STUDIES OF MECHANISMS CONTROLLING DNA METHYLTRANSFERASE 1

CARINA FRAUER

SIGILL: UNIVERSITAT: LUDOVIC: MAXIM:

MÜNCHEN 2010

STUDIES OF MECHANISMS CONTROLLING DNA METHYLTRANSFERASE 1

CARINA FRAUER

Dissertation
an der Fakultät für Biologie
der Ludwig-Maximilians-Universität
München

vorgelegt von
Carina Frauer
aus Worms

München, den 14. Juni 2010

Erstgutachter: Prof. Dr. Heinrich Leonhardt

Zweitgutachter: Prof. Dr. Peter Becker

Tag der mündlichen Prüfung: 01.10.2010

CONTENT

Summary 1

1 Introduction 3

1.1 DNA methylation in eukaryotes 3

1.1.1 Epigenetic gene regulation 3
1.1.2 Discovery of 5-methylcytosine in DNA 4
1.1.3 Eukaryotic DNA methyltransferases 6
1.1.4 Distribution of 5-methylcytosine in the mammalian genome 9
1.1.5 Regulation of gene expression by DNA methylation 14
1.1.6 DNA methylation dynamics 19

1.2 The Dnmt1 enzyme 23

1.2.1 Structure 23
1.2.2 Mechanism of the methyl transfer reaction 25
1.2.3 Substrate specificity 26

1.3 Regulation of Dnmt1 28

1.3.1 Transcriptional and translational Regulation 28
1.3.2 Regulation of Dnmt1 by post-translational modifications 29
1.3.3 DNA binding and allosteric activation 29
1.3.4 Regulation by interaction with other factors 31

1.4 Aims of this work 33

2 Results 35

2.1 A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding 35

2.2 Modulation of protein properties in living cells using nanobodies 49

2.3 Different binding properties and functional relevance of CXXC zinc finger domains of enzymes involved in cytosine modification 71

2.4 The multi-domain protein Np95 connects DNA methylation and histone modification 101

2.5 Single molecule fluorescence spectroscopy of Dnmt1:DNA complexes 121

2.6 DNA methylation 133

3 Discussion 135
3.1 Development of a versatile assay to measure DNA binding and methyltransferase activity 136
3.1.1 Potential of the developed assay 136
3.1.2 GBP binding enhances GFP fluorescence 138
3.1.3 Using the newly established tools for further applications 139
3.2 Maintenance methylation – a clear but demanding task 143
3.2.1 How Dnmt1 recognizes its substrate 143
3.2.2 The functional role of the Dnmt1 CXXC domain 145
3.2.3 The role of Uhrf1 in maintenance methylation and regulation of Dnmt1 148
3.2.4 Dnmt1:DNA complex stoichiometry 150
3.3 Outlook 152
3.3.1 Maintenance methylation - more than a copy process? 152
3.3.2 On the role of Uhrf1 in maintenance methylation 153
3.3.3 The need for a three-dimensional Dnmt1 structure 155
4 Annex 157
4.1 References 157
4.2 Contributions 172
4.3 Declaration 174
4.4 Acknowledgements 175
5 Curriculum Vitae 177

SUMMARY

DNA methylation plays a central role in the epigenetic control of mammalian gene expression. DNA methyltransferase 1 (Dnmt1) is the most ubiquitously expressed DNA methyltransferase and responsible for maintenance of DNA methylation patterns during semi-conservative DNA replication. The fidelity of this process is crucial for genome stability and is based on the recognition of hemimethylated CpG sites emerging at the replication fork. Indeed, it is well established that Dnmt1 has a preference for substrates containing hemimethylated over unmethylated CpG sites *in vitro*. However, it remained elusive how and at which step of the methyl transfer reaction the substrate discrimination occurs, and also if or how intrinsic or interacting factors regulate Dnmt1 activity.

To investigate the mechanistic basis of Dnmt1's maintenance function we developed a versatile non-radioactive assay for methyltransferase activity and DNA binding. This assay not only allows to rapidly screen for active methyltransferases, but also to determine substrate specific DNA binding activity by testing up to four DNA substrates in direct competition. With this assay, we showed that Dnmt1 does not discriminate between different methylation states at the step of DNA binding, but rather at an early step of the methyl transfer reaction, the covalent complex formation between enzyme and target cytosine residue.

Furthermore, we systematically analyzed the DNA binding properties of the Dnmt1 CXXC domain and the Dnmt1 interacting cofactor Uhrf1 in order to characterize their respective functional role in the regulation of Dnmt1. We could show that the CXXC domain, although specifically binding to unmethylated DNA, is dispensable for DNA binding, enzymatic activity and substrate specificity of Dnmt1. For Uhrf1, we detected and confirmed preferential binding to DNA containing hemimethylated CpG sites, which is however very low compared to the intrinsic preference of Dnmt1 for methylation of these sites. These data together with the evidence for specific histone tail binding of Uhrf1 lead to the conclusion that the role of Uhrf1 in maintenance methylation is more complex than previously suggested, exceeding the mere recruitment of Dnmt1 to hemimethylated CpG sites.

At last, we addressed the question of Dnmt1 dimerization and its functional impact on Dnmt1 activity by single molecule investigation of Dnmt1:DNA complexes using fluorescence intensity distribution analysis (FIDA) and fluorescence cross-correlation spectroscopy (FCCS). Surprisingly, we obtained first evidence that Dnmt1 might be able to covalently bind two DNA substrates simultaneously, indicating another possibility of enzyme regulation and supporting the hypothesis of major structural differences between the catalytic domain of Dnmt1 and prokaryotic DNA methyltransferases.

In conclusion, this work contributes to further elucidation of the mechanisms of Dnmt1 regulation as the basis for stable inheritance of epigenetic information.

1 INTRODUCTION

1.1 DNA METHYLATION IN EUKARYOTES

1.1.1 EPIGENETIC GENE REGULATION

All cells of a multi-cellular organism contain the same genetic information, but they differ in structure and function. The basis of cellular differentiation is the establishment of differential and stable tissue-specific gene expression patterns during development. The decision on which set of genes is expressed at any specific time point, and in which cell, can in principle be taken at different steps of gene expression. However, transcriptional regulation is the most important control mechanism. Transcriptional control is accomplished by complex protein networks, which finally affect gene transcription by RNA polymerase. Components of these regulatory networks are i) general and sequence-specific transcription factors, and ii) cis-acting regulatory elements (promoters, enhancers, silencers and insulators). The resulting networks can dynamically respond to environmental changes and signals, but also initiate developmental programs by generation of stable feedback loops. For example, the transcription factor MyoD triggers differentiation of myoblasts into mature muscle cells via activation of its own and a battery of other muscle-specific genes, inducing a series of positive feedback loops. However, due to the complexity of mammalian gene expression programs, cellular differentiation requires an additional level of cellular memory for stable changes in gene expression patterns. Thus, in order to allow long-term stability of specific transcriptional states, transcriptional regulation is accompanied by DNA methylation, histone modification and chromatin remodeling (Reik, 2007). These chromatin modifications are by definition epigenetic, since they affect gene expression and chromatin structure without alteration of the underlying genomic sequences.

Epigenetic mechanisms are thought to regulate gene expression by controlling the condensation and accessibility of genomic DNA. Importantly, DNA within the nucleus is packed into higher ordered chromatin structures consisting of DNA, histone and non-histone proteins. The first level of packaging is achieved solely by the core histones H2A, H2B, H3 and H4 forming an octamer, around which 146 base pairs of DNA are wrapped. These histone core particles are connected via linker DNA of variable length. One histone core particle plus one adjacent DNA linker are together referred to as nucleosome. Through nucleosome formation, DNA is compacted by one third and adopts a ‘beads on a string’-like structure visible by electron microscopy. The position of nucleosomes is controlled by DNA binding proteins as well as DNA sequence and flexibility; AT-rich sequences for example are easier to compress. Condensation of DNA into a 30 nm fiber is achieved by generation of regular arrays, in which nucleosomes are packed on top of each other, involving binding of histone H1 to both core particle and linker DNA. Levels of chromatin organization beyond the 30 nm fiber are poorly understood, but certainly involve the formation of various loops and coils. Importantly, one can distinguish different levels of condensation of interphase DNA. The less condensed and transcriptionally active form is called euchromatin and the highly condensed form is called heterochromatin. Constitutive heterochromatin makes up around 10 % of the genome and is concentrated in centromeric and telomeric regions.

As mentioned above, chromatin structure and activity are controlled by a distinct pattern of epigenetic modifications. In this respect, methylation, as the only modification of DNA, is set by DNA methyltransferases and is generally associated with gene silencing. Histones however are subject to various post-translational modifications by histone modification enzymes and distinct modification states of histone tails correlate with chromatin activity. Transcriptionally active regions are for example associated with histone H3 methylation at lysine 4 (H3K4) and H3K9 acetylation, whereas transcriptionally inactive regions are associated with H3K9 trimethylation. Interestingly, histone modifications seem to have only little effect on nucleosome stability, but to rather affect the formation of higher order chromatin structures. Still, nucleosomal arrangements are highly dynamic and in this respect, an important role is played by ATP-dependent chromatin remodeling complexes, which can modulate the distribution and positioning of nucleosomal structures.

Epigenetic mechanisms that regulate mammalian gene expression at the transcriptional level are very complex and often work in concert. Besides the already mentioned DNA methylation, histone modification and chromatin remodeling factors, they also involve regulatory proteins of the Polycomb/Trithorax group and non-coding RNA. Some of these epigenetic marks are replicated in each cell division cycle and propagated through successive cell generations and, in the case of genomic imprinting, even passed on to the offspring. The work of this thesis focuses on aspects of DNA methylation, which occurs as 5-methylcytosine in mammals. This important modification is required for genome integrity and the stable repression of genes and transposable elements. It is involved in X chromosome inactivation (the mechanism of X-chromosome dosage compensation in female mammals), genomic imprinting and silencing of endogenous retroviral sequences.

1.1.2 DISCOVERY OF 5-METHYLCYTOSINE IN DNA

Already in 1904, Wheeler and Johnson anticipated the natural occurrence of 5-methylcytosine in DNA and chemically synthesized this pyrimidine. However, it was not until 1925 that 5-methylcytosine was claimed to be found by Johnson and Coghill as hydrolysis product of tuberculinic acid, the DNA of tubercle bacillus. The authors identified the new substance by comparing the optical properties of its crystalline picrate with those from cytosine picrate of the same DNA source as well as with synthetic 5-methylcytosine crystals (Johnson and Coghill, 1925). Another two decades later, in 1948, Hotchkiss established a paper chromatographic method for quantitative separation of purines, pyrimidines and nucleosides from hydrolyzed DNA samples. From a calf thymus DNA preparation, he obtained a small fraction of a substance, which was proposed to be 5-methylcytosine based on its chromatographic behavior and ultraviolet absorption characteristics (Hotchkiss, 1948). This finding was confirmed in 1950, when Wyatt used a more sensitive method for chromatographic and spectral analyses and unambiguously identified 5-methylcytosine in calf thymus DNA (Wyatt, 1950, 1951b). The same author additionally analyzed preparations from other animal and a plant species and showed that the 5-methylcytosine content varies with the source, but is very constant for a particular DNA source (Wyatt, 1951a). This observation suggested the percentage of 5-methylcytosine in DNA to be biologically and functionally relevant. Today, we know that DNA methylation is present in all kingdoms of life and that the level of DNA methylation is indeed species-specific. In eukaryotes, methylation levels are ranging from undetectable or far below 1 % in some insects to very high levels in plants with up to 50 % of all cytosine bases being modified (Montero et al., 1992). DNA methylation levels of mammals are intermediate, and it was suggested that approximately 1% of all DNA bases in humans are 5-methylcytosine (Ehrlich et al., 1982; Kriaucionis and Bird, 2003). Moreover, also the distribution and patterns of cytosine methylation in the genome differ in between different species (Suzuki and Bird, 2008). For example, some invertebrate genomes show a mosaic methylation pattern where large, heavily methylated domains are interspersed with equivalent lengths of unmethylated DNA regions. In contrast, vertebrate methylation is distributed over the entire genome and constitutes a pattern of global DNA methylation. It has been suggested that the pattern of 5-methylcytosine distribution in the genome reflects its functions in different organisms (Colot and Rossignol, 1999). Moreover, not only the C^5 position of cytosine can be methylated in DNA, but also the N^4 position of cytosine and the N^6 position of adenine, giving rise to N^4-methylcytosine and N^6-methyladenine, respectively (Dunn and Smith, 1958; Ehrlich et al., 1985). Both cytosine and adenine methylation occur in bacteria and plants. In eukaryotes, however, 5-methylcytosine is the dominant DNA modification with some exceptions of mainly unicellular organisms showing also low levels of adenine methylation (Gorovsky et al., 1973; Hattman, 2005).

The role of DNA methylation in bacteria was already discussed in the 1960s, where methylation of DNA was proposed as mechanism to protect (bacterial) host DNA. The modification of DNA with methyl groups was suggested to alter its biochemical properties and to induce structural changes, which would protect against restriction endonucleases that are directed against foreign (bacterial or viral) unmethylated DNA (Arber and Linn, 1969; Srinivasan and Borek, 1964). About a decade later, two independent publications indicated a role of DNA methylation in transcriptional regulation in mammals (Holliday and Pugh, 1975; Riggs, 1975). Hollliday and Pugh suggested that the enzymatic modification of specific bases in repeated DNA sequences might be the basis for developmental clocks and regulate gene activity during development and cellular differentiation. Furthermore, Riggs proposed that DNA methylation affects the DNA binding ability of regulatory proteins and that DNA methylation is essential for initiation and maintenance of X inactivation. The major questions are how such a tiny modification, the attachment of a methyl group, can contribute to development and cellular differentiation, which factors set these methylation marks, and how methylation patterns are maintainend during cell division.

1.1.3 EUKARYOTIC DNA METHYLTRANSFERASES

In mammalian cells, DNA methylation occurs at cytosine residues mainly of CpG dinucleotides, where a methyl group is covalently attached to the C^5 position of the nucleobase. This modification is set and maintained by the DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b (Goll and Bestor, 2005; Rottach et al., 2009b). Another methyltransferase, Dnmt2, has been shown to methylate tRNA (Goll et al., 2006); the DNA methyltransferase activity of Dnmt2, however, is very low *in vitro* and its functional relevance in mammals is still controversially discussed (Hermann et al., 2003). All mammalian Dnmts contain a highly conserved catalytic domain, which is also very similar to prokaryotic methyltransferases (Figure 1). Thus, prokaryotic and mammalian DNA methyltransferases have also been proposed to use a similar catalytic mechanisms ((Bestor and Verdine, 1994; Cheng et al., 1993; Klimasauskas et al., 1994; Wu and Santi, 1987), see also chapter 1.2.2).

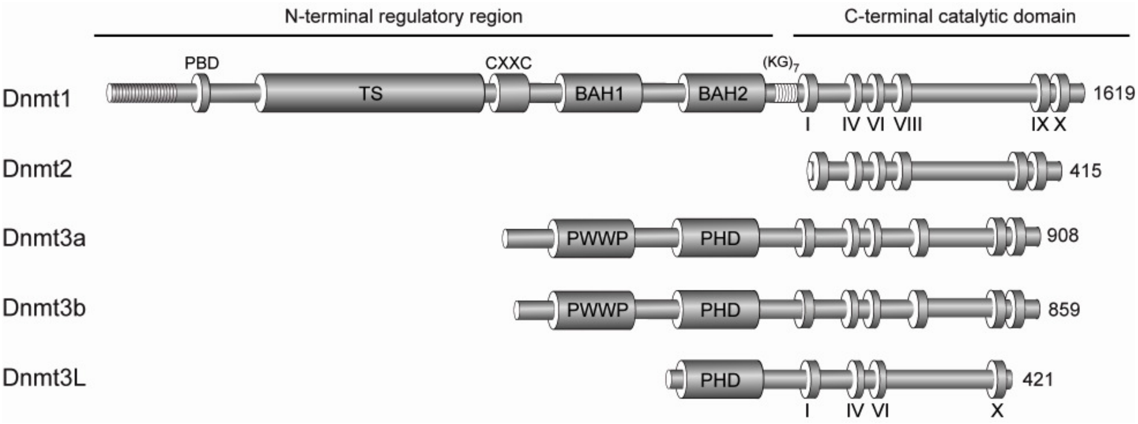


Figure 1. Domain structure of the mammalian Dnmt family members. All but Dnmt2 possess a regulatory N-terminal region in addition to the highly conserved catalytic domain. Conserved domains in the N-terminal parts of Dnmt1 and Dnmt3a/b/L are highlighted and described in the text (chapter 1.2.1 and 1.1.3, respectively). PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC zinc finger domain; BAH: Bromo adjacent homology domain; PWWP: domain containing Pro-Trp-Trp-Pro motif; PHD: plant homeodomain. (modified from (Rottach et al., 2009b))

1.1.3.1 DNMT1

Dnmt1 was the first mammalian DNA methyltransferase to be identified and cloned (Bestor et al., 1988). It is constitutively expressed in proliferating as well as post-mitotic cells making it the most ubiquitous DNA methyltransferase. Dnmt1 is considered to be the maintenance methyltransferase, responsible for copying the defined methylation pattern during semi-conservative DNA replication by specific methylation of hemimethylated CpG sites occurring at the replication fork. Consistently, Dnmt1 is transcribed mostly during S phase and has a strong preference for hemimethylated DNA substrates (Robertson et al., 2000b). Moreover, in mammalian cells, Dnmt1 has been shown to be targeted to sites of DNA replication (Chuang et al., 1997a; Leonhardt et al., 1992), and also to be recruited to DNA repair sites (Mortusewicz et al., 2005).

Evidence for the importance of Dnmt1 in mammals came from genetic studies in mice targeting the Dnmt1 gene. Homozygous embryos deficient in Dnmt1 were delayed in development and did not survive mid-gestation (Li et al., 1992). Compound heterozygous mice carrying a hypomorphic and a null allele expressed Dnmt1 to only 10 % of wild-type levels and possess a globally hypomethylated genome, leading to chromosomal instability and the development of aggressive tumors (Gaudet et al., 2003). More specifically, Dnmt1 was shown to be essential for X chromosome inactivation and to be required for the maintenance of genomic imprints (Howell et al., 2001). Furthermore, mouse embryos lacking Dnmt1 show increased transcription of intracisternal A-particle (IAP) retrotransposons, the most aggressive parasitic DNA sequence in the mammalian genome, which suggests Dnmt1 to also play a role in suppression of retroviral and transposable elements (Gaudet et al., 2004; Walsh et al., 1998). Importantly, these processes are not directly controlled by Dnmt1, but are rather indirectly affected as a consequence of global DNA hypomethylation due to the failure of maintenance methylation. Somatic and cancer cells require Dnmt1 for survival and hematopoietic and epidermal stem cells require Dnmt1 for self-renewal (Chen et al., 2007; Sen et al., 2010; Spada et al., 2007; Trowbridge et al., 2009). Surprisingly, *dnmt1*$^{-/-}$ embryonic stem cells are viable and show no obvious abnormalities, although global methylation levels are significantly reduced in these cells (Lei et al., 1996; Li et al., 1992).

The Dnmt1 enzyme comprises a regulatory N-terminal domain connected to its C-terminal catalytic domain, by a linker of seven glycine-lysine repeats (Figure 1). Although containing all conserved motifs identified to be involved in the methyl transfer reaction of prokaryotic methyltransferases (Bestor and Verdine, 1994; Kumar et al., 1994), the catalytic domain of Dnmt1 is not active by itself, but has to be activated by intramolecular interaction with the N-terminal part of the enzyme ((Fatemi et al., 2001; Margot et al., 2003; Zimmermann et al., 1997), see also chapter 1.2.2). The N-terminal part of Dnmt1, which comprises two thirds of the enzyme, not only regulates the activity of the catalytic domain, but also determines the cell-cycle specific subcellular localization of Dnmt1. The exact mechanistic basis of Dnmt1's maintenance function for stable inheritance of epigenetic information remains to be elucidated. In addition to the intrinsic preference of Dnmt1 for DNA substrates containing hemimethylated CpG sites, also interaction with regulatory cofactors such as PCNA and Uhrf1 (see also chapter 1.3.4) have been proposed to target Dnmt1 activity to sites of replication. These factors likely contribute to the faithful maintenance of methylation patterns during replication and thereby to the inheritance of this important epigenetic mark.

1.1.3.2 THE DNMT3 FAMILY

Dnmt3a and Dnmt3b are known to establish methylation patterns during early embryonic development acting as de novo methyltransferases (Kaneda et al., 2004; Okano et al., 1999). Accordingly, Dnmt3a and 3b are highly expressed in embryonic stem cells, early embryos and developing germ cells, but down-regulated in non-embryonic somatic tissues. They methylate both unmethylated and hemimethylated DNA substrates with the same efficiency. The third member of the Dnmt3 family, Dnmt3L, lacks some crucial catalytic motifs and is not able to catalyze the methyl group transfer. However, Dnmt3L serves as cofactor for Dnmt3a and 3b and stimulates their activity (Suetake et al., 2004).

