was checked on an 1% agarose gel.

2.2.4.3 Standard PCR

Standard PCR was performed in an GeneAmp PCR System 9700 cycler (Applied-Biosciences, Darmstadt, Germany) with Perkin-Elmer Taq polymerase (Applied-Biosystems). If not indicated otherwise, the following PCR conditions were used:

Step	Temperature	Time	Cycles	PCR mix
Initial denaturation	94°C	5 min	1	2 µl template
Denaturation	94°C	45 sec	35	5 µl 10x reaction buffer
Annealing	Primer-specific	1 min		1 µl dNTPs (10 mM each)
Elongation	72°C	1 min		1 µl forward primer (1 µM)
Final elongation	72°C	7 min	1	1 µl reverse primer (1 µM)
Storage	4°C	infinite	1	0.5 µl PE-Taq polymerase 39.5 µl H ₂ O

Table 11. Standard PCR temperature profile and PCR mix.

Touchdown PCR was applied for products that proved difficult to amplify:

Step	Temperature	Time	Cycles	PCR mix
Initial denaturation	94°C	5 min	1	2 µl template
Denaturation	94°C	45 sec	10	5 µl 10x reaction buffer
Annealing	(T _A + 10°C) - 1°C/cycle	1 min		1 µl dNTPs (10 mM each)
Elongation	72°C	1 min		1 µl forward primer (1 µM) 1 µl reverse primer (1 µM)
Denaturation	94°C	45 sec	30	0.5 µl PE-Taq polymerase
Annealing	Primer-specific	1 min		39.5 µl H ₂ O
Elongation	72°C	1 min		
Final elongation	72°C	7 min	1	
Storage	4°C	infinite	1	

Table 12. Touchdown PCR temperature profile and PCR mix. T_A is the primer specific annealing temperature.

2.2.4.4 cDNA synthesis and subsequent gene-specific PCR (RT-PCR)

RNA (20 µg) isolated from total mouse brain, CCRF-CEM cells or lymphoblastoid cells using the Trizol reagent (Invitrogen) (see section 2.4.2) was subjected to a reverse transcriptase reaction in the presence of 25 ng/ml oligo(dT) and 2.5 mM dA/C/G/TTP with 10 U/ml SSII reverse transcriptase and without enzyme in negative control reactions. The RNA in 17 µl RNase-free water was mixed with the oligo-dTs and the dNTPs to a final volume of 23 µl and incubated at 65°C for 5 min followed by 5 min on ice. After addition of 8 µl 5 x first-strand Buffer, 4 µl 0.1 M DDT, 1 µl RNAsin, and 4 µl SSII reverse transcriptase (200 U/µl), the mixture was incubated for 55 min at 42°C and for 10 min at 70°C.

2 µl of the mixture were used in gene-specific PCR reactions or alternatively, the resulting cDNA was purified with the QIAquick PCR purification kit and 200ng of the purified cDNA were subject to gene-specific PCRs.

Primer name	Forward primer	Reverse primer	Annealing temperature	Annealing time
mFkbp5_exon1_exp_for	5'-gcggcgacaggtcttcta-3'	5'-gcttgataacctggccttg-3'	56°C	1 min
KIAA1887_exon_6	5'-cagaccccctactgtatttc-3'	5'-caaaaggtaaagcttccat-3'	49°C	1 min
KIAA1461	5'-ctagaccatgggaaaatgt-3'	5'-acttggagactgctcctcta-3'	51°C	1 min
β-actin	5'-tgaaccctaaggccaacctgt-3'	5'-gctcatagctcttccagg-3'	57°C	45 sec

Table 13. Primers used for the amplification of gene specific PCR products from cDNA templates.

Standard PCR conditions were applied except for β-actin where the annealing time was only 45 sec.

2.2.4.5 Agarose gel electrophoresis

1% agarose gels were used to separate DNA or RNA samples by gel electrophoresis.

Gel composition: 1% w/v ultra pure agarose in TBE buffer (0.1 M Tris, 0.1 M boric acid, 2 mM EDTA). 0.5 µl/ml EtBr was added to the gels to make visualization of the DNA via UV-light possible. Before loading, samples were supplemented with at least 0.5 volumes of gel loading buffer composed of 0.25 % bromphenol blue, 0.25 % xylene cyanol FF, and 30 % glycerol. 2 different DNA size markers were used depending on the size of the fragment sizes expected: pUC mix 8 (MBI Fermentas, St. Leon-Rot, Germany) with a range from 110 bp - 1116 bp and Lambda DNA / EcoR1+HindIII 3 (MBI Fermentas) with a range from 564 bp to 21226 bp. Voltage (80 -150 V) and running times (15 - 45 min) depended on the expected fragment sizes. For visualization, the E.A.S.Y Win32 gel documentation system (Herolab, Wiesloch, Germany) was used.

2.2.4.6 Sequencing

For sequencing of PCR products, the fragments were purified by gel extraction using the QIAquick gel extraction kit. Using the TOPO TA cloning kit, PCR products were then ligated into a pCR2.1-TOPO vector. For transformation, 'One shot competent *E. coli* cells' were used according to the manufacturer's instructions. *E. coli* cells were grown overnight at 37°C on selective plates (containing 50 µg/ml ampicillin) and supplemented with 50 µl 4% X-Gal and 50 µl IPTG (0.1M). White clones were picked and grown in 2 ml LB medium (Tryptone 10

g/l, NaCl 10 g/l, yeast extract 5 g/l) with 50 µg/ml ampicillin overnight. Plasmid DNA for sequencing was isolated using the Qiagen plasmid mini kit according to the manufacturer's instructions. Per 100bp to be sequenced 10 ng DNA were required and subject to sequencing PCR with M13rev primers:

Step	Temperature	Time	Cycles	PCR mix
Initial denaturation	94°C	2 min	1	2 µl Dye Terminator (made in house)
Denaturation	94°C	20 sec	30	10 pmol primer
Annealing	54°C	15 sec		7 µl DNA solution
Elongation	60°C	4 min		

Table 14. Temperature profile and PCR mix for sequencing reactions with M13 primers. After the final elongation step, the PCR plates are stored on ice.

The PCR-reaction product was precipitated by addition of 25 µl EtOH. The solutions were mixed gently and the incubated in the dark for 10 min at RT. After a centrifugation step for 45 min at 4000 rpm at RT the obtained pellet was washed with 100 µl 70% EtOH, mixed gently and centrifuged for 30 min at 4000 rpm at RT. The supernatant was discarded, the pellet dried and the obtained fragments separated and analyzed on an ABI377 DNA sequencer.

2.2.4.7 Northern blotting

A cDNA fragment of KIAA1461 was PCR-amplified with exon specific primers KIAA1461 (see section 2.2.4.4) and human genomic DNA from mouse primary fibroblasts as template using standard conditions. For KIAA1887 a cDNA fragment was PCR-amplified with exon 6 specific primers KIAA1887_exon_6 (see section 2.2.4.4) and cDNA from a lymphoblastoid cell line as template using standard conditions. A cDNA fragment of β -actin amplified with primers β -actin (see section 2.2.4.4) was used as a loading control. These probes were radioactively labeled with ^{32}P -dCTP in a random prime reaction and hybridized in ExpressHyb solution (CLONTECH, Palo Alto, CA, USA) to a human multiple tissue Northern blot (CLONTECH, Palo Alto, CA, USA) for 16 h at 65°C. Washing was performed in $2 \times \text{SSC} / 0.1\% \text{SDS}$ at 65°C for 10 min. Signals were detected with a PhosphorImager (Amersham Biosciences, Freiburg, Germany).

2.2.4.8 Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed with NEURO-2A cells to establish the procedure and test primers pairs. Final experiments were carried out with brain tissue and primary neurons.

NEURO-2A cells and primary neurons (generally one 150cm² flask was used per condition or treatment) were crosslinked by addition of 1% formaldehyde (Roth) and incubation at 37°C for 10 min to the culture medium. Subsequent quenching was achieved by adding glycine (Merck) to a final concentration of 125 mM.

Frozen total brain tissue was pulverized with a mortar, taken up in 10 ml DMEM (BioWhittaker) with protease inhibitors (1 tablet Complete Mini Cocktail (Roche) per 10 ml solution as well as 2 mM PMSF) and incubated for 10 min at 37°C in the presence of 1% formaldehyde. Formaldehyde treatment was quenched by addition of glycine to a final concentration of 125 mM.

The Chromatin immunoprecipitation protocol was based on the Upstate Biotechnology ChIP kit protocol with the following modifications:

Sonication was carried out with a Branson Sonifier 450 (Branson, Danbury, CT, USA) was used at output control level 5 with 100% duty and pulses of 30 sec. The number of pulses depended on the cell type and the amount of chromatin in the sample. After sonication, DNA fragment sizes 5 µl of the sheared material was loaded on a 1% agarose gel to check the fragment sizes. If the smear on the gel was > 200 - 1000 bp, additional shearing was necessary.

DNA concentration in the sheared chromatin was determined with an Ultrospec 3100pro UV/VIS photometer (Amersham) and samples were diluted to 100 µg/ml DNA and split into portions of 1 ml in 1.5 ml eppendorf tubes. The tubes were either subjected to immunoprecipitation or stored at -20°C until further use.

The chromatin was then subjected to immunoprecipitation using antibodies specific to acetylated histone H3, mSin3A, the MeCP2 C-terminus or the DNA binding domain of GR according to the Upstate ChIP kit protocol. To reduce unspecific binding in the immunoprecipitation, the pre-clearing was performed with 75 µl (instead of 60 µl) Protein A agarose / Salmon sperm DNA and all washing steps described in the Upstate protocol were carried out twice. After reverse crosslinking, protein digestion, and DNA isolation, fragments were dissolved in 50 µl aqua ad iniectabilia and subjected to a two-step DOP-PCR. In the first

step 20 μ l template were ligated to degenerated SR1 primers in a PCR reaction. In the SR2 PCR the fragments ligated to the SR1 primer were amplified using the SR2 primer.

	SR1	SR2
Primer	5'-gccgtcgacgaattcnnnnnnnn-3' (where n can be any base a, g, t or c)	5'-gccgtcgacgaattc-3'
Template	20 μ l purified ChIP product	2 μ l SR1 -PCR product
PCR mix	20 μ l template 1 μ l dNTPs (10 mM each) 5 μ l 10x reaction buffer (Perkin-Elmer) 2 μ l SR1-primer (10pMole/ul) 2 μ l PE-Taq (Perkin-Elmer) 20 μ l ultra pure H ₂ O	2 μ l template 5 μ l 10 x reaction buffer (Perkin-Elmer) 1 μ l dNTPs (10 mM each) 2 μ l primer SR2 2 μ l PE-Taq (Perkin-Elmer) 38 μ l ultra pure H ₂ O

Table 15. PCR mix for SR1 and SR2 reaction.

	SR1			SR2			
	SR1 temperature profile			SR2 temperature profile			
Step	Temperature	Time	Cycles	Temperature	Time	Cycles	
1	94°C	5 min	1	94°C	5 min	1	
2	94°C	30 sec	5	94°C	30 sec	60	
3	15°C	60 min		50°C	45 sec		
4	20°C	45 min		72°C	1 min		
5	25°C	30 min		72°C	4 min		1
6	30°C	30 min		4°C	infinite		1
7	35°C	20 min					
8	40°C	1 min					
9	45°C	30 sec					
10	50°C	30 sec					
11	55°C	20 sec					
12	60°C	20 sec					
13	60°C	7 min		1			
14	4 °C	forever		1			

Table 16. Temperature profile for SR1 and SR2 PCR reaction.

2 μ l of the DNA samples amplified by SR1 and SR2 reactions were subject to standard PCR amplifications with primers specific for binding sites of MECP2 and GR with the following annealing temperatures and cycle numbers.

Primer name	Forward primer	Reverse primer	Template	Annealing temperature
m1_0	5'-tgctcccttagattcatcccacac-3'	5'-ccactggctccgatacacattctc-3'	2 µl	55°C
m1_4	5'-agtaccaacagaggtcaga-3'	5'-tgtggatacaaaatattcca-3'	2 µl	49°C
m2_1	5'-tcttggccttaccttaat-3'	5'-agttctcagggactttcag-3'	2 µl	49°C
m2_2	5'-acacagaaacaataacaaaagc-3'	5'-tcaacaatatggctgtagga-3'	2 µl	49°C
m2_4	5'-gcacaatgctggactagata-3'	5'-ttaaagtgtcctcagtgcttc-3'	2 µl	49°C
Fkp	5'-agccacggctcctagatgagagc-3'	5'-gtgtgtgaaggagagtgccagaac-3'	2 µl	60°C
Fkp1	5'-tgctcccttagattcatcccacac-3'	5'-ccactggctccgatacacattctc-3'	2 µl	60°C
GRE_2	5'-tccaaagtcctatgtacc-3'	5'-tggccacaacatacacg-3'	2 µl	53°C

Table 17. Primer used to amplify DNA fragments precipitated by ChIP.

Three buffers, not included the Upstate ChIP kit, were prepared in the lab:

	Nuclei extraction buffer	Nuclei wash buffer	Re-suspension buffer
HEPES	10 mM	10 mM	
Sucrose	320 mM	320 mM	
MgCl ₂	5 mM	5 mM	
Tris pH= 7.5			15 mM
EDTA			1 mM
NaCl			150 mM
Triton X-100	1%		0.5%
Millipore H ₂ O	in millipore H ₂ O pH=7.4 adjusted with NaOH	in millipore H ₂ O pH=7.4 adjusted with NaOH	in millipore H ₂ O

Table 18. Composition of the ChIP buffers not included in the Upstate ChIP Kit.

2.2.4.9 Bisulfite sequencing

Genomic DNA from brain was isolated using the DNeasy tissue kit (see section 2.2.4.1). Conversion of genomic DNA by sodium bisulfite was performed as described in the CpGenome DNA Modification Kit manual (Chemicon, Temecula, USA). Regions of interest were amplified from the bisulfite treated genomic DNA with specific primers. Amplification was performed using standard PCR conditions or touchdown PCR (TD) and fragments were cloned and sequenced.

Primer name	Forward primer	Reverse primer	Template
FKB5musP1_4bis	5'-aataccaacaaaaatcaaaaaaa-3'	5'-tgtggttataaaatattttataggtaat-3'	3 µl
FKB5_2_1meth_new	5'-cacacttctataaccttacctttaat-3'	5'-gttttagggatttttaggtgtgtt-3'	3 µl
FKB5_2_2meth_new	5'-aaattaaacacaaaaacaataacaaaaac-3'	5'-ttgtttttttaataataggtgtagga-3'	3 µl
mFKB5_2_4_me4for/ LH1rev	5'-gtataatgttgtaggataattaggaattag-3'	5'-ttaataactcaatactgtatcctattata-3'	3 µl
mFKB5_2_4_me5for/ LH2rev	5'-atttagtttagtataatgttgtaggata-3'	5'-cctaataccttaataactcaatacttc-3'	3 µl

Table 19. Primers used for bisulfite sequencing of the regions 1_4, 2_1, 2_2, and 2_4. The touchdown PCR protocol was applied with an annealing temperature of 53°C and an annealing time of 1 min and 15 sec.

3. Results

3.1 MBD protein family

3.1.1 Objective

Bioinformatics approaches to find gene products with certain properties have become fashionable since the publication of the complete human genome in 2001 (Lander *et al.*, 2001, Venter *et al.*, 2001) and its free availability to researchers. This is particularly true for families of proteins. A protein family is a group of proteins that share at least a part of their amino acid sequence – usually a domain. Protein domains are defined as independent and often globular folding units within a three-dimensional protein structure. Many databases have been created that group proteins according to the domains they contain (see Table 20).

Database	Internet address
Pfam	http://www.sanger.ac.uk/Pfam
Smart	http://smart.embl-heidelberg.de
Prosite	http://www.expasy.org/prosite
Systems	http://systems.molgen.mpg.de
ProDom	http://prodes.toulouse.inra.fr/prodom/current/html/home.php
InterPro	http://www.ebi.ac.uk/interpro
Blocks	http://blocks.fhrc.org
PROTOMAP	http://protomap.cornell.edu
SBASE	http://hydra.icgeb.trieste.it/sbase
TIGRFAM	http://www.tigr.org/TIGRFAMs

Table 20. Protein domain and protein family databases. Of these databases, Pfam, Prosite, and Smart proved to be very useful for this study.

To this end, the aa sequence of known domains are defined as motifs using sophisticated mathematical algorithms such as Hidden Markov Models (HMMs). Proteins are grouped into a protein family if they comprise a certain motif, that is to say, contain an aa sequence similar (each database has its own criteria for these similarities) to a motif in the database. This implies, that one protein can belong to several protein families if the corresponding motifs are present in its sequence. Most of these databases use an alignment of polypeptide sequences known to represent a certain motif in order to create a standard profile that is then applied to

screen other databases for the occurrence of a similar sequence.

In this project, the MBD motif from the Pfam database was used to look for additional proteins that could potentially bind methylated CpGs.

3.1.2 Human MBD proteins

The MBD of human MECP2 served as query sequence to search for new members of the MBD protein family. Initial standard BLAST searches of the NCBI, Celera, and SwissProt databases resulted only in five MBD proteins (MECP2, MBD1, MBD2, MBD3, and MBD4) which had previously been described and studied intensively.

However, the search of protein domain family databases (CDD, Pfam, Prosite and Smart) revealed six additional proteins, i.e. BAZ2A/TIP5, BAZ2B, CLLD8/SETDB2, SETDB1, KIAA1461, and KIAA1887 with similarities to the MBD of MECP2. Due to the results of this study, KIAA1461 has been named MBD5 and KIAA1887 is now called MBD6. The Pfam, Smart, and Prosite database use HMMs to detect motifs in amino acid sequences. This method is more sensitive than standard sequence similarity searches such as BLAST and FASTA, which explains why our BLAST searches did not detect these additional proteins. An MBD has been described in CLLD8, SETDB1, and BAZ2A/TIP5 earlier (Mabuchi *et al.*, 2001, Schultz *et al.*, 2002, Strohner *et al.*, 2001) without including these proteins into the MBD protein family.

Nine of the eleven MBD-containing protein sequences could also be detected by screening the Sequence Similarity DataBase (SSDB) (Kanehisa *et al.*, 2002) at GenomeNet. The cDNAs for KIAA1461/MBD5 and KIAA1887/MBD6 were not found since the KEGG database underlying the SSDB contains only confirmed but not predicted protein sequences. Table 21 summarizes all proteins identified, their domains, the position of the MBDs and the search method by which the protein was found.

The MBD amino acid sequences of the five previously published proteins as well as the MBD sequence of the six newly described human MBD protein family members were aligned to analyze the conservation of amino acids across the different proteins (Fig. 9).

Accession No. ^{a)}	Name	Search method /database ^{b)}	Domains ^{c)}	MBD position ^{d)}
P51608	MECP2	a, b, c, d, e, f, g, ac, bc, cc		96-149
Q9UIS9	MBD1	a, b, c, d, e, f, g, ac, bc, cc	zf-CXXC	7-59
Q9UBB5	MBD2	a, b, c, d, e, f, g, ac, bc, cc		151-204
O95983	MBD3	a, b, c, d, e, f, g, ac, bc, cc		8-60
Q9Z2D7	MBD4	a, b, c, d, e, f, g, ac, bc, cc	HhH-GPD	82-135
Q9UIF9	BAZ2A/TIP5	e, f, g	AT-hook, DDT, PHD, bromodomain	526-577
Q9UIF8	BAZ2B	e, f, g	DDT, PHP, bromodomain	549-600
Q15047	SETDB1	e, f, g	SET	597-653
Q96T68	SETDB2	e, f, g	SET	162-216
Q9P267	MBD5	e, f	PWWP	21-79
Q96Q00	MBD6	e, f		304-456

Table 21 Summary of the human MBD polypeptides. ^{a)} Accession numbers according to the SwissProt database. ^{b)} a = BLASTP, b = TBLASTN, c = TBLASTX, d = PSI-BLAST, e = Pfam, f = Ncbi, g = SSDB, ac = BLASTP in Celera database, bc = TBLASTN in Celera database, cc = TBLASTX in Celera database. ^{c)} domain nomenclature according to the Pfam database, all sequences contain an MBD in addition. ^{d)} position in amino acid sequence.

