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 *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 form 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 *et al.*, 2001	CNS-specific deletion of exon 4 by cre/loxP system with a nestin promoter	***Mecp2*-null mice (normal until ~5 weeks):** nervousness body trembling pila erection hard breathing *at later stages:* overweight physical deterioration hypoactive death at ~10 weeks reduced brain size and weight ***Mecp2*$^{+/-}$ females (normal for ~4 months):** weight gain reduced activity ataxic gait
Guy *et al.*, 2001	Excision of exons 3 and 4 by cre/loxP system	***Mecp2*-null mice (normal until ~3-8 weeks)**: stiff, uncoordinated gate hind limb clasping 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 ***Mecp2*$^{+/-}$ females:** inertia and hindlimb clasping after 3 months
Shahbazian *et al.*, 2002	Premature stop codon at aa 308 leaves MBD and TRD intact	***Mecp2*$^{308/y}$ mice (normal until ~6 weeks):** first symptom is a subtle tremor when suspended by tail tremor worsens with age stereotypic forelimb motions and clasping 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 ***Mecp2*$^{+/-}$ females:** milder and variable phenotype
Collins *et al.*, 2004	Slight over-expression of human *MECP2* from a PAC clone	normal until ~10-12 weeks forepaw clasping when hung by tail aggressiveness hypoactivity seizures spasticity kyphosis premature death

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 overexpresed, 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, TFIIE, TFIIF, and TFIIH the pre-initiation complex (PIC) is formed and transcription can take place (Fig. 1).

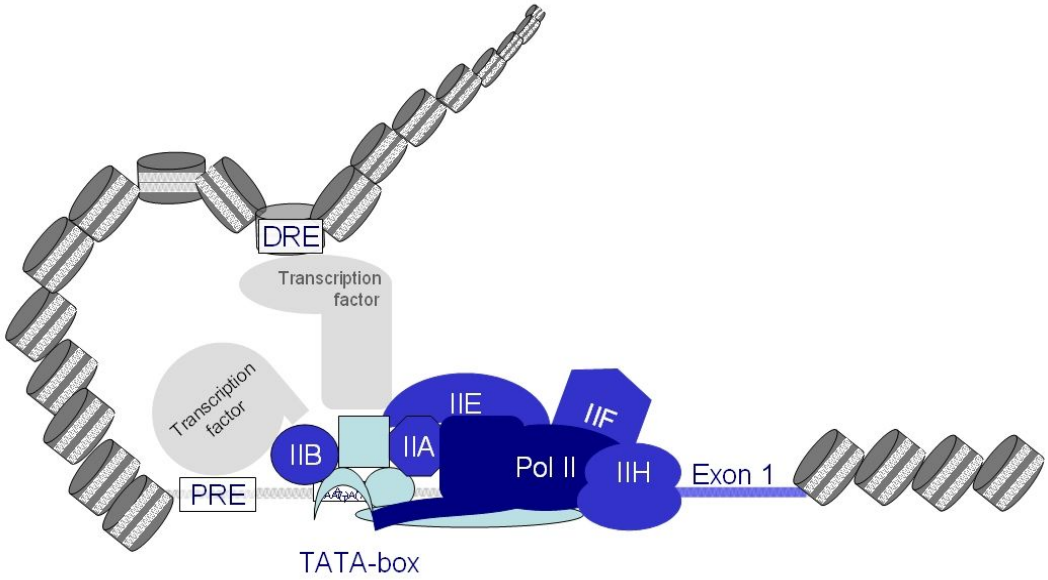

Fig. 1. Transcription initiation. This scheme shows the factors build up a pre-initiation complex at the promoter of a gene. The polymerase is shown in dark blue (Pol II), the general transcription factors (IIA, IIB, IIE, IIF, IIH) are depicted in blue. The TATA-binding protein and associated factors have a turqouis 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).

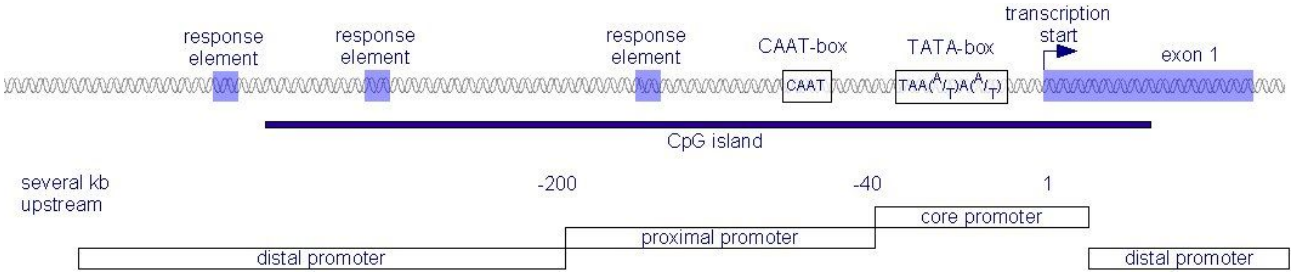

Fig. 2. Representation of an eukaryotic promoter. The relative position of transcription start, TATA-box (-20 to -30), CAAT-box (-70 to -80), and the CpG island can be seen. The existence and position of response elements varies from promoter to promoter.