Dnmt3b knock-out mice die at late embryonic stage and lack methylation in centric minor satellite repeats. Dnmt3a knock-out mice show developmental abnormalities and die a few weeks after birth, due to deficient methylation of single-copy genes, retrotransposons and genomic imprints during germ cell development (Okano et al., 1999). Human patients with mutations in *DNMT3B* suffer from the ICF syndrome (immunodefiency, centromere instability and facial abnormalities). They show methylation defects at pericentric heterochromatin and at CpG-rich regions (CpG islands) on the inactive X chromosome (Hansen et al., 1999; Miniou et al., 1994; Xu et al., 1999). Dnmt3L is together with Dnmt3a necessary for the establishment of genomic imprints (Bourc'his et al., 2001; Kaneda et al., 2004). Thus, all three members of the Dnmt3 family are required for de novo methylation during development and have overlapping as well as distinct functions.

Like Dnmt1, also Dnmt3a and 3b possess an N-terminal regulatory domain in addition to the C-terminal catalytic domain (Figure 1). However, in contrast to Dnmt1, the N-terminal part of Dnmt3a and 3b is not necessary for catalysis. In both cases, the N-terminal part contains a PHD domain (ATRX-like Cys-rich domain) as well as a PWWP domain and mediates a variety of protein interactions as well as association with heterochromatin (Chen et al., 2004; Fuks et al., 2001; Qiu et al., 2002). Whereas the substrate of Dnmt1 is unambiguously identified as hemimethylated CpG sites with very low, if any, impact of neighboring sequences, the important question of target specificity remains for Dnmt3a and 3b. Indeed, Dnmt3a and 3b do have flanking sequence preferences, YNCGY and RCGY, respectively (Handa and Jeltsch, 2005; Lin et al., 2002), which might help in targeting Dnmt3s to their respective sites of action. However, considering the complexity of DNA methylation patterns, there have to be additional mechanisms, possibly involving non-coding RNA and interactions with specific transcription and chromatin factors, that contribute to the establishment of specific DNA methylation patterns.

1.1.3.3 COOPERATIVE FUNCTION OF DNMTS

It was initially thought that the functions of Dnmt1 as the maintenance methyltransferase and of Dnmt3a and Dnmt3b as the de novo methyltransferases were clearly separated. Currently, there is increasing evidence that all methyltransferases cooperatively set and maintain methylation patterns. For instance, Dnmt1, Dnmt3a and Dnmt3b have been shown to interact with each other (Kim et al., 2002; Margot et al., 2003). On the one hand, some de novo methylation activity was reported in embryonic stem cells lacking Dnmt3a and Dnmt3b (Lorincz et al., 2002). On the other hand, Dnmt3s seem to be required for proper maintenance of DNA methylation patterns in both somatic and embryonic stem cells. Conditional *dnmt3b*$^{-/-}$ mouse embryonic fibroblasts show hypomethylation of minor satellite and type C retroviral elements (Dodge et al., 2005). Mouse embryonic stem cells lacking Dnmt3a and Dnmt3b show altered methylation of imprinted genes and repeats and moreover, they gradually lose DNA methylation with increasing cell divisions down to an undetectable level (Chen et al., 2003; Okano et al., 1999). These data provide convincing evidence for cooperativity between Dnmt1 and Dnmt3a/Dnmt3b at least in maintenance methylation for methylated CpG rich regions like endogenous repetitive sequences (Jones and Liang, 2009; Liang et al., 2002).

1.1.4 DISTRIBUTION OF 5-METHYLCYTOSINE IN THE MAMMALIAN GENOME

1.1.4.1 SPATIAL DISTRIBUTION OF 5-METHYLCYTOSINE

As it was introduced before, DNA methylation in mammals exclusively occurs at the C^5 position of cytosine residues and mainly within CpG dinucleotides (Sinsheimer, 1955). CpG sites are distributed throughout the genome including all types of sequences: promoter regions, gene bodies, intergenic sequences and repetitive elements. However, they are unevenly distributed and preferentially localize to gene rich loci (Lander et al., 2001). CpG dinucleotides are methylated to approximately 60-80% in mouse and human (Ehrlich et al., 1982; Gruenbaum et al., 1981) and, although the mammalian genome displays genome-wide methylation, CpG methylation is like CpG dinucleotides unevenly distributed. Whereas gene bodies, repetitive sequences and some intergenic sequences are highly methylated, there are some largely unmethylated regions including mainly regulatory promoter sequences and enhancers, but also the first exons of genes (Lister et al., 2009; Rollins et al., 2006; Schmidl et al., 2009). Interestingly, the CpG site density of genomic DNA sequences is anti-correlated with their CpG methylation level. In other words, sequences of high CpG density (also called CpG islands) are often unmethylated, whereas sequences of low CpG content are generally highly methylated.

Notably, CpG dinucleotides are significantly underrepresented in the mammalian genome with only 21% of their statistically expected occurrence (Bird, 1980; Lander et al., 2001). The main (but not only) reason for this underrepresentation is spontaneous deamination [reviewed in (Pfeifer, 2006) and (Walsh and Xu, 2006)]. Deamination of unmethylated cytosine to uracil generates a UG mismatch, which can be readily recognized and corrected by the DNA repair system involving uracil DNA glycosylase. In contrast, if methylated cytosine (that mostly occurs within CpG dinucleotides) is deaminated to thymine, the emerging TG base pair is not as efficiently repaired (see also chapters 1.1.5.2 and 1.1.6). In consequence, methylated cytosines tend to mutate to thymines over the evolutionary time course if methylated in the germ line, leading to the underrepresentation of CpG sites in the mammalian genome. CpG depletion is very pronounced within the repetitive sequences of transposable elements: LINE transposons and LTRs of endogenous retroviruses (18-19% of expected) as well as SINE transposons (41% of expected, with mainly quite young Alu SINE transposons) (Lander et al., 2001). In contrast, CpG islands show the lowest depletion levels resulting in more than 50% of expected CpG sites. These numbers suggest that CpG islands are largely unmethylated in the germ line, whereas transposons are methylated (Rollins et al., 2006).

Early gene expression studies suggested that promoter methylation can lead to stable gene silencing, and that intragenic methylation represses transposable elements and reduces transcriptional noise (Bird, 1995). Recently, a genome-wide study explored the relationship between promoter methylation and gene expression using data from the ENCODE project (Birney and consortium, 2007). They found that highly expressed genes indeed show a pattern of low promoter methylation and higher gene body methylation, whereas the weakly expressed genes were moderately methylated over both regions (Ball et al., 2009). Moreover, this study revealed that gene body methylation is a general feature of the human genome and it was suggested to reflect the ancestral function of DNA methylation in animals, i.e. the reduction of transcriptional noise (Ball et al., 2009; Suzuki and Bird, 2008).

1.1.4.2 CPG ISLANDS AND PROMOTER REGULATION

CpG islands are DNA stretches of approximately 1 kb in length characterized by an elevated C/G content and the overrepresentation of CpG sites by 10 to 20 times their average density (Illingworth and Bird, 2009). Consistently, CpG islands represent only 0.68 % of the genome but contain 6.8 % of all CpG sites (Rollins et al., 2006). Not only in the germ line as mentioned above, but also in somatic cells, CpG islands are often unmethylated and overlap with promoter regions. Interestingly, they occur at promoters of most constitutively expressed genes (housekeeping genes) and at some promoters of tissue-specific genes (Bird, 1986; Gundersen et al., 1992; Larsen et al., 1992; Zhu et al., 2008). Exceptions to the rule of unmethylated CpG islands promoters include the inactive X chromosome, some silent imprinted genes, and some tissue-specific genes.

It was suggested that CpG island promoters could define a class of transcription start sites, which can, in contrast to non-CpG island promoters, initiate transcription from multiple positions (Sandelin et al., 2007). However, not all CpG islands are localized to annotated transcription start sites. Besides the possibility of actual CpG island occurrence outside of transcription start sites, this could be due to the fact that not all existing transcription start sites are identified or that the prediction of CpG islands is not sufficiently accurate or both. Indeed, CpG island annotation has led to the discovery of additional, previously not annotated transcription start sites. Additionally, CpG island prediction is somehow arbitrary, since the results of prediction algorithms highly depend on the chosen parameters for C/G content, CpG site frequency and island length (Illingworth and Bird, 2009). The most recent computational predictions in accordance with experimental evidence suggest an overall number of 24000 to 27000 CpG islands distributed throughout the human genome (Illingworth et al., 2008; Illingworth and Bird, 2009).

Recently, a definition of promoter classes was proposed based on CpG content and methylation levels, distinguishing between non-CpG island promoters and CpG island promoters (Weber et al., 2005). Their chromosome-wide analysis revealed one striking difference between these promoter classes: whereas non-CpG island promoters were highly methylated in most cases, strong CpG island promoters were methylated to only 3 % (Figure 2). It is interesting to note that for this latter class of CpG island promoters, methylation levels differ between autosomal and X-inactivated genes suggesting a role for CpG island promoters in X inactivation (Weber et al., 2005). Moreover, it has been shown that a fraction of CpG islands is differentially methylated between different somatic tissues and cell types (Illingworth et al., 2008; Illingworth and Bird, 2009). This suggests that a few CpG island promoters become methylated during normal development. Indeed, some CpG island promoters of developmental or germ-line specific genes have been shown to be methylated during embryogenesis, thereby leading to persistent silencing of these genes in somatic tissues (Weber et al., 2007). Thus, CpG island promoter methylation plays a profound role in differentiation and development, X inactivation and genomic imprinting by transcriptional silencing of associated genes. Moreover, there is also substantial evidence for differential methylation of promoters and enhancers of low CpG density during development and differentiation (Fouse et al., 2008; Meissner et al., 2008; Mohn et al., 2008).

CpG island promoter methylation always results in stable transcriptional repression of the associated genes (Bird, 2002), whereas the transcriptional state of genes associated with non-CpG island promoters does not reflect the methylation state of the promoter (Figure 2). This suggests that low concentrations of methylated CpG sites do not preclude gene activity and that transcriptional repression by DNA methylation in promoter regions requires high levels and density of 5-methylcytosine (Weber et al., 2007). Notably, DNA methylation is sufficient but not necessary to inactivate CpG island promoters, as silencing of CpG island promoters is not always associated with promoter methylation. This means that promoter methylation is not the only route towards gene inactivation and that there are other mechanisms for transcriptional silencing. Whether and how CpG islands distal to transcription start sites contribute to transcriptional regulation of gene expression is still poorly understood (Illingworth and Bird, 2009).

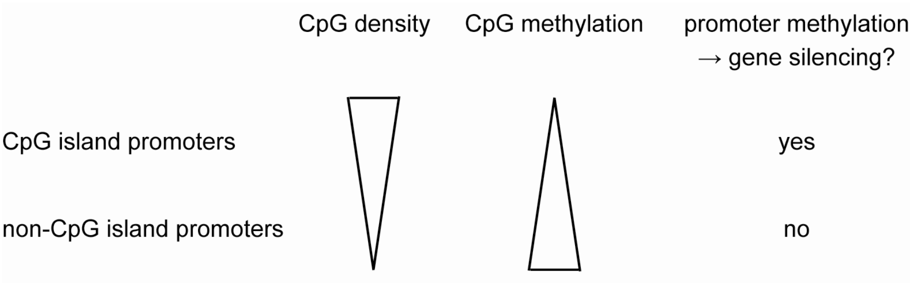


Figure 2. Regulation of gene expression by promoter methylation. Promoters can be subdivided into two principle classes: CpG-rich CpG island promoters and CpG-poor non-CpG island promoters. Characteristically, CpG density and CpG methylation levels negatively correlate within these promoter sequences. Interestingly, this anti-correlation is also observed for the total genomic DNA sequence. Whereas CpG island promoter methylation leads to gene silencing, non-CpG island promoter methylation does not preclude gene expression.

As mentioned above, the characteristic clustering of CpG sites within CpG islands is thought to be a consequence of two factors: global loss of CpG sites due to mutagenic deamination of 5-methylcytosine and resistance to de novo methylation during early development. Still, very little is known about how CpG islands are maintained unmethylated during the wave of global de novo methylation in early development. However, protection against methylation seems to involve active chromatin marks like H3K4 dimethylation (Weber et al., 2007) and H3K9/14 acetylation (Roh et al., 2005). Furthermore, there is emerging evidence for a role of the zinc finger protein VEZF1 binding to G-rich methylation protection elements, which has been shown to maintain the APRT CpG island promoter unmethylated (Dickson et al., 2010). A recent study also showed that CXXC finger protein 1 (CXXC1/Cfp1) binds to unmethylated CpG islands and thereby influences chromatin structure (Thomson et al., 2010). Interestingly, whereas the majority of inactive strong CpG island promoters remains unmethylated, a much higher proportion of inactive weak CpG promoters, with lower CpG density, becomes methylated. This implies that protection of CpG islands against de novo methylation might be based on their CpG density (reviewed and discussed in (Illingworth and Bird, 2009)).

1.1.4.3 REPETITIVE DNA SEQUENCES

Repetitive elements including transposable elements and satellites make up almost 50 % of the mouse and human genome, and the majority of 5-methylcytosine is found within these sequences (Ikegami et al., 2009). The methylation of these sequences has been suggested to be responsible for transposon suppression in order to accomplish genome stability (Yoder et al., 1997) and to reduce transcriptional noise (Bird, 1995). Indeed, endogenous retroviruses become transcriptionally silenced during early embryogenesis and aberrant expression of retroviral sequences has been shown to induce cancerous transformations in somatic cells (reviewed in (Maksakova et al., 2008). Significantly, DNA hypomethylation in mice leads to development of aggressive tumors associated with activation of endogenous retroviral elements (Gaudet et al., 2003; Gaudet et al., 2004; Howard et al., 2008; Walsh et al., 1998). These studies provide strong evidence for a role of DNA methylation in retroviral silencing in somatic cells and tissues. Recently however, a methylation-independent pathway for silencing of IAP retrotransposons has been proposed to occur in embryonic stem cells (Deng et al., 2009). The protein KAP1 has been shown to control this process (Rowe et al., 2010) by recruiting the histone methyltransferase SETDB1, heterochromatin binding protein HP1 and the NuRD histone deacetylase complex (Matsui et al., 2010).

1.1.4.4 TEMPORAL DISTRIBUTION OF 5-METHYLCYTOSINE

Global changes of methylation levels and patterns occur at two stages of mammalian development: in the early embryo shortly after fertilization and in primordial germ cells during gametogenesis. These changes include a phase of global demethylation to almost complete 5-methylcytosine depletion, followed by a phase of sequence-specific de novo methylation. However, the demethylation process does not affect all sequences. The first phase of demethylation occurs during preimplantation development (Monk, 1987; Rougier et al., 1998) and, importantly, it does not affect imprinted genes and repetitive sequences. However, all other epigenetic marks have to be erased in the early embryo in order to reset cellular gene expression programs and to generate a pluripotent state. In primordial germ cells, methylation patterns are almost completely erased including the methylation marks of imprinting genes (Monk, 1987). Remarkably, a few, primarily IAPs, but not all transposons remain methylated at this step (Reik, 2007). DNA methylation in male and female germ cells is subsequently re-established with sex-specific patterns (Chaillet et al., 1991). To date, it is not completely understood how methylation of some sequences is selectively maintained against a background of widespread demethylation.

Furthermore, as introduced above, methylation of specific promoter sequences changes throughout development, and DNA methylation patterns differ markedly between different cell types. On the one hand, differentiation-specific genes have to be kept inactive in pluripotent cells until differentiation is triggered and on the other hand, pluripotency–associated genes become inactivated in differentiated cells (Reik, 2007). For example, key transcription factors as oct4 and nanog form a regulatory network for specific gene transcription in pluripotent ES cells. Their promoter regions become methylated and silenced upon differentiation and this methylation-associated inactivation of pluripotency-associated genes is very stable in differentiated cells. Likewise, transposons have to be stably silenced to insure genome stability and also the transcriptional activity of imprinted genes and the inactive X chromosome has to be permanently repressed. Significantly, all processes that require stable and irreversible gene silencing involve DNA methylation. In contrast, the inactivation of differentiation-specific genes in pluripotent cell has to be reversible, and is mostly accomplished by rather dynamic epigenetic mechanisms involving histone modifications and Polycomb group proteins (Reik, 2007).

1.1.5 REGULATION OF GENE EXPRESSION BY DNA METHYLATION

DNA methylation is essential for viability and genome integrity and it can be associated with regulation of gene expression. Two principle molecular mechanisms have been shown to mediate transcriptional control by DNA methylation: either the methyl group directly prevents binding of transcriptional regulators, or it leads to specific binding of methyl-CpG binding proteins.

1.1.5.1 REGULATION VIA TRANSCRIPTIONAL REGULATOR BINDING

It is well established that the methylation of 5-methylcytosine can directly interfere with binding of transcriptional regulators (Becker et al., 1987). In general, DNA methylation is associated with gene silencing. However, this is not the only possible consequence of DNA methylation. In fact, it is interesting to note that in one well studied example, DNA methylation leads to activation of gene expression at the imprinted H19/Igf2 locus. The protein CTCF usually functions as insulator by blocking the action of an enhancer signal downstream the Igf2 gene promoter, and thereby represses gene expression. On the paternally imprinted gene copy however, where the CTCF binding site is methylated, CTCF binding and function are impaired resulting in active Igf2 transcription (Bell and Felsenfeld, 2000; Hark et al., 2000).

1.1.5.2 REGULATION VIA METHYL-CPG BINDING PROTEINS

Methylated CpG sites are recognized by a family of methyl-CpG binding proteins (MBPs) that translate DNA methylation marks into specific chromatin states by recruitment of chromatin modifiers and remodeling complexes. Generally, these MBPs are thought to work as transcriptional repressors by binding to DNA and recruiting corepressors. The complexity of this process and the resulting protein interaction network are enormous and it is not clear whether corepressors work simultaneously as complex or individually on different targets at different cell-cycle or developmental stages. Up to date, there are three families of MBPs known, which recognize methylated CpG sites via distinct domains: the methyl-CpG binding domain (MBD) family, the Uhrf family and the Kaiso protein family (Figure 3).