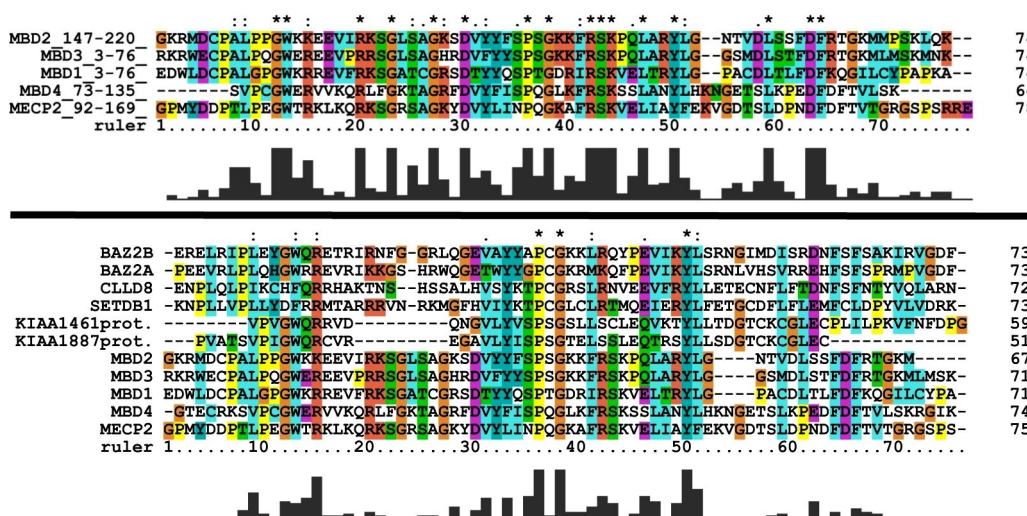


Fig. 9. Alignment of the human methyl-CpG-binding proteins according to Swissprot. ClustalX alignment of the original human MBDs (above) and all MBDs (below). Columns are colored by conservation and property (Thompson *et al.*, 1997). Residue conservation above each column indicates: “*” completely conserved; “:” favored substitutions; “.” weakly favored substitutions. A quality graph is depicted below each alignment.

The analyses of all 11 polypeptides implicate a small number of highly conserved and apparently essential amino acids within the MBD domain, especially the proline at position 36 (P36) as well as glycine 38 (G38) and tyrosine 50 (Y50). Five positions with conservative substitutions can be found (indicated by ":").

Fig. 10 depicts a sequence logo derived from the alignment of all eleven human MBD sequences.

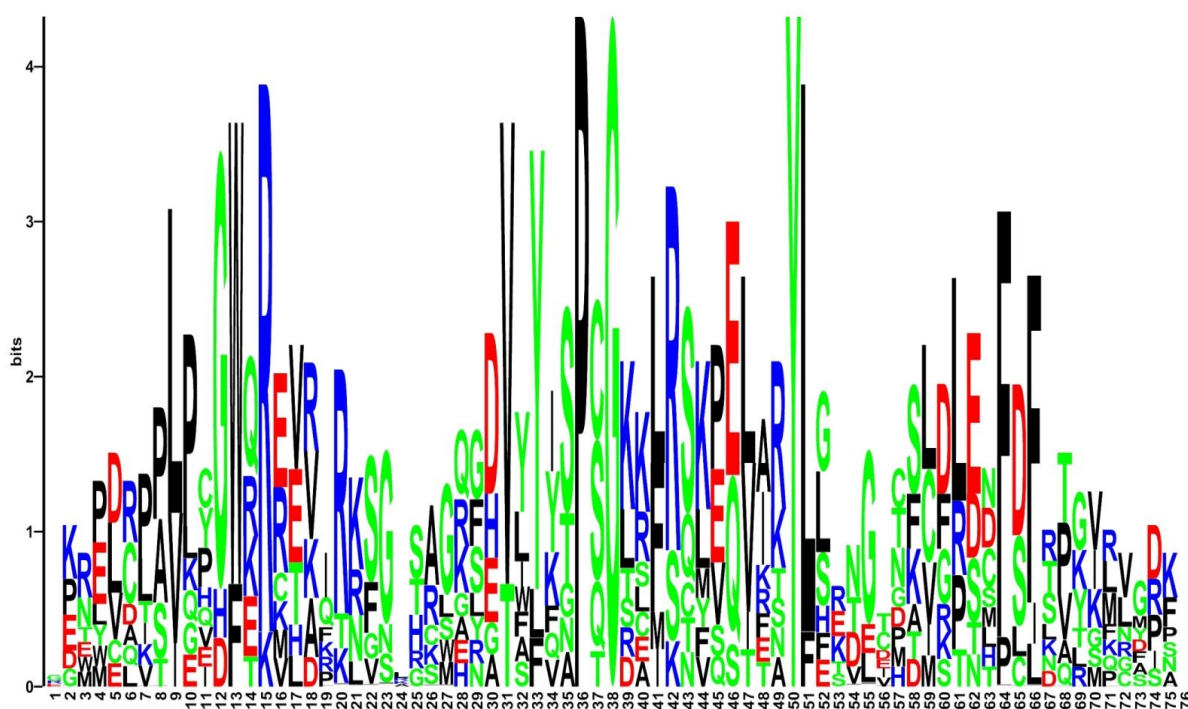


Fig. 10. Sequence logo of the eleven human MBD sequences. The height of the letters relative to the other aa, corresponds to the frequency of the amino acid at its position. The total height of each stack stands for the information present at this position, measured in bits. Top letters represent the consensus sequence. Grey bars indicate gaps in some of the aligned sequences.

A phylogenetic tree of the MBD amino acid sequences of all eleven polypeptides was computed using the ClustalW software at GenomeNet and is shown in Fig. 11.

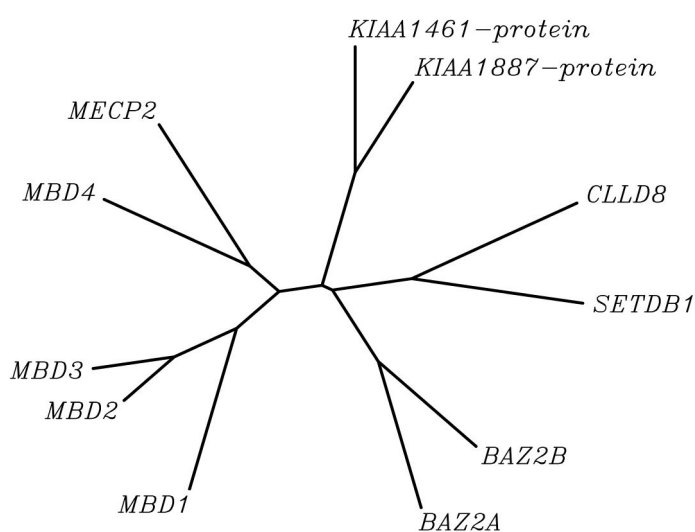


Fig. 11. Unrooted dendrogram depicting methyl-CpG-binding domains of the human protein family. A tree representation shows the similarity between the human MBD proteins. The left branch clusters the original 5 MBD proteins. Branch lengths are proportional to the amount of inferred evolutionary change.

Four major MBD subsets are indicated in Fig. 11. The MBDs of the originally described proteins (MBD1, MBD2, MBD3, MBD4, and MECP2) are found as one group besides a second (BAZ2A/TIP5, BAZ2B) and a third subset (CLLD8/SETDB2 and SETDB1) which are joined by a very short branch. KIAA1461/MBD5 and KIAA1887/MBD6 appear in a fourth branch. MBDs of the original five proteins are more similar to each other than to the novel ones, which explains why BLAST analyses with the MBD of MECP2 as query failed to identify the second, third, or fourth class.

3.1.3 Domain analysis

An analysis of the amino acid sequences revealed that the MBD was the only domain shared by all eleven sequences.

The MBD of MECP2, MBD1, MBD2, MBD4, and BAZ2A/TIP5 mediates binding to DNA, in case of MECP2, MBD1, and MBD2 preferentially to methylated CpG (Ng *et al.*, 2000, Strohner *et al.*, 2001, Lewis *et al.*, 1992, Hendrich and Bird, 1998, Bird, 2002). MBD4 has a special role acting as a DNA repair enzyme as described in the introduction.

In case of human MBD3 and SETDB1, the MBD has been shown to mediate protein-protein interactions (Schultz *et al.*, 2002, Saito and Ishikawa, 2002). *Xenopus* MBD3 is exceptional in its binding to methylated CpGs which can be explained by the difference of an amino acid residue within the MBD (Lys30) important for DNA binding (Saito and Ishikawa, 2002). It remains to be determined whether the MBDs of BAZ2B, CLLD8/SETDB2, KIAA1461/MBD5, and KIAA1887/MBD6 mediate DNA binding or protein-protein

interactions.

Additional domains found in seven of the eleven polypeptides indicate that they are associated with chromatin and function in epigenetic mechanisms of gene regulation. Some of the proteins are already known to be involved in transcriptional repression and the domains of the remainder strongly suggest a comparable function.

MECP2 recruits the Sin3A co-repressor complex and MBD2 the Mi-2/NuRD co-repressor complex, which itself contains MBD3. Both complexes contain HDACs, and MBD1 is also associated with HDAC activity although the identity of the deacetylase remains unknown (Ng *et al.*, 2000). Within the C-terminal part of MECP2 a histidine and proline-rich region is present which is conserved in certain neural-specific transcription factors (Vacca *et al.*, 2001). BAZ2A/TIP5 is part of the nucleolar remodeling complex (NoRC) which represses rDNA transcription by recruiting histone methyltransferases, HDACs, and DNA methyltransferases (Santoro *et al.*, 2002).

BAZ2B has a domain structure similar to BAZ2A/TIP5. Both contain a DDT domain (DNA binding homeobox and Different Transcription factors) and a tandem PHD-bromodomain. The PHD domain is a C4HC3 zinc-finger-like motif and the bromodomain consists of 110 amino acids and is found in many chromatin-associated proteins that can interact specifically with acetylated lysine. Tandem PHD-bromodomains have been found in several transcriptional co-repressors (Schultz *et al.*, 2002). The DDT domain is exclusively associated with nuclear domains in other proteins and was found in different transcription and chromatin remodeling factors (Doerks *et al.*, 2001). An AT-hook motif (which allows binding to the minor groove of AT-rich DNA regions) was found in BAZ2A/TIP5 but not in BAZ2B.

SETDB1 is a H3-K9 histone methyltransferase (Schultz *et al.*, 2002). Its mouse homologue, ESET, has furthermore been shown to interact with the mSin3A/B co-repressor complex (Yang *et al.*, 2002). The SET domain is a signature motif for lysine-specific histone methyltransferases (Wang *et al.*, 2001, Tachibana *et al.*, 2001). This domain is also present in CLLD8/SETDB2 to which no function has yet been assigned. A recent study showed the formation of an S-phase-specific complex including SETDB1, MBD1, and CAF-1 that facilitates methylation of H3-K9 during replication-coupled chromatin assembly (Sarraf and Stancheva, 2004).

The predicted protein sequence of KIAA1461/MBD5 harbors a PWWP motif (Stec *et al.*, 1998) named after the conserved amino acids Pro-Trp-Trp-Pro. It was first described in the

WHSC1 protein, encoded by a gene within the Wolf-Hirschhorn syndrome critical region. The PWWP domain of Dnmt3b, a DNA methyltransferase, has recently been shown to bind to DNA (Qiu *et al.*, 2002). A common feature of PWWP containing proteins is the presence of additional domains known to be associated with chromatin (Qiu *et al.*, 2002). Furthermore, KIAA1461/MBD5 has been shown to interact with the KIAA1549 protein in a yeast-two-hybrid experiment (<http://www.kazusa.or.jp/huge/ppi>). This interaction partner has not been studied in detail so far, but contains a serine-rich stretch as well as a helix-turn-helix motif (PS00622, LuxR family) according to Prosite. Helix-turn-helix motifs can be found in many transcription regulation proteins. Interestingly, the *KIAA1461/MBD5* gene lies in a region that was found to be deleted in a mentally retarded patient (Koolen *et al.*, 2004).

In the predicted protein sequence of KIAA1887/MBD6 only a proline-rich extension but no protein motif as such could be found.

The co-existence of MBDs and domains involved in histone modification (e.g. the SET domain and the bromodomain) in SETDB1, SETDB2, BAZ2A, and BAZ2B indicates a potential connection between the recognition of methylated DNA and histone modifications. This link is especially tempting for SETDB1, which has been shown to methylate histones (Schultz *et al.*, 2002), while its mouse homologue ESET binds to mSinA co-repressor complex (Yang *et al.*, 2002) that is also bound by MECP2 (Nan *et al.*, 1998).

3.1.4 MBD proteins in other species

In the mouse homologues were found for all human MBD proteins. Sequence identity scores range from 63.8% to 94.0% (Tab. 22) indicating a conserved function in both species. Human and mouse MBD1, MBD2, MBD3, MBD4 and MECP2 are curated orthologues (Hendrich and Bird, 1998, Hendrich *et al.*, 1999, Quaderi *et al.*, 1994, D'Esposito *et al.*, 1996).

Human protein	Mouse protein	% identity
MECP2	MECP2 (541 aa)	93.4% in 542 aa
MBD1	MBD1 (713 aa)	66.8% in 698 aa
MBD2	MBD2 (454 aa)	93.2% in 545 aa
MBD3	MBD3 (362 aa)	85.1% in 355 aa
MBD4	MBD4 (631 aa)	63.8% in 647 aa
BAZ2A	BAZ2A (1972 aa)	80.1% in 2039 aa
BAZ2B	ENSMUSP00000028367 (2065 aa)	82.6% in 2027 aa
SETDB1	ESET (1457 aa)	91.0% in 1465 aa
SETDB2	ENSMUSP00000022552 (701 aa)	65.2% in 715 aa
MBD5	ENSMUSP00000036847 (1518 aa)	94.0% in 1521 aa
MBD6	ENSMUSP00000026476 (101 aa)	93.5% in 216 aa

Table 22. Mouse homologues to human MBD proteins. Comparison of the human MBD proteins and their mouse homologues. Identity scores were calculated with the LALIGN program (http://www.ch.embnet.org/software/LALIGN_form.html). The numbers of aa residues are indicated. ^{a)} The number of aligned amino acids can exceed the residue number of the sequences due to gaps inserted by the algorithm.

Homology searches in the Ensembl database revealed the following murine homologues of BAZ2B, CLLD8, *KIAA1461/MBD5* protein, and *KIAA1887/MBD6* protein. The mouse homologue of BAZ2B, the Ensembl protein ENSMUSP00000028367 (gene ENSMUSG00000026987), shows a 82.6% amino acid sequence identity. The predicted mouse gene ENSMUSG00000021980 coding for Ensembl protein ENSMUSP00000022552 has a 65.2% amino acid sequence identity to the human CLLD8/SETDB2. It was furthermore found, that the predicted gene ENSMUSG00000036792 coding for the Ensembl protein ENSMUSP00000036847 is a mouse homologue of human KIAA1461/MBD5 protein with a 94.0% amino acid sequence identity. For the KIAA1887/MBD6 protein, the mouse gene sequence ENSMUSG00000025409 (Ensembl protein ENSMUSP00000026476) with 93.5% sequence identity was present in the database. However, the latter database entry consists of only 216 amino acids.

The existence of corresponding translated proteins remains to be determined for all four mouse genes.

DNA methylation as a mechanism of gene expression regulation exists also in plants. In the database searches plant MBD proteins were detected as well. The Pfam database contains polypeptides from *Arabidopsis thaliana* (thale cress) and *Triticum aestivum* (bread wheat).

BLAST analyses revealed additional proteins in *Zea Mays* (maize), *Hordeum vulgare* (two-rowed barley) and *Lycopersicon esculentum* (tomato). Entries for MBD containing proteins from plants over *C. elegans* (nematode) to mouse and human are present in Pfam.

3.1.5 Expression patterns of MBD genes

Expression analyses had been carried out previously for all genes of the mouse/human MBD family except for *KIAA1461/MBD5* and *KIAA1887/MBD6* (only the abundance of *KIAA1887/MBD6* ESTs in different tissues had been reported (Nagase *et al.*, 2001a)). The results of published Northern blot experiments are summarized in Tab. 22. Since expression levels of *MBD4* were too low to be detected by Northern blots, only results of RT-PCR studies in three tissues are shown. However the presence of *MBD4* EST sequences from numerous tissues points to a ubiquitous expression (Unigene).

Tissue	MECP2 ^{a)}	MBD1 ^{b)}	MBD2 ^{b)}	MBD3 ^{b)}	MBD4 ^{b)}	BAZ2A ^{d)}	BAZ2B ^{d)}	SETDB1 ^{e)}	SETDB2 ^{c)}	MBD5 ^{f)}	MBD6 ^{g)}
Brain	+	+	+	+	+	+	(+)	+	(+)	(+)	(+)
Heart	+	+	+	+		+	+	+	+	+	+
Kidney	+	+	+	+		+	-	+	+	(+)	+
Liver	+	(+)	(+)	+		(+)	-	+	+	(+)	+
Lung	+	+	+	+		+	-	+	+	(+)	(+)
Skeletal muscle	+	+	+	+		+	+	+	(+)	+	+
Spleen		+	+	+		+	-	+	+		+
Testis		+	+	+	+	+	+	+	+		+
ES cells		-	(+)	+	+						
Placenta	+					+	+	+	(+)	+	+
Pancreas	+					+	+	+	+	+	+

Table 23. Expression patterns of the mouse/human MBD genes. The table shows the expression of the eleven MBD genes in major tissues as detected by Northern blot. Empty spaces denote lacking information for that tissue. (+) indicates very low or doubtful expression, - no expression. ^{a)} (D'Esposito *et al.*, 1996, Coy *et al.*, 1999). ^{b)} (Hendrich and Bird, 1998) For these genes, comprehensive expression data was only available from mouse. ^{c)} (Mabuchi *et al.*, 2001) Expression in additional tissues has been reported. ^{d)} (Jones *et al.*, 2000) Expression in additional tissues has been reported. ^{e)} (Nomura *et al.*, 1994) *SETDB1* was originally called *KIAA0067*. ^{f)} Northern blot results of this study. ^{g)} (Nagase *et al.*, 2001a) and Northern blot results of this study.

Northern blot analyses for *KIAA1461/MBD5* and *KIAA1887/MBD6* were performed during

this thesis. Strong signals, representing a transcript of ~8 kb were detected for KIAA1461/MBD5 in skeletal muscle, heart, pancreas, kidney and placenta. A faint band could be detected in brain, lung and liver. For KIAA1887/MBD6 a strong band corresponding to a transcript of ~5 kb was present in heart, kidney, liver, skeletal muscle, placenta, and pancreas, and weaker signals could be seen for brain and lung tissue (Fig. 12).