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 lead 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.

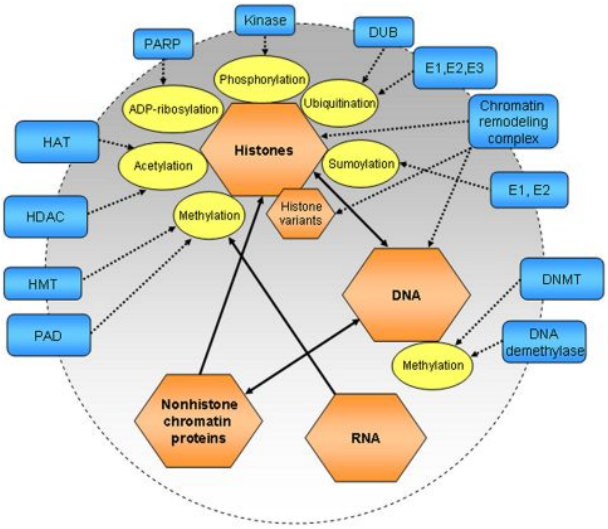

Fig. 3. Summary of epigenetic mechanisms. The basic components of epigenetic gene regulation are depicted in orange, possible modifications to DNA or proteins are shown in yellow, and enzyme classes mediating the modifications are colored in blue. (Adopted form Roloff and Nuber, 2005)

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.

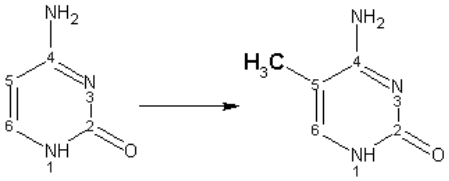

Fig. 4. Cytosines can get methylated at the 5' position .

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 m^5CpG.

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 m^5CpGs, 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 m^5CpGs 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).

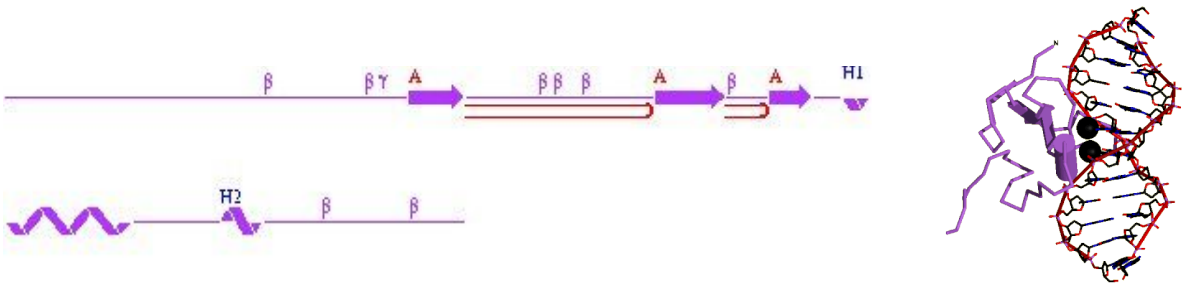

Fig. 5. Structure of the MBD domain of human MECP2. The MBD domain consists of three β-twisted sheets (indicated by arrows) and one main α-helix (indicated by H1 and H2) (protein chain from PDB protein data bank). On the right, the 3D structure of a methyl-CpG binding domain interacting with DNA is shown. 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 N *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 m^5CpGs 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$^{-/-}$/*Mecp*$^{-/y}$ 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 m^5CpGs *in vitro* (Jiang *et al.*, 2002). MBD3L1 is a trancriptional 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 m^5CpG/GpT mismatches instead of symmetrically methylated m^5CpGs. 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.