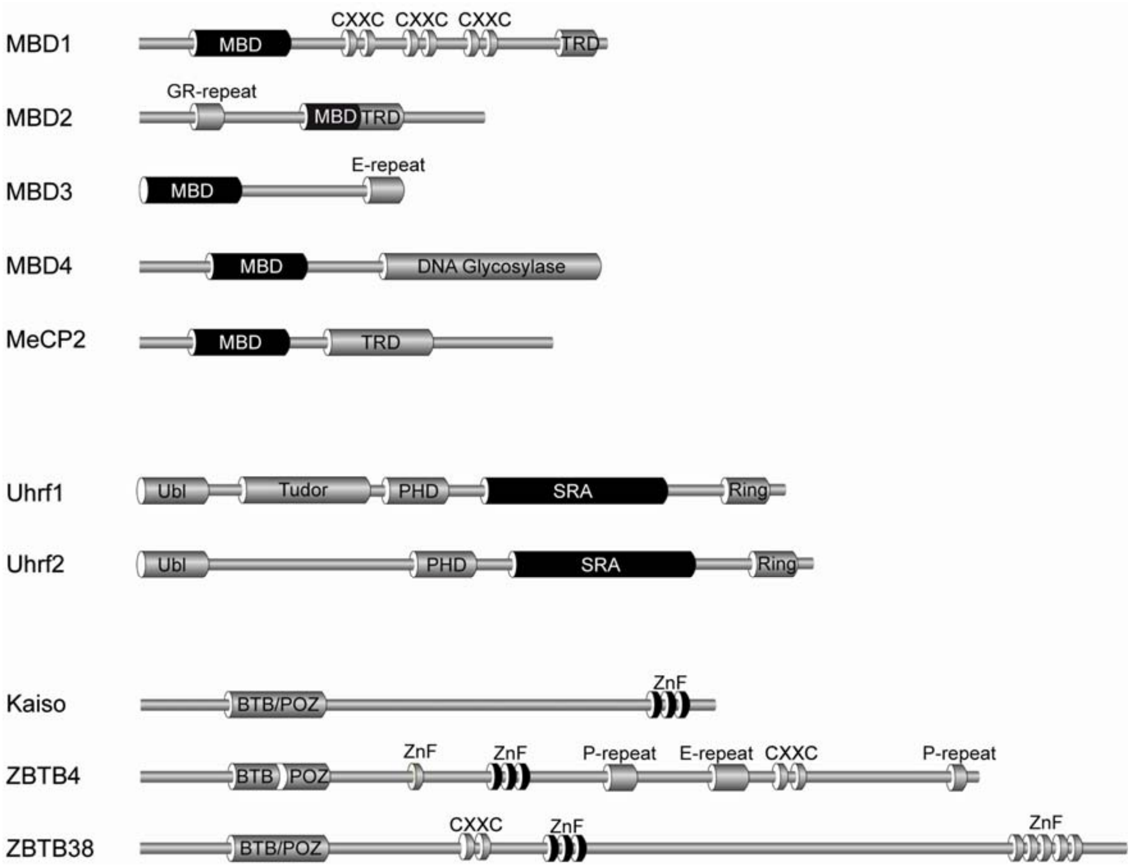

MBD1
MBD
CXXC CXXC CXXC
TRD
MBD2
GR-repeat
MBD TRD
MBD3
MBD
E-repeat
MBD4
MBD
DNA Glycosylase
MeCP2
MBD
TRD
Uhrf1
Ubl
Tudor
PHD
SRA
Ring
Uhrf2
Ubl
PHD
SRA
Ring
Kaiso
BTB/POZ
ZnF
ZBTB4
BTB POZ
ZnF
ZnF
P-repeat
E-repeat
CXXC
P-repeat
ZBTB38
BTB/POZ
CXXC
ZnF
ZnF

Figure 3. Domain structures of methyl-CpG binding proteins. Members of the MBD family bind to methylated DNA via their MBD domain, Uhrf proteins via their SRA and Kaiso(-like) proteins via their Krüppel-like C_2H_2 ZnF domain. These DNA binding motifs are highlighted in black. MBD: methyl-CpG binding domain; CXXC: CXXC zinc finger domain; TRD: transcriptional repression domain; Ubl: ubiquitin-like domain; SRA: SET and Ring-associated domain; Ring: really interesting new gene; BTB/POZ: broad complex, tramtrack and bric à brac/poxvirus and zinc finger domain; ZnF: zinc finger domain, GR-, E- and P-repeat: Gly-Arg, Glu and Pro amino acid repeats, respectively. (modified from (Rottach et al., 2009b))

The MBD protein family comprises five members (MBD1, MBD2, MBD3, MBD4 and MeCP2) and, with the exception of MBD3, all specifically bind to methylated CpG sites via their MBD domain (Figure 3). Furthermore, all MBD family members interact with and recruit histone deacetylases (HDACs) and all but MBD4 recruit nucleosome remodeling complexes (NuRD), both associated with an inactive chromatin state and transcriptional silencing (Hendrich and Tweedie, 2003; Kondo et al., 2005; Nan et al., 2007). Furthermore, MBD1 and MeCP2 interact with and recruit histone H3K9 methyltransferases, which again set modification marks characteristic for silent chromatin (Fujita et al., 2003; Sarraf and Stancheva, 2004). In addition, MBD1 and MeCP2 interact with the heterochromatin binding protein HP1 again inducing transcriptional silencing (Agarwal et al., 2007;

Fujita et al., 2003). Importantly, MBD1 and MeCP2 have been shown to bind DNA and compact chromatin also independently of DNA methylation (Brero et al., 2005; Georgel et al., 2003; Jorgensen et al., 2004; Nikitina et al., 2007). Moreover, MBD proteins do not always mediate transcriptional silencing. The majority of MeCP2 target genes in neurons are transcriptionally active (63%) and only a minority of MeCP2 target promoters is highly methylated (6%) (Yasui et al., 2007). Interestingly, MBD4 contains a thymine glycosylase domain and functions as a DNA repair enzyme for TG mismatches generated by deamination of 5-methylcytosine (Walsh and Xu, 2006) (see also chapter 1.1.4.1). MBD4 was also suggested to function in active DNA demethylation (discussed in chapter 1.1.6). Considering the severe effects of DNA hypomethylation, it is surprising that mice lacking MBD family members show only very mild phenotypes. This finding suggests functional redundancy, but, the hypothesis of redundancy contrasts not only with the evidence for different functions of MBD proteins due to different knock-out phenotypes, but also with the diversity of sequence and structure between MBD members outside their MBD domain (reviewed in (Sasai and Defossez, 2009), Figure 3). In other words, the precise targets and functions of individual MBD family members remain elusive.

Uhrf1 (also called Np95 or ICBP90) has been shown to bind DNA containing hemimethylated CpG sites via its SET- and Ring-associated (SRA) domain (Figure 3). In analogy to MBD family members, Uhrf1 was reported to interact with both, histone H3K9 methyltransferase G9a and HDAC1, and it was suggested to be involved in the silencing of tumor suppressor genes in breast cancer cells (Kim et al., 2009a; Unoki et al., 2004). In addition, Uhrf1 was shown to bind to histones via three distinct domains. First, Uhrf1 binds histone H3 and functions as E3 ubiquitin ligase (Citterio et al., 2004). Second, crystallographic data indicated specific binding to histone trimethylated H3K9 tails via the tandem Tudor domain (PBD 3DB3). Third, Uhrf1 contains a plant homeodomain (PHD) domain that has been implicated in binding to histone H3. The PHD domain also seems to be required for large-scale rearrangements of chromocenter structures by Uhrf1 (Papait et al., 2008). Thus, Uhrf1 was proposed to provide a direct link between DNA methylation and histone modification. Recently, Uhrf1 has been shown to also directly interact with Dnmt1 (Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007b; Rottach et al., 2009a; Sharif et al., 2007). Its genetic ablation leads to remarkably similar phenotypes to those of Dnmt1 ablation including severely reduced DNA methylation levels (Sharif et al., 2007). These data strongly suggest that Uhrf1 is an essential cofactor for maintenance methylation (see also chapter 1.3.4). Moreover, Uhrf1 has recently been shown to interact with all three Dnmts and to be required for transgene silencing (Meilinger et al., 2009). Interestingly, Uhrf1 has a homolog in mammals, Uhrf2 (also called Np97), but so far it is not known whether Uhrf2 has a role in controlling or reading DNA methylation.

The protein Kaiso was identified in HeLa cell nuclear extracts and was shown to bind methylated DNA via a Krüppel-like C_2H_2 zinc finger motif ((Yoon et al., 2003), Figure 3). In contrast to other methyl-CpG binding proteins of the MBD and Uhrf families, stable DNA binding of Kaiso requires two consecutive methylated CpG sites. Kaiso has been implicated in methylation-dependent repression of the MTA2 gene locus by recruiting the nuclear corepressor (N-CoR) complex, which leads to hypoacetylation and H3K9 methylation at the promoter region (Yoon et al., 2003). Kaiso belongs to the family of BTB/POZ motif containing transcription factors that includes two further members with a very similar zinc finger motif: Kaiso-like proteins ZBTB4 und ZBTB38 (Figure 3). These proteins have also been suggested to bind to DNA, but unlike Kaiso, to single methylated CpG sites. They have been implicated in silencing of the imprinted H19/Igf2 locus (Filion et al., 2006). In contrast to Kaiso, which recruits the N-CoR complex, transcriptional repression by the two Kaiso-like proteins ZBTB38 and ZBTB4 involves the CtBP and Sin3/HDAC corepressor complexes, respectively. Furthermore, Kaiso and ZBTB4 have been shown to bind a specific unmethylated DNA sequence (Kaiso binding sequence, KBS), which, in the case of Kaiso, leads to transcriptional repression of factors involved in Wnt signaling (Kim et al., 2004). Ultimately, all remodeling complexes, which are recruited by Kaiso proteins, contain HDACs and other remodeling activities, leading to transcriptional silencing. Whether Kaiso and Kaiso-like proteins are functionally redundant is not clear yet (reviewed in (Sasai and Defossez, 2009)).

1.1.5.3 INTERCONNECTION OF DNA METHYLATION WITH OTHER EPIGENETIC PATHWAYS

As mentioned above, DNA methylation is translated by methyl-CpG binding proteins into specific chromatin states by recruiting chromatin modification and remodeling factors. In addition, direct links have been identified between Dnmts with proteins of histone modification and chromatin remodeling pathways (Figure 4). More specifically, Dnmts have been shown to interact with the H3K9 methyltransferases Suv39h1 (Fuks et al., 2003), SetDB1 (Li et al., 2006), and G9a (Esteve et al., 2006), histone deacetylases (HDACs) (Fuks et al., 2000; Fuks et al., 2001; Geiman et al., 2004; Robertson et al., 2000a) and SNF2H (Geiman et al., 2004), the ATPase subunit of several chromatin remodeling complexes. Notably, the interaction of Dnmt1 with G9a at replication foci could allow coordinated replication of DNA and H3K9 methylation (Esteve et al., 2006). Another interaction possibly helping to link the replication of DNA and histone modifications is the interaction of MBD1 with SetDB1 and Suv39h1. As mentioned above MBD1 also interacts with HP1 (Fujita et al., 2003; Sarraf and Stancheva, 2004). HP1 in turn binds to Dnmt1, H3K9 methyltransferases, methylated H3K9 and MeCP2 and induces transcriptional silencing (Agarwal et al., 2007; Fuks et al., 2003; Lachner et al., 2001; Nielsen et al., 2002). Moreover, Dnmt3s have been shown to specifically bind particular histone modification marks. Dnmt3a binds specifically to methylated H4R3, a modification set by arginine methyltransferase PRMT5 (Zhao et al., 2009). Dnmt3a, Dnmt3b and Dnmt3L bind specifically to unmethylated H3K4 (Jia et al., 2007; Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010). This implies a mechanism for controlling de novo methylation during early development, whereby CpG island promoter occupancy by RNA polymerase II recruits H3K4 methyltransferases, leading to H3K4 methylation and prevention of de novo methylation by Dnmt3s (Guenther et al., 2007). Also, a role of non-coding RNA and transcription factors in directing Dnmt3 enzymes has been suggested (Aravin and Bourc'his, 2008; Hervouet et al., 2009). Furthermore, Dnmts have been shown to interact with components of the Polycomb repressive complex 2 (PRC2). Polycomb group proteins represent besides DNA methylation a second essential epigenetic system that heritably represses transcription. This important interaction might thus establish feedback loops and an interaction network that stabilizes and spreads silent chromatin states (Vire et al., 2005).

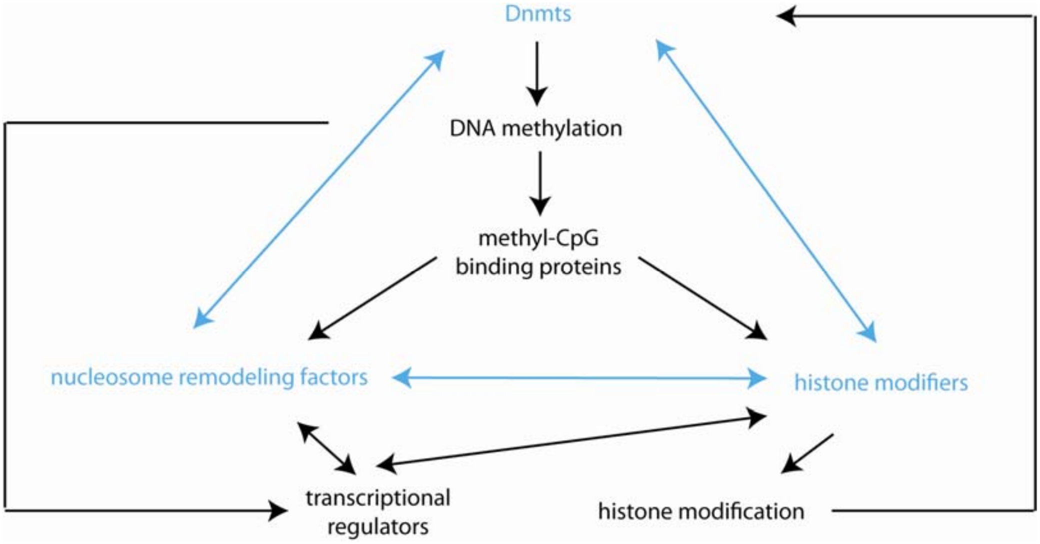

Dnmts
DNA methylation
methyl-CpG binding proteins
nucleosome remodeling factors
histone modifiers
transcriptional regulators
histone modification

Figure 4. Interconnection of DNA methylation with other epigenetic pathways. The chromatin modifiers Dnmts, histone modifiers and nucleosome remodeling factors directly and indirectly interact. These epigenetic pathways work in concert to generate chromatin activity states which finally control gene expression. In addition, gene regulatory proteins of the Polycomb/Trithorax group and non-coding RNA are suggested to interact with Dnmts and implicated in transcriptional regulation (not shown). Bidirectional arrows in the network indicate reciprocal ‘interaction’, unidirectional arrows indicate ‘generation’ if starting from a blue node or ‘recruitment’ if starting from a black node.

Regarding the high level of interconnection of DNA methylation with other epigenetic pathways, it is not surprising that the correct establishment and maintenance of DNA methylation patterns in the mammalian genome requires several factors in addition to active Dnmts. Genetic targeting of the following factors in mice or murine cells led to hypomethylation and/or defects in methylation patterns: Dnmt3L (Bourc'his et al., 2001), Uhrf1 (Sharif et al., 2007), the unmethylated CpG binding protein CXXC1 (CXXC finger protein 1/Cfp1, (Carlone et al., 2005)), the Zinc finger protein 57 (ZFP57, (Li et al., 2008)), the histone methyltransferases G9a (Ikegami et al., 2007) and Suv39h1/2 (Lehnertz et al., 2003), the lysine-specific H3K4 and H3K9 demethylase LSD1 (which also demethylates Dnmt1, (Wang et al., 2009)), histone H1 (Fan et al., 2005), the SNF2 related chromatin-remodelers ATRX (Gibbons et al., 2000) and Lsh2 (Dennis et al., 2001), and the Argonaute proteins MILI, MIWI2 (Aravin et al., 2007; Carmell et al., 2007). For many of these factors, it is not clear yet, whether they cause methylation defects due to effects on establishment or maintenance of methylation marks (reviewed in (Ooi and Bestor, 2008b)).

The functional hierarchy and sequence of events leading to gene silencing is often difficult to establish. However, there is increasing evidence, that DNA methylation might be a late rather than the initiating event. For example in teratocarcinoma cells, complete methylation of retroviral DNA is observed only 15 days after induction of differentiation, whereas transcription of these sequences is already repressed at day 2 (Gautsch and Wilson, 1983; Niwa et al., 1983). Moreover, although de novo methylation of proviral DNA in embryonic cells depends on both Dnmt3a and Dnmt3b (Okano et al., 1999), knock-out of both methyltransferases does not interfere with initial retroviral silencing (Pannell et al., 2000). Also, expression of the Xist RNA triggers silencing of one X chromosome before methylation of the CpG island promoter or hyperacetylation of histone H4 occur (Keohane et al., 1996; Lock et al., 1987; Wutz and Jaenisch, 2000). However, DNA methylation stably and irreversibly represses the inactive X chromosome as X inactivation can be reversed by silencing of Xist RNA expression within the first 72 hours of differentiation, but not at a later time point, when CpG island promoters are methylated. These observations lead to the assumption that the role of DNA methylation is primarily to lock genes in a silent state, which were initially repressed by other mechanisms.

1.1.6 DNA METHYLATION DYNAMICS

As described in chapter 1.1.4.4, there are two waves of global DNA demethylation during mammalian development: one shortly after fertilization and one in primordial germ cells. In contrast to the maternal genome, which is thought to undergo passive demethylation, the paternal genome is thought to be actively demethylated after fertilization (Mayer et al., 2000; Oswald et al., 2000). Likewise, DNA demethylation in primordial germ cells was suggested to be an active process (Hajkova et al., 2008; Hajkova et al., 2002). Passive demethylation is achieved by simply preventing maintenance of methylation after DNA replication. Indeed, demethylation of the genome in the early embryo occurs when Dnmt1, present in a shorter oocyte-specific isoform (Dnmt1o), is retained in the cytoplasm. Dnmt1o only shortly enters the nucleus at the 8-cell stage, which has been suggested to maintain genomic imprints ((Cardoso and Leonhardt, 1999; Carlson et al., 1992; Hirasawa et al., 2008; Howell et al., 2001; Mertineit et al., 1998), see also chapter 1.3.1). For active DNA demethylation there are primarily two possibilities: i) Breakage of the carbon-carbon bond between the C^5 of the nucleobase and the carbon of the methyl group, ii) Excision of the methylated base or nucleotide possibly with involvement of the repair machinery, either the base excision repair (BER) or the nucleotide excision repair (NER). Proposed mechanisms for active DNA demethylation are discussed in the following paragraphs and summarized in Figure 5.

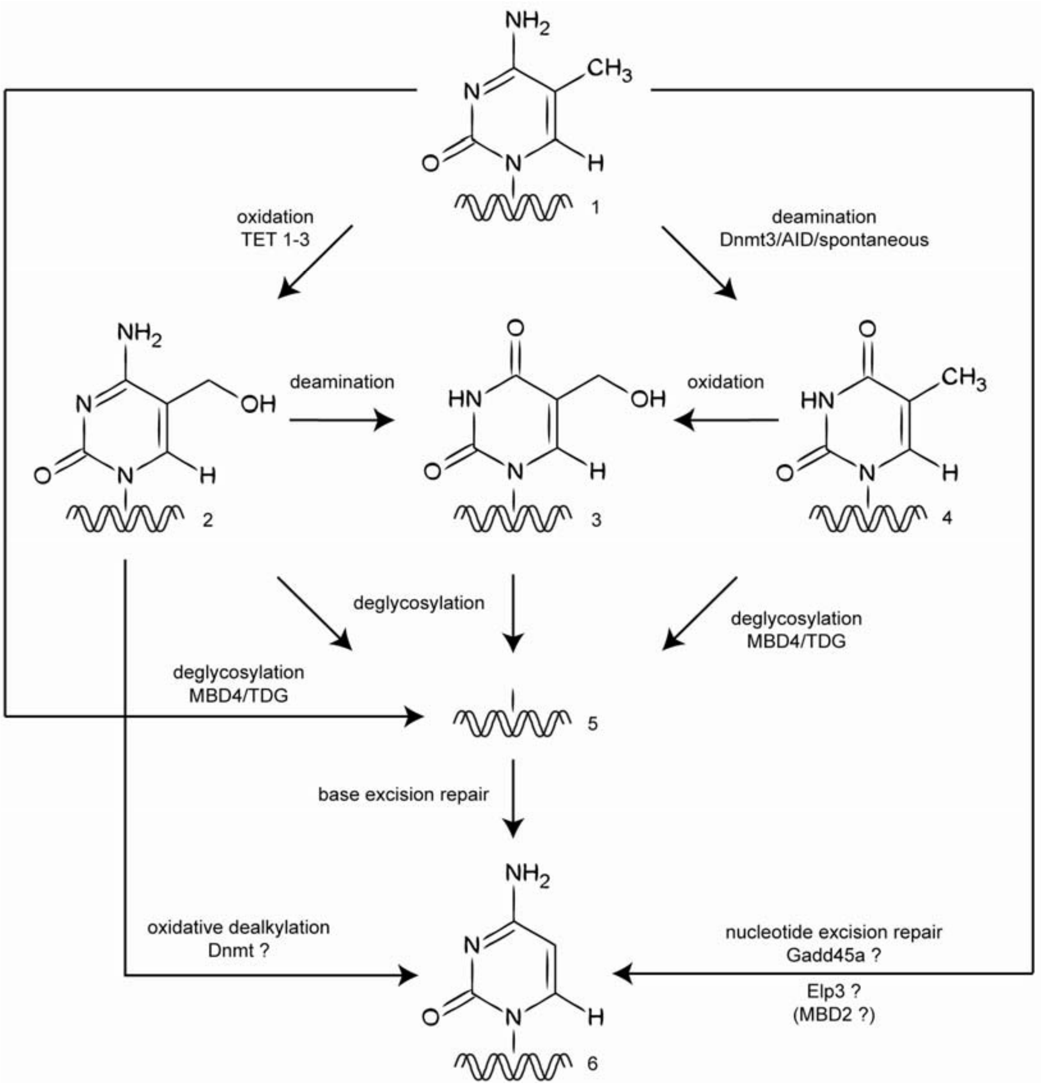

NH2
CH3
N
O
N
H
1
oxidation
TET 1-3
deamination
Dnmt3/AID/spontaneous
NH2
OH
2
deamination
O
HN
OH
3
oxidation
O
HN
CH3
4
deglycosylation
deglycosylation
MBD4/TDG
deglycosylation
MBD4/TDG
5
base excision repair
oxidative dealkylation
Dnmt ?
nucleotide excision repair
Gadd45a ?
Elp3 ?
(MBD2 ?)
6

Figure 5. Proposed mechanisms for active DNA demethylation. Demethylation of 5-methylcytosine (1) to cytosine (6) possibly involves the intermediates 5-hydroxymethylcytosine (2), 5-hydroxymethyluracil (3), thymine (4), and/or an abasic site (5) generated by oxidation, deamination and/or deglycosylation, which activates base excision repair. Moreover, it was suggested that 5-hydroxymethylcytosine can be dealkylated to cytosine. Alternatively, 5-methylcytosine was suggested to be converted to cytosine by induction of the nucleotide excision repair pathway or involving the elongator complex. Further information about the factors potentially involved in these mechanisms can be found in the text.

Whereas methyltransferases, the enzymes that set and maintain methylation marks, are well characterized and known since years, there is no clear evidence for a demethylase enzyme (Ooi and Bestor, 2008a). However, early evidence for a mechanism of active demethylation in the mammalian genome came from a study in 1982 (Gjerset and Martin, 1982), in which the authors describe an enzymatic demethylation activity in the nucleoplasm of murine erythroleukemia cells. Since then, there have been other studies showing demethylase activities in cell extracts and even providing evidence for several very different candidates to be involved in active demethylation. Of all proteins suggested to be involved in active demethylation, exclusively MBD2 (Bhattacharya et al., 1999; Ramchandani et al., 1999) has been claimed to remove methyl groups without the need for any other protein or pathway to be involved. However, the demethylation activity of MBD2 could not be reproduced by several groups (Ooi and Bestor, 2008a).

Processes involving the repair machinery have been suggested to start either directly with a deglycosylation event (base excision) or alternatively, with a deamination event resulting in a TG mismatch, which can be recognized by a thymine DNA glycosylase. Deglycosylation events generate an abasic site, which can be subsequently replaced with cytosine via BER involving an endonuclease, polymerase and ligase. Interestingly, the thymine DNA glycosylases TDG (Zhu et al., 2000b) and MBD4 (Zhu et al., 2000a) have been suggested to possess in addition to their thymine DNA glycosylase also 5-methylcytosine glycosylase activity. The 5-methylcytosine glycosylase activity of both TDG and MBD4 could be enhanced by other factors of the glycosylase complex or sequences flanking the methylated CpG site. The MBD4 5-methylcytosine glycosylase activity is very low *in vitro*, however it has been recently shown to be enhanced by phosphorylation (Kim et al., 2009b). Still, it is not known whether these activities are sufficient for a role of these glycosylases in global demethylation or rather in active demethylation of tissue-specific genes during development. In case of initial deamination of 5-methylcytosine to thymine, TDG and MBD4 could be involved in demethylation employing their thymine DNA glycosylase activity again producing an abasic site, which could be subsequently repaired by BER. Alternatively, thymine can be oxidized to 5-hydroxymethyluracil and subsequently excised by a DNA glycosylase activity (Cannon-Carlson et al., 1989).

In addition to spontaneous deamination of 5-methylcytosine to thymine, deaminase enzymes could catalyze this reaction. Interestingly, both Dnmt3a and Dnmt3b have been suggested to possess deaminase activity and have been reported to be responsible for dynamic transcriptional regulation in human cells by cyclic changes in the methylation status of the pS2/TFF1 gene promoter (Kangaspeska et al., 2008; Metivier et al., 2008). In these studies, Dnmt3a and Dnmt3b have been reported to be responsible for both methylation and demethylation of the involved gene promoters by induction of DNA repair. Intriguingly, it has been shown that limiting concentrations of the methyl donor AdoMet can lead to deamination of both cytosine and 5-methylcytosine by prokaryotic methyltransferases, since in this case the enamine form as intermediate of the methyl transfer reaction destabilizes the exocyclic C4 amine and increases the rate of oxidative deamination (Shen et al., 1992). The deamination of 5-methylcytosine would result in a TG mismatch, which could be again recognized by TGD and MBD4 leading to thymine excision and replacement with cytosine by repair (BER). In addition, activation–induced cytidine deaminase (AID) has been shown to deaminate 5-methylcytosine in DNA creating a C to T transition (Morgan et al., 2004). Recently, this AID-dependent DNA demethylation has been suggested to be required for reprogramming and generation of pluripotent IPS cells (Bhutani et al., 2010).

Gadd45 protein family members have been suggested as candidates for induction of NER-dependent demethylation in order to avoid the accumulation of C to T transitions. Gadd45α (growth arrest and DNA-damage-inducible protein 45α) is a nuclear protein and was shown to interact with the NER repair endonuclease XPG leading to demethylation by DNA repair upon deamination. TAF12 was later shown to recruit Gadd45α and the repair machinery to these sites of active DNA demethylation (Barreto et al., 2007; Ma et al., 2009; Rai et al., 2008; Schmitz et al., 2009) . However, the involvement of Gadd45α is still controversially discussed (Jin et al., 2008) and has rather been suggested to be involved in the demethylation of tissue-specific gene promoters.

An additional possibility for active demethylation would be the modification of 5-methylcytosine to a base that allows breakage of the carbon-carbon bond, by shifting chemical properties. Indeed, the mammalian genome was recently shown to contain 5-hydroxymethylcytosine in certain cell types (Kriaucionis and Heintz, 2009). This modification has been reported to result from oxidation of 5-methylcytosine by TET proteins, and has been suggested as a route towards active demethylation (Tahiliani et al., 2009). In this regard, 5-hydroxymethylcytosine has been proposed to be dealkylated directly resulting in cytosine (Liutkeviciute et al., 2009) and also to induce demethylation by a process involving deamination, deglycosylation and BER (Cannon et al., 1988).

Recently, live cell imaging monitoring the paternal DNA methylation state in zygotes in combination with siRNA knock-down (Okada et al., 2010) identified the elongator complex component Elp3 (or KAT9) as candidate for active demethylation. Moreover, knock-down of other elongator components (Elp1 and Elp4) impaired demethylation, suggesting the elongator complex to be involved in active DNA demethylation.

1.2 THE DNMT1 ENZYME

DNA methylation plays a profound role in epigenetic gene regulation and Dnmt1 is the most ubiquitous DNA methyltransferase in mammals, responsible for maintenance methylation during semi-conservative DNA replication (see chapter 1.1.3.1). The murine enzyme was first cloned and sequenced by Bestor and coworkers in 1988 (Bestor et al., 1988). Four years later, in 1992, Yen cloned and isolated the cDNA for human Dnmt1 (Yen et al., 1992). The precise extension of the N-terminus of the murine enzyme, however, was only defined in 1996 by Tucker and coworkers, showing that the N-terminal 171 amino acids, which were not included in the first sequence, are necessary for stable Dnmt1 expression and function (Tucker et al., 1996). In total, the somatic form of murine Dnmt1 comprises 1619 amino acids, has a molecular weight of about 180 kDa and comprises several distinct domains (Figure 6). Human Dnmt1 is 3 amino acids shorter. Mouse and human Dnmt1 protein sequences share 78 % identity.

1.2.1 STRUCTURE

Dnmt1 is a large enzyme with a complex domain structure and likely evolved by fusion of at least three genes (Margot et al., 2000). To date, there is no three-dimensional structure of the complete Dnmt1 enzyme known; only a crystal structure of the TS domain is available (PBD: 3EPZ, (Fellinger et al., 2009)). All structural information presented in the following paragraphs is thus based on biochemical studies and/or the comparison with structures of prokaryotic enzymes.

The Dnmt1 enzyme comprises a regulatory N-terminal domain, which is connected to its C-terminal catalytic domain by a linker of seven glycine-lysine repeats. As mentioned above, the C-terminal domain of Dnmt1 shows high sequence similarity to prokaryotic DNA methyltransferases and was thereby early identified as the domain responsible for catalysis of the methyl transfer reaction (Figure 6). Sequence alignments identified up to ten conserved motifs for the family of DNA (cytosine-5) methyltransferases, with motifs I, IV, VI, VIII, IX and X being highly conserved ((Cheng et al., 1993; Kumar et al., 1994)). The crystal structure of the prokaryotic enzyme M.HhaI revealed that single residues of motifs I, II, III, IV, V and X contribute to the binding pocket for the cofactor AdoMet, and that the combined region of motifs I, II and III strongly resembles the Rossman fold of a dinucleotide-binding motif. Importantly, the invariant Pro-Cys dipeptide in motif IV is part of the catalytic site (Kumar et al., 1994). Between motifs VIII and IX resides the target recognizing domain (TRD), which is significantly longer in Dnmt1 than in prokaryotic DNA methyltransferases. This region has been shown to determine the sequence specificity of the prokaryotic enzymes (Lauster et al., 1989; Wilke et al., 1988), whereas the functional role of the TRD in mammalian enzymes is still unclear.

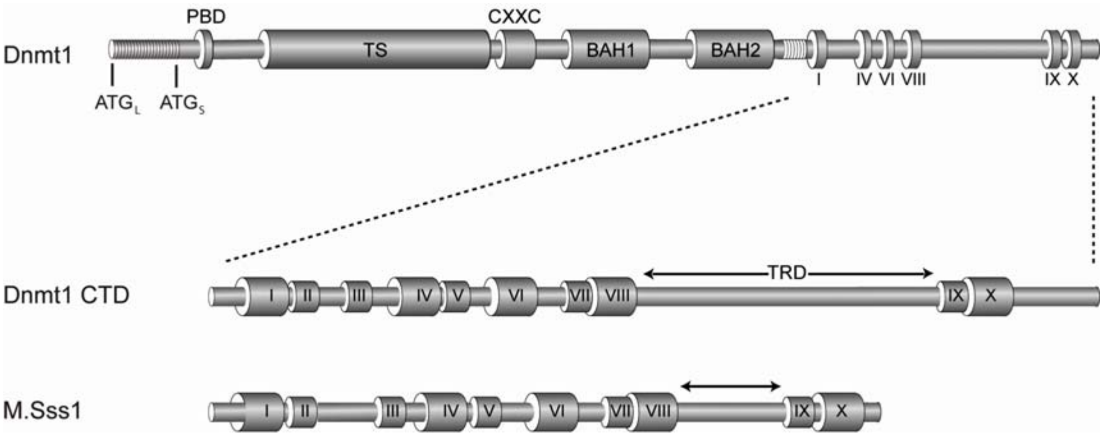


Figure 6. Dnmt1 domain structure and sequence similarity of the catalytic C-terminal domain to prokaryotic DNA methyltransferase (here: M.Sss1). The catalytic motifs I-X and the target recognizing domain, as well as the domains included in the N-terminal part of Dnmt1 are highlighted (TRD: target recognizing domain; PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC zinc finger domain; BAH: Bromo adjacent homology domain). (modified from (Leonhardt and Bestor, 1993))

Surprisingly, although the catalytic domain of Dnmt1 contains all conserved motifs identified to be necessary for the methyl transfer reaction ((Bestor and Verdine, 1994; Kumar et al., 1994)), it is not active by itself, but has to be activated by interaction with the N-terminal part of the enzyme (Fatemi et al., 2001; Margot et al., 2003; Zimmermann et al., 1997). In order to reveal which part of the N-terminal domain is needed for activation of enzyme activity, numerous attempts have been undertaken to find the minimal active Dnmt1 construct (Bacolla et al., 2001; Margot et al., 2000; Zimmermann et al., 1997). The shortest peptide that has been shown to still possess enzymatic activity is human Dnmt1Δ1-580. In contrast, Dnmt1Δ672 was shown to be inactive (Pradhan and Esteve, 2003) suggesting that the first half of the N-terminal domain of Dnmt1 is dispensable for enzymatic activity.

The N-terminal part of Dnmt1 contains the following well defined domains (Figure 6): PCNA binding domain (PBD), targeting sequence (TS), CXXC zinc finger domain (CXXC), and two Bromo adjacent homology domains (BAH1 and BAH2). The PBD mediates the interaction of Dnmt1 with PCNA (proliferating cell nuclear antigen), which has been shown to serve as dynamic loading platform for factors involved in replication (Sporbert et al., 2005). Indeed, this interaction is required for accumulation of Dnmt1 at replication foci and directly couples the replication of DNA sequence and DNA methylation. Furthermore, the TS domain directs association of Dnmt1 with heterochromatin. Consequently, these two interactions determine the cell-cycle specific subcellular localization of Dnmt1 (Chuang et al., 1997a; Easwaran et al., 2004; Leonhardt et al., 1998). Finally, the function of the BAH domains is completely unknown and the function of the CXXC domain is still controversially discussed (see also chapter 1.3.3).

1.2.2 MECHANISM OF THE METHYL TRANSFER REACTION

Based on the striking sequence similarity of the catalytic domain of mammalian Dnmt1 with prokaryotic type II DNA cytosine methyltransferases, prokaryotic and mammalian Dnmts were proposed to function catalytically in a similar multi-step mechanism ((Bestor and Verdine, 1994; Klimasauskas et al., 1994; Wu and Santi, 1987), Figure 7). Upon DNA binding, the target cytosine is flipped out of the DNA double helix and a covalent complex with the C^6 position of the target cytosine is formed. After transfer of a methyl group from the methyl group donor *S*-Adenosyl-*L*-Methionine (AdoMet) to the C^5 position of the nucleobase, the covalent enzyme-DNA complex is released by β-elimination. This mechanism was first described for the prokaryotic methyltransferase M.HhaI, but kinetic investigations of mouse and human Dnmt1 (Flynn et al., 1996; Flynn and Reich, 1998; Pradhan et al., 1999) also revealed that the methyl transfer reaction is a sequential Bi Bi mechanism including first binding of AdoMet and DNA followed by release of *S*-Adenosyl-*L*-Homocysteine (AdoHcy) and methylated DNA.

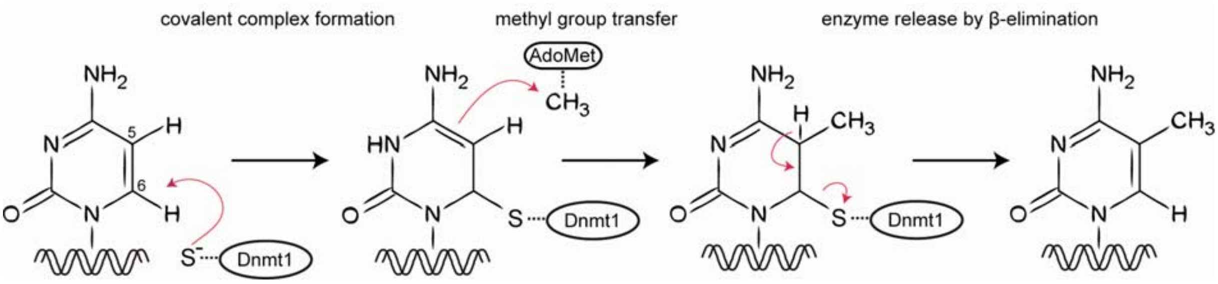


Figure 7. Mechanism of the methyl transfer reaction. Upon DNA binding and flipping the target cytosine out of the double helix, the methylation reaction includes covalent attachment of the enzyme to the C^6 position of the cytosine base, methyl group transfer from the methyl-donor AdoMet to the activated C^5 position and release of the enzyme by ß-elimination.

The co-crystal structure of the prokaryotic methyltransferase M.HhaI with DNA containing 5-fluoro-dC revealed that the conserved cysteine 81 of the invariant Pro-Cys dipeptide in motif IV is involved in covalent complex formation (Cheng et al., 1993; Wu and Santi, 1987). Accordingly, single base mutations of this conserved cysteine residue (C->S or C->W) result in catalytically inactive proteins (Wyszynski et al., 1993). Interestingly, whereas also DNA binding activity was reduced for the C->W mutation, the C->S mutation did not diminish specific DNA binding, but rather enhanced affinity of M.HhaI for DNA (Wyszynski et al., 1993). Consistently, the C->S mutant of human Dnmt1 was also proposed to bind tighter to DNA as the wild-type enzyme (Araujo et al., 2001).

Based on the positive correlation between methylation rate and DNA substrate length, it was early hypothesized that Dnmt1 might be a processive enzyme sliding along the DNA molecule by one-dimensional diffusion (Bestor and Ingram, 1983). Later on, Dnmt1 was suggested to methylate hemimethylated DNA substrates processively ignoring fully or unmethylated sites in between. Thereby, Dnmt1 was suggested to keep its orientation with respect to the DNA (Hermann et al., 2004). Other reports specified by direct analysis of methylation patterns that Dnmt1 methylates hemimethylated DNA with 95-99% fidelity processively, but unmethylated DNA with much less or no processivity (Goyal et al., 2006; Vilkaitis et al., 2005). Interestingly, the first 290 amino acids including the PCNA binding domain were not required for this behavior (Vilkaitis et al., 2005). Still, the hypothesis of processivity is subject of discussion (Bacolla et al., 1999).

1.2.3 SUBSTRATE SPECIFICITY

The substrate for Dnmt1 is DNA containing CpG sites. *In vitro*, Dnmt1 has been shown to have a strong preference for DNA substrates with hemimethylated over substrates with unmethylated CpG sites. It has been suggested that optimal binding sites for Dnmt1 carry the CpG site within a G/C-rich context (Flynn et al., 1998), and that high levels of supercoiling increase the low de novo activity (Bacolla et al., 2001; Bestor, 1987). However, there is no evidence for further sequence prerequisite for Dnmt1 methyl transfer activity other than a cytosine in the context of a CpG site.

In contrast to the clear picture of Dnmt1's preference for hemimethylated substrate in accordance to its role in maintenance methylation, the question remains at which step of the methyl transfer reaction substrate discrimination occurs. On the one hand, it has been suggested that Dnmt1 binds equally well to un- and hemimethylated DNA, indicating that a later step of the reaction is responsible for preferential methylation of hemimethylated DNA (Flynn et al., 1996). On the other hand, it was claimed that Dnmt1 preferentially binds to hemimethylated DNA (Araujo et al., 2001; Bacolla et al., 2001). Dnmt1 is unique among other DNA methyltransferases not only in terms of its preference for hemimethylated DNA, but also for its large regulatory N-terminal domain. Thus, to address the question whether the N-terminal domain is involved in substrate recognition, a hybrid mouse-prokaryotic DNA methyltransferase, consisting of the mouse Dnmt1 N-terminus and the M.HhaI sequence, was generated and tested for substrate specificity. Remarkably, the hybrid enzyme retained the nucleotide sequence specificity of M.HhaI (GCGC sites), but gained a 2.5 fold preference for hemi- over unmethylated substrate compared to an around 2 fold preference for un- over hemimethylated DNA of the wild-type M.HhaI enzyme (Pradhan and Roberts, 2000). These results suggest a role for the N-terminal part of Dnmt1 in the preference for hemimethylated substrate. If this is the case, the responsible sequences seem to be located within the second half of the N-terminus, since human Dnmt1Δ1-580 still prefers hemimethylated substrate (Pradhan and Esteve, 2003). However, another publication claims that the Dnmt1 target recognition domain resides in the N-terminal amino acids 122-417 (Araujo et al., 2001). In addition, the linker region between N- and C-terminal domains has been implicated in substrate recognition. Cleavage of Dnmt1 with protease V8, 10 amino acids downstream the alternating KG linker and 6 residues upstream methyltransferase motif 1, was suggested to decrease the preference for hemimethylated DNA by stimulation of de novo activity (Bestor, 1992). However, we have recently shown that linker cleavage as well as modification of linker length and charge, does not affect Dnmt1 activity or substrate preference (our unpublished data, Weihua Qin). In conclusion, the step at which substrate recognition occurs and the exact interplay of domains responsible for substrate discrimination remain elusive.

Hence, in order to further elucidate the mechanism of substrate recognition, we developed an assay with which we can easily test and compare Dnmt1 mutants and domains for both DNA binding and covalent complex formation (chapter 2.1).

1.3 REGULATION OF DNMT1

1.3.1 TRANSCRIPTIONAL AND TRANSLATIONAL REGULATION

Dnmt1 is controlled by alternative splicing of sex-specific 5’ exons in mammalian germ cells. In oocytes, this leads to a shorter Dnmt1 transcript and expression of a shorter Dnmt1 isoform, whereas in pachytene spermatocytes, a longer Dnmt1 transcript is produced, which is post-transcriptionally downregulated and reported as not translatable (Carlson et al., 1992; Mertineit et al., 1998). The oocyte-specific isoform of Dnmt1 (Dnmt1o) is lacking the first 118 amino acids of the somatic full-length protein. Dnmt1o was shown to be fully active *in vitro* (Pradhan et al., 1997) and *in vivo,* as it restored DNA methylation in *dnmt1*$^{-/-}$ embryonic stem cells (Gaudet et al., 1998). Interestingly, Dnmt1o is actively retained in the cytoplasm during early embryonic development until implantation, possibly in order to allow global DNA demethylation. Responsible for this cytoplasmatic retention is a broad region in the N-terminus (amino acids 426-972), with amino acids 545-756 being crucial as determined by deletion analysis. This region includes the CXXC domain of Dnmt1 (amino acids 651-698) (Cardoso and Leonhardt, 1999). Dnmt1o only transiently enters the nucleus in 8-cell embryos suggesting that this variant of Dnmt1 provides maintenance methylation at imprinted loci during the fourth embryonic S phase (Howell et al., 2001). Interestingly, an alternative Dnmt1 transcript was identified in skeletal muscle, specifically expressed in differentiated myotubes when the ubiquitously expressed Dnmt1 isoform is down-regulated. This longer transcript is identical to the one previously reported to be unique for sperm cells and to be untranslatable (Mertineit et al., 1998). However, a later work suggested that the transcript is after all translated and results in a protein identical to Dnmt1o (Aguirre-Arteta et al., 2000).

In addition to the 5’ alternative splice variants, there is a one-codon alternative splice variant in mouse somatic cells leading to a protein differing by 2 amino acids shortly before the PBD domain (amino acid 145 F -> SV) (Lin et al., 2000). Similar abundance was suggested for both isoforms. Analogously, in human somatic cells, a Dnmt1 variant occurs due to alternative splicing of the primary Dnmt1 gene transcripts in intron 4, where a copy of Alu repeats is inserted. This variant leads to expression of a Dnmt1 isoform, which has 16 additional amino acids between PBD and TS domain (Bonfils et al., 2000; Hsu et al., 1999). In this case, the mRNA level of the longer transcript makes up 6-25 % of the total Dnmt1 transcripts, whereas the translated protein product makes up only 2-5 % of the total Dnmt1 protein. Like Dnmt1o, also the somatic splice variants are active DNA methyltransferases and seem to behave like the major ubiquitous Dnmt1 enzyme.

1.3.2 REGULATION OF DNMT1 BY POST-TRANSLATIONAL MODIFICATIONS

Dnmt1 is subject to several posttranslational modifications, namely phosphorylation, methylation, SUMOylation and ubiquitylation, which were reported to regulate Dnmt1 activity or stability. Phosphorylation of the SV variant of mouse Dnmt1 at serine 146 (Lin et al., 2000) has been recently suggested to reduce DNA binding activity of Dnmt1 (Sugiyama et al., 2010). Moreover, serine 514 phosphorylation within the TS domain has been reported to regulate enzyme activity (Glickman et al., 1997; Goyal et al., 2007), but this hypothesis could not be confirmed in our laboratory (unpublished data, Karin Fellinger and Andrea Rottach). Methylation at lysine 1096 and other lysine residues has been shown to destabilize Dnmt1. Dnmt1 levels are thus controlled by the action of the enzymes that set and remove these methyl groups, that is the methyltransferase Set7/9 and the demethylase LSD1, respectively (Wang et al., 2009). SUMOylation of Dnmt1 has been reported to enhance Dnmt1 activity; however, the exact position of this modification could not be mapped (Lee and Muller, 2009). At last, two studies proposed that Dnmt1 levels are regulated by proteasomal degradation pathways triggered by ubiquitylation of Dnmt1 (Agoston et al., 2005; Ghoshal et al., 2005). One of these studies suggested that the first 120 amino acids are crucial for proteasomal degradation and thus to be the region targeted by ubiquitylation (Agoston et al., 2005).

1.3.3 DNA BINDING AND ALLOSTERIC ACTIVATION

Dnmt1 was proposed to be activated by DNA binding to an allosteric site outside the catalytic center (Bacolla et al., 2001; Bacolla et al., 1999). Binding of methylated DNA to this allosteric site was suggested to increase the affinity for AdoMet, which again facilitates substrate DNA binding and finally leads to acceleration of the reaction rate (Bacolla et al., 1999). Notably, methylated DNA was proposed to play a dual role, on the one hand allosterically activating Dnmt1, and on the other hand inhibiting the methyl transfer reaction by competition for binding to the catalytic site (Bacolla et al., 1999). The allosteric effect of methylated DNA was proposed to be important for methylation of unmethylated CpG sites in proximity to methylated sites by a mechanism called methylation spreading. Furthermore, it was suggested that the enzyme is allosterically inhibited by unmethylated DNA affecting both enzyme activity and processivity (Svedruzic and Reich, 2005). This might be important for regulation of Dnmt1 activity *in vivo* and contribute to the fidelity of Dnmt1's maintenance function.

Steady-state kinetic analyses indicated that the methyl transfer reaction of Dnmt1 might be repressed by the first 501 N-terminal amino acids, as Dnmt1Δ1-501 was suggested to possess higher activities than the wild-type enzyme on both un- and hemimethylated substrate (Bacolla et al., 2001)). Interestingly, the repression was proposed to be relieved by binding of methylated DNA to an allosteric site within this region (Bacolla et al., 2001; Svedruzic and Reich, 2005).

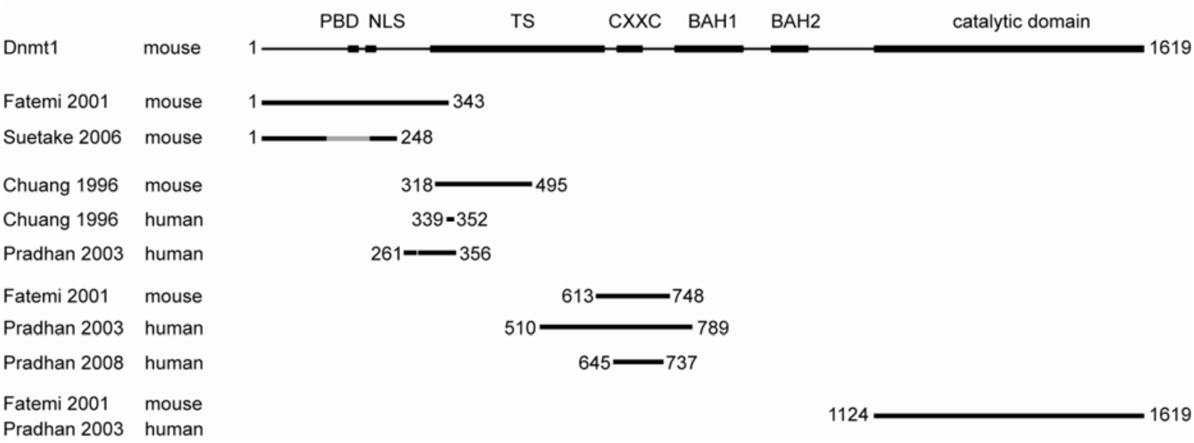

PBD
NLS
TS
CXXC
BAH1
BAH2
catalytic domain
Dnmt1 mouse 1 1619
Fatemi 2001 mouse 1 343
Suetake 2006 mouse 1 248
Chuang 1996 mouse 318 495
Chuang 1996 human 339 352
Pradhan 2003 human 261 356
Fatemi 2001 mouse 613 748
Pradhan 2003 human 510 789
Pradhan 2008 human 645 737
Fatemi 2001 mouse
Pradhan 2003 human
1124 1619

Figure 8. Proposed DNA binding domains of Dnmt1. The DNA binding domains that were reported for mouse and human Dnmt1 are summarized. Besides the C-terminal catalytic domain, the very N-terminal part of the enzyme, the region including the first part of the TS domain and the CXXC domain have been shown to possess DNA binding activity.

Moreover, numerous DNA binding domains of murine and human Dnmt1 have been described, which could possibly regulate enzyme activity. Proposed DNA binding domains include besides the catalytic domain itself (Fatemi et al., 2001), the very N-terminal region of mouse Dnmt1 (Fatemi et al., 2001; Suetake et al., 2006), sequences shortly before and/or within the first part of the TS domain (Chuang et al., 1996; Pradhan and Esteve, 2003) and peptides including the CXXC domain (Figure 8). Thus, there reside at least three DNA binding regions within the N-terminal region of both the murine and the human Dnmt1 enzymes. First, the very N-terminal part of murine Dnmt1 was shown to bind to DNA (Fatemi et al., 2001; Suetake et al., 2006). In this respect, amino acids 1-248 were suggested to form an independent domain based on limited digestion of Dnmt1 with different proteases. This domain was shown to bind to DNA via amino acids 119-197 including the PCNA binding domain, and to preferentially bind to the minor groove of AT rich sequences (Suetake et al., 2006). Second, the N-terminal domain of human Dnmt1 comprises a DNA binding motif DB1 adjacent to a region for zinc binding within the TS sequence. The DNA binding properties of different GST fusion constructs comprising both DB1 and the zinc binding region were suggested to be dependent on the length of DNA substrate, but not the DNA sequence. Thereby, the Zn-loop(s) adjacent to the DB1 motif appeared to inhibit the binding of DB1 to smaller DNA duplexes, leading to the speculation that this could lead to a delay between replication and methylation *in vivo* (Chuang et al., 1996). In accordance to the human protein, a peptide sequence corresponding to this region in the murine enzyme was also shown to bind to DNA (Chuang et al., 1996; Pradhan and Esteve, 2003). In addition, for the human Dnmt1, a short sequence including the motif KKHR (amino acids 284-287) shortly before the TS domain was shown to bind to DNA. Third, three peptides including the CXXC domain were shown to possess DNA binding activity: amino acids 510-789 (Pradhan and Esteve, 2003) or amino acids 645-737 (Pradhan et al., 2008) of the human enzyme, and amino acids 613-748 (Fatemi et al., 2001) of the murine enzyme.

Two of these DNA binding domains have been independently suggested to be involved in allosteric activation of Dnmt1. The KKHR motif of the human enzyme was proposed to be involved in methylated DNA-mediated allosteric activation based on analysis of point mutants which showed reduced activation upon addition of methylated DNA (Pradhan and Esteve, 2003). Moreover, the CXXC domain of mouse Dnmt1 has been suggested to interact with the catalytic domain upon binding of methylated DNA and thereby to allosterically activate enzyme activity (Fatemi et al., 2001). However, methylation spreading on unmethylated CpG sequences by allosteric activation was reported to be not affected for human Dnmt1Δ121, but for deletion of the first 501, 540, or 580 amino acids, which indicates that the allosteric DNA binding site lies between amino acids 121 and 501 (Pradhan and Esteve, 2003), and thereby argues against a role of the CXXC domain of Dnmt1 in methylation spreading, at least for the human enzyme.

The DNA binding specificity of the CXXC domain and its role in enzyme regulation is controversially discussed. However, the observation that deletion of the first 580 amino acids of the Dnmt1 does not affect its activity, whereas deletion of a few additional amino acids including parts of the CXXC domain completely abolishes enzyme activity, provides strong evidence for an important role of this domain. Therefore, we addressed the role of the CXXC domain of mouse Dnmt1 for DNA binding and enzymatic activity of Dnmt1 *in vitro* and *in vivo* (chapter 2.3).

1.3.4 REGULATION BY INTERACTION WITH OTHER FACTORS

In addition to the intrinsic preference of Dnmt1 for hemimethylated substrate and the allosteric regulation of Dnmt1 activity by DNA or intramolecular interaction, other factors control the reliable maintenance of DNA methylation pattern *in vivo* (see also chapter 1.1.5.3). As mentioned above, Dnmt1 interacts with PCNA via its PBD domain and thereby locates to site of DNA replication during S phase. This association with the replication machinery could be an efficient mechanism to couple the replication of genetic and epigenetic information (Chuang et al., 1997b). Importantly, although being very transient and not strictly required for restoring global methylation level *in vivo*, the interaction with PCNA enhances Dnmt1 activity about two-fold (Schermelleh et al., 2007; Spada et al., 2007). The reason for this enhancement is still subject of speculation and results either from elevation of local Dnmt1 concentrations at the site of replication or from the induction of structural changes upon Dnmt1 binding to PCNA and thus allosteric activation. Notably, evidence for the latter hypothesis was provided by a study showing that Dnmt1 has a higher activity on PCNA associated DNA than on free DNA (Iida et al., 2002).

From late S until early G1 phase, Dnmt1 is associated with (preferentially constitutive) heterochromatin via its TS domain in a replication-independent mechanism (Easwaran et al., 2004; Leonhardt et al., 1992). This prolonged association of Dnmt1 with chromatin after DNA replication was speculated to allow time for the maintenance of densely methylated regions as pericentric heterochromatin, eventually even serving as loading platform for subsequent chromatin modifiers or remodelers (Easwaran et al., 2004).

Recently, Uhrf1 (also known as Np95) emerged as essential cofactor for maintenance methylation as Uhrf1 ablation leads to global DNA hypomethylation, a phenotype similar to that obtained upon Dnmt1 ablation (Bostick et al., 2007a; Sharif et al., 2007). To this end, Uhrf1 has been shown to preferentially bind to DNA with hemimethylated CpG sites via its SRA domain, but also to directly interact with Dnmt1 and to colocalize with Dnmt1 at replication foci ((Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007b; Sharif et al., 2007), see also chapter 1.1.5.2). Based on these data, it was suggested that Uhrf1 recruits Dnmt1 to hemimethylated CpG sites at the replication fork. However, the exact mechanism and the hierarchy of events necessary for Uhrf1-dependent maintenance methylation are far from being understood. The high intrinsic preference of Dnmt1 even in the absence of Uhrf1 and the emerging evidence that Uhrf1 also strongly binds to histone tails can lead to the assumption that Uhrf1 might play a more complex role in maintenance methylation than simply recruiting Dnmt1 to hemimethylated DNA. We thus further investigated the functional role of Uhrf1 in maintenance methylation in this work (chapter 2.4).

Finally, it has been recently suggested that Dnmt1 forms a stable dimer (Fellinger et al., 2009). This self-interaction would allow another level of enzymatic regulation. To further test this hypothesis and to investigate whether Dnmt1 needs to dimerize for enzymatic activity, we performed single molecule spectroscopy of Dnmt1 and Dnmt1:DNA complexes (chapter 2.5).

1.4 AIMS OF THIS WORK

Although the biological importance and the sequence of the most ubiquitous mammalian DNA methyltransferase Dnmt1 are known for two decades, its mechanistic regulation still remains elusive. Therefore, the main objective of this PhD thesis was to study the molecular basis of Dnmt1 regulation with respect to stable maintenance of DNA methylation patterns.

At first, I aimed at elucidating the functional role of the CXXC domain of Dnmt1. This domain has been reported to allosterically activate Dnmt1 by intramolecular interaction with the catalytic C-terminal domain. Moreover, the observation that CXXC domains of numerous other proteins bind preferentially to unmethylated DNA substrates suggests that the CXXC domain of Dnmt1 might be involved in substrate discrimination by blocking enzymatic activity of Dnmt1 upon binding to unmethylated DNA. To address this hypothesis, I developed a non-radioactive assay for DNA methyltransferase activity and DNA binding (chapter 2.1). Besides the advantage of avoiding the use of isotope labeled molecules, this assay was designed to distinguish different steps of the methyl transfer reaction. Furthermore, by using distinct fluorescent DNA labels, I aimed at comparing up to four different DNA substrates in direct competition. Using this assay, I tested the binding properties of the CXXC domain of Dnmt1 and its functional relevance for enzymatic activity and substrate specificity (chapter 2.3).

It is well established that Dnmt1 requires other factors for reliable maintenance of methylation patterns *in vivo*. Notably, Uhrf1 recently emerged as an essential cofactor for maintenance methylation. Although Uhrf1 has been proposed to function by recruiting Dnmt1 to hemimethylated CpG sites, there is evidence that this might not be the exclusive role of Uhrf1. To further investigate the exact mechanism by which Uhrf1 contributes to maintenance methylation, I tested DNA binding activity and specificity of Uhrf1 and a variety of Uhrf1 domains with the newly established assay (chapter 2.4).

Moreover, it was recently suggested that Dnmt1 forms a stable dimer (Fellinger et al., 2009), which potentially represents a further process that regulates Dnmt1 activity. Thus, I also addressed the question whether Dnmt1 is active as a monomer or dimer and investigated the stoichiometry of Dnmt1:DNA complexes upon covalent complex formation (chapter 2.5).

During assay development, I observed an interesting feature of a GFP binding protein (GBP). The GBP is the active part of the GFP-Trap® (ChromoTek GmbH), which was used for purification of the tested DNA binding proteins. Interestingly, the GBP as well as other nanobodies were able to modulate GFP fluorescence and spectral properties *in vitro* and *in vivo* (chapter 2.2).

2 RESULTS

2.1 A VERSATILE NON-RADIOACTIVE ASSAY FOR DNA METHYLTRANSFERASE ACTIVITY AND DNA BINDING

Published online 7 January 2009

Nucleic Acids Research, 2009, Vol. 37, No. 3 e22
doi:10.1093/nar/gkn1029

A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding

Carina Frauer and Heinrich Leonhardt*

Department of Biology, Center for Integrated Protein Science Munich (CIPS^M), Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany

Received October 21, 2008; Revised December 3, 2008; Accepted December 10, 2008

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

ABSTRACT

We present a simple, non-radioactive assay for DNA methyltransferase activity and DNA binding. As most proteins are studied as GFP fusions in living cells, we used a GFP binding nanobody coupled to agarose beads (GFP nanotrap) for rapid one-step purification. Immobilized GFP fusion proteins were subsequently incubated with different fluorescently labeled DNA substrates. The absolute amounts and molar ratios of GFP fusion proteins and bound DNA substrates were determined by fluorescence spectroscopy. In addition to specific DNA binding of GFP fusion proteins, the enzymatic activity of DNA methyltransferases can also be determined by using suicide DNA substrates. These substrates contain the mechanism-based inhibitor 5-aza-dC and lead to irreversible covalent complex formation. We obtained covalent complexes with mammalian DNA methyltransferase 1 (Dnmt1), which were resistant to competition with non-labeled canonical DNA substrates, allowing differentiation between methyltransferase activity and DNA binding. By comparison, the Dnmt1^{C1229W} catalytic site mutant showed DNA-binding activity, but no irreversible covalent complex formation. With this assay, we could also confirm the preference of Dnmt1 for hemimethylated CpG sequences. The rapid optical read-out in a multi-well format and the possibility to test several different substrates in direct competition allow rapid characterization of sequence-specific binding and enzymatic activity.

INTRODUCTION

The modification of DNA by DNA methyltransferases is widespread and has a variety of biological functions (1). In bacteria, DNA methylation is involved in host defense mechanisms and strand discrimination during mismatch repair. In eukaryotic cells, DNA methylation is part of a highly complex epigenetic network regulating genome structure and activity (2,3). In contrast to the bacterial enzymes, eukaryotic DNA methyltransferases contain large regulatory domains that are involved in numerous intermolecular interactions and control enzyme activity through a largely unknown mechanism (4). The biochemical and cell biological characterization of DNA methyltransferases is pivotal for the understanding of epigenetic network regulation.

The basic biochemistry of the 5-methyl cytosine (5mC) methylation reaction is by now well understood. In a post-replicative reaction, DNA methyltransferases catalyze the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to the C5 position of the nucleobase. During this multi-step reaction, the target cytosine is flipped out of the double helix (base flipping) and the recipient C5 position is activated by a transient, covalent complex formation with the enzyme at the C6 position (5,6). After methyl group transfer, the enzyme is released by β-elimination together with the proton at the C5 position. This last and crucial step of the enzymatic reaction can be exploited for a specific and mechanism-based inhibition with DNA substrates containing nucleotide analogs like 5-aza-dC or zebularine that are missing the essential proton at the C5 position (7–9). Although the catalytic mechanism of the 5mC DNA methyltransferases is known, the crucial question how eukaryotic enzymes recognize and discriminate target sites for methylation remains elusive.

Over the past decades, a variety of biochemical assays has been developed to determine the activity of DNA methyltransferases. The most commonly used methyltransferase activity assays measure the transfer of radioactively labeled methyl groups from the cofactor AdoMet to DNA substrates (10–14). Alternatively, DNA methylation by active methyltransferases can be monitored as protection against nucleolytic cleavage by restriction enzymes. The amount of methylated DNA can be measured as

---

*To whom correspondence should be addressed. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

release or retention of terminal affinity probes of DNA substrates (15,16). Another indirect approach uses bisulfite treatment followed by incorporation and detection of hapten-labeled dCTPs at non-converted sites (17). Also direct detection of methylated cytosine residues by MALDI-TOF mass spectrometry (18) or monitoring of conversion of AdoMet to *S*-adenosyl-homocysteine (AdoHcy) by liquid chromatography and mass spectroscopy has been used (19). All these methods depend on either radioisotopes, expensive and demanding equipment, and/or multiple-step protocols.

Here, we present a simple, non-radioactive and versatile method to measure DNA methyltransferase activity. The assay measures methyltransferase activity as irreversible covalent complex formation with fluorescently labeled DNA substrates containing the mechanism-based inhibitor 5-aza-dC. The variation of DNA sequence and fluorescent label allows detection of DNA sequence specificity and discrimination of methyltransferase activity from DNA binding. We tested this assay using mammalian DNA methyltransferase 1 and mutants thereof.

MATERIALS AND METHODS

Expression vectors

The eukaryotic expression vectors for enhanced GFP (pEGFP-C1, Clontech, USA) and fusions with mouse Dnmt1 and its catalytically inactive mutant Dnmt1^{C1229W} were previously described (7). For GFP expression in bacteria, the pRSET-EGFP vector was generated. The GFP-coding sequence was amplified from pEGFP-C1 by PCR to add flanking XbaI/EcoRI restriction sites and a C-terminal His$_6$-tag. The PCR fragment was digested with XbaI and EcoRI and subsequently ligated into the bacterial expression vector pRSET (Clontech, USA).

Cell culture and transfection

Human embryonic kidney (HEK) 293T cells were cultured in DMEM supplemented with 10% fetal calf serum and 50 μg/ml gentamycine (PAA, Germany). HEK 293T cells were transiently transfected with expression plasmids for GFP, GFP-Dnmt1 and GFP-Dnmt1^{C1229W} using polyethylenimine as transfection reagent (Sigma, Germany) (20). After 48 h, about 80–90% of the cells were expressing GFP as determined by fluorescence microscopy. Cells were harvested, washed twice with PBS and stored at −80°C.

GFP purification

A 2 l culture of BL21 (DE3) *Escherichia coli* transformed with pRSET-EGFP was grown to OD 0.6 and induced with 1 mM IPTG for 20 h at RT. Bacteria were harvested and resuspended in 20 ml of binding buffer (500 mM NaCl, 20 mM imidazole, 1 mM PMSF in PBS). Lysis of *E. coli* was performed by sonification in the presence of 1 μg/ml lysozyme and 25 μg/ml DNase I. After centrifugation, 10 ml of soluble *E. coli* protein extract was loaded onto a His-Trap HP column containing 1 ml of Ni-NTA resin (GE Healthcare, Germany) using an ÄKTA purifier (GE Healthcare, Germany). After extensive washing of the bound material, the protein was eluted with elution buffer (500 mM NaCl, 250 mM imidazole in PBS) and 1 ml fractions were collected. Aliquots of elution fractions were subjected to SDS–PAGE and coomassie brilliant blue staining. Pure fractions of GFP were pooled and dialyzed three times against 1 l of PBS. The GFP concentration was determined by an analytical SDS–PAGE and coomassie brilliant blue staining with carbonic anhydrase as concentration standard.

Preparation of DNA substrates

DNA oligonucleotides were purchased from Metabion (Germany) or from IBA (Germany) and the sequences are listed in Table 1. Double-stranded DNA substrates were synthesized by primer extension using the large (Klenow) fragment of *E. coli* DNA polymerase I (Figure 1, Supplementary Figure 1A).

To prepare the DNA substrates, one upper (CG-up or MG-up) and one lower strand (Fill-In, Fill-In-550 or Fill-In-647N) oligonucleotide were denatured in NEB2 buffer (50 mM NaCl, 10 mM Tris–HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol) for 2 min at 95°C and annealed by slowly cooling down to 37°C. Upon addition of 0.05 u/μl Klenow fragment (NEB, Germany), dTTP, dGTP, dATP (PeqLab, Germany) at 1 mM final concentration, and either CTP at 1 mM, 5-aza-dCTP or 5-methyl-dCTP at 50 μM (Jena Bioscience, Germany), the Fill-In oligonucleotide was extended to produce either unmethylated, hemimethylated or fully methylated canonical DNA substrates or un- or hemimethylated suicide DNA substrates containing 5-aza-dC at the CpG site. 5-aza-dC containing suicide DNA substrates are referred to as 'trapping substrates' and DNA substrates not containing 5-aza-dC as 'binding substrates'. The design of the oligonucleotides allows the preparation of 15 different unlabeled, ATTO550 or ATTO647N labeled substrates with only five different oligonucleotides (Supplementary Figure 1B). Hemimethylated ATTO550 labeled and

Table 1. Sequences of DNA oligonucleotides used for preparation of double-stranded DNA substrates (M, 5-methylcytosine)

CG-up	5′-CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG-3′
MG-up	5′-CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG-3′
Fill-In-550	5′-ATTO550-CCATGATGACTCTTCTGGTC-3′
Fill-In-647N	5′-ATTO647N-CCATGATGACTCTTCTGGTC-3′
Fill-In	5′-CCATGATGACTCTTCTGGTC-3′

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

unmethylated ATTO647N labeled binding and trapping substrates were therefore prepared as described earlier, using MG-up and Fill-In-550 or CG-up and Fill-In-647N oligos. Unlabeled hemimethylated competitor DNA substrate was prepared using MG-up and Fill-In oligos.

Calibration measurements for GFP, ATTO550 and ATTO647N

Calibration curves for the fluorescent DNA substrates and proteins were determined by measuring the fluorescence signal of known concentrations of the DNA-coupled fluorophores and purified GFP and calculated by linear regression. For this, we used the PolarStarOptima fluorimeter (BMG, Germany) and the following excitation/emission band path filter sets: 485 ± 8 nm/520 ± 17 nm for GFP, 545 ± 5 nm/575 ± 5 nm for ATTO550 and 645 ± 5 nm/675 ± 5 nm for ATTO647N. The beads do not cause fluorescence background, and within the measurement error, no change of fluorescence intensity of the ATTO dyes was observed upon addition of beads. Interestingly, the GFP fluorescence signal is enhanced by binding to the GFP-binding protein (GBP), which is the active part of the GFP nanotrap. With the indicated filter set for GFP detection, the fluorescence signal is about 1.7 times enhanced (Supplementary Figure 3). This effect was taken into account for later conversion of the fluorescent signal into fluorophore concentration and calculation of binding and trapping rates as the ratio of ATTO and GFP signal.

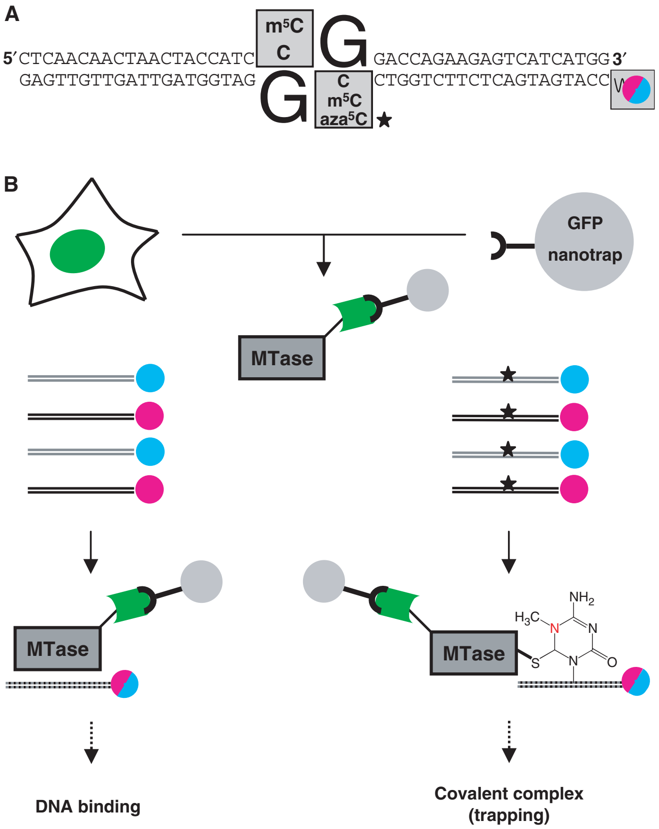


Figure 1. Outline of the binding and activity assay. The covalent complex formation is the first and crucial step of the methylation reaction. The incorporation of the mechanism-based inhibitor 5-aza-dC (depicted as a star) in DNA substrates leads to an irreversible complex formation with catalytically active DNA methyltransferase (trapping). Capture and detection of this reaction intermediate thus serves as a measure of enzyme activity. (**A**) Un-, hemi- or fully methylated canonical or 5-aza-dC containing double-stranded DNA substrates (binding and trapping substrates, respectively) are 42 base pairs long including one central CpG site and can be unlabeled, labeled with ATTO550 or labeled with ATTO647N. The asterisk marks 5-aza-dC. (**B**) The GFP fusion protein of interest, e.g. a DNA methyltransferase (MTase), is purified from cell lysates using a GFP nanotrap and incubated with binding or trapping DNA substrates. After pull-down of protein–DNA complexes, unbound DNA substrate is removed by two washing steps. Protein and DNA substrate amounts are calculated from fluorescence measurements of GFP, ATTO550 and ATTO647N, respectively.

Pull-down of GFP or GFP fusion proteins

Extracts from $\sim 1 \times 10^7$ cells were prepared by resuspension and incubation of the cell pellet in 200 µl lysis buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 2 mM PMSF, 0.5% NP40, 1× mammalian protease inhibitor mix) for 30 min on ice. After centrifugation, supernatants were diluted to 500 or 1000 µl with immunoprecipitation buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA). Extracts were incubated with 1 µg of a GBP coupled to agarose beads (GFP nanotrap; Chromotek, Germany) (21) for 1–2 h at 4°C with constant mixing. GFP or GFP fusion proteins were pulled down by centrifugation at 540*g*. The beads were washed twice with 1 ml of wash buffer (20 mM Tris–HCl pH 7.5, 300 mM NaCl, 0.5 mM EDTA). The amount of protein on the beads was determined with the PolarStarOptima fluorimeter after resuspension in 100 µl wash buffer or by western blot. In the latter case, beads were resuspended in 2× Laemmli buffer (22) and 25% was loaded onto a 6% SDS–PAGE. After blotting to a nitrocellulose membrane, GFP-Dnmt1 was detected with a specific antibody against Dnmt1 (kindly provided by Nowak, D. and Cardoso, M.C.) and an HRP-labeled secondary antibody.

Binding and trapping assay

The pull-down of GFP or GFP fusion protein was performed as described earlier. After the second washing step, beads were equilibrated with assay buffer (100 mM KCl, 10 mM Tris–HCl pH 7.6, 1 mM EDTA, 1 mM DTT). For determination of binding and trapping rates, the beads were resuspended in 500 or 1,000 µl of assay buffer supplemented with 160 ng/µl BSA and 100 µM *S*-adenosyl-L-methionine (AdoMet), and 0.1 µM binding or trapping DNA substrate, unless indicated otherwise. For qualitative determination of DNA methyltransferase activity, binding (with canonical-binding substrates) and trapping (with suicide trapping substrates) were performed at 37°C for 90 min, unless indicated otherwise. After washing twice with assay buffer to remove unbound substrate, beads were resuspended in 100 µl assay buffer and transferred into a 96-well microplate. The amounts of protein and DNA were determined by fluorescence measurements and comparison to a calibration curve.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

Binding competition assay

Trapping and binding assays were performed as described earlier, except that for binding competition, referred to as binding or trapping with competitor, a further incubation step with 1 μM hemimethylated unlabeled binding DNA was performed for 45 min at 37°C to compete for binding of labeled non-covalently bound substrate in the binding and trapping sample. Before fluorescence measurement, two final washing steps with assay buffer were performed.

RESULTS AND DISCUSSION

Assay design

We previously generated a set of fluorescent Dnmt1 fusions and mutants thereof and characterized their cell-cycle dependent dynamics in living cells (23,24). To complement these data and to gain further insights into the structure, function and regulation of DNA methyltransferases, it is crucial to determine their sequence specific DNA binding and methyltransferase activity. For fast biochemical characterization of these GFP fusion proteins, we developed a simple, non-radioactive assay.

The assay is based on immunoprecipitation of fusion proteins with a GBP coupled to agarose beads [GFP nanotrap (21)]. Bound GFP fusion proteins were incubated with fluorescently labeled double-stranded DNA substrates. After removal of unbound substrate, the concentrations of fluorescent protein and bound DNA substrate were measured with a filter-based fluorescence spectrometer in a multi-well format (Figure 1).

The design of DNA oligonucleotides combined with a primer extension method allows preparation of a variety of substrates (Figure 1A, Supplementary Figure 1). Canonical DNA substrates (binding substrates) were used for binding studies and suicide DNA substrates containing 5-aza-dC at the CpG site (trapping substrates) for monitoring irreversible covalent enzyme–DNA complex formation as the first and crucial step of the DNA methylation reaction. The capture of these reaction intermediates serves as a measure of enzyme activity, although the final step of the methylation reaction, the methyl group transfer, is not detected. As DNA substrates can be labeled with different fluorophores, several different sequences, containing, e.g. un-, hemi- or fully methylated CpG sites, can be tested in direct competition. The fluorescence of protein and substrate allows direct determination of concentrations, molar ratios and specific activity.

Linear GFP-Dnmt1 pull-down with the GFP nanotrap

The GFP nanotrap allows fast and efficient one-step purification of GFP or GFP fusion proteins. For demonstration of linearity, we incubated a constant amount of the GFP nanotrap with different lysate volumes of GFP-Dnmt1 overexpressing HEK 293T cells and determined the concentration of GFP fusion protein bound by the beads. The amount of bound GFP-Dnmt1 did indeed increase linearly with the amount of lysate added, as quantified by fluorescence detection and western blot. Importantly, quantification with a fluorescence plate reader was very sensitive and showed a larger linear range than the corresponding western blot (Figure 2). This demonstrates the strength of the fluorescence-based readout of this assay. The exact quantification of the protein input allows the comparison of different samples and takes into account possible differences in pull-down efficiency.

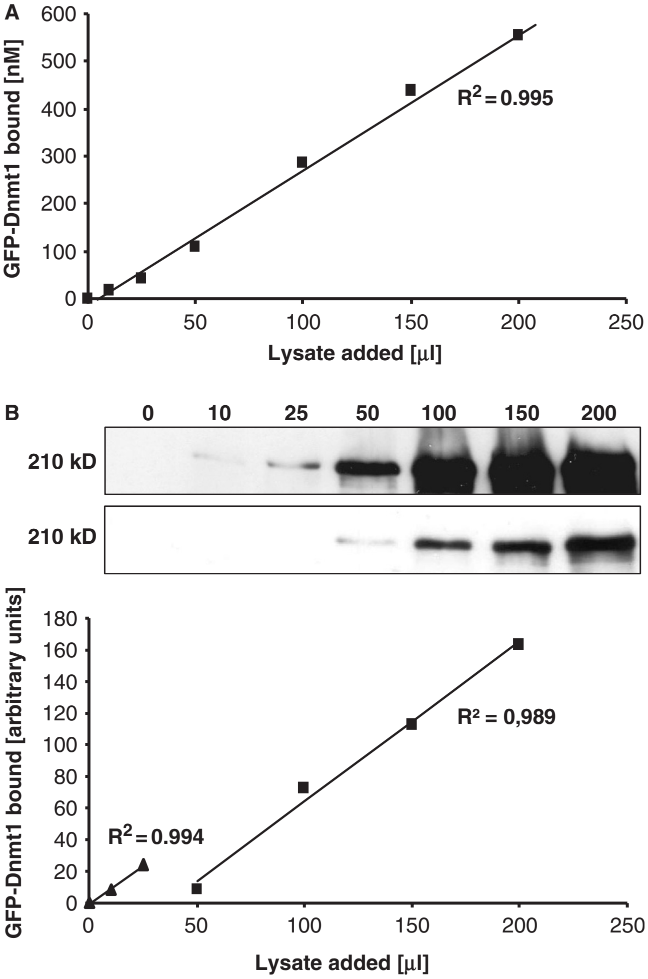


Figure 2. Linear GFP-Dnmt1 pull-down. Different amounts of cell lysate (0, 10, 20, 25, 50, 100, 150 and 200 μl) from GFP-Dnmt1 overexpressing HEK 293T cells were incubated with constant aliquots of the GFP nanotrap. (**A**) The concentration of precipitated GFP-Dnmt1 was calculated from the measured intensity of the GFP fluorescence signal. (**B**) Aliquots of the same samples were analyzed by western blot with an anti-Dnmt1 antibody. Shown are two different exposure times (2 min and 15 s). The band intensities were quantified with the Image J software using the higher exposure time for data points 0, 10, 25 and the lower exposure time for data points 50, 100, 150 and 200.

Characterization and optimization of assay conditions

To optimize assay conditions, we first determined the time course of DNA binding and irreversible covalent complex formation (trapping) of GFP-Dnmt1 with hemimethylated DNA substrates. The time course of GFP-Dnmt1 binding to hemimethylated substrate followed the classical-binding kinetics with an observed rate constant of $k = 0.034 \pm 0.002\ \mathrm{min}^{-1}$ (Supplementary Figure 2A).

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

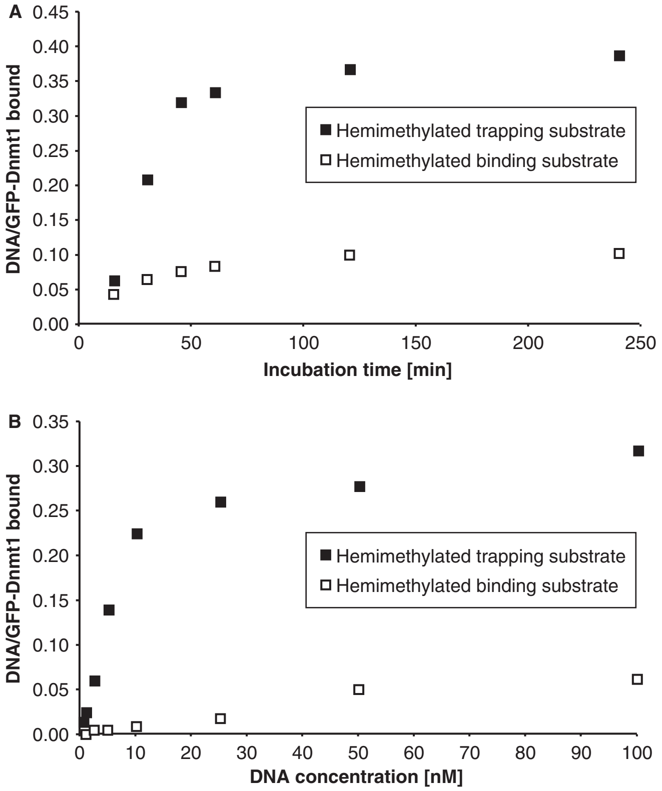


Figure 3. Optimization of trapping assay conditions. (**A**) Time course of binding and trapping reaction. GFP-Dnmt1 (25 nM) was incubated with 100 nM hemimethylated ATTO550 labeled binding (open square) or trapping substrate (filled square). The reactions were stopped by washing after 15, 30, 45, 60, 120 and 240 min, respectively. (**B**) Dependence of binding and trapping rate on the initial DNA substrate concentration. GFP-Dnmt1 (20 nM) was incubated with increasing amounts of hemimethylated ATTO550 labeled binding (open square) or trapping substrate (filled square). Binding and trapping rates are shown for initial substrate concentrations of 0.5, 1, 2.5, 5, 10, 25, 50 and 100 nM.

The trapping rate (ratio of bound suicide DNA substrate per protein) increased linearly within the first 50 min of reaction and reached a plateau at about 90 min (Figure 3A). For substrate specificity and qualitative methyltransferase activity assays, we chose 90-min incubation time to obtain maximal signals. For determination of initial reaction velocities, shorter incubation times were used to stay within the linear range of this assay.

To test the dependence of binding and trapping rate on the initial DNA substrate concentration, we incubated a constant amount of GFP-Dnmt1 with hemimethylated trapping substrate at different concentrations (Figure 3B). The fitting of binding data is shown in Supplementary Figure 2B. For substrate concentrations below the concentration of methyltransferase molecules, the trapping rate increased linearly with the substrate concentration until a plateau was reached at excess concentration of DNA substrate. Likewise, in the presence of an excess of DNA substrate, the concentration of bound fluorescent DNA increased with the amount of precipitated methyltransferase (Supplementary Figure 4), indicating that the trapping rate is constant in this range.

To test for unspecific DNA binding, we incubated a constant amount of the GFP nanotrap with increasing volumes of cell lysate from GFP overexpressing HEK 293T cells followed by incubation with trapping substrate. The concentration of precipitated GFP increased linearly with the amount of lysate added. In contrast, the minor unspecific binding of substrate was shown to be independent of the amount of precipitated protein (Supplementary Figure 5). The unspecific binding to the agarose beads was below the detection limit for DNA coupled ATTO647N (Supplementary Figure 5B and D) and negligible for DNA coupled ATTO550 (Supplementary Figure 5A and C), when compared with the values obtained for binding to GFP-Dnmt1 and its mutant GFP-Dnmt1^{C1229W}. Thus, the minor unspecific binding is attributable to the agarose beads rather than to the protein indicating that different amounts of precipitated GFP fusions can be compared reliably. The trapping rates were slightly dependent on the lysate preparation likely reflecting the percentage of active enzyme, but highly reproducible results were obtained with independent samples from the same experimental setup.

Discrimination of enzymatic activity-dependent trapping from DNA binding

To evaluate the possibility to distinguish between DNA binding and covalent complex formation, the crucial first step of the methyl transfer reaction, we incubated GFP-Dnmt1 and the catalytic site mutant GFP-Dnmt1^{C1229W} with DNA binding and trapping substrates and measured the fluorescence after precipitation (Figure 4A). Interestingly, wild-type and mutant protein showed similar specific DNA-binding activity. However, GFP-Dnmt1 showed a higher trapping than binding rate, whereas GFP-Dnmt1^{C1229W} did not. The difference between binding and trapping rate is due to the accumulation of covalent protein-DNA complexes over time, and thus confirms previously published results on Dnmt1 and its catalytic site mutant (7).

The trapping rate obtained for the active methyltransferase GFP-Dnmt1 after 90 min at excess initial substrate concentration reflects almost exclusively covalently bound DNA substrate. This was demonstrated by an additional competition step with unlabeled binding substrate to compete with non-covalently bound labeled substrate (Figure 4B). The maximal trapping rate after this binding competition step did not change, whereas the maximal binding rate decreased proportionally. These results show that the combination of DNA binding and trapping substrates with non-fluorescent competitors allows the distinction between DNA binding and enzyme activity dependent covalent complex formation of DNA methyltransferases.

Cofactor dependence of covalent complex formation

Covalent complex formation of cytosine methyltransferases with DNA has been shown to be independent from the cofactor AdoMet. In the absence of AdoMet, the activated cytosine undergoes hydrogen exchange

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

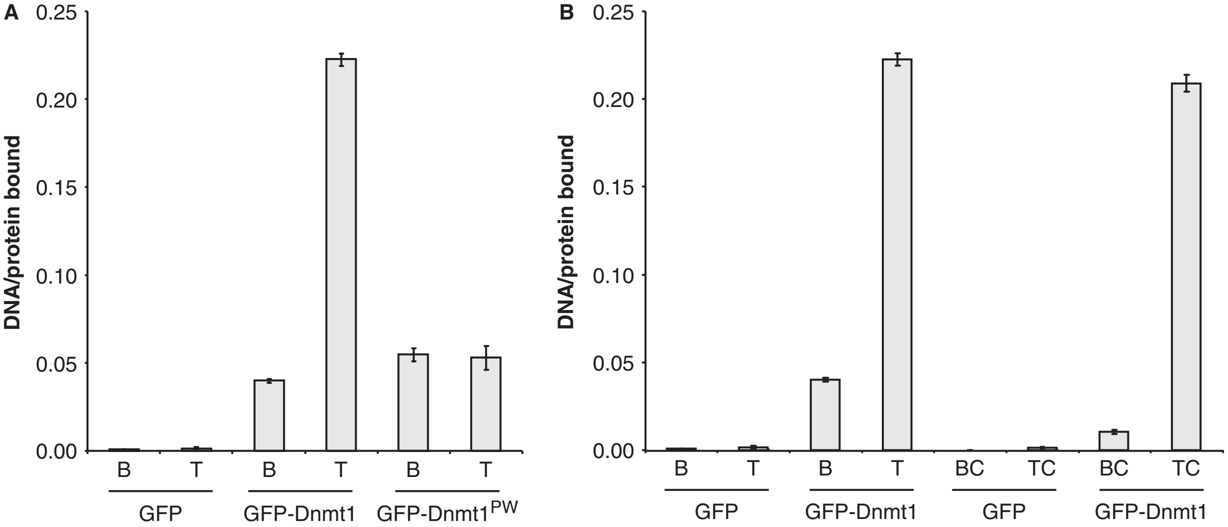


Figure 4. Binding and trapping assay with competitors. (**A**) Binding and trapping assays were performed with GFP-Dnmt1 and GFP-Dnmt1^{C1229W} and hemimethylated ATTO550 labeled DNA. Shown are the means of maximal binding and trapping rates with standard error bars from three independent experiments for the GFP control and GFP-Dnmt1 and two independent experiments for GFP-Dnmt1^{C1229W}. (**B**) Assays with substrates for binding [B], trapping [T], binding with competitor [BC] and trapping with competitor [TC] were performed with GFP-Dnmt1 as described earlier. Shown are the means with standard error bars from three independent experiments. GFP was used as negative control.

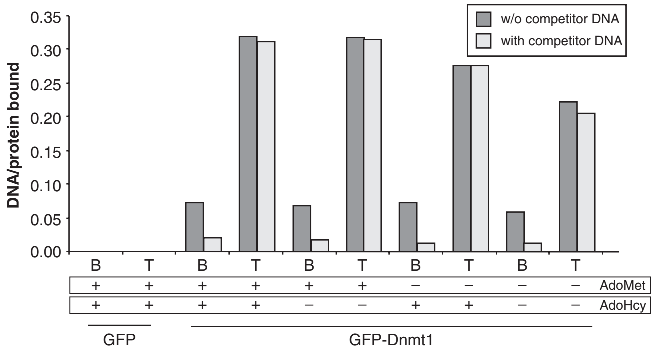


Figure 5. Covalent complex formation in dependence on AdoMet and AdoHcy. Maximal binding and trapping rate were determined for GFP-Dnmt1 and hemimethylated ATTO550 labeled DNA substrate with or without unlabeled competitor DNA. The assay buffer was supplemented with 10 µM AdoMet or AdoHcy as indicated. GFP was used as negative control.

instead of methylation at position 5. AdoMet as well as its analog and competitor *S*-Adenosyl-L-homocysteine (AdoHcy) significantly bind to the enzyme only after the DNA substrate is bound (25–27). We tested GFP-Dnmt1 binding and trapping with hemimethylated DNA substrate and compared maximal rates at different conditions (Figure 5). An additional competition step with unlabeled competitor DNA to compete for non-covalently bound labeled DNA was included to monitor irreversible covalent complex formation. In accordance with the prior biochemical studies (25–27), we found that GFP-Dnmt1 forms a covalent complex with DNA in the presence and absence of AdoMet and AdoHcy, albeit at different efficiencies. Similarly, this assay could be used for inhibitor studies and to screen for small molecules that prevent covalent enzyme-DNA complex formation.

Competition assay to directly determine substrate preference

A unique feature of this method is the possibility to compare different DNA substrates in direct competition. The trapping rates of GFP-Dnmt1 with either un- or hemimethylated DNA trapping substrate or with both substrates in direct competition clearly showed a preference for hemimethylated DNA (Figure 6A). This result demonstrates that substrate preference can be detected in a single measurement by direct competition. Interestingly, the preference for hemimethylated DNA was only pronounced in the rate of covalent complex formation (trapping assay) and not in the DNA-binding assay. The direct competition of un- and hemimethylated DNA-binding substrates revealed even a slight preference of GFP-Dnmt1 for unmethylated substrate (Figure 6B). The substrate preference of GFP-DNMT1 was tested in four independent experiments and revealed on average about 15-fold higher activity on hemimethylated than on unmethylated DNA substrate (Figure 6C). These results are consistent with data obtained with previous biochemical activity assays measuring the transfer of radioactively labeled methyl groups by purified Dnmt1 or GFP-Dnmt1 and catalytic site mutants (28–31).

In summary, we present a novel, non-radioactive assay for fast characterization of DNA methyltransferase activity and DNA binding. We show that the DNA binding, substrate specificity and activity of DNA methyltransferases fused with GFP can reliably be measured with this method. The simplicity and versatility of this assay allows fast and inexpensive screening of enzymes, complexes and mutants. By careful selection of fluorophores with distinct excitation and emission spectra, multiple fluorescent substrates can be analyzed simultaneously in direct competition. We applied the assay to the mammalian Dnmt1 and confirmed its preference for DNA

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

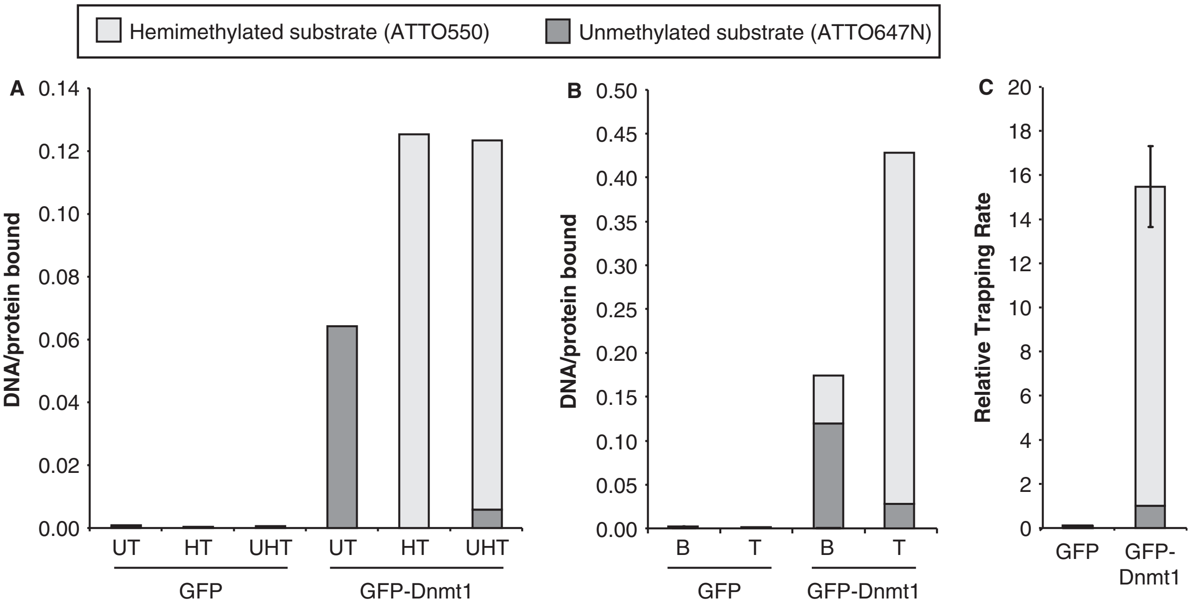


Figure 6. Substrate preference of GFP-Dnmt1. (**A**) Maximal trapping rates were determined by incubation of constant concentrations of GFP-Dnmt1 with unmethylated ATTO647N labeled DNA trapping substrate [UT], hemimethylated ATTO550 labeled DNA trapping substrate [HT] or unmethylated ATTO647N and hemimethylated ATTO550 labeled DNA trapping substrate in competition [UHT]. (**B**) Maximal binding and trapping rates for GFP-Dnmt1 are given for incubation with either unmethylated ATTO647N or hemimethylated ATTO550 labeled binding substrate (B), or unmethylated ATTO647N and hemimethylated ATTO550 labeled trapping substrate [T] in competition. (**C**) The trapping rates for GFP-Dnmt1 on unmethylated ATTO647N and hemimethylated ATTO550 labeled trapping substrate in competition were determined in four independent experiments. The value for unmethylated substrate was set to one and the relative rate for hemimethylated substrate was calculated accordingly. Means of the relative trapping rates are shown with standard error bars. GFP was used as negative control.

substrates containing hemimethylated CpG sites. In addition, we could show that the active site mutation (C1229W) abolishes covalent complex formation, but not DNA binding. The usage of GFP fusion proteins allows a direct link of biochemical data to cell biological data on subcellular localization and mobility of the very same molecule obtained by fluorescence microscopy and photobleaching experiments. However, endogenous DNA methyltransferases could analogously be assayed by incubation with fluorescent binding and/or trapping substrates and subsequent precipitation with specific antibodies. Alternatively, samples incubated with fluorescent trapping substrates could also be separated by SDS–PAGE and catalytically active methyltransferases could be detected in gel and identified by western blot or mass spectrometry. This assay can easily be adapted for general DNA- and RNA-binding studies providing a time-saving alternative to electrophoretic gel shift assays (32).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Ulrich Rothbauer for preparation of the GFP nanotrap and members of the Leonhardt group for critical discussions and proofreading of the manuscript.

FUNDING

This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG), the Nanosystem Initiative Munich (NIM) and the Center for NanoScience (CeNS) to H.L. C.F. acknowledges support from the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS) and by the International Doctorate Program 'NanoBioTechnology' of the Elite Network of Bavaria. Funding for open access charge: the Deutsche Forschungsgemeinschaft (DFG).

Conflict of interest statement. H.L. is a co-founder of Chromotek.

REFERENCES

1. Goll,M.G. and Bestor,T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
2. Hermann,A., Gowher,H. and Jeltsch,A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol. Life Sci.*, **61**, 2571–2587.
3. Robertson,K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
4. Spada,F., Rothbauer,U., Zolghadr,K., Schermelleh,L. and Leonhardt,H. (2006) Regulation of DNA methyltransferase 1. *Adv. Enzyme Regul.*, **46**, 224–234.
5. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
6. Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) Crystal structure of the HhaI DNA methyltransferase complexed with *S*-adenosyl-L-methionine. *Cell*, **74**, 299–307.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

7. Schermelleh,L., Spada,F., Easwaran,H.P., Zolghadr,K., Margot,J.B., Cardoso,M.C. and Leonhardt,H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods*, **2**, 751–756.
8. Kuch,D., Schermelleh,L., Manetto,S., Leonhardt,H. and Carell,T. (2008) Synthesis of DNA dumbbell based inhibitors for the human DNA methyltransferase Dnmt1. *Angew Chem. Int. Ed. Engl.*, **47**, 1515–1518.
9. Yoo,C.B. and Jones,P.A. (2006) Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.*, **5**, 37–50.
10. Roth,M. and Jeltsch,A. (2000) Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biol. Chem.*, **381**, 269–272.
11. Yokochi,T. and Robertson,K.D. (2004) DMB (DNMT-magnetic beads) assay: measuring DNA methyltransferase activity in vitro. *Methods Mol. Biol.*, **287**, 285–296.
12. Kim,B.Y., Kwon,O.S., Joo,S.A., Park,J.A., Heo,K.Y., Kim,M.S. and Ahn,J.S. (2004) A column method for determination of DNA cytosine-C5-methyltransferase activity. *Anal. Biochem.*, **326**, 21–24.
13. Hubscher,U., Pedrali-Noy,G., Knust-Kron,B., Doerfler,W. and Spadari,S. (1985) DNA methyltransferases: activity minigel analysis and determination with DNA covalently bound to a solid matrix. *Anal. Biochem.*, **150**, 442–448.
14. Margot,J.B., Aguirre-Arteta,A.M., Di Giacco,B.V., Pradhan,S., Roberts,R.J., Cardoso,M.C. and Leonhardt,H. (2000) Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J. Mol. Biol.*, **297**, 293–300.
15. Woo,Y.H., Rajagopalan,P.T. and Benkovic,S.J. (2005) A non-radioactive DNA methyltransferase assay adaptable to high-throughput screening. *Anal. Biochem.*, **340**, 336–340.
16. Tamura,T., Kataoka,A., Shu,L.Y., Ashida,A., Tanaka,H. and Inagaki,K. (2002) An in vitro screening method for DNA cytosine-C5-methylase inhibitor. *Nat. Prod. Lett.*, **16**, 25–27.
17. Yamamoto,T., Nagasaka,T., Notohara,K., Sasamoto,H., Murakami,J., Tanaka,N. and Matsubara,N. (2004) Methylation assay by nucleotide incorporation: a quantitative assay for regional CpG methylation density. *Biotechniques*, **36**, 846–850, 852, 854.
18. Humeny,A., Beck,C., Becker,C.M. and Jeltsch,A. (2003) Detection and analysis of enzymatic DNA methylation of oligonucleotide substrates by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Anal. Biochem.*, **313**, 160–166.
19. Salyan,M.E., Pedicord,D.L., Bergeron,L., Mintier,G.A., Hunihan,L., Kuit,K., Balanda,L.A., Robertson,B.J., Feder,J.N., Westphal,R. *et al.* (2006) A general liquid chromatography/mass spectroscopy-based assay for detection and quantitation of methyltransferase activity. *Anal. Biochem.*, **349**, 112–117.
20. Boussif,O., Lezoualc'h,F., Zanta,M.A., Mergny,M.D., Scherman,D., Demeneix,B. and Behr,J.P. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl Acad. Sci. USA*, **92**, 7297–7301.
21. Rothbauer,U., Zolghadr,K., Muyldermans,S., Schepers,A., Cardoso,M.C. and Leonhardt,H. (2007) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell Proteomics.*, **7**, 282–289.
22. Laemmli,U.K., Beguin,F. and Gujer-Kellenberger,G. (1970) A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.*, **47**, 69–85.
23. Easwaran,H.P., Schermelleh,L., Leonhardt,H. and Cardoso,M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.*, **5**, 1181–1186.
24. Schermelleh,L., Haemmer,A., Spada,F., Rosing,N., Meilinger,D., Rothbauer,U., Cardoso,M.C. and Leonhardt,H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.*, **35**, 4301–4312.
25. Svedruzic,Z.M. and Reich,N.O. (2004) The mechanism of target base attack in DNA cytosine carbon 5 methylation. *Biochemistry*, **43**, 11460–11473.
26. Svedruzic,Z.M. and Reich,N.O. (2005) DNA cytosine C5 methyltransferase Dnmt1: catalysis-dependent release of allosteric inhibition. *Biochemistry*, **44**, 9472–9485.
27. Wu,J.C. and Santi,D.V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.*, **262**, 4778–4786.
28. Jeltsch,A. (2006) On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics*, **1**, 63–66.
29. Pradhan,M., Esteve,P.O., Chin,H.G., Samaranayke,M., Kim,G.D. and Pradhan,S. (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry*, **47**, 10000–10009.
30. Pradhan,S., Bacolla,A., Wells,R.D. and Roberts,R.J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J. Biol. Chem.*, **274**, 33002–33010.
31. Hermann,A., Goyal,R. and Jeltsch,A. (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.*, **279**, 48350–48359.
32. Man,T.K. and Stormo,G.D. (2001) Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res.*, **29**, 2471–2478.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding

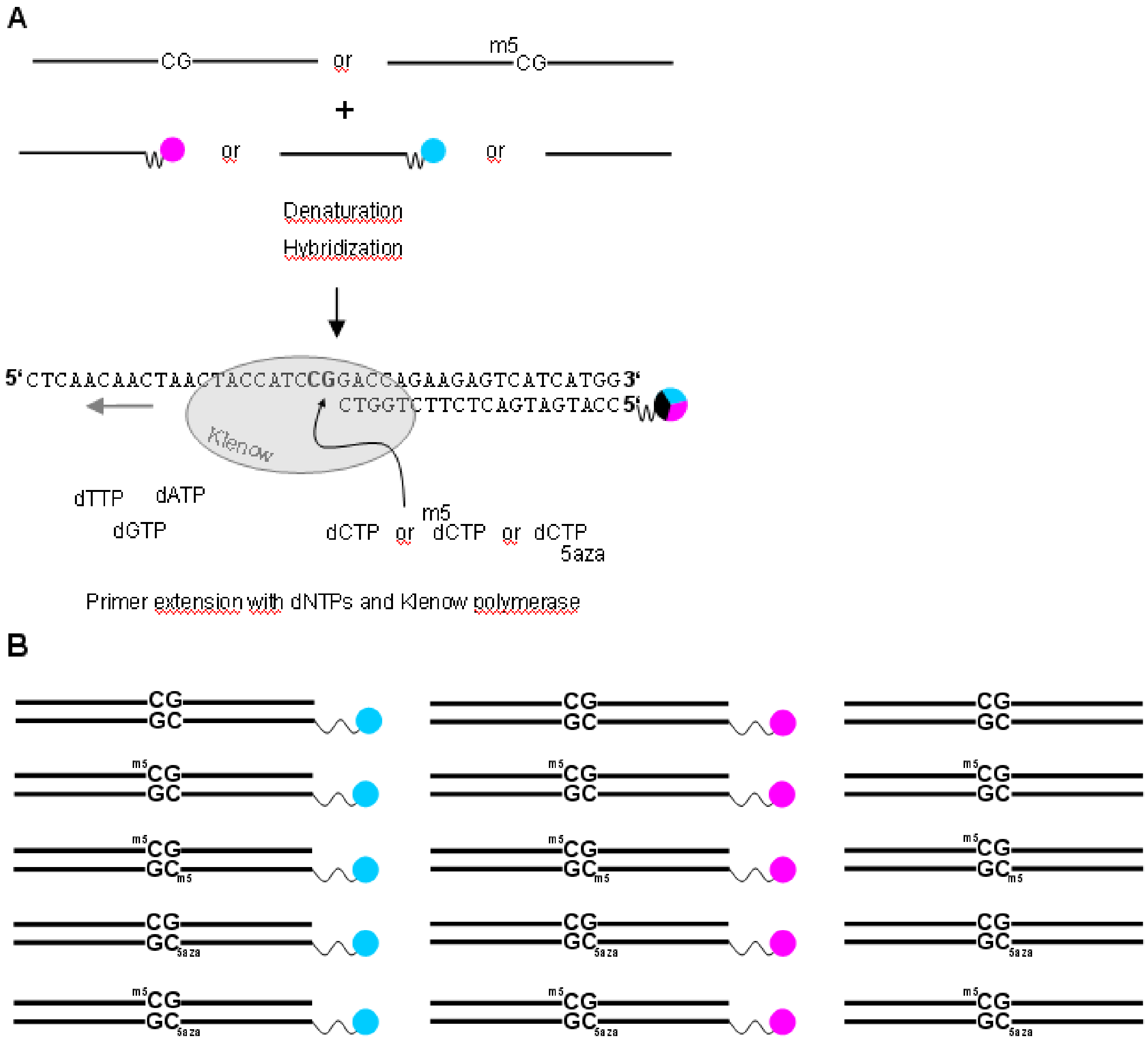
Carina Frauer and Heinrich Leonhardt*

Ludwig Maximilians University Munich, Department of Biology, Center for Integrated Protein Science Munich (CIPS^{M}), 82152 Planegg-Martinsried, Germany.

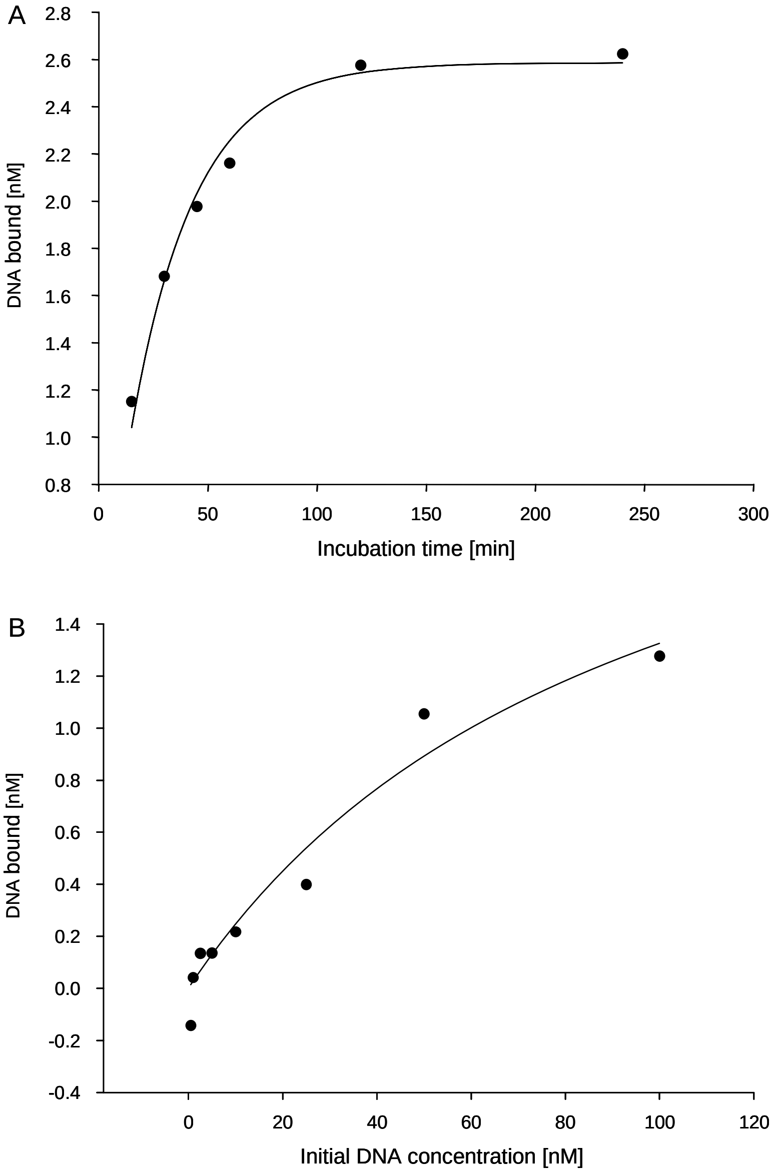
*Corresponding author: e-mail: h.leonhardt@lmu.de; Fax +49 89 2180-74236; Tel. +49 89 2180-74232.

Keywords: DNA methyltransferase activity, DNA binding, 5-aza-dC, Dnmt1.

SUPPLEMENTARY DATA:


A
CG
or
m5
CG
+
or
or
Denaturation
Hybridization
5‘CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG3‘
CTGGTCTTCTCAGTAGTACC5‘
Klenow
dTTP
dATP
dGTP
dCTP or m5 dCTP or dCTP 5aza
Primer extension with dNTPs and Klenow polymerase
B
CG
GC
m5
m5
5aza

Supplementary Figure 1. DNA substrate preparation. (A) Experimental strategy for preparation of different DNA substrates from a small set of oligonucleotides. (B) Outline of the structure of the 15 possible different DNA substrates.



Supplementary Figure 2. Fitting of DNA binding data for GFP-Dnmt1 and hemimethylated ATTO550 labeled binding substrate in dependence on time and initial DNA substrate concentration. (A) The data of DNA bound over time were fitted using the equation

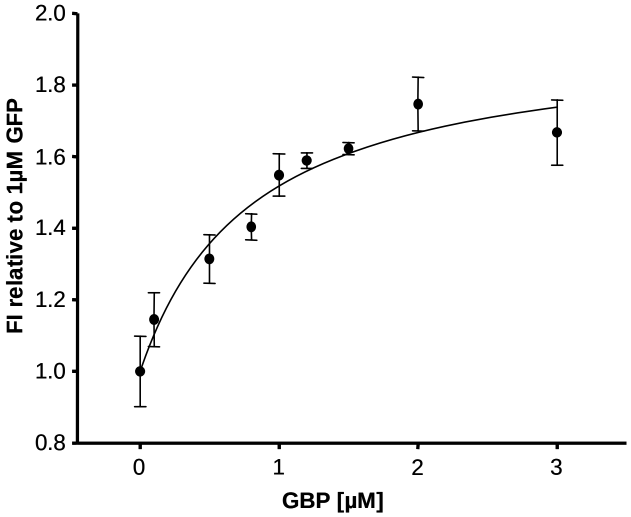
$$y = Y_{\max} \cdot (1 - e^{-kx}),$$

resulting in Y_{max}= 2.59 +/- 0.06 nM and an observed rate constant of k=0.034 +/-0.002 min^{-1} (R^2=0.983, SigmaPlot).

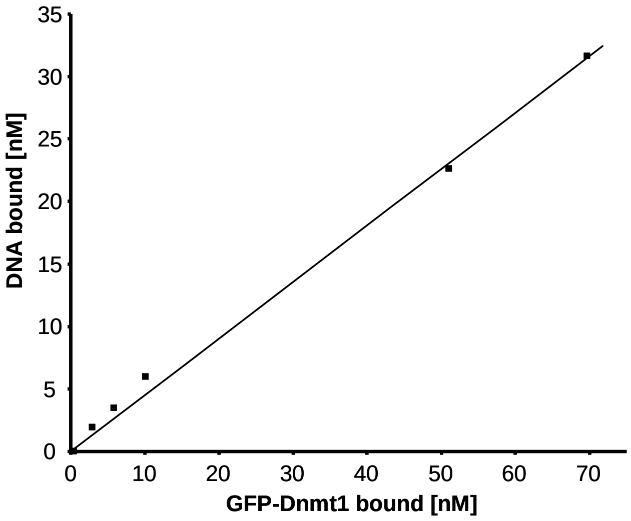
The data of DNA bound over initial DNA concentration were fitted with

$$y = \frac{Y_{\max} \cdot x}{K_d + x},$$

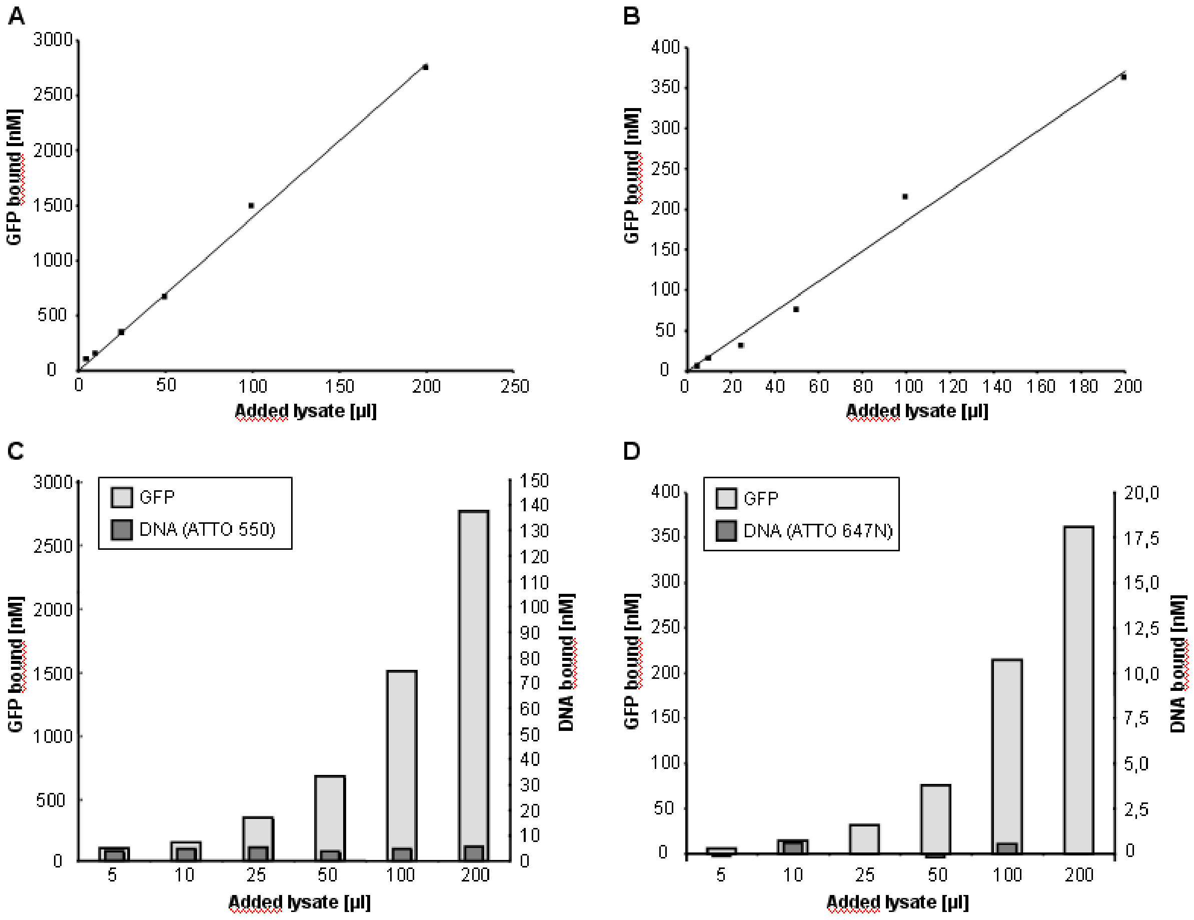
resulting in Y_{max}= 2.58 +/- 0.84 nM and a dissociation constant K_d= 94.63 +/- 52.77 nM (R^2=0.956, SigmaPlot). The data for this analysis were taken from Figure 3.



Supplementary Figure 3. Determination of GFP fluorescence upon GBP binding. GFP (1 µM) was incubated with an increasing concentration of GBP (0, 0.05, 0.1, 0.3, 0.5, 0.8, 1, 1.2 and 1.5 µM) and the fluorescence signal was detected with a 485±8 nm excitation filter and a 520±17 nm emission filter. The fluorescence signal of GFP alone was set to 1. The mean relative fluorescence signal with standard deviation bars of three measurements was plotted against the amount of GBP added. The results show that GBP binding increases GFP fluorescence under these assay conditions about 1.7 fold, which was taken into account for all quantifications throughout.



Supplementary Figure 4. Linearity of the trapping assay. Trapping was performed with an increasing concentration of GFP-Dnmt1 at a concentration of 100 nM hemimethylated ATTO550 labeled trapping substrate. The concentration of DNA that was pulled down with the beads was plotted against the concentration of pulled down GFP-Dnmt1.



Supplementary Figure 5. Determination of unspecific DNA binding. Increasing amounts of GFP lysate were added to equal aliquots of beads and incubated with ATTO550 (A, C) or ATTO647N labeled DNA substrate (B, D). In (A) and (B) the concentration of precipitated GFP was plotted against the volume of lysate added to the beads. (C) and (D) show the concentration of GFP and DNA precipitated relative to the volume of lysate added. The results show that unspecific binding of labeled DNA substrates to either the beads or GFP is negligible.

2.2 MODULATION OF PROTEIN PROPERTIES IN LIVING CELLS USING NANOBODIES

TECHNICAL REPORTS

nature
structural &
molecular biology

Modulation of protein properties in living cells using nanobodies

Axel Kirchhofer[1–3], Jonas Helma[2,4], Katrin Schmidthals[2,4], Carina Frauer[2,4], Sheng Cui[1–3], Annette Karcher[1–3], Mireille Pellis[5,6], Serge Muyldermans[5,6], Corella S Casas-Delucchi[7], M Cristina Cardoso[7], Heinrich Leonhardt[2,4,8], Karl-Peter Hopfner[1–3] & Ulrich Rothbauer[2,4,8,9]

© 2010 Nature America, Inc. All rights reserved.

Protein conformation is critically linked to function and often controlled by interactions with regulatory factors. Here we report the selection of camelid-derived single-domain antibodies (nanobodies) that modulate the conformation and spectral properties of the green fluorescent protein (GFP). One nanobody could reversibly reduce GFP fluorescence by a factor of 5, whereas its displacement by a second nanobody caused an increase by a factor of 10. Structural analysis of GFP–nanobody complexes revealed that the two nanobodies induce subtle opposing changes in the chromophore environment, leading to altered absorption properties. Unlike conventional antibodies, the small, stable nanobodies are functional in living cells. Nanobody-induced changes were detected by ratio imaging and used to monitor protein expression and subcellular localization as well as translocation events such as the tamoxifen-induced nuclear localization of estrogen receptor. This work demonstrates that protein conformations can be manipulated and studied with nanobodies in living cells.

Green fluorescent protein (GFP) is a barrel-shaped protein with a central *p*-hydroxybenzylidene-imidazolidone chromophore. The formation of the chromophore results from an oxidative backbone cyclization involving residues Ser65, Tyr66 and Gly67 (refs. 1–3). The original wild-type GFP (wtGFP) is characterized by a dual-peak excitation spectrum with a major absorption maximum at 395 nm and a minor one at 477 nm. Excitation at either wavelength results in the emission of green fluorescence at ~507 nm. This dual absorption of GFP stems from the existence of two interconvertible alternative states of the chromophore. The neutral phenol state of the chromophore absorbs at 395 nm, whereas the deprotonated phenolate anion absorbs at 477 nm[4]. During the past decade, the fluorescence properties of GFP have been successfully modified by mutagenesis[5–7]. For example, the most widely used mutant, enhanced GFP (eGFP), features increased brightness, improved photostability and a single excitation peak at 488–490 nm[5]. Additional types of bioimaging applications became possible with the photoactivatable (paGFP) variant[8]. Recently, new permuted GFP derivates were described as molecular sensors to monitor the presence of calcium, which induces structural rearrangements that block solvent access to the chromophore[9,10].

Here we investigated whether spectral properties of fluorescent proteins can be modulated with antibody derivatives. For this purpose, we tested so-called 'nanobodies', which are small, antigen-binding, single-domain polypeptides derived from the variable heavy chain (VHH) of the heavy chain–only antibodies of camelids[11]. Nanobodies are potent alternatives to conventional antibodies, with enhanced stability and reduced size but similar antigen-binding characteristics[12]. Applications described thus far include targeting and tracing of antigens in live cells and targeted modulation of enzymes as well as their usage as immobilized nanotraps to precipitate protein complexes *in vivo* and *in vitro*[13–17].

RESULTS

Generation of GFP-binding nanobodies

To isolate and characterize GFP-binding nanobodies, we generated a phagemid library by cloning the VHH repertoire from the heavy-chain antibodies of GFP-immunized camelids. Next, we displayed the VHH repertoire on phage particles and selected individual GFP-specific binders after panning followed by a solid-phase ELISA screening. Seven unique GFP-specific binders were determined by DNA sequence analysis of the clones. We termed the resulting proteins GFP-binding proteins (GBPs) 1–7. The GBPs were cloned with a C-terminal hexahistidine (His_6) tag, expressed in *Escherichia coli* and purified by immobilized metal affinity chromatography (IMAC). All GBPs coelute with wtGFP in an apparent 1:1 complex in gel filtration chromatography, verifying their stable binding to wtGFP (data not shown).

Nanobodies affecting GFP fluorescence intensity

To screen for nanobodies that alter GFP fluorescence properties, we added increasing amounts of GBP1–7 to wtGFP and measured the fluorescence intensity. We identified two nanobodies, GBP1 and

---

[1]Gene Center at the Department of Chemistry and Biochemistry, Ludwig-Maximilians University Munich, Munich, Germany. [2]Center for Integrated Protein Science, Munich, Germany. [3]Munich Center for Advanced Photonics, Munich, Germany. [4]Biocenter at the Department of Biology II, Ludwig-Maximilians University Munich, Planegg-Martinsried, Germany. [5]Department of Molecular and Cellular Interactions and [6]Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Brussels, Belgium. [7]Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany. [8]Center for NanoScience, Ludwig-Maximilians University Munich, Munich, Germany. [9]ChromoTek GmbH, Planegg-Martinsried, Germany. Correspondence should be addressed to H.L. (H.Leonhardt@lmu.de), K.-P.H. (Hopfner@lmb.uni-muenchen.de) or U.R. (U.Rothbauer@chromotek.com).

Received 4 June; accepted 21 October; published online 13 December 2009; doi:10.1038/nsmb.1727

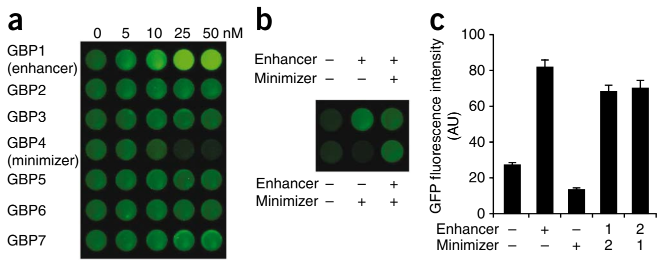


Figure 1 Identification of nanobodies modulating the fluorescence of GFP. (**a**) Fluorescence *in vitro* binding assay. Titration of seven unique GFP binding proteins (GBP1–7) from 0–50 nM on 50 nM purified wtGFP. The fluorescence signal intensity of wtGFP was quantified using a laser scanner. (**b**) Minimizer can be displaced by Enhancer but not vice versa. Upper row, GFP was either mock incubated or incubated with equimolar amounts of Enhancer, or Enhancer was added followed immediately (5–15 s) by equimolar amounts of Minimizer. Lower row, same experimental setup as above but with Minimizer being added first. GFP emission was detected as described for **a**. (**c**) Quantification of GFP fluorescence as shown in **b**. The order of addition of Enhancer or Minimizer is indicated by numbers 1 and 2. Means and s.d. (error bars) of three independent experiments are shown.

© 2010 Nature America, Inc. All rights reserved.

GBP4, that had a pronounced effect on the fluorescence emission of wtGFP (**Fig. 1a** and **Supplementary Fig. 1a**). Whereas binding of GBP1 leads to a fourfold fluorescence enhancement, binding of GBP4 reduces the fluorescence by a factor of 5. Overall, there is a remarkable 20-fold difference in fluorescence intensity between the two GFP–nanobody complexes under the conditions used. According to their observed impact on GFP fluorescence, we termed GBP1 and GBP4 'Enhancer' and 'Minimizer', respectively. The augmented fluorescence of the GFP–Enhancer complex is comparable to the improved spectral properties of eGFP. This raised the question of whether Enhancer might be able to further increase the optimized fluorescence of eGFP. Indeed, binding of Enhancer to recombinantly purified eGFP resulted in an additional fluorescence increase of about 1.5-fold. In contrast, binding of Minimizer reduced the fluorescence intensity of eGFP by a factor of 8 (**Supplementary Fig. 1b**,**c**). A comparable fluorescence modulation was also observed after addition of Enhancer or Minimizer to soluble cell extract derived from human embryonic kidney (HEK) 293T cells expressing eGFP (**Supplementary Fig. 1c**). To investigate whether selected nanobodies recognize different epitopes, we performed sandwich-binding assays. Additive binding to already-constituted GFP–Enhancer complexes could be detected for GBP2, GBP5, GBP6 and GBP7 but not for Minimizer, suggesting that Enhancer and Minimizer compete for overlapping epitopes of GFP (**Supplementary Fig. 2a**).

Based on the opposite effects of Enhancer and Minimizer binding, we investigated the wtGFP fluorescence modulation in the presence of both nanobodies (**Fig. 1b**,**c**). Notably, after a primary addition of Enhancer, wtGFP fluorescence increases and is only slightly reduced by the consecutive addition of Minimizer (**Fig. 1c**). In contrast, when Minimizer is added first, the reduction of fluorescence can be completely reversed, and fluorescence further enhanced, by subsequent addition of Enhancer (**Fig. 1c**). Although the ability of Enhancer to displace Minimizer at equimolar concentrations suggests that Enhancer has a higher affinity for wtGFP, no substantial difference in binding constants (K_d) was detected (**Supplementary Fig. 2b** and **Supplementary Table 1**).

Structure of GFP–Enhancer and GFP–Minimizer complexes

To elucidate the molecular mechanism underlying the observed fluorescence modulation, we determined crystal structures of the GFP–Enhancer and GFP–Minimizer complexes to 2.15 and 1.6 Å resolution, respectively (**Table 1**). Both nanobodies recognize two different, slightly overlapping epitopes on the GFP surface (**Fig. 2a**,**b**). Thus, the observed competition for binding seems to result from a steric clash between the nanobodies. Enhancer binds wtGFP in a frontwise manner at an exposed loop region between GFP β-strands 6 and 7 as well as parts of β-strand 8, making specific contacts with all three complementarity-determining regions (CDR) of the nanobody (**Fig. 2a**). Previous structural studies of nanobodies have shown that CDR3 normally folds over the framework 2 region, which in the case of classical antibodies binds to the variable domain of the light chain (VL)[18]. In contrast, the extremely short CDR3 of Enhancer is stretched out, thereby making the framework 2 region accessible to solvent in the antigen-free form. Unexpectedly, the entire framework 2 area participates in GFP recognition, in contrast to the structure of classical antibodies where the framework 2 area would contact the VL domain. The majority of the specific contacts are formed between CDR3 and GFP, whereas CDR1 and 2 remain exposed to the solvent. Notably, the interaction between GFP and Enhancer is predominantly electrostatic, spanning an interface of 672 Å^2 (**Supplementary Table 2**). An additional nonpolar contact is mediated by Phe98$^{\mathrm{Enhancer}}$, which binds a hydrophobic surface patch on GFP formed by Ala206$^{\mathrm{GFP}}$, Leu221$^{\mathrm{GFP}}$ and Phe223$^{\mathrm{GFP}}$.

In contrast, the Minimizer nanobody has its CDR3 folded over the framework 2 region and binds wtGFP in a sideways orientation, using its elongated CDR3 to target β-strands 6 and 7 of GFP (**Fig. 2c**). The interaction with GFP is quite remarkable, since the nanobody targets the rigid and flat surface rather than the more flexible and easily accessible loops at the top and the bottom of the β-can. In comparison to Enhancer, Minimizer occupies a smaller surface area (652 Å^2) on GFP and the overall number of contacts