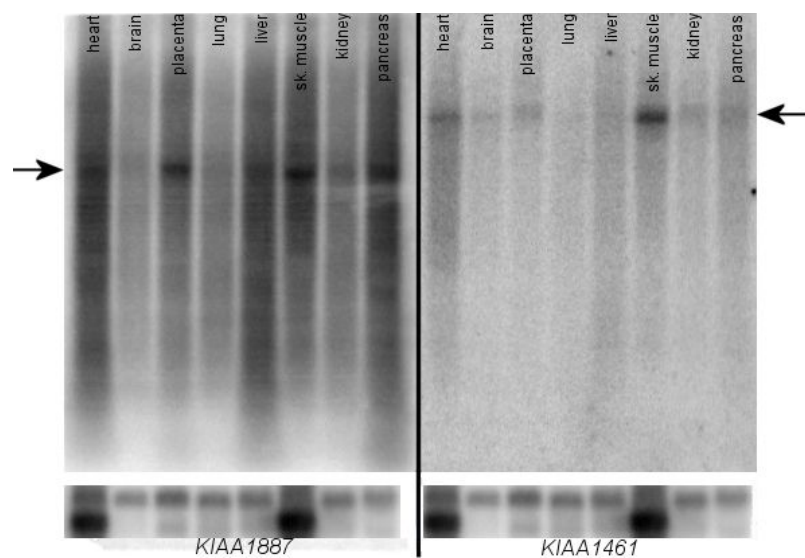


Fig. 12. Northern blots for KIAA1461/MBD5 and KIAA1887/MBD6. Human multiple tissue Northern blot showing the expression of the KIAA1461/MBD5 and KIAA1887/MBD6 genes. The calculated sizes of the transcripts are ~5 kb for KIAA1887/MBD6 and ~8 kb for KIAA1461/MBD5, indicated by arrows. A β -actin probe was hybridized as loading control, shown at the bottom.

Taken together, *MBD1*, *MBD2*, *MBD3*, and *MECP2* as well as *SETDB1*, *CLLL8/SETDB2*, *BAZ2A*, *KIAA1461/MBD5*, and *KIAA1887/MBD6* show a broad tissue distribution. In contrast the expression of *BAZ2B* is more restricted according to Northern blot results (Table 23) (Jones *et al.*, 2000). It is noteworthy, that *CLLL8/SETDB2*, *BAZ2A*, *KIAA1461/MBD5*, and *KIAA1887/MBD6* are expressed at very low levels in brain.

3.2 Search for MECP2 paralogues

3.2.1 Objective

The function of MECP2 is not only dependent on the MBD, but also on other domains of the protein and on its overall structure (Klose and Bird, 2004). As yet, only the 3D-structure of the MBD has been resolved while the exact shape of the entire protein remains to be elucidated.

A bioinformatics approach was therefore used to find human proteins with a global similarity to MECP2. Such paralogue proteins are thought to arise through gene duplication and can

give clues about the tertiary structure and about related functions of the protein of interest.

3.2.2 Proteins with an overall sequence similarity to MECP2

To avoid false positives in this search, MECP2 was first checked for repetitive sequences. A dotplot analysis revealed three such regions with low sequence complexity (aa 20 - aa 90, aa 160 - aa 200, and aa 340 - aa 400, see also Fig. 13). These regions were masked and BLASTP searches against the NCBI protein sequence database were performed.

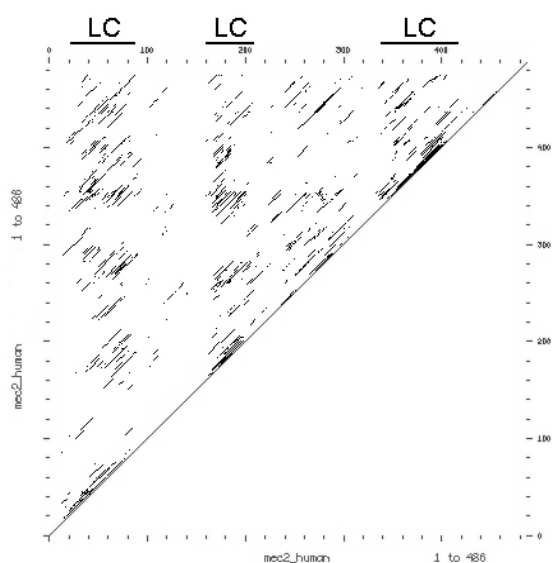


Fig. 13. DotPlot output of a comparison of MECP2 to itself. Every dot in the graph represents a similarity. Clusters of dots that do not lie on the diagonal indicate potential repetitive sequences. Such regions with low complexity are indicated by LC.

The BLASTP searches were followed by database queries of the sequence similarity database (SSDB) at GenomeNet and the BLink database at NCBI using the search terms hsa:4204 and P51608 respectively.

The BLink search revealed a total of 30 matches in the *Homo sapiens* proteome, but due to redundancy there were only 12 unique hits, four of which were MBD1, MBD2, MBD3 and MBD4. Since similarities to the MBD protein family members were mainly due to consensus in the MBD domains and therefore not of interest for overall structural similarity, they were disregarded. The remaining eight proteins are summarized in Table 24. The SSDB query with a SW-score cut-off at 200 yielded 18 non-redundant hits, including MBD1, MBD2 and MBD4.

The list resulting from the BLink search (8 proteins) and the SSDB query (15 proteins) was compared to the 100 best hits of the protein Blast with the masked MEC2_HUMAN as query sequence. Only one protein gave a positive result with all three search strategies.

Gene symbol	Protein name	SSDB	BLink	Blast
<i>NEFH</i>	Neurofilament heavy polypeptide	x	x	x
<i>MLLT2/AF4</i>	MLL/AF1p fusion - Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>)	x	x	
<i>FOXG1B</i>	Forkhead box G1B, Oncogene QIN		x	
<i>TRDN</i>	Triadin		x	
<i>PRG4</i>	Proteoglycan	x		
<i>SRRM1</i>	Serine/arginine repetitive matrix 1	x		
<i>SRRM2</i>	Serine/arginine repetitive matrix 2	x		
<i>MDC1</i>	Mediator of DNA damage checkpoint 1, KIAA0170 gene product	x		
<i>NOLC1</i>	Nucleolar and coiled-body phosphoprotein 1	x		
<i>BAT2</i>	HLA-B associated transcript 2	x		
<i>IRS2</i>	Insulin receptor substrate 2	x		
<i>SRCAP</i>	Snf2-related CBP activator protein	x		
<i>CRK7</i>	CDC2 related protein kinase 7	x		
<i>AIM1</i>	Absent in melanoma 1	x		
	Hypothetical protein BC001584	x		
<i>ANK2</i>	Ankyrin 2, neuronal	x		
<i>ZNF469</i>	Zinc finger protein 469	x		

Table 24. Results of the database queries. Crosses indicate positive results in the searches. SSDB = Sequence similarity database, BLink = Blast Link database, Blast = Blast search with masked MEC2_HUMAN as query. Only NEFH could be found in all 3 search methods.

The human neurofilament NEFH (Mattei *et al.*, 1988) belongs to the intermediate filament family and has been associated with amyotrophic lateral sclerosis (ALS). Overexpression of *Nefh* causes motor neuron degeneration in transgenic mice (Collard *et al.*, 1995) and deletions in the C-terminal part of *NEFH* can be found in ALS patients (Figlewicz *et al.*, 1994), supporting the link between NEFH and ALS. The alignment with MECP2 shows a similarity in the C-terminal part of NEFH as depicted in Fig. 14 and Annex 6.1.

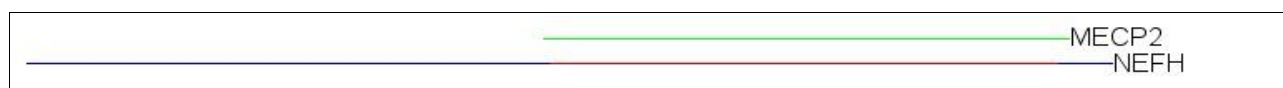


Fig. 14. Alignment of MECP2 (486 aa) and NEFH (1020 aa) derived from the SSDB. The overlap of 494 aa includes gaps and has an overall similarity of 0.235 (with 1 being a perfect similarity). Rather than having an overlap in only one domain, the whole MECP2 protein has a similarity to the C-terminal part of NEFH.

The product of the *MLLT2* gene, also called AF4 for "ALL1-fused gene from chromosome 4"

is a serine- and proline-rich putative transcription factor with a glutamine-rich carboxy terminus that results from translocation of ALL1 in acute leukemias. Fig. 15 represents an alignment of MLLT2 and MECP2.

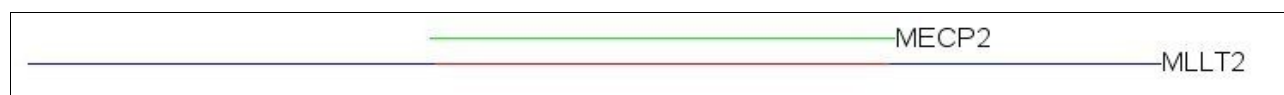


Fig. 15. Alignment of MECP2 (486 aa) and MLLT2 (2311 aa) derived from the SSDB. The overlap of 528 aa includes gaps and has an overall similarity of 0.225 (with 1 being a perfect similarity).

FOXG1B was described as the *QIN* gene product and is a transcription factor with a forkhead domain (Li and Vogt, 1993). It controls the fate of Cajal-Retzius cells (Hanashima *et al.*, 2004).

Triadin (TRDN) is a protein thought to be involved in the calcium release from muscle contraction (Brandt *et al.*, 1993). It is also the founder of the poorly understood triadin protein family (for review see Marty, 2004). Proteoglycan 4 (PRG4) also known as SZP and CACP can act as megakaryocyte-stimulating factor (Merberg *et al.*, 1993). It is expressed by chondrocytes in bovine articular cartilage (Schumacher *et al.*, 1994). *PRG4* is thought to be the disease gene for camptodactyly-arthropathy-coxavara-pericarditis (CACP) syndrome and has an isoform known as HAPO for hemangiopoietin. Finally, it is a growth factor acting on hematopoietic and endothelial progenitor cells (Liu *et al.*, 2004).

SRRM1 is a co-activator of pre-mRNA splicing. It binds different splicing factors, associates with the nuclear matrix, and is involved in mRNA export (Wagner *et al.*, 2003). SRRM2 is also involved in splicing, but was shown to be less important than SRRM1 for this process *in vitro* (Blencowe *et al.*, 2000).

MDC1, also termed NFDB1, works with H2AX to promote the recruitment of repair proteins to DNA breaks and it controls damage-induced cell-cycle arrest checkpoints (Stewart *et al.*, 2003).

NOLC1, formerly known as p130 or Nopp140, is a putative snoRNP assembly factor (Yang *et al.*, 2000).

Little is known about BAT2, but Lehner *et al.* in 2004 proposed a role in the regulation of pre-mRNA splicing.

IRS2 links cell surface receptors to the intracellular insulin/IGF signaling cascade and has an important role in both peripheral insulin response and pancreatic beta-cell growth and

function. Dysregulation of *Irs2* signaling in mice causes the failure of compensatory hyperinsulinemia during peripheral insulin resistance (for a review see Lee and White, 2004). In humans it has been associated with type II diabetes (Mammarella *et al.*, 2000), severe obesity, and glucose intolerance (Lautier *et al.*, 2003).

SRCAP is an ATPase binding to the CREB-binding protein (CBP) and a transcription activator (Johnston *et al.*, 1999). Furthermore, it is also a co-activator for the androgen receptor and can enhance glucocorticoid receptor-mediated transcription (Monroy *et al.*, 2003).

CRK7, also termed CrkRS, is a protein kinase found in nuclear SC35 speckles (Ko *et al.*, 2001).

A single-nucleotide polymorphism (SNP) in *AIM1* has been associated with differences in skin color in humans (Fukamachi *et al.*, 2001).

ANK2 is a spectrin-binding protein that is required for localization of inositol 1,4,5-trisphosphate receptor and ryanodine receptor in neonatal cardiomyocytes (Mohler *et al.*, 2004). It was furthermore linked to long QT syndrome 4 (sick sinus syndrome with bradycardia) (Mohler *et al.*, 2004).

ZNF469 is a zinc finger protein that was first described as the *KIAA1858* gene product (Nagase *et al.*, 2001b).

None of the proteins could be found in the protein structure database PDB, which means that their 3D structure has not been determined as yet. Sequence comparison of MECP2 with proteins for which structural data is available, could have helped to get an idea of the 3D structure of MECP2 itself.

3.3 MECP2 target genes

3.3.1 Objective

Even though *MECP2* has been identified as disease gene for Rett syndrome in 1999 (Amir *et al.*, 1999), the disease is still poorly understood. Especially knowledge on the exact function of MECP2, i.e. the genes regulated by MECP2 as a transcriptional repressor is missing. In addition, it is unknown whether MECP2 has functions other than transcriptional regulation.

A mouse model created by the Bird group (Guy *et al.*, 2001) was used to find genes regulated

by MECP2 in the brain. The approach consisted of a microarray hybridization of total brain RNA from a *Mecp2^{-ly}* mouse versus total brain RNA from a healthy litter mate control. The hybridization results were verified by northern blot and real-time PCR performed by our collaborators in Edinburgh. Subsequently, chromatin immunoprecipitation was carried out to prove direct binding of MECP2 to the genomic region of genes that might be involved in the disease.

Four microarray studies on Rett syndrome have been published since the start of this thesis. Two studies used cultured cells, i.e. primary fibroblasts (Traynor *et al.*, 2002) and lymphoblastoid cells (Ballestar *et al.*, 2005). Colantuoni and colleagues isolated RNA from post mortem human brain tissue (Colantuoni *et al.*, 2001) and Tudor and colleagues isolated RNA from whole brains of MECP2 mutant mice (Tudor *et al.*, 2002).

Apart from the study by Tudor and colleagues, the biological material used in these experiments does not seem to be well suited to find the neuronal target genes of MECP2 responsible for the phenotype (see 4.3.6 DNA microarray studies and chromatin immunoprecipitation).

3.3.2 Genes differentially expressed in *Mecp2*-null mice – a microarray study

Gene expression levels in the brains of a symptomatic *Mecp2^{-ly}* mouse were compared to the expression levels of a *wt* litter mate control brain. Isolated RNA was sent by our collaborators from Edinburgh. The RNA was fluorescently labeled with Cy3 and Cy5 in a reverse transcription reaction and co-hybridized to a microarray with 13,627 cDNA clones.

After scanning, image analysis and quality control the data was subjected to ANOVA. From the resulting clones, those with an intensity ratio of 2.00 or higher were considered relevant. A total of 17 clones were found to be either up- or down-regulated at least 2-fold in the brain RNA of the mutant mouse. Due to redundancy, the 17 clones correspond to 11 transcripts of which three were down- and eight were up-regulated (Table 25).

Fold change	Up / down regulated	Gene description	UniGene cluster	Accession number
3.4412	+	<i>Sgk1</i> Serum/glucocorticoid regulated kinase	Mm.28405	AA273540
3.2992	+	<i>Sgk1</i> Serum/glucocorticoid regulated kinase	Mm.28405	AI527833
2.6313	+	<i>Fkbp5</i> FK506 binding protein 5 (51 kDa)	Mm.276405	BC015260
2.5815	+	<i>Fkbp5</i> FK506 binding protein 5 (51 kDa)	Mm.276405	BC015260
2.3051	+	<i>Fkbp5</i> FK506 binding protein 5 (51 kDa)	Mm.276405	BC015260
2.4131	+	<i>Cirbp</i> Cold inducible RNA binding protein	Mm.17898	NM_007705
2.254	+	<i>Cirbp</i> Cold inducible RNA binding protein	Mm.17898	NM_007705
2.1947	+	<i>Cirbp</i> Cold inducible RNA binding protein	Mm.17898	NM_007705
2.2748	+	<i>Sult1a1</i> Sulfotransferase family 1A, phenol-preferring, member 1	Mm.17339	AB029487
2.2651	-	<i>Sorcin</i>	Mm.96211	AK008970
2.2386	-	<i>Hsp105</i> Heat shock protein, 105 kDa	Mm.270681	AA105012
2.0727	-	<i>Hsp105</i> Heat shock protein, 105 kDa	Mm.270681	BC018378
2.2012	+	<i>Pomc1</i> Pro-opiomelanocortin-alpha	Mm.277996	BC061215
2.1584	+	<i>Scya17</i> Small inducible cytokine subfamily A17	Mm.41988	NM_011332
2.1477	-	<i>Gja12</i> -pending Gap junction membrane channel protein alpha 12	Mm.40016	AW742272
2.1375	+	<i>S3-12</i> -pending Plasma membrane associated protein, S3-12	Mm.347924	NM_020568
2.0019	+	RIKEN cDNA 4930546H06 gene	Mm.227456	AK016052

Table 25. Gene expression changes in the brain of an *Mecp2*^{-y} animal compared to a *wt* litter mate control detected by microarray hybridizations. Transcripts with more than 2-fold expression difference are shown. Some genes are represented by more than one differential clone on the array and therefore listed several times. The identity of all 17 clones was confirmed by sequencing of the spotted DNA. Up-regulation (+) and down-regulation (-) in the *Mecp2*^{-y} sample are indicated.

At least 5 of the 11 differentially expressed genes (*Fkbp5*, *Sgk*, *Pomc*, *Sult1A1*, and *Hsp105*) are known to be regulated by the stress hormones, glucocorticoids. Stress provokes release of corticotropin-releasing hormone (CRH) by the hypothalamus, which stimulates synthesis of adrenocorticotrophic hormone (ACTH). ACTH in turn causes the adrenal cortex to produce circulating glucocorticoids (cortisol in humans, corticosterone in rodents) which bind to glucocorticoid receptors and coordinate the transcriptional response (Reichardt and Schütz, 1998). Negative feedback by glucocorticoids on the hypothalamus and pituitary ensures that the stress response is transient under normal conditions.

The POMC polypeptide is the precursor of ACTH. The POMC gene was found to be higher

expressed in the brain of the *Mecp2*^{-y} mouse as compared to the brain of its *wt* litter mate. *Sgk* and *Sult1A1* are reportedly induced by glucocorticoids (Lang and Cohen, 2001, Duanmu *et al.*, 2001) and they were found to be up-regulated in the brain of the *Mecp2*^{-y} animal. Another up-regulated glucocorticoid-inducible gene, *Fkbp5*, encodes a peptidyl-prolyl cis-trans-isomerase associated with glucocorticoid receptor complexes (Yoshida *et al.*, 2002). *Hsp105*, also called *Hsp110*, was down-regulated in the brain of the *Mecp2*-deficient animals (Fig. 29 A). Its expression is inhibited by the glucocorticoid dexamethasone (Wadekar *et al.*, 2001). *Cirp* (the cold-inducible RNA binding protein gene) is induced by low temperature or low oxygen tension (Wellmann *et al.*, 2004) but is not known to be induced by glucocorticoids.

3.3.3 Localization of MECP2, FKBP5, and SGK in mouse brain

Next, the immunolocalization of FKBP5, SGK (=SGK1) and MECP2 proteins in selected brain regions of adult female mice were examined. The results revealed a significant colocalization of FKBP5 with cells that also synthesize MECP2 (Fig. 16 A, B). A few MECP2-positive cells showed low or undetectable levels of FKBP5 staining (e.g. cortical cells in the position of Cajal-Retzius cells; Fig. 16 A). Many cells positive for SGK and MECP2 were found, though fewer than positive for FKBP5 and MECP2 (Fig. 16 C-E). In addition, cells producing either MECP2 or SGK are present in the brain (Fig. 16 D, E). As these animals had not been stressed or glucocorticoid-treated, the data suggest that presence of MECP2 in a neuronal cell is compatible with FKBP5 and SGK synthesis as proposed in Fig. 16. Therefore, MECP2 does not act as a transcriptional silencer on *Fkbp5* of *Sgk* in the brain.

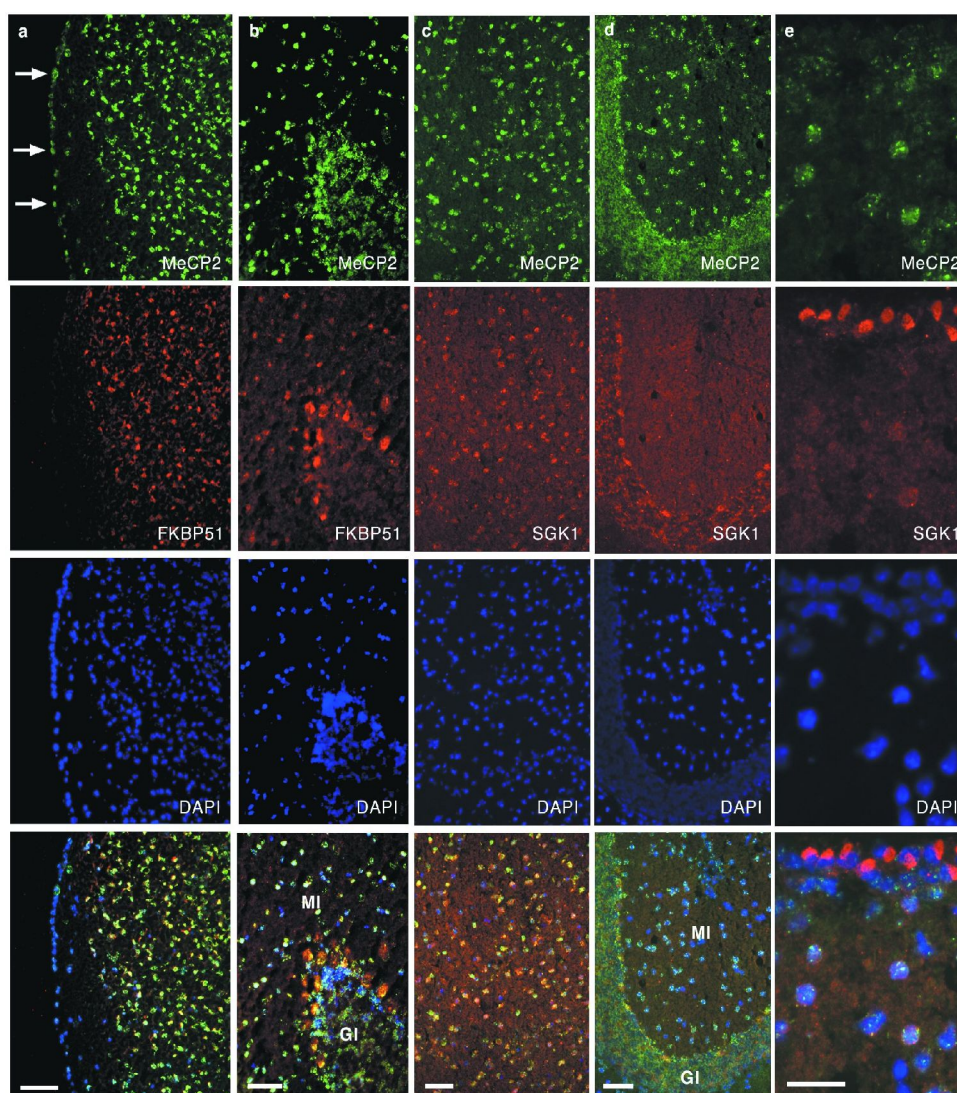


Fig. 16. Expression of MECP2, FKBP5, and SGK (=SGK1) in the brain of adult female mice. Immunolocalization of MECP2 (green signals, FITC-coupled secondary antibody), FKBP5 (red signals in A and B, Cy3-coupled secondary antibody), and SGK (red signals in C-E, Cy3-coupled secondary antibody) on cryostat sections of mouse brain. Panels show the temporal cortex (A), the cerebellum (B, D), the region immediately above the corpus callosum (C), and the lateral ventricle (E). Blue signals: DAPI-stained nuclei. GI: granular cell layer. MI: molecular cell layer. A clear cellular overlap of the staining with anti-MECP2 and anti-FKBP5 antibodies can be seen for the majority of cells in all brain regions. There are few MECP2-positive cells with an absent or weak FKBP5 expression: the most superficial cortical layer is FKBP5-negative (arrows in A) and some cells with weak FKBP5 signals can be seen in the molecular and granular cell layer of the cerebellum. Many cells co-express MECP2 and SGK in the brain, for example cells above the corpus callosum (C) and cells in the granular cell layer of the cerebellum (D). Distinct MECP2-positive cells in the molecular layer show only a weak homogeneous staining with the SGK antibody (D). The strongest SGK signals were detected in cells lining the ventricles (top cell layer in E). These cells show absent or very weak MECP2-staining in contrast to cells at the subluminal site. Scale bars in A equal 100 μm , in B, C, D 50 μm , and in E 25 μm .

3.3.4 Establishment of chromatin immunoprecipitation (ChIP)

To test the direct interaction of MECP2 and *Fkbp5*, the chromatin immunoprecipitation method was established and applied in this work. ChIP is the technique of choice to study the *in vivo* binding of DNA-associated proteins.

Even though laborious optimization of many parameters is required to obtain proper results with ChIP (for a review, see Das *et al.*, 2004), the technique has been widely used in recent years. Among the conditions that have to be adjusted are the time and intensity of fixation, the purification of the input material (i.e. purification of cells, or even isolation of nuclei or chromatin) the shearing of the chromatin, the washing of the immunoprecipitated complexes, and the final detection.

For the experiments carried out in this thesis, fixation with 1% formaldehyde in the medium for 10 minutes at 37°C proved to be the most efficient way to bind the proteins of interest to the DNA. Using these settings, the MECP2 polypeptide was properly coupled to the DNA while most of the epitopes recognized by the polyclonal anti-MECP2 antibody were still accessible. Differences between batches of anti-MECP2 antibodies were however observed. Chromatin immunoprecipitation with antibodies for histones was generally more reliable and resulted in more precipitated DNA. This is probably due to the higher abundance of the histone proteins as compared to MECP2 or GR in the chromatin regions studied.

The shearing parameters for the fragmentation of the chromatin had to be optimized for the two types of biological material used (primary neurons and grounded brain tissue). Different sonication times and intensities were tested, using a Branson sonifier. Optimal results were achieved with 5 pulses of 30 seconds at 100 % duty and output level 5. The sonication was carried out in 2 ml buffer containing sodium dodecyl sulfate (SDS). The samples were kept in ice cold salt water to prevent the disruption of the DNA-protein complexes by heat as well as to hinder the formation of foam during the shearing procedure.

The described settings resulted in a smear of DNA ranging from 200 bp to about 1000 bp. These fragment sizes allowed a proper amplification after immunoprecipitation and isolation of the DNA.

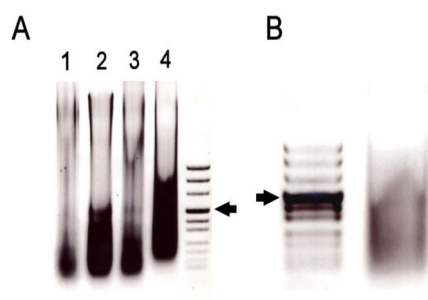


Fig. 17. DNA fragment distribution obtained by sonication of chromatin. (A) Lanes 1 to 4 depict the smears obtained by sonication with 2, 3, 4 and 5 pulses of 30 sec at output level 5, respectively. Lane 4 represents the optimal DNA fragment size distribution for amplification with specific primers after ChIP. (B) DNA smear resulting from sonication (5 pulses of 30 sec, output level 5) of chromatin from brain tissue. Arrowheads indicate the pUC mix marker band at 501 bp.

3.3.5 MECP2 binding sites in the genomic regions of *Fkbp5*

Since the only DNA motif necessary for MECP2 binding is the dinucleotide $m^5\text{CpG}$ (Nan *et al.*, 1993), it is not trivial to predict binding sites from the genomic sequence alone. Therefore, a comparative genomics approach was used to find genomic fragments conserved between the mouse *Fkbp5* and the human *FKBP5* (see methods section 2.2.1.1). This data was compared to binding sites of known transcription factors which were mapped to the *Fkbp5* genomic region. To do so, all motifs of the Transfac database were blasted against the *Fkbp5* region of the mouse genome. Finally, the CpG content of the sequence was considered, since CpG islands are often found in promoters of genes, but are usually unmethylated and hence not a promising binding region for MECP2. Primer pairs were designed covering regions of about 1 kb at the two most promising target sites (region 1 with primer pairs 1_0 to 1_4 and region 2 with primer pairs 2_1 to 2_4, see Fig. 18).

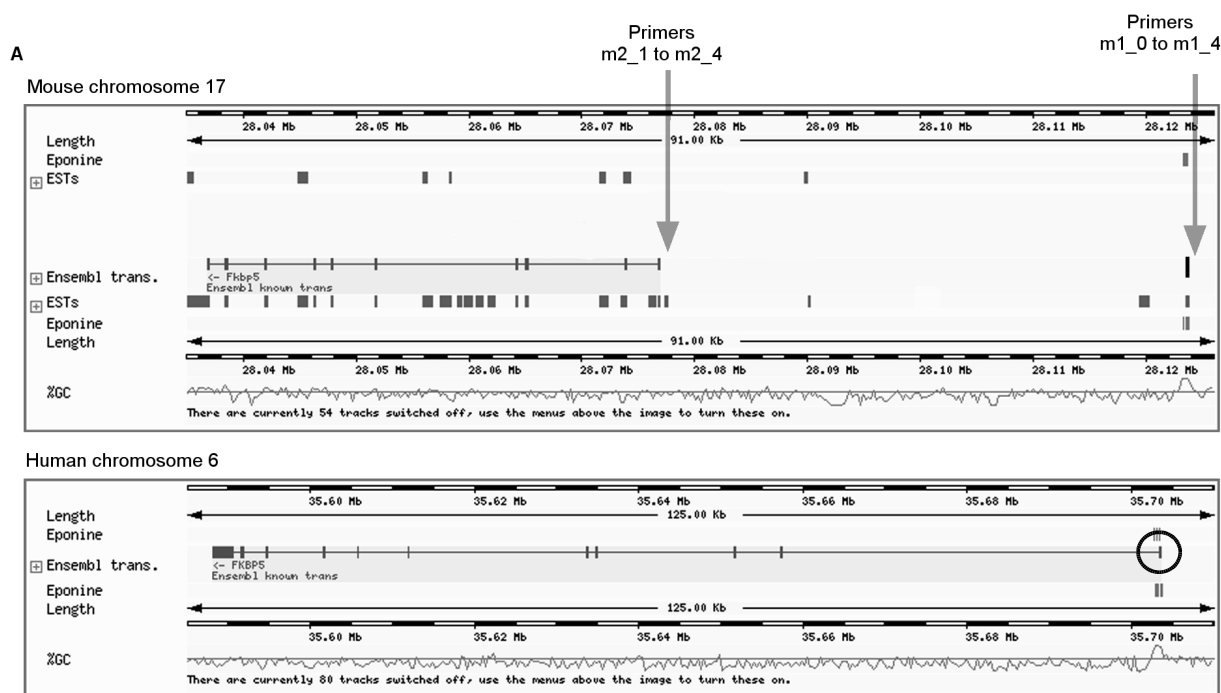


Fig. 18. Comparison of human *FKBP5* and mouse *Fkbp5* genomic regions. Ensembl representation of sections of human chromosome 6 and mouse chromosome 17 depicts the known transcripts, 5' to 3', right to left. Eponine denotes regions predicted to contain transcription start sites and promoters. The % GC curves show the percentage of GC base pairs in the sequence. Arrows indicate the regions covered with primers to find MECP2 binding sites. The black circle depicts the human exon 1 that was missing in the mouse transcript.

In the Ensembl database (release 24.33.1) the mouse *Fkbp5* transcript was missing the first exon present in the human transcript (Fig. 18). Thus, an RT-PCR experiment was performed with one primer corresponding to a sequence in the fourth exon and the other primer corresponding to a sequence in the first and second exon of the predicted transcript. The size of the PCR band confirmed the existence of a mouse transcript that contains an exon (similar to the first exon of *FKBP5* in the human genome) upstream of the transcription start in the Ensembl database (Fig. 19).

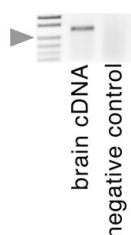


Fig. 19. PCR using mouse brain cDNA with primers specific for mouse *Fkbp5*. The first exon of *Fkbp5* is not part of the Ensembl known transcript (ENSMUST00000062167), but was present as EST in the database. RT-PCR revealed a cDNA fragment that corresponds to a sequence from exon 1 to exon 4 in brain tissue. Negative control: no reverse transcriptase was added to the reaction. Arrowhead: 242 bp DNA marker.

To confirm the binding of MECP2 to the two regions 1 and 2 of *Fkbp5*, ChIP was performed with whole mouse brain tissue (Fig. 20). Specific interactions could be detected in both regions in *wt* brains. As a negative control, *Mecp2*-null mouse brain tissue was used. In this case, no PCR products were obtained. This argues for the specificity of the antibody and against a cross-reaction with a protein other than MECP2. An antibody against acetylated histone H3 was used as positive control. An antibody against acetylated histone H3 was used as positive control.

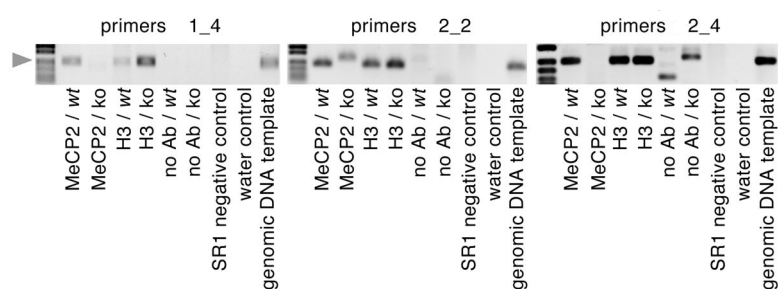


Fig. 20. Chromatin immunoprecipitation shows binding of MECP2 to the genomic region of *Fkbp5*. ChIP with antibodies against MECP2 and acetylated histone H3 using total brain tissue. MECP2 binds to region 1_4 and region 2 (2_2 and 2_4) of the *Fkbp5* gene in *wt* brain. Arrowhead: 242 bp DNA marker. SR1 negative control: no template was used in the SR1 PCR reaction. Water control: H₂O instead of template was used in the PCR with specific primers.

Reliable binding of MECP2 to the *Fkbp5* genomic regions 1_4, 2_2, and 2_4 could be shown in this thesis, as well as to the regions *fkp* and *fkp1* detected by our collaborators in Edinburgh. The relative position of the primers can be seen in Fig. 21. The regions 1_4, 2_1 and 2_2 are poor in CpG content while the regions *fkp1* and 2_4 contain many CpGs (Fig. 21). 1_4 and *fkp1* flank a CpG island.

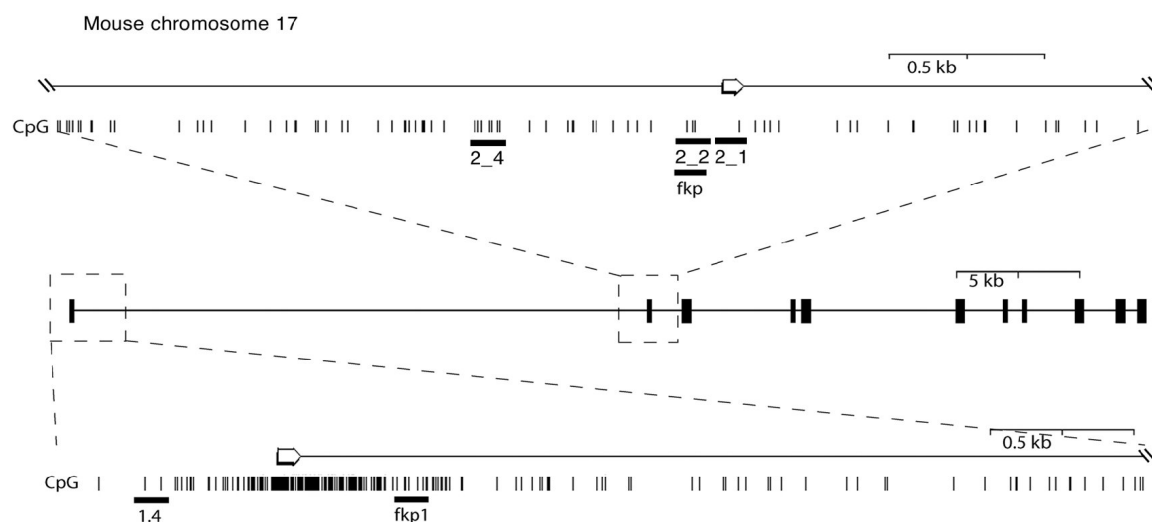


Fig. 21. Schematic representation of the mouse *Fkbp5* gene. Positions of primer pairs used for the amplification of MECP2 antibody precipitated *Fkbp5* genomic DNA fragments are indicated by black horizontal bars. The positions of CpGs present in the region are visualized by vertical bars.

MECP2 only binds to methylated DNA. To determine whether the potential binding sites of MECP2 determined by ChIP are methylated, bisulfite sequencing was performed. Genomic DNA from *wt* mouse brain was treated with sodium bisulfite, amplified with primers specific for the genomic regions 1_4, 2_1, 2_2, and 2_4 after bisulfite conversion, cloned into a vector, and sequenced. Since sodium bisulfite deaminates unmethylated cytosine residues which results in an uracil residue, only m⁵C still produced a cytidine in the sequencing output. The CpGs in the regions studied were mainly methylated, supporting the idea that these are binding sites of MECP2 (Fig. 22).

1_4		2_1	2_2			2_4						
27060946	27061064	27034766	27035122	27035077	27035067	27036337	27036345	27036385	27036402	27036452	27036474	27036493
•	•	•	•	•	•	•	•	•	•	•	•	•
○	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•
•	•	•	•	•	•	•	•	•	•	○	•	•

Fig. 22. Methylation status of CpGs in *Fkbp5* genomic DNA determined by bisulfite sequencing. The four genomic regions, 1_4 and 2_1, 2_2 (= fkp), and 2_4 were analyzed. DNA was derived from the brain of a *wt* mouse. For each region, 10 clones were sequenced. Numbers correspond to the location on Ensembl mouse genome chromosome 17 (release 24.33.1) Methylated CpGs are depicted by filled circles; unmethylated CpGs by open ones.

3.3.6 MECP2 and the glucocorticoid receptor compete for binding to the GRE_2 locus

In response to glucocorticoids, the glucocorticoid receptor has been shown to bind to human *FKBP5* and induce its expression (U *et al.*, 2004). So far, no studies on GR binding sites in the mouse *Fkbp5* genomic region were done.

Human and mouse *Fkbp5* genomic sequences were compared to identify the region in the mouse genome corresponding to the GR binding site in the human genome. Such a region, however, could not be detected in the mouse *Fkbp5* gene. Therefore, the Transfac database was used to search for all regions in the *Fkbp5* gene that show similarities to the GR binding site consensus sequence stored in the database. Six such loci, called glucocorticoid response elements (GRE), could be identified in this thesis (Fig. 23).

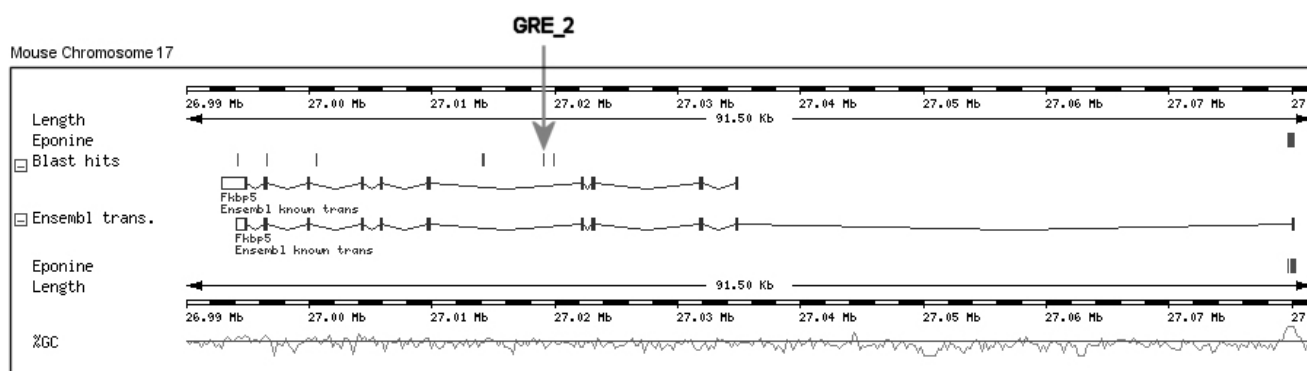


Fig. 23. *Fkbp5* intron/exon structure and the localization of GR response elements. Two Ensembl transcripts (starting at the two promoters, respectively) are represented by black squares (exons), 5' to 3', right to left. Eponine denotes regions predicted to contain transcription start sites and promoters. The % GC curves shows the percentage of GC base pairs in the sequence. Glucocorticoid receptor binding sites are indicated as blast hits and the binding site common to GR and MECP2 (GRE_2) is marked by a grey arrow.

In order to test whether MECP2 binds to the same genomic regions as GR, chromatin immunoprecipitation experiments were performed with antibodies against MECP2 and GR. Primary neurons were treated with a synthetic glucocorticoid (dexamethasone), with a glucocorticoid inhibitor (RU-486), with both, dexamethasone and RU-486, or with ethanol as solvent control.



Fig. 24. MECP2 and GR both bind to the GRE_2 genomic region. Chromatin immunoprecipitation results reveal MECP2 binding to the GRE_2 region under normal conditions (Ethanol) and in the presence of both, RU-486 (RU) and dexamethasone (Dex). In contrast, GR only binds when ethanol or dexamethasone are present, but not if RU-486 is added to the medium. Reactions without antibody (No Ab) served as negative controls and genomic DNA as positive control.

Fig. 24 illustrates, that under normal conditions MECP2 and GR bind to the GRE_2 region. MECP2 binding is abolished in the presence of the glucocorticoid dexamethasone. In contrast, GR binding can be reversed by addition of RU-486 which restores MECP2 binding. These results suggest a model in which MECP2 and GR compete for binding to the GRE_2 region and the repressor function of MECP2 is fine-tuned by glucocorticoids. Addition of RU-486 alone led to dissociation of MECP2 as well as GR from the GRE_2 region. This suggests, that excessive amounts of RU-486 disturb the gene expression regulation of *Fkbp5* in that neither the repressor MECP2 nor the activator GR binds to GRE_2 under this condition.

4. Discussion

4.1 MBD protein family

4.1.1 Expression of new MBD protein family members

For many years, MECP2 was thought to be a global transcriptional repressor. This assumption was based on *MECP2* expression in almost all tissues studied (D'Esposito *et al.*, 1996, Coy *et al.*, 1999) and the fact that MECP2 is essential for embryonic development in mice (Tate *et al.*, 1996). However, mutations in *MECP2* lead to Rett syndrome with an exclusively neuronal phenotype. This could be explained by a greater need of long lived, non-dividing neuronal cells for a special chromatin state that involves MECP2 and tightly suppresses transcription of genes not needed in that tissue. Another explanation would be that the loss of function of MECP2 in non-neural tissues is compensated by another protein with similar properties.

In this study, six proteins (BAZ2A, BAZ2B, KIAA1461/MBD5, KIAA1887/MBD6, SETDB1, and SETDB2) with a methyl-CpG-binding domain in mouse and man were identified as new members of the MBD protein family. Transcripts of *SETDB1*, *SETDB2*, *BAZ2A*, *KIAA1461/MBD5*, and *KIAA1887/MBD6* are found in all adult tissues studied. *SETDB2*, *BAZ2B*, and *KIAA1461/MBD5* show a low expression in brain and the mRNA levels of *KIAA1887/MBD6* are particularly lower than in other tissues. According to RT-PCR data from the HUGE database, *KIAA1461/MBD5* expression is basically confined to the cerebellum and the corpus callosum. The exact cell types that express *BAZ2A/TIP5*, *SETDB1*, and *SETDB2* as well as *KIAA1461/MBD5* and *KIAA1887/MBD6* in the brain are not known. MECP2 is found preferentially in mature neurons of the brain (Shahbazian *et al.*, 2002, LaSalle *et al.*, 2001). Should one of the other MBD proteins compensate for the loss of function of MECP2 in Rett syndrome it would be expected that the compensating gene shows an expression pattern complementary to that of *MECP2*, in particular in peripheral tissues that are not obviously affected in the disease. More detailed studies will be needed on the expression patterns of the MBD protein family members to confirm this hypothesis.

Based on gene expression studies with *Mecp2* knockout mice and on biochemical evidence, it has been suggested that the essential function of MECP2 in the brain might not be transcriptional regulation (Tudor *et al.*, 2002). In view of this aspect and of the protein–

protein interaction property of the methyl-CpG-binding domain of human MBD3 and SETDB1, functional compensation would not necessarily require a DNA binding capacity.

Except for KIAA1887/MBD6 all presented polypeptides are known or predicted to be involved in mechanisms of gene expression regulation. In order to understand the higher-order interplay of MBD proteins and associated complexes, it will be a major task to identify interacting proteins as well as regulated targets. This will help to solve the question whether some of these proteins can functionally complement MECP2 in tissues other than the brain.

4.1.2 The MBD domain of the new family members

The sequence comparison of the MBD domains of all MBD proteins showed only few amino acids conserved in all proteins. This raises the question whether the newly identified MBD sequences do really bind m⁵CpGs.

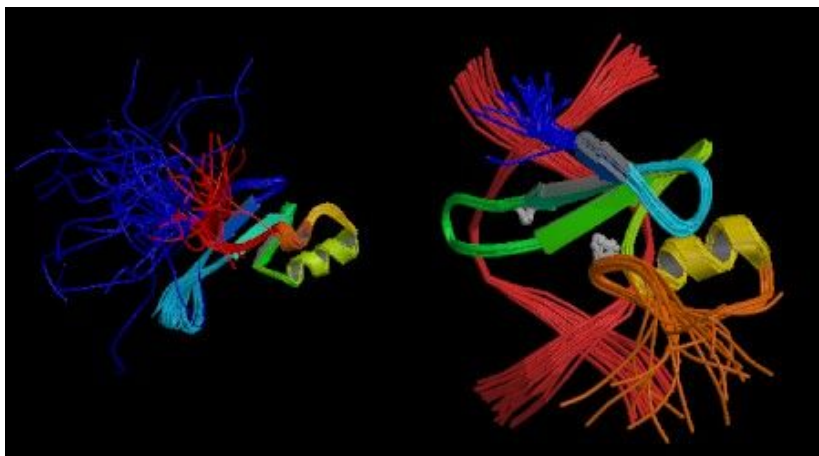


Fig. 25. Solution structure of the MBDs of MBD1 (left) and MECP2 (right). The α -helix (yellow) opposite the three β -sheets (arrows in green and blue), that form the 3D structure, can be seen. Loop 2, important for DNA binding, is turquoise in the MECP2 and green in the MBD1 structure. (Images adopted from PDB)

The solution structures of the methyl-CpG binding domains of three of the human MBD proteins, namely of MECP2 (Wakefield et al., 1999), MBD1 (Ohki *et al.*, 2001), and MBD4 (Wu *et al.*, 2003) have been published. A 3-D model of the MBDs of MECP2 and MBD1 are shown in Fig. 25. The structural data in these publications revealed residues important for DNA binding.

The following amino acids and structures have been found to be important for DNA and more specifically for m⁵CpG binding of MECP2: Loop 2 is inserted in the major groove of the DNA while the α -helix interacts with the sugar-phosphate backbone. The methyl groups of the cytosines on the forward strand interact with hydrophobic side chains of Val-18, Tyr-32, and

Arg-20. The residues Arg-42 and Ser-43 interact with the methyl groups on reverse strand. Arg-20 and Arg-42 form hydrogen bonds with the guanines of the m⁵CpG dinucleotides.

Amino acid	Residue present in
Val-18	MBD1, MBD2, and MBD3
Arg-20	MECP2, MBD1, MBD2, MBD3, MBD4, SETDB1, and BAZ2B
Tyr-32	MECP2, MBD1, MBD2, and MBD4
Arg-42	All MBD proteins except for MBD5, and MBD6
Ser-43	MECP2, MBD1, MBD2, MBD3, MBD4, and MBD6

Table 26. The existence of amino acids important for m⁵CpG binding in MBD1 in other MBD protein family members. This summary shows that in the new MBD protein family members many of the residues are not conserved. Also, in MBD4 and MECP2 Val-18 is not present.

Important residues of the structure of MECP2 are summarized in Table 27. In contrast to the methyl-CpG binding domain of MBD1 where Val-18, Arg-20, Tyr-32, Arg-42, and Ser-43 are necessary for m⁵CpG binding, in MECP2 also Ile-35 interacts with a methylated cytosine. This amino acid only exists in MECP2 and MBD4. Such structural differences probably cause the different binding specificities of the MBD proteins (see 1. Introduction).

The unrooted tree in Fig. 11 (Results 3.1.2) shows that the MBD domain of MECP2 is structurally most closely related to MBD4. However, MBD4 has a specificity for m⁵CpG/GpT mismatches. This suggests, that either other domains present in the protein modify the binding specificity of the MBD domain or that minor changes in the MBD can lead to an altered binding behavior. Fig. 9 (Results 3.1.2) clearly shows that the methyl-CpG binding domains of all MBD proteins vary considerably in their polypeptide sequence. It is even more intriguing, that at least three of the proteins (MECP2, MBD1, and MBD2) specifically bind the very small m⁵CpG motif. This suggests, that even a great variation in the sequence of the different MBDS still allows the same specialized m⁵CpG binding function

Amino acid	Function in structure
Leu-9	Hydrophobic core
Trp-13	Hydrophobic core
Arg-15	Hydrophobic core
Leu-17	Hydrophobic core
Tyr-29	Hydrophobic core
Val-31	Hydrophobic core
Leu-33	Hydrophobic core
Phe-41	Hydrophobic core
Leu-47	Hydrophobic core
Phe-51	Hydrophobic core
Leu-59	Hydrophobic core
Phe-64	Hydrophobic core
Phe-66	Hydrophobic core
Val-68	Hydrophobic core
Tyr-33	Interaction with methylated C
Ile-35	Interaction with methylated C

Table 27. Residues of the MBD of MECP2 and their function. Amino acids of the hydrophobic core form the three β -sheets and the α -helix.

From the sequence alone it is not possible to predict whether the new MBD protein family members will bind m⁵CpG or not. As far as the conservation of important residues in the Tables 26 and 27 are concerned, a binding is not very likely since many of the amino acids important for m⁵CpG-binding of MBD1 are not conserved. To get certainty about the binding capabilities, DNA binding assays with the individual domain sequences will have to be carried out. Since the MBDs of MBD3 and SETDB1 have been shown to bind to other polypeptides, protein-protein interactions should also be studied (e.g. by yeast-two-hybrid screens).

This of course raises the question, whether the newly identified MBD proteins merit the name MBD. However, it is nowadays common to name proteins according to protein family members they show similarity to. A recent example are the proteins MBD3L1 and MBD3L2 that have homologies with MBD3 but do not even contain an MBD motif (Jiang *et al.*, 2002).

4.2 Search for MECP2 paralogues

A total of 17 proteins with an overall sequence similarity to MECP2 were found in the different searches applied. Only two proteins were detected in the SSDB as well as the BLink search, namely "myeloid/lymphoid or mixed-lineage leukemia (thrithorax homolog, Drosophila); translocated to, 2" (MLLT2) and "neurofilament, heavy polypeptide 200kDA" (NEFH). NEFH was detected in all three analyses.

Proteins that were found in the database queries but not in the BLAST search have their similarity probably due to regions of MECP2 with low complexity such as the histidine-rich region from aa 266 to aa 372 or the proline-rich region from aa 376 to aa 405 (Fig. 26). Polypeptides similar to each other only due to matches in these domains are not likely to be paralogues since such local resemblances will have arisen independently during evolution of the proteins and are not remains of the sequence of a common ancestor. Therefore only NEFH and MLLT2 were considered to have an overall sequence similarity to MECP2 since they were found in two database searches and NEFH is the most promising candidate as it was also detected in the BLAST search in which the low complexity regions were masked.

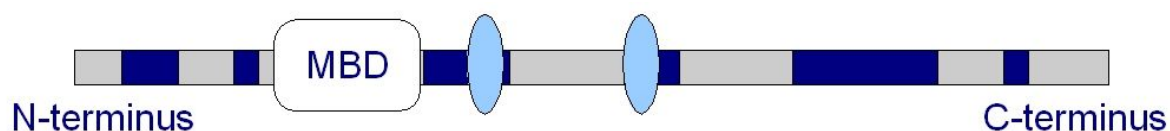


Fig. 26. Graphical representation of MECP2 with the low complexity regions depicted in blue. The MBD is shown in white, while the AT-hook motifs are indicated as turquoise ovals. The low complexity region next to the C-terminal AT-hook corresponds to the histidine-rich region, while the low complexity region further to the right is the proline-rich region in MECP2.

4.2.1 NEFH

NEFH is a member of the neurofilament protein family and is as such composed of a highly conserved alpha-helical core region of approximately 310 amino acids that form double stranded coiled-coils flanked by head (amino)- and tail (carboxy)-domains (Fig. 27).



Fig. 27. Structure of the NEFH protein. The C-terminal tail domain has sequence similarities to MECP2.

NEFH coassembles with "neurofilament, light polypeptide 68kDa" (NEFL) via its core region *in vivo* to form heteropolymers while it binds to neighboring filaments via the tail domain. The amino acid sequence lysine-serine-proline (KSP) is repeated 51 times in the central part of the tail domain. This aa triplet represents a major protein kinase recognition site (Julien *et al.*, 1998). This explains the high degree of phosphorylation of the protein and its large apparent M_r of 200 kDa, as determined by SDS-polyacryl amide gel electrophoresis (Liem *et al.*, 1978).

The tail region in particular has similarities to MECP2. The complete sequence alignment (Annex 6.1) shows a consensus for three KSP motifs between NEFH and MECP2, while the poly-histidine and poly-proline segments present in MECP2 do not occur in NEFH. Also the MBD domain of MECP2 at position 90-165 has few matching aa and two large gaps. Therefore, the KSP phosphorylation sites are the only motifs present in both sequences. Interestingly, the exact phosphorylation site of MECP2 as well as the kinase that adds the phosphate groups remain unknown. Based on this bioinformatics study, the KSP motifs represent candidate phosphorylation sites of MECP2.

Phosphorylation however, as the literature shows, has different effects on the two proteins. While MECP2 becomes inactivated by phosphorylation and consequently dissociates from the *Bdnf* promoter (Chen *et al.*, 2003), phosphorylation of NEFH has been shown to slow down filament transport in the axon (Ackerley *et al.*, 2003).

Finally, a recent publication (Klose and Bird, 2004) suggests, that MECP2 has an elongated shape which is supported by studies on its chicken homologue ARBP (von Kries *et al.*, 1994). This would correspond to the elongated form of NEFH.

4.2.2 MLLT2

There are currently no data on the structure of the *MLLT2* protein AF-4 in the literature. Also database searches did not result in any helpful hints. The only motif according to Pfam is the AF-4 motif (comprising the whole AF-4 protein) which is shared by FMR2. As yet, no

function has been assigned to this AF-4 domain.

Therefore no assumptions on the shape of MECP2 can be made with the help of AF-4. However, AF-4 has been reported to cause ataxia when mutated (Isaacs *et al.*, 2003), a symptom, that is also found in RTT.

4.3 MECP2 target genes

4.3.1 MECP2 regulates *Fkbp5* expression

Our microarray study revealed genes potentially regulated by MECP2 in the mouse brain. The expression of five of these differentially expressed genes was known to be regulated by glucocorticoids. Based on these results, two major questions arose: A) Are these genes direct targets of MECP2 or are the changes in gene expression due to secondary effects of loss of MECP2 function in the brain? B) How do glucocorticoids, known to act as transcription factors, interfere with transcription repression by MECP2?

To address these questions, chromatin immunoprecipitation was established, which allowed to analyze direct interactions of MECP2 with the genomic DNA of these potential targets. *Fkbp5* was chosen to be studied in further detail. Three MECP2 binding regions were revealed (region 1, region 2, and GRE_2), one of which (GRE_2) can also be bound by the GR. Addition of the glucocorticoid dexamethasone to the culture medium of primary neurons induced binding of GR to *Fkbp5* while binding of MECP2 was abolished. In contrast, the glucocorticoid inhibitor RU-486 reverses the effect of dexamethasone and binding of MECP2 to the GRE_2 region is re-established.

A model can therefore be proposed, in which GR and MECP2 control the expression of *Fkbp5* with glucocorticoids increasing and MECP2 decreasing the expression levels of *Fkbp5*. GR and MECP2 compete for the binding to at least on locus in the *Fkbp5* genomic region (i.e. GRE_2). If the concentration of either nuclear GR or MECP2 changes, the expression of *Fkbp5* will be altered (Fig. 28), which in turn leads to changes in GR signaling. This is due to the fact, that FKBP5 itself is involved in GR signaling (Davies *et al.*, 2002). FKBP5 is bound to the GR receptor in the cytoplasm and replaced by FKBP4 after binding of GR to glucocorticoids. This then leads to translocation of the receptor complex into the nucleus and consequent gene expression regulation by GR.

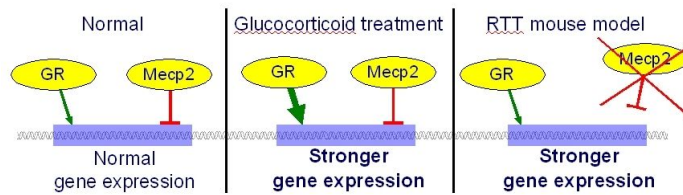


Fig. 28. Model of *Fkbp5* gene regulation. The GR and MECP2 regulate the same gene. Stronger activation due to higher GR levels after glucocorticoid treatment or missing repression due to loss of function of MECP2, both situations lead to a stronger expression of the target gene (blue bar).

It is noteworthy, that MECP2 not only has one binding site in the *Fkbp5* genomic region, but seems to regulate *Fkbp5* gene expression by binding to several loci. The first binding site is in the proximal promoter, the second one near the start of the second exon. The third binding site (GRE_2), however, is in the fifth intron, about 50 kb away from the core promoter. This raises the question how a binding far downstream of the transcription start can influence the expression of the gene. Nevertheless, this regulatory element probably is important since at least for the human *FKBP5* gene it has been shown that expression is regulated by GR via a binding site 2 kb downstream from the promoter (U *et al.*, 2004). In addition, the results of ChIPs with a GR antibody in this thesis prove that the GRE_2 region is a binding site for GR in mouse.

An antibody specific for the C-terminal part of MECP2 was used to perform ChIP. Therefore, binding of both MECP2 isoforms (MECP2e1 and MECP2e2) was detected simultaneously. In follow-up studies the use of antibodies specific for the N-terminal parts of MECP2e1 or MECP2e2 should allow to determine if the two variants bind to the same targets or have different specificities. Antibodies specific for an N-terminal antigen might, however, be difficult to use in chromatin immunoprecipitation since the antigen would be close to the MBD and therefore probably poorly accessible after binding to the DNA and formaldehyde treatment.

4.3.2 Northern blot and quantitative real-time PCR results support the microarray findings

Our collaborators in Edinburgh carried out experiments that substantiate our findings:

To confirm the results obtained by our microarray experiment, Northern blot analyses and

quantitative real-time PCRs were carried out in Edinburgh. Their Northern blot results show the differential expression of these transcripts in *wt* as compared to symptomatic *Mecp2*^{-/-} mice.

For real-time PCR analysis, *Mecp2*-null mice were grouped into three categories: pre-symptomatic, early-symptomatic and late-symptomatic. Male *Mecp2*-null mice acquire neurological symptoms at ~ 6 weeks of age and die at ~ 10 weeks. The postnatal onset of symptoms resembles RTT, as the affected girls develop normal until the age of 6 – 18 months. This raised the question whether the de-regulated expression of the *Sgk* and *Fkbp5* genes seen in late symptomatic mice was also present in animals that had yet to develop symptoms. As there is considerable heterogeneity in the timing of symptom-onset and progression, *Mecp2*-null animals were classified according to symptoms rather than age using the criteria shown in Table 28.

Phenotype	Clasping	Inertia	Tremor	Weight loss	LOC*	Average age
Presymptomatic	-	-	-	-	-	~30d
Early symptomatic	-/+	+	+	-	-	~55d
Late symptomatic	+	+	+	+	+	~ 70d

Table 28. Classification of *Mecp2*-null mice according to the manifestation of phenotype. The last column shows the average age of mice displaying the symptoms. *LOC – loss of condition.

Real-time PCR, relative to *Gapd* as an internal control, revealed that *Sgk* and *Fkbp5* are both up-regulated in pre-symptomatic and early symptomatic mice in comparison to *wt* controls (Fig. 29B). Further analysis of the same cDNA samples established that many other genes show indistinguishable expression levels in mutant and *wt* mice at all three stages (data not shown) in agreement with previous results (Tudor *et al.*, 2002). The de-regulation of *Sgk* and *Fkbp5* genes in mice with no obvious symptoms supports the possibility that they are direct targets for MECP2.

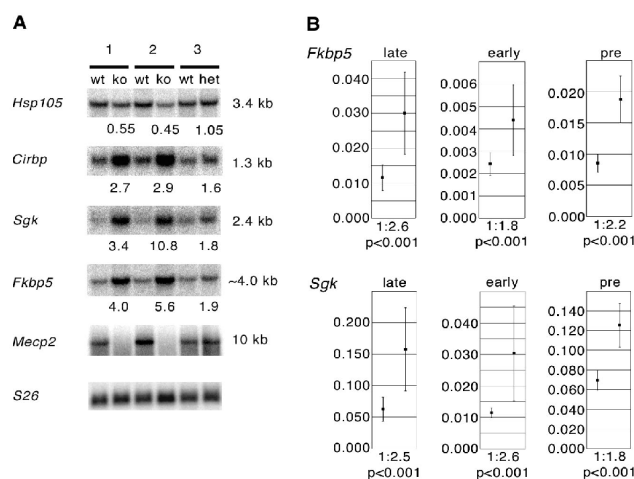


Fig. 29. Northern blot and real-time PCR results confirm microarray data. A) Northern blot analyzes of *Hsp105*, *Cirbp*, *Sgk* and *Fkbp5* confirm the microarray results for these genes. Total brain RNA isolated from *Mecp2*^{-/-} (ko), heterozygote *Mecp2*^{+/-} (het) and wild-type (*wt*) control animals was used. Three litter mate pairs (1, 2, 3) were investigated. The *Mecp2*^{-/-} animals showed advanced neurological symptoms. Intensities of the *Hsp105*, *Cirbp*, *Sgk*, and *Fkbp5* bands were normalized against a loading control

(*S26* ribosomal protein cDNA probe). Expression levels of *wt* animals were arbitrarily set to 1.0 and relative expression levels of the respective ko and het litter mate samples are indicated below the lanes. (B) Real time PCR analyzes of *Fkbp5* and *Sgk* in pre-symptomatic (pre) animals, and animals with early or late symptoms as defined in Table 28. Nine *wt* and nine *Mecp2*-null animals were investigated per stage. A significant up-regulation of both genes was detected in all three investigated stages (see p-values). The mean is shown as a dot; whiskers indicate \pm the standard deviation. The y axes represent arbitrary threshold cycle number converted into expression levels relative to *Gapd* expression.

4.3.3 Up-regulation of glucocorticoid-inducible *Fkbp5* and *Sgk* in *Mecp2*-null animals is not due to elevation of glucocorticoid levels

As hyper-secretion of glucocorticoids has not been shown to induce expression of *Fkbp5* and *Sgk* in the brain, pumps were implanted into *wild-type* mice by Dr. Holmes in Edinburgh. These pumps delivered a corticosterone solution at a constant rate. The effect on transcription of *Fkbp5* and *Sgk* was measured. The experiments were performed by Northern analyses of the mRNA levels in the brains of cortisol treated mice and vehicle treated control animals and the results showed that two days of hormone exposure induced expression of both *Sgk* and *Fkbp5* compared to animals that received the solvent vehicle alone (Fig. 30 A). To determine whether the comparable level of induction of *Sgk* and *Fkbp5* in *Mecp2*-null mouse brain is due to high levels of circulating glucocorticoids, basal and stressed hormone levels were measured in Edinburgh. There was no significant difference between the basal plasma glucocorticoid levels of *wt* and *Mecp2*-null mice (Fig. 30 B), suggesting that elevated hormone levels are not responsible.

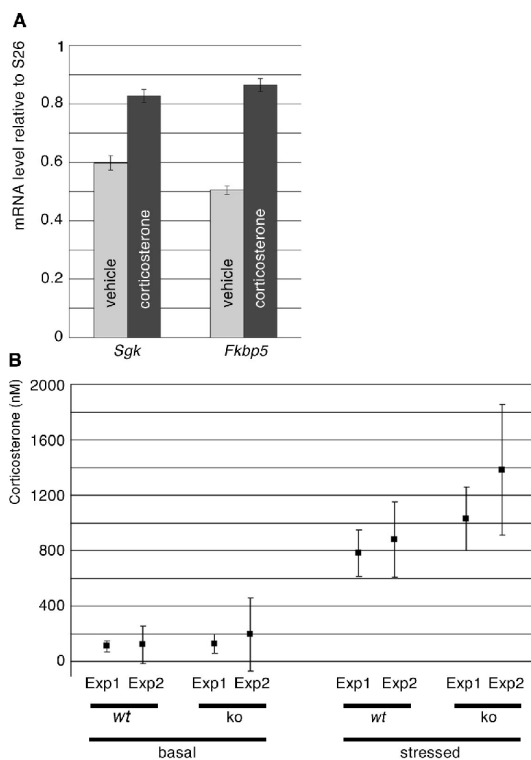


Fig. 30. Hormone-dependent and hormone-independent induction of *Fkbp5* and *Sgk* in mouse brain.

(A) Corticosterone administration induces *Fkbp5* and *Sgk* expression in wt mouse brain. Hormone (n = 6) or hormone-free vehicle (n = 7) were delivered continuously to wt mice for 2 days, after which *Fkbp5* and *Sgk* mRNA levels in mouse brain were measured by Northern blotting. Results were normalized against S26 ribosomal protein mRNA levels on the same blots. (B) Corticosterone levels are not significantly elevated in *Mecp2*-null mice, although *Fkbp5* and *Sgk* are induced. Resting and stressed levels of plasma glucocorticoid were measured in symptomatic *Mecp2*-null mice (ko) and *wild-type* litter mates. Experiment 1 (n = 4) and experiment 2 (n = 5) are shown separately. Boxes depict mean plasma glucocorticoid values and whiskers show \pm the standard deviation. Statistical analysis using Students t-test gave p-values for the comparison *wt* versus ko under basal conditions as 0.6673

(experiment 1 Exp1) and 0.5891 (Exp2). Under stressed conditions, p values were 0.141 (Exp1) and 0.1037 (Exp2). It can be concluded, that there are no significant differences between *wt* and ko hormone levels.

By ChIP analysis - carried out in Edinburgh - no changes in MECP2 binding to the regions fkp and fkp1 could be shown after implantation of corticosterone pumps (see Fig. 31). This suggests, that MECP2 and GR either regulate *Fkbp5* via independent sites or that there is another binding site of MECP2 that can also be bound by GR (see Fig. 23, 24). As shown by the experiments with glucocorticoid-treated primary neurons in this thesis, the GRE_2 region probably is such a binding site. This region has not been studied in mice, treated with corticosterone, so far.

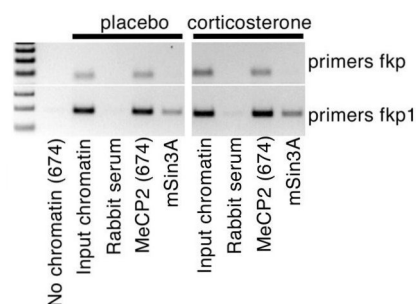


Fig. 31. MECP2 binding to the *Fkbp5* gene is not abolished in *wt* mouse brain upon corticosterone infusion as shown by ChIP with placebo and corticosterone-treated mice. Rabbit serum is used a negative control. mSin3A is a co-repressor complex interacting with MECP2 and appears to bind to the fkp1 site but not to the fkp region.

The data presented in the Results section and the results obtained by our collaborators in Edinburgh support the suggestion that certain downstream consequences of the glucocorticoid pathway are constitutively activated in this mouse model, even though glucocorticoid levels are not increased. Instead, the activation seems to be due to the missing repressor function of MECP2.

The activation of genes that are physiologically induced by glucocorticoids is in line with the observation that mice with different mutations in the *Mecp2* gene reportedly display seizures and heightened anxiety that may be attributed to the inappropriate activation of stress response genes (Chen *et al.*, 2001; Shahbazian *et al.*, 2002).

4.3.4 Rett syndrome: a stress pathway disease - supporting evidence from the literature

Several pieces of evidence suggest that constitutive activation of glucocorticoid-responsive genes may be relevant to our understanding of human RTT. Cortisol levels in Rett syndrome patients appear normal (Echenne *et al.*, 1991; Huppke *et al.*, 2001), but the over-alertness and agitation that may result from hormone-independent activity of stress pathways, including prolonged fits of screaming and crying, are characteristic features of RTT.

Elevated glucocorticoid exposure has profound, deleterious effects on the development of the fetus and neonate, altering cell proliferation, migration, and differentiation (de Kloet *et al.*, 1988). Early postnatal administration of glucocorticoids has been shown to result in an adverse outcome at school age with some similarities to Rett syndrome. Affected children show significantly reduced height and head circumference, poor motor skills and motor coordination, and significantly lower IQ scores (Yeh *et al.*, 2004). Prenatal glucocorticoid exposure also increases anxiety behavior in adult offspring (Welberg *et al.*, 2001) and postnatal exposure alters cerebellar development and function in rodents (Bohn *et al.*, 1980). The decreased dendritic branching and shortened dendrite arbors that have been described in Rett syndrome patients (Armstrong, 2002) are also mimicked by chronic administration of glucocorticoids, leading to dendritic atrophy of hippocampal neurons and dendritic reorganization of prefrontal cortical neurons (Wellman, 2001). In addition, osteopenia frequently occurs in Rett syndrome patients and is a known side-effect of glucocorticoid exposure (Haas *et al.*, 1997; Leonard *et al.*, 1999; Budden and Gunness, 2001).

In summary, many of the phenotypic features of Rett syndrome can be explained by altered expression of the glucocorticoid-regulated genes. Of course *Fkbp5* will not be the only gene

regulated by MECP2 that contributes to the phenotype of the Rett mouse model we used. Rather the sum of many expression changes will constitute the disease.

4.3.5 Repression by MECP2: global and total versus local and partial

The results of the Northern blot, RT-PCR, immunostainings, and ChIP experiments for *Fkbp5* show, that MECP2 does not regulate *Fkbp5* in an "all-or-none" manner, but is rather responsible for the fine tuning of *Fkbp5* expression. Binding of MECP2 therefore might not condense the chromatin completely, but result in a chromatin state that makes transcription more difficult. Such an attenuating function of MECP2 has already been shown for the *Bdnf* gene (Chen *et al.*, 2003, Martinowich *et al.*, 2003).

If this model holds true for many other MECP2 targets, this would explain the difficulties scientists have had with microarray studies for Rett syndrome. An attenuating function of MECP2 would lead to small changes in gene expression which are hard to measure by microarray experiments. Tudor and colleagues could only detect subtle transcriptional changes using different brain samples (Tudor *et al.*, 2002) and the number of differentially expressed genes found in this study is much lower than what would be expected for a transcription factor with an extremely abundant binding site such as MECP2.

4.3.6 DNA microarray studies and chromatin immunoprecipitation

The results of this thesis show, that the combination of DNA microarray analysis and chromatin immunoprecipitation are extremely useful tools to find target genes of transcription factors as well as the corresponding binding sites. The disadvantage is, that both techniques are not trivial require laborious optimization. Some critical factors for chromatin immunoprecipitation have been outlined in the results part.

This study also highlights, that the appropriate biological material has to be chosen to obtain meaningful results. Only by using RNA from brain tissue of severely symptomatic *Mecp2^{-y}* mice for the microarray study and appropriate brain tissue as well as primary neurons for the ChIP, *Fkbp5* could be detected as a direct target of MECP2. Microarray studies carried out with primary fibroblasts, lymphoblastoid cells and postmortem human brain did not find *Fkbp5* to be differentially expressed. The binding pattern of MECP2 in fibroblasts and lymphoblastoid cells might be totally different from that in the brain. Even the more so if MECP2 only has an attenuating function.

In that perspective it would be interesting to hybridize the DNA fragments obtained by chromatin immunoprecipitation to a high-resolution genomic microarray covering the whole genome. This would allow to identify many more, possibly all MECP2 targets in one single experiment and enable a comparison of its binding pattern in different cell types and cellular states possible.

4.3.7 Future studies

This study suggests that enhanced glucocorticoid signaling occurs in RTT patients and may be therapeutically addressed by antagonists of the glucocorticoid pathway, e.g. at the GR level. To further analyze the interplay of MECP2 and the glucocorticoid receptor target genes, studies in animal models should be carried out in the future:

1. Administration of glucocorticoid pathway inhibitors to Rett syndrome animal models should reverse at least part of the phenotype.
2. If the activation of the glucocorticoid genes is due to loss of function of MECP2, a mouse GR-null mutant crossed to the *Mecp2*^{-y} strain should reverse the phenotype.
3. It will be interesting to see at what time point in development the regulation of *Fkbp5* by MECP2 starts. The answer should shed some light on the course of the disease and how long it takes from the start of the dysregulation until the first symptoms appear.
4. Hybridization of the products of a ChIP with an anti-MECP2 antibody to a high-resolution whole-genome microarray would allow to find additional genes and pathways involved in the pathogenesis of Rett syndrome and might also resolve the questions why MECP2 mutations have a brain-specific phenotype.

4.4 Conclusions

Three aspects of MECP2, its function and its role in Rett syndrome have been studied in this thesis.

- Six new MBD family proteins with possible functional redundancy to MECP2 were found. Five of these proteins (SETDB1, SETDB2, BAZ2A, MBD5, and MBD6) are candidates to compensate for the loss of function of MECP2 outside the brain in Rett patients.
- Database analysis revealed homology of MECP2 to several polypeptides in the human proteome. The best matches were NF-H and MLLT2/AF4. A closer investigation of the homologous regions indicated, that overall MECP2 might have an elongated structure and revealed putative phosphorylation sites of MECP2.
- Finally, data presented in this thesis suggest, that downstream targets of the stress pathway might be involved in the pathology of Rett syndrome. More specifically, MECP2 regulates some genes that can be induced by glucocorticoids. Loss of function therefore leads to an increase in expression of these genes, which in turn may explain many features of the Rett syndrome phenotype. This gives rise to the hope, that inhibitors of the glucocorticoid pathway could constitute a treatment for Rett syndrome patients.

5. Zusammenfassung

Rett Syndrom, eine schwere Form der geistigen Behinderung, wird durch Mutationen in dem X-chromosomalen Gen des Methyl-CpG bindenden Proteins 2 (*MECP2*) verursacht und kommt fast ausschließlich bei Mädchen vor. Mutationen in *MECP2* wurden erstmals 1999 beschrieben, die molekularen Mechanismen, welche der Krankheit zugrunde liegen, sind bisher aber unbekannt. Eine bekannte Funktion von *MECP2* ist die der Transkription-Repression.

Lediglich zwei neuronale Zielgene (*Bdnf* und *Dlx5*) wurden jedoch bisher beschrieben. Obwohl *MECP2* ubiquitär exprimiert ist, besteht bei Rett Syndrom Patienten ein primär neuronaler Phänotyp. Diese Beobachtung lässt die Vermutung zu, dass *MECP2* im Gehirn eine wichtige Funktion erfüllt, während in peripheren Geweben der Funktionsverlust von *MECP2* durch ein funktionell redundantes Protein kompensiert werden könnte.

Um Proteine zu finden, die eine solche kompensatorische Funktion übernehmen könnten, wurden zwei Strategien angewandt. Im ersten Projekt wurde mit bioinformatischen Mitteln nach Proteinen gesucht, die ebenso wie *MECP2* eine methyl-CpG bindende Domäne (MBD) besitzen. Sechs solcher Proteine wurden gefunden und auf ihre Expression und Domänenstruktur hin untersucht.

Das zweite Projekt zielte darauf ab, Proteine zu identifizieren, die eine globale Ähnlichkeit zu *MECP2* aufweisen. Solche Proteine sollten Rückschlüsse auf die Form und eventuell sogar auf bisher unbekannte Funktionen von *MECP2* zulassen. Zwei solche paraloge Polypeptide wurden gefunden und die Struktur eines dieser, NEFH, lässt vermuten, dass *MECP2* eine längliche Form hat. Der Sequenzvergleich zwischen NEFH und *MECP2* deutet zudem auf eine potentielle Phosphorylierungsstelle in *MECP2* hin.

Um Zielgene von *MECP2* im Gehirn zu finden, wurde die Chromatinimmunpräzipitation (ChIP) etabliert und mit cDNS-Mikroarray Untersuchungen eines Mausmodells für Rett Syndrom kombiniert. Die Auswertung der Mikroarray-Daten zeigte, dass in diesen Tieren mehrere Gene differenziell exprimiert sind, die normalerweise während der Stressantwort durch Glukokortikoide reguliert werden. Erhöhte mRNA-Werte konnten für die "Plasma Glukokortikoid-induzierbare Kinase 1" (*Sgk*) sowie für das "FK506-bindende Protein 51" (*Fkbp5*) nachgewiesen werden. Durch Immunfärbung histologischer Gehirnschnitte, konnte gezeigt werden, dass *Mecp2* und *Fkbp5* sowie *Sgk* in denselben Zellen des Gehirns

synthetisiert werden. Dies legt nahe, dass MECP2 eher eine Modulation der Expression bewirkt, als eine komplette Repression.

Mittels ChIP konnten drei Bindungsstellen von MECP2 in der genomischen Region von *Fkbp5* gefunden werden. Eine dieser Bindungsstellen kann zudem auch vom Glukokortikoid-Rezeptor (GR) gebunden werden. Aufgrund dieser Ergebnisse wurde eine Hypothese formuliert, nach der MECP2 und GR um die Bindung an eine Region im *Fkbp5* Gen konkurrieren und beide die Expression von *FKBP5* regulieren. In Rett Syndrom Patienten ist diese Regulation auf Grund des Funktionsverlusts von MECP2 gestört. Dies würde zu einer Überexpression von Glukokortikoid-regulierten Genen führen, wodurch mehrere Merkmale des Rett Syndrom-Phänotyps erklärt werden könnten.

6. Annex

6.1 Sequence alignment of MECP2 and NEFH

```

MECP2 -----
NEFH MMSFGGADALLGAPFAPLHGGSLLHYALARKGGAGGTRSAAGSSSGFHSWTRTSVSSVSA

MECP2 -----
NEFH SPSRFRGAGAASSTDSLDTLSNGPEGCMVAVATSRSEKEQLQALNDRFAGYIDKVRQLEA

MECP2 -----
NEFH HNRSLGEAAALRQQQAGRSAMGELYEREVREMRGAVLRLGAARGQLRLEQEHLLEDIAH

MECP2 -----
NEFH VRQRLDDEARQREEEAARALARFAQEAEAAARVDLQKKAQALQECCGYLRRHHQEEVGE

MECP2 -----
NEFH LLGQIQGSGAAQMQAETRDALKCDVTSALREIRAQLEGHAVQSTLQSEEFWRVRLDRL

MECP2 -----MVAGMLGLREEKSEDQLQGLKDKPLKFKKVKKD-----
NEFH SEAAKVNTDAMRSAQEEITEYRRQLQARTTELEALKSTKDSLERQRSELEDRHQADIASY
      . * . : ** : * :      : . * : * . **

MECP2 -----
NEFH QEAIQQLDAELRNTKWEMAAQLREYQDLLNVKMALDIEIAAYRKLLEGECCRIGFGPIPF

MECP2 -----KKEEKEGKH
NEFH SLPEGLPKIPSVSTHIKVKSEEKIKVVEKSEKETVIVEEQTEETQVTEEVTEEEKEAKE
      : : ** * . * .

MECP2 -----
NEFH EPVQPSAHSAPAEAGKAETSESGSAPAVPEASAS--PKQRRSIIIRDRGPMYDDPTL
      * : . . * ** * : ** . . . . * ** . * . : : * . : . * : . .

MECP2 -----
NEFH PEGWTRKLRKQKRSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVT
      PAEVKSPKAKSPAKEEAKSPPEAKSPEKEEAKSPAEVKSPKAKSPAKEEAKSPA
      * . * : . . . . * . * : : * * : : : . : . : . . : :

MECP2 -----
NEFH GR-----GSPSRREQPKPKPKSPKAP--
      PEKAKSPVKEEAKSPAEEAKSPVKEEAKSPAEVKSPKAKSPKAKSPAEEAKSPA
      . * : : * * : * : * : * : * : * : * : * : * : * : * : * : * :

MECP2 -----
NEFH GTGRGRGRPKGSGTTRPKAATSEGVQKRVLE--KSPGKLLVKMP-FQTSPPGKAEGGGA
      KSPEKAKSPVKAEAKSPEKAKSPVKAEAKSPEKAKSPVKEEAKSPEKAKSPVKEEAKSPE
      : . * : : . * : * * : * * * * * * * * * * * * : .

MECP2 -----
NEFH TTSTQVMVIKRPGRKRKAEADPQAI PKRGRKPGSVVAAAAAEAKKAVKKESSIRSVQET
      KAKSPVKEEAKTPEKAKSPVKEEAKSPEKAKSPEKAKTLDVKSPEAKTPEAKKEARS
      . : . : * . . . * * : . . : * . : . . . * . . : . . : * : . . . * * : .

MECP2 -----
NEFH VLPKIKRKTRETVSIEVKEVVKPLLVLSTLGEKSGKGLKTKSPGR-----KSKESSP
      FPEKAKSPVKEEAKSPEKAKSPLKEDAKAPEKEIPKKEEVKSPVKEEAKPQEVKVKPEPP
      . * . : * * . * . : . . * * . : * * : * * : * * : * * .

MECP2 -----
NEFH KGRSSSASSPPKKEHHHHHHHSESPKAPVP---LLPPLPPPPPEPESSEDPTSPPEPQDL
      KAVEEKAPATPKTEKKDSKKEEAPKKEAPKPKVEEKKEPAVEKPKESKVEAKKEAEDK
      * . . . . * . : . * * . : . : * : * * . * : * . : * : * : . . * : *

MECP2 -----
NEFH SSSVCKEEKMPRGGSLES DGCPEPAKTQPAVATAATAAEKYKHRGEG---ERKDIVSS
      KKVPTPEKEAPAKVEVKEDAKPKKEKTEVAKKEPDDAKAKEPSKPAEKKEAAPEKDKTKEE
      . . * : : * . : . . * . * * * : . . * * * * * : * : * * .

MECP2 -----
NEFH SMPRPNRE-----EPVDSRTPVTERVS-----
      KAKKPEEKPKTEAKAKEDDKTLSKEPSKPKAEKAEKSSSTDQKDSKPKPEKATEDKAAKGG
      . : * : : * * . . : : * : *

```

7. References

- Ackerley, S., Thornhill, P., Grierson, A. J., Brownlees, J., Anderton, B. H., Leigh, P. N., Shaw, C. E. and Miller, C. C. (2003) *J Cell Biol*, 161, 489-95.
- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U. and Zoghbi, H. Y. (1999) *Nat Genet*, 23, 185-8.
- Amir, R. E., Van den Veyver, I. B., Schultz, R., Malicki, D. M., Tran, C. Q., Dahle, E. J., Philippi, A., Timar, L., Percy, A. K., Motil, K. J., Lichtarge, O., Smith, E. O., Glaze, D. G. and Zoghbi, H. Y. (2000) *Ann Neurol*, 47, 670-9.
- Andersen, J. B., Strandbygard, D. J., Hartmann, R. and Justesen, J. (2004) *Eur J Biochem*, 271, 628-36.
- Antequera, F. and Bird, A. (1993) *Proc Natl Acad Sci U S A*, 90, 11995-9.
- Armstrong, D. D. (2002) *Ment Retard Dev Disabil Res Rev*, 8, 72-6.
- Bakker, J., Lin, X. and Nelson, W. G. (2002) *J Biol Chem*, 277, 22573-80.
- Ballestar, E. and Wolffe, A. P. (2001) *Eur J Biochem*, 268, 1-6.
- Ballestar, E., Ropero, S., Alaminos, M., Armstrong, J., Setien, F., Agrelo, R., Fraga, M. F., Herranz, M., Avila, S., Pineda, M., Monros, E. and Esteller, M. (2005) *Hum Genet*, 116, 91-104.
- Bhattacharya, S. K., Ramchandani, S., Cervoni, N. and Szyf, M. (1999) *Nature*, 397, 579-83.
- Birch, H. G., Richardson, S. A., Baird, D., Horobin, G. and Ilesley, R. (1970) Mental subnormality in the community: a clinical and epidemiological study. Baltomier MD: Williams and Wilkins.
- Bird, A. (2002) *Genes Dev*, 16, 6-21.
- Blencowe, B. J., Bauren, G., Eldridge, A. G., Issner, R., Nickerson, J. A., Rosonina, E. and Sharp, P. A. (2000) *Rna*, 6, 111-20.
- Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U. and Renkawitz, R. (2000) *J Biol Chem*, 275, 34963-7.
- Bohn, M. C. and Lauder, J. M. (1980) *Dev Neurosci*, 3, 81-9.
- Brackertz, M., Boeke, J., Zhang, R. and Renkawitz, R. (2002) *J Biol Chem*, 277, 40958-66.
- Brandt, N. R., Caswell, A. H., Carl, S. A., Ferguson, D. G., Brandt, T., Brunschwig, J. P. and Bassett, A. L. (1993) *J Membr Biol*, 131, 219-28.

- Budden, S. S. and Gunness, M. E. (2001) *Brain Dev*, 23 Suppl 1, S133-7.
- Campbell, P. M., Bovenzi, V. and Szyf, M. (2004) *Carcinogenesis*, 25, 499-507.
- Cedar, H., Stein, R., Gruenbaum, Y., Naveh-Many, T., Sciaky-Gallili, N. and Razin, A. (1983) *Cold Spring Harb Symp Quant Biol*, 47 Pt 2, 605-9.
- Cervoni, N. and Szyf, M. (2001) *J Biol Chem*, 276, 40778-87.
- Chandler, S. P., Guschin, D., Landsberger, N. and Wolffe, A. P. (1999) *Biochemistry*, 38, 7008-18.
- Cheadle, J. P., Gill, H., Fleming, N., Maynard, J., Kerr, A., Leonard, H., Krawczak, M., Cooper, D. N., Lynch, S., Thomas, N., Hughes, H., Hulten, M., Ravine, D., Sampson, J. R. and Clarke, A. (2000) *Hum Mol Genet*, 9, 1119-29.
- Chen, R. Z., Akbarian, S., Tudor, M. and Jaenisch, R. (2001) *Nat Genet*, 27, 327-31.
- Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R. and Greenberg, M. E. (2003) *Science*, 302, 885-9.
- Clayton-Smith, J., Watson, P., Ramsden, S. and Black, G. C. (2000) *Lancet*, 356, 830-2.
- Colantuoni, C., Jeon, O. H., Hyder, K., Chenchik, A., Khimani, A. H., Narayanan, V., Hoffman, E. P., Kaufmann, W. E., Naidu, S. and Pevsner, J. (2001) *Neurobiol Dis*, 8, 847-65.
- Collard, J. F., Cote, F. and Julien, J. P. (1995) *Nature*, 375, 61-4.
- Collins, A. L., Levenson, J. M., Vilaythong, A. P., Richman, R., Armstrong, D. L., Noebels, J. L., David Sweatt, J. and Zoghbi, H. Y. (2004) *Hum Mol Genet*, 13, 2679-89.
- Coy, J. F., Sedlacek, Z., Bachner, D., Delius, H. and Poustka, A. (1999) *Hum Mol Genet*, 8, 1253-62.
- D'Esposito, M., Quaderi, N. A., Ciccodicola, A., Bruni, P., Esposito, T., D'Urso, M. and Brown, S. D. (1996) *Mamm Genome*, 7, 533-5.
- Das, P. M., Ramachandran, K., vanWert, J. and Singal, R. (2004) *Biotechniques*, 37, 961-9.
- Davies, T. H., Ning, Y. M., Sanchez, E.R. (2002) *J Biol Chem*, 277, 4597-600.
- De Bona, C., Zappella, M., Hayek, G., Meloni, I., Vitelli, F., Bruttini, M., Cusano, R., Loffredo, P., Longo, I. and Renieri, A. (2000) *Eur J Hum Genet*, 8, 325-30.
- De Kloet, E. R., De Kock, S., Schild, V. and Veldhuis, H. D. (1988) *Neuroendocrinology*, 47, 109-15.
- Detich, N., Theberge, J. and Szyf, M. (2002) *J Biol Chem*, 277, 35791-4.
- Di Fiore, B., Palena, A., Felsani, A., Palitti, F., Caruso, M. and Lavia, P. (1999) *Nucleic Acids*

- Res*, 27, 2852-9.
- Doerks, T., Copley, R. and Bork, P. (2001) *Trends Biochem Sci*, 26, 145-6.
- Drewell, R. A., Goddard, C. J., Thomas, J. O. and Surani, M. A. (2002) *Nucleic Acids Res*, 30, 1139-44.
- Duanmu, Z., Kocarek, T. A. and Runge-Morris, M. (2001) *Drug Metab Dispos*, 29, 1130-5.
- Echenne, B., Bressot, N., Bassir, M., Daures, J. P. and Rabinowitz, A. (1991) *J Child Neurol*, 6, 257-62.
- El-Osta, A., Wolffe, A. P. (2001) *Biochem Biophys Res Commun*, 289, 733-7.
- El-Osta, A., Kantharidis, P., Zalberg, J. R. and Wolffe, A. P. (2002) *Mol Cell Biol*, 22, 1844-57.
- Evans, J. C., Archer, H. L., Whatley, S. D., Kerr, A., Clarke, A., Butler, R. (2005) *Eur J Hum Genet*, 13, 124-6.
- Figlewicz, D. A., Krizus, A., Martinoli, M. G., Meininger, V., Dib, M., Rouleau, G. A. and Julien, J. P. (1994) *Hum Mol Genet*, 3, 1757-61.
- Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A. and Esteller, M. (2003) *Nucleic Acids Res*, 31, 1765-74.
- Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H. and Nakao, M. (1999) *Mol Cell Biol*, 19, 6415-26.
- Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T. and Nakao, M. (2003a) *J Biol Chem*, 278, 24132-8.
- Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H. and Nakao, M. (2003b) *Mol Cell Biol*, 23, 2834-43.
- Fujita, H., Fujii, R., Aratani, S., Amano, T., Fukamizu, A. and Nakajima, T. (2003) *Mol Cell Biol*, 23, 2645-57.
- Fukamachi, S., Shimada, A., Shima, A. (2001) *Nat Genet*, 28, 381-5.
- Fuks, F., Hurd, P. J., Deplus, R. and Kouzarides, T. (2003a) *Nucleic Acids Res*, 31, 2305-12.
- Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P. and Kouzarides, T. (2003b) *J Biol Chem*, 278, 4035-40.
- Fukue, Y., Sumida, N., Nishikawa, J. and Ohyama, T. (2004) *Nucleic Acids Res*, 32, 5834-40.
- Ghoshal, K., Majumder, S., Datta, J., Motiwala, T., Bai, S., Sharma, S. M., Frankel, W. and Jacob, S. T. (2004) *J Biol Chem*, 279, 6783-93.
- Gregory, R. I., Randall, T. E., Johnson, C. A., Khosla, S., Hatada, I., O'Neill, L. P., Turner, B.

- M. and Feil, R. (2001) *Mol Cell Biol*, 21, 5426-36.
- Guy, J., Hendrich, B., Holmes, M., Martin, J. E. and Bird, A. (2001) *Nat Genet*, 27, 322-6.
- Haas, J. P., Hoper, K., Leipold, G., Dorr, H. G. and Hoper, J. (1997) *Adv Exp Med Biol*, 428, 409-13.
- Hagberg, B., Aicardi, J., Dias, K. and Ramos, O. (1983) *Ann Neurol*, 14, 471-9.
- Hagberg, B. A. and Skjeldal, O. H. (1994) *Pediatr Neurol*, 11, 5-11.
- Hamel, B. C., Chiurazzi, P. and Lubs, H. A. (2000) *Am J Med Genet*, 94, 361-3.
- Hanashima, C., Li, S. C., Shen, L., Lai, E. and Fishell, G. (2004) *Science*, 303, 56-9.
- Hattori, N., Abe, T., Suzuki, M., Matsuyama, T., Yoshida, S., Li, E. and Shiota, K. (2004) *Genome Res*, 14, 1733-40.
- Hendrich, B. and Bird, A. (1998) *Mol Cell Biol*, 18, 6538-47.
- Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S. and Bird, A. (1999) *Mamm Genome*, 10, 906-12.
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. and Bird, A. (2001) *Genes Dev*, 15, 710-23.
- Hendrich, B. and Tweedie, S. (2003) *Trends Genet*, 19, 269-77.
- Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., Samid, D., Duan, D. S., Gnarr, J. R., Linehan, W. M. and et al. (1994) *Proc Natl Acad Sci U S A*, 91, 9700-4.
- Herman, J. G., Jen, J., Merlo, A. and Baylin, S. B. (1996) *Cancer Res*, 56, 722-7.
- Hermann, A., Schmitt, S. and Jeltsch, A. (2003) *J Biol Chem*, 278, 31717-21.
- Holmgren, C., Kanduri, C., Dell, G., Ward, A., Mukhopadhyaya, R., Kanduri, M., Lobanenkova, V. and Ohlsson, R. (2001) *Curr Biol*, 11, 1128-30.
- Honeycutt, A., Dunlap, L., Chen, H. and al Homsy, G. (2003) In *Mortality and morbidity weekly report* CDC, Centers for Disease Control and Prevention, Atlanta.
- Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. and Vingron, M. (2002) *Bioinformatics*, 18 Suppl 1, S96-104.
- Huppke, P., Laccone, F., Kramer, N., Engel, W. and Hanefeld, F. (2000) *Hum Mol Genet*, 9, 1369-75.
- Huppke, P., Roth, C., Christen, H. J., Brockmann, K. and Hanefeld, F. (2001) *Acta Paediatr*, 90, 1257-61.
- Isaacs, A. M., Oliver, P. L., Jones, E. L., Jeans, A., Potter, A., Hovik, B. H., Nolan, P. M., Vizor, L., Glenister, P., Simon, A. K., Gray, I. C., Spurr, N. K., Brown, S. D., Hunter,

- A. J. and Davies, K. E. (2003) *J Neurosci*, 23, 1631-7.
- Jiang, C. L., Jin, S. G., Lee, D. H., Lan, Z. J., Xu, X., O'Connor, T. R., Szabo, P. E., Mann, J. R., Cooney, A. J. and Pfeifer, G. P. (2002) *Genomics*, 80, 621-9.
- Jiang, C. L., Jin, S. G. and Pfeifer, G. P. (2004) *J Biol Chem*, 279, 52456-64.
- Johnston, H., Kneer, J., Chackalaparampil, I., Yaciuk, P. and Chrivia, J. (1999) *J Biol Chem*, 274, 16370-6.
- Jones, M. H., Hamana, N., Nezu, J. and Shimane, M. (2000) *Genomics*, 63, 40-5.
- Jones, P. L. and Shi, Y. B. (2003) *Curr Top Microbiol Immunol*, 274, 237-68.
- Julien, J. P. and Mushynski, W. E. (1998) *Prog Nucleic Acid Res Mol Biol*, 61, 1-23.
- Kaludov, N. K. and Wolffe, A. P. (2000) *Nucleic Acids Res*, 28, 1921-8.
- Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) *Nucleic Acids Res*, 30, 42-6.
- Kano, H., Arakawa, Y., Takahashi, J. A., Nozaki, K., Kawabata, Y., Takatsuka, K., Kageyama, R., Ueba, T. and Hashimoto, N. (2004) *Biochem Biophys Res Commun*, 317, 902-8.
- Kim, G. D., Ni, J., Kelesoglu, N., Roberts, R. J. and Pradhan, S. (2002) *Embo J*, 21, 4183-95.
- Kimura, H. and Shiota, K. (2003) *J Biol Chem*, 278, 4806-12.
- Kishino, T., Lalonde, M. and Wagstaff, J. (1997) *Nat Genet*, 15, 70-3.
- Klose, R. J. and Bird, A. P. (2004) *J Biol Chem*, 279, 46490-6.
- Ko, T. K., Kelly, E. and Pines, J. (2001) *J Cell Sci*, 114, 2591-603.
- Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C. and Ishii, S. (2001) *J Biol Chem*, 276, 34115-21.
- Koolen, D. A., Vissers, L. E., Nillesen, W., Smeets, D., van Ravenswaaij, C. M., Sistermans, E. A., Veltman, J. A. and de Vries, B. D. (2004) *Clin Genet*, 65, 429-32.
- Kriaucionis, S. and Bird, A. (2004) *Nucleic Acids Res*, 32, 1818-23.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard,

- T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissole, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) *Nature*, 409, 860-921.
- Lang, F. and Cohen, P. (2001) *Sci STKE*, 2001, RE17.
- LaSalle, J. M., Goldstine, J., Balmer, D. and Greco, C. M. (2001) *Hum Mol Genet*, 10, 1729-40.
- Lautier, C., El Mkaïdem, S. A., Renard, E., Brun, J. F., Gris, J. C., Bringer, J. and Grigorescu, F. (2003) *Hum Genet*, 113, 34-43.
- Laxova, R., Ridlre, M. A. C. and Bowen-Bravery, M. (1977) *Am J Med Genet*, 1, 75-86.
- Lee, Y. H. and White, M. F. (2004) *Arch Pharm Res*, 27, 361-70.
- Lehner, B., Semple, J. I., Brown, S. E., Counsell, D., Campbell, R. D. and Sanderson, C. M. (2004) *Genomics*, 83, 153-67.
- Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C. B. and Chiariotti, L. (2003) *Mol Cell Biol*, 23, 1656-65.
- Leonard, H., Thomson, M. R., Glasson, E. J., Fyfe, S., Leonard, S., Bower, C., Christodoulou, J. and Ellaway, C. (1999) *Dev Med Child Neurol*, 41, 323-8.
- Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) *Cell*, 69, 905-14.
- Li, J. and Vogt, P. K. (1993) *Proc Natl Acad Sci U S A*, 90, 4490-4.
- Liem, R. K., Yen, S. H., Salomon, G. D. and Shelanski, M. L. (1978) *J Cell Biol*, 79, 637-645.
- Liu, Y. J., Lu, S. H., Xu, B., Yang, R. C., Ren, Q., Liu, B., Li, B., Lu, M., Yan, F. Y., Han, Z. B. and Han, Z. C. (2004) *Blood*, 103, 4449-56.
- Lubs, H., Chiurazzi, P., Arena, J., Schwartz, C., Tranebjaerg, L. and Neri, G. (1999) *Am J Med Genet*, 83, 237-47.
- Luikenhuis, S., Giacometti, E., Beard, C. F. and Jaenisch, R. (2004) *Proc Natl Acad Sci U S A*, 101, 6033-8.

- Mabuchi, H., Fujii, H., Calin, G., Alder, H., Negrini, M., Rassenti, L., Kipps, T. J., Bullrich, F. and Croce, C. M. (2001) *Cancer Res*, 61, 2870-7.
- Macleod, D., Ali, R. R. and Bird, A. (1998) *Mol Cell Biol*, 18, 4433-43.
- Mammarella, S., Romano, F., Di Valerio, A., Creati, B., Esposito, D. L., Palmirota, R., Capani, F., Vitullo, P., Volpe, G., Battista, P., Della Loggia, F., Mariani-Costantini, R. and Cama, A. (2000) *Hum Mol Genet*, 9, 2517-21.
- Marhold, J., Kramer, K., Kremmer, E. and Lyko, F. (2004) *Development*, 131, 6033-9.
- Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. and Sun, Y. E. (2003) *Science*, 302, 890-3.
- Marty, I. (2004) *Cell Mol Life Sci*, 61, 1850-3.
- Mattei, M. G., Dautigny, A., Pham-Dinh, D., Passage, E., Mattei, J. F. and Jolles, P. (1988) *Hum Genet*, 80, 293-5.
- Meloni, I., Bruttini, M., Longo, I., Mari, F., Rizzolio, F., D'Adamo, P., Denvriendt, K., Fryns, J. P., Toniolo, D. and Renieri, A. (2000) *Am J Hum Genet*, 67, 982-5.
- Merberg, D. M., Fitz, L. J., Temple, P., Giannotti, J., Murtha, P., Fitzgerald, M., Scaltreto, H., Kelleher, K., Preissner, K., Kriz, R., Jacobs, K., and Turner, K. (1993) *Elsevier Science, B. V., Biology of Vitronectins and Their Receptors*, 45-53.
- Mnatzakanian, G. N., Lohi, H., Munteanu, I., Alfred, S. E., Yamada, T., MacLeod, P. J., Jones, J. R., Scherer, S. W., Schanen, N. C., Friez, M. J., Vincent, J. B. and Minassian, B. A. (2004) *Nat Genet*, 36, 339-41.
- Mohler, P. J., Splawski, I., Napolitano, C., Bottelli, G., Sharpe, L., Timothy, K., Priori, S. G., Keating, M. T. and Bennett, V. (2004) *Proc Natl Acad Sci U S A*, 101, 9137-42.
- Monroy, M. A., Schott, N. M., Cox, L., Chen, J. D., Ruh, M. and Chrivia, J. C. (2003) *Mol Endocrinol*, 17, 2519-28.
- Morgan, H. D., Dean, W., Coker, H. A., Reik, W. and Petersen-Mahrt, S. K. (2004) *J Biol Chem*, 279, 52353-60.
- Nagase, T., Kikuno, R. and Ohara, O. (2001a) *DNA Res*, 8, 179-87.
- Nagase, T., Nakayama, M., Nakajima, D., Kikuno, R. and Ohara, O. (2001b) *DNA Res*, 8, 85-95.
- Nan, X., Meehan, R. R. and Bird, A. (1993) *Nucleic Acids Res*, 21, 4886-92.
- Nan, X., Tate, P., Li, E. and Bird, A. (1996) *Mol Cell Biol*, 16, 414-21.
- Nan, X., Campoy, F. J. and Bird, A. (1997) *Cell*, 88, 471-81.

- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A. (1998) *Nature*, 393, 386-9.
- Neul, J. L. and Zoghbi, H. Y. (2004) *Neuroscientist*, 10, 118-28.
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. and Bird, A. (1999) *Nat Genet*, 23, 58-61.
- Ng, H. H., Jeppesen, P. and Bird, A. (2000) *Mol Cell Biol*, 20, 1394-406.
- Nomura, N., Nagase, T., Miyajima, N., Sazuka, T., Tanaka, A., Sato, S., Seki, N., Kawarabayasi, Y., Ishikawa, K., Tabata, S. (1994) *DNA Res*, 1, 251-262.
- Ohki, I., Shimotake, N., Fujita, N., Nakao, M. and Shirakawa, M. (1999) *Embo J*, 18, 6653-61.
- Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M. and Shirakawa, M. (2001) *Cell*, 105, 487-97.
- Okano, M., Xie, S. and Li, E. (1998) *Nucleic Acids Res*, 26, 2536-40.
- Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999) *Cell*, 99, 247-57.
- Orrico, A., Lam, C., Galli, L., Dotti, M. T., Hayek, G., Tong, S. F., Poon, P. M., Zappella, M., Federico, A. and Sorrentino, V. (2000) *FEBS Lett*, 481, 285-8.
- Patra, S. K., Patra, A., Zhao, H., Carroll, P. and Dahiya, R. (2003) *Biochem Biophys Res Commun*, 302, 759-66.
- Percy, A. K. and Lane, J. B. (2004) *Curr Opin Pediatr*, 16, 670-7.
- Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., Yeung, A. T., Matsumoto, Y. and Bellacosa, A. (2000) *J Biol Chem*, 275, 32422-9.
- Prader, A., Labhart, A., Willi, H. (1956) *Schweiz Med Wschr*, 86, 1260-1261.
- Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A. and Prokhortchouk, E. (2001) *Genes Dev*, 15, 1613-8.
- Qiu, C., Sawada, K., Zhang, X. and Cheng, X. (2002) *Nat Struct Biol*, 9, 217-24.
- Quaderi, N. A., Meehan, R. R., Tate, P. H., Cross, S. H., Bird, A. P., Chatterjee, A., Herman, G. E. and Brown, S. D. (1994) *Genomics*, 22, 648-51.
- Ramchandani, S., Bhattacharya, S. K., Cervoni, N. and Szyf, M. (1999) *Proc Natl Acad Sci U S A*, 96, 6107-12.
- Ramsahoye, B. H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A. P. and Jaenisch, R. (2000) *Proc Natl Acad Sci U S A*, 97, 5237-42.
- Reichardt, H. M. and Schutz, G. (1998) *Mol Cell Endocrinol*, 146, 1-6.

- Reichwald, K., Thiesen, J., Wiehe, T., Weitzel, J., Poustka, W. A., Rosenthal, A., Platzer, M., Stratling, W. H., Kioschis, P. (2000) *Mamm Genome*, 11, 182-90.
- Renieri, A., Meloni, I., Longo, I., Ariani, F., Mari, F., Pescucci, C. and Cambi, F. (2003) *J Mol Med*, 81, 346-54.
- Rett, A. (1966) *Wien Med Wochenschr*, 116, 723-6.
- 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) *Nature*, 416, 552-6.
- Roloff, T. C., Nuber, U. A. (2005) *Eur J Cell Biol*, (in press)
- Roloff, T. C., Ropers, H. H. and Nuber, U. A. (2003) *BMC Genomics*, 4, 1.
- Ropers, H. H. and Hamel, B. C. (2005) *Nat Rev Genet*, 6, 46-57.
- Russo, V. E. A., Martienssen, R. A. and Riggs, A. D. (1996) *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press, Plainview, NY.
- Saito, M. and Ishikawa, F. (2002) *J Biol Chem*, 277, 35434-9.
- Santoro, R., Li, J. and Grummt, I. (2002) *Nat Genet*, 32, 393-6.
- Sarraf, S. A. and Stancheva, I. (2004) *Mol Cell*, 15, 595-605.
- Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J., 3rd (2002) *Genes Dev*, 16, 919-32.
- Schumacher, B. L., Block, J. A., Schmid, T. M., Aydelotte, M. B. and Kuettner, K. E. (1994) *Arch Biochem Biophys*, 311, 144-52.
- Schwartzman, J. S., De Souza, A. M., Faiwichow, G. and Hercowitz, L. H. (1998) *Arq Neuropsiquiatr*, 56, 824-8.
- Sekimata, M. and Homma, Y. (2004) *Nucleic Acids Res*, 32, 590-7.
- Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R. and Zoghbi, H. (2002) *Neuron*, 35, 243-54.
- Stec, I., Wright, T. J., van Ommen, G. J., de Boer, P. A., van Haeringen, A., Moorman, A. F., Altherr, M. R. and den Dunnen, J. T. (1998) *Hum Mol Genet*, 7, 1071-82.
- Stevenson, R. E. and Schwartz, C. E. (2002) *Cytogenet Genome Res*, 99, 265-75.
- Stewart, G. S., Wang, B., Bignell, C. R., Taylor, A. M. and Elledge, S. J. (2003) *Nature*, 421, 961-6.
- Stoughton, R. B. (2004) *Annu Rev Biochem*.
- Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Langst, G. and

- Grummt, I. (2001) *Embo J*, 20, 4892-900.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Suyama, A., Sakaki, Y., Morishita, S., Okubo, K. and Sugano, S. (2001) *Genome Res*, 11, 677-84.
- Tachibana, M., Sugimoto, K., Fukushima, T., Shinkai, Y. (2001) *J Biol Chem*, 276, 25309-17.
- Takai, D. and Jones, P. A. (2002) *Proc Natl Acad Sci U S A*, 99, 3740-5.
- Tang, L. Y., Reddy, M. N., Rasheva, V., Lee, T. L., Lin, M. J., Hung, M. S. and Shen, C. K. (2003) *J Biol Chem*, 278, 33613-6.
- Tate, P., Skarnes, W. and Bird, A. (1996) *Nat Genet*, 12, 205-208.
- Tatematsu, K. I., Yamazaki, T. and Ishikawa, F. (2000) *Genes Cells*, 5, 677-88.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) *Nucleic Acids Res*, 25, 4876-82.
- Traynor, J., Agarwal, P., Lazzeroni, L. and Francke, U. (2002) *BMC Med Genet*, 3, 12.
- Tudor, M., Akbarian, S., Chen, R. Z. and Jaenisch, R. (2002) *Proc Natl Acad Sci U S A*, 99, 15536-41.
- U, M., Shen, L., Oshida, T., Miyauchi, J., Yamada, M. and Miyashita, T. (2004) *Leukemia*, 18, 1850-6.
- Utsch, B., Becker, K., Brock, D., Lentze, M. J., Bidlingmaier, F. and Ludwig, M. (2002) *Hum Genet*, 110, 488-94.
- Vacca, M., Filippini, F., Budillon, A., Rossi, V., Mercadante, G., Manzati, E., Gualandi, F., Bigoni, S., Trabanelli, C., Pini, G., Calzolari, E., Ferlini, A., Meloni, I., Hayek, G., Zappella, M., Renieri, A., D'Urso, M., D'Esposito, M., MacDonald, F., Kerr, A., Dhanjal, S. and Hulten, M. (2001) *J Mol Med*, 78, 648-55.
- Vairapandi, M. (2004) *J Cell Biochem*, 91, 572-83.
- 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 Francesco, 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., et al. (2001) *Science*, 291, 1304-51.
- Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P. and et al. (1991) *Cell*, 65, 905-14.
- von Kries, J. P., Buhrmester, H. and Stratling, W. H. (1991) *Cell*, 64, 123-35.
- von Kries, J. P., Rosorius, O., Buhrmester, H. and Stratling, W. H. (1994) *FEBS Lett*, 342, 185-8.
- Vorsanova, S. G., Demidova, I. A., Ulas, V., Soloviev, I. V., Kazantzeva, L. Z. and Yurov Yu, B. (1996) *Neuroreport*, 8, 187-9.
- Vorsanova, S. G., Demidova, I. A., Ulas, V., Solov'ev, I. V., Kravets, V. S., Kazantseva, L. Z. and Iurov Iu, B. (1998) *Zh Nevrol Psikhiatr Im S S Korsakova*, 98, 53-6.
- Waddington, C. H. (1942) *Endeavour*, 1, 18-20.
- Wade, P. A., Geggion, A., Jones, P. L., Ballestar, E., Aubry, F. and Wolffe, A. P. (1999) *Nat Genet*, 23, 62-6.
- Wadekar, S. A., Li, D., Periyasamy, S. and Sanchez, E. R. (2001) *Mol Endocrinol*, 15, 1396-410.
- Wagner, S., Chiosea, S. and Nickerson, J. A. (2003) *Proc Natl Acad Sci U S A*, 100, 3269-74.
- Wakefield, R. I., Smith, B. O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A. P. and Barlow, P. N. (1999) *J Mol Biol*, 291, 1055-65.
- Wan, M., Lee, S. S., Zhang, X., Houwink-Manville, I., Song, H. R., Amir, R. E., Budden, S., Naidu, S., Pereira, J. L., Lo, I. F., Zoghbi, H. Y., Schanen, N. C. and Francke, U. (1999) *Am J Hum Genet*, 65, 1520-9.
- Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P. and Zhang, Y. (2001) *Mol Cell*, 8, 1207-17.
- Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R. G. and Zhang, Y. (2003) *Mol Cell*, 12, 475-87.

- Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M. and Nakao, M. (2003) *Proc Natl Acad Sci U S A*, 100, 12859-64.
- Welberg, L. A., Seckl, J. R. and Holmes, M. C. (2001) *Neuroscience*, 104, 71-9.
- Wellman, C. L. (2001) *J Neurobiol*, 49, 245-53.
- Wellmann, S., Buhrer, C., Moderegger, E., Zelmer, A., Kirschner, R., Koehne, P., Fujita, J. and Seeger, K. (2004) *J Cell Sci*, 117, 1785-94.
- Wu, P., Qiu, C., Sohail, A., Zhang, X., Bhagwat, A. S. and Cheng, X. (2003) *J Biol Chem*, 278, 5285-91.
- Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J. and Viegas-Pequignot, E. (1999) *Nature*, 402, 187-91.
- Yang, Y., Isaac, C., Wang, C., Dragon, F., Pogacic, V. and Meier, U. T. (2000) *Mol Biol Cell*, 11, 567-77.
- Yang, X., Zhang, F. and Kudlow, J. E. (2002) *Cell*, 110, 69-80.
- Yeh, T. F., Lin, Y. J., Lin, H. C., Huang, C. C., Hsieh, W. S., Lin, C. H. and Tsai, C. H. (2004) *N Engl J Med*, 350, 1304-13.
- Yoder, J. A., Yen, R. W., Vertino, P. M., Bestor, T. H. and Baylin, S. B. (1996) *J Biol Chem*, 271, 31092-7.
- Yoshida, N. L., Miyashita, T., U, M., Yamada, M., Reed, J. C., Sugita, Y. and Oshida, T. (2002) *Biochem Biophys Res Commun*, 293, 1254-61.
- Yu, F., Thiesen, J. and Stratling, W. H. (2000) *Nucleic Acids Res*, 28, 2201-6.
- Yusufzai, T. M. and Wolffe, A. P. (2000) *Nucleic Acids Res*, 28, 4172-9.
- Zalfa, F. and Bagni, C. (2004) *Curr Issues Mol Biol*, 6, 73-88.
- Zappella, M. (1992) *Brain Dev*, 14, 98-101.
- Zappella, M., Gillberg, C. and Ehlers, S. (1998) *J Autism Dev Disord*, 28, 519-26.
- Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A. and Reinberg, D. (1999) *Genes Dev*, 13, 1924-35.
- Zhao, X., Ueba, T., Christie, B. R., Barkho, B., McConnell, M. J., Nakashima, K., Lein, E. S., Eadie, B. D., Willhoite, A. R., Muotri, A. R., Summers, R. G., Chun, J., Lee, K. F. and Gage, F. H. (2003) *Proc Natl Acad Sci U S A*, 100, 6777-82.
- Zhu, Y., Spitz, M. R., Zhang, H., Grossman, H. B., Frazier, M. L. and Wu, X. (2004) *Cancer*, 100, 1853-8.

Acknowledgments

I would like to thank Professor Dr. Ropers for the opportunity to carry out my thesis in his department at the Max Planck Institute for molecular genetics. I am indebted to Professor Dr. Hucho for accepting to be my second referee.

A special thank you goes to Dr. Ulrike Nuber, for all the invested time, the excellent supervision, and the inspiring discussions we had during these three years.

This work would not have been possible without the superb technical assistance by Ralph Schulz and Bettina Lipkowitz as well as their social support during my time in the lab.

Many thanks to Ulf Gurok for very helpful comments on this thesis as well as the fruitful debates and the possibility to share challenging moments.

Thank you also to the other members in the lab, Ines Müller, Falk Hertwig, Georg Wiczorek, Annkatrin Wernstedt, Fikret Erdogan and Rogier Versteeg for the motivating spirit.

I am very grateful for all the patience and comprehension Anne had during the last three years.

Finally, many thanks to my parents. Only their continuous support over all these years enabled me to get this far.

Curriculum Vitae and Publications

Name: Tim-Christoph Roloff
Address: Department of Human Molecular Genetics
Max-Planck Institute for Molecular Genetics,
Berlin, Germany
Phone: ++49 (0) 30 8413 1290
E-mail: rolof_t@molgen.mpg.de
Date of birth: 26/09/1976 (Mülheim a.d. Ruhr, Germany)
Nationality: Swiss
Marital status: single

Education

Oct 2001 - 2005 PhD studies in biochemistry at the
Max Planck Institute for Molecular Genetics (MPIMG), Berlin, Germany

Sep 1998 - Sep 2001 Studies at the ESBS (Ecole Supérieure de Biotechnologie à Strasbourg; European
School of the Upper Rhine Universities),
Diploma in Biotechnology
Louis Pasteur University, Strasbourg, France

Jan 2001 - Sep 2001 Diploma Thesis
Kyoto University Bioinformatics Center (KUBIC)
Kyoto University, Japan

Oct 1996 - Jul 1998 Studies of Biology 2
Vordiplom (equal to a 2 year undergraduate degree course in Maths,
Physics, Physical Chemistry, Chemistry, Microbiology, Cell Biology and Genetics)
University of Basel, Switzerland

Jul 1992 - Dec 1995 Gymnasium
Matura C (equivalent to A-levels in Maths, Physics and Chemistry)
Gymnasium MuttENZ, Switzerland

Scientific publications

2004 **Roloff TC**, Nuber UA

Chromatin, epigenetics and stem cells.

European Journal of Cell Biology. 2005 (in press)

2004 Nuber UA*, Kriaucionis S*, **Roloff TC***, Guy J, Selfridge J, Steinhoff C, Schulz R, Lipkowitz B, Ropers HH, Holmes MC, Bird A

Disturbance of the stress response pathway in a mouse model of Rett syndrome.

Submitted to Human Molecular Genetics

* These authors contributed equally to this work

2003 Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gecz J, Sefiani A, Hoffmann K, Moser B, Haas S, Gurok U, Haesler S, Aranda B, Nshedjan A, Tzschach A, Hartmann N, **Roloff TC**, Shoichet S, Hagens O, Tao J, van Bokhoven H, Turner G, Chelly J, Moraine C, Fryns JP, Nuber U, Hoeltzenbein M, Scharff C, Scherthan H, Lenzner S, Hamel BCJ, Schweiger S, Ropers HH

Mutations in the polyglutamine-binding protein 1 gene cause X-linked mental retardation.

Nature Genetics. 2003 Nov 23;35(4):p313-315

2003 **Roloff TC**, Ropers HH, Nuber UA

Comparative study of methyl-CpG-binding domain proteins.

BMC Genomics. 2003 Jan 16;4(1):1

Conferences and speaking engagements

2004 Gurok U, **Roloff TC**, Steinhoff C, Lipkowitz B, Ropers HH, Scharff C, Nuber UA

Dynamic gene expression changes underlying neural progenitor differentiation.

Poster Keystone Symposium on Stem cells: Asymmetric cell division & cell fate, Keystone, USA

2003 **Roloff TC**

The methyl-CpG-binding domain protein family - a comparative study.

Workshop speaker and poster at the HUGO meeting HGM2003, Cancun, Mexico

2002 **Roloff TC**, Ropers HH, Nuber UA

Identification of proteins containing a Methyl-CpG-binding domain (MBD).

Poster at Gesellschaft für Humangenetik (GfH), Jahrestagung 2002, Leipzig, Germany