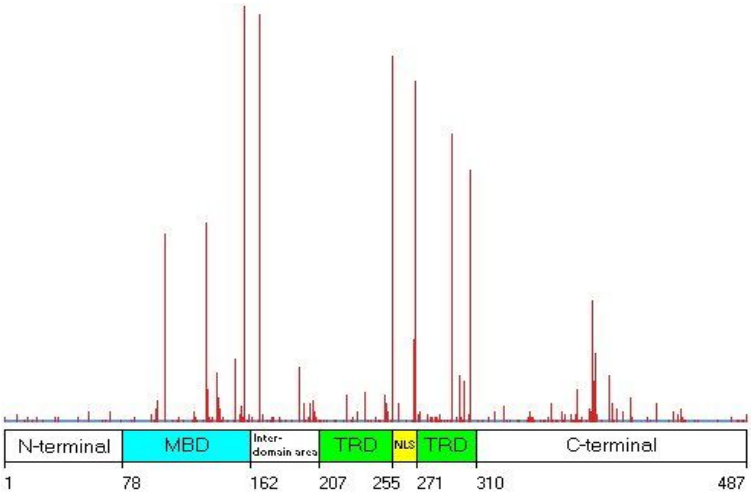

Fig. 6. Distribution and frequency of mutations in *MECP2e2*. The height of the lines indicates how many individuals carrying a mutation at this site have been found. Mutation hot spots are seen in the MBD and the TRD domain. (adopted from RettBASE).

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 U2afl-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.

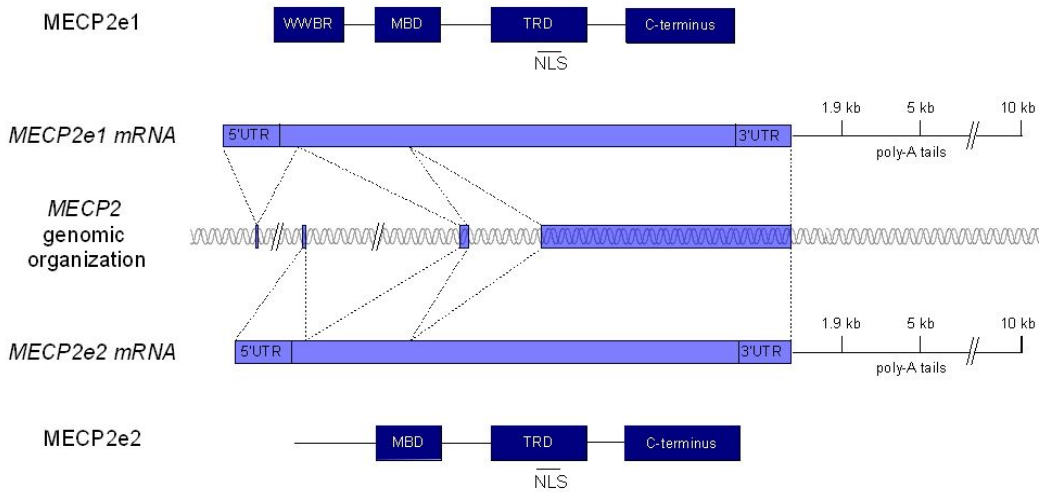

Fig. 7. *MECP2* genomic, transcript, and protein structure. Human MECP2: genomic organization (in the middle), mRNAs and domain organization of MECP2e1 (top) and MECP2e2 (bottom) proteins. The three different transcripts for each splice variant are denoted by the different poly-A tail sizes that are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

AT-hooks are DNA binding motifs of eleven residues that allow binding to the minor grove 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 prolin-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 containg 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 unkown.

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).

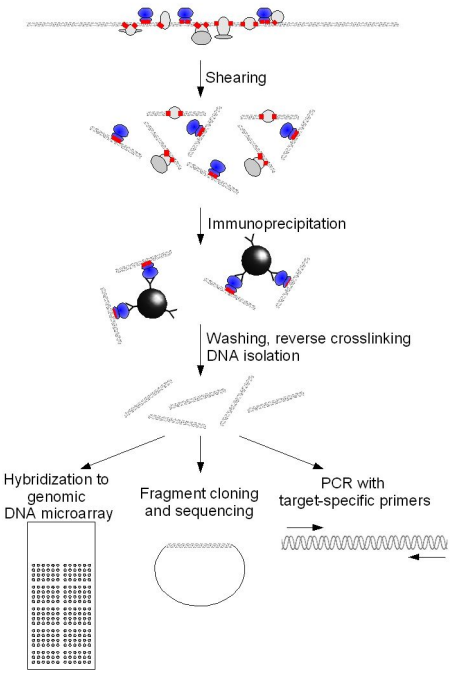

Fig. 8. Scheme of chromatin immunoprecipitation. Protein-DNA complexes are linked (red dots) in cells by the use of formaldehyde. The chromatin is then sheared into small fragments and the protein of interest (blue) is immunoprecipitated with a specific antibody bound to protein A-agarose (black). After 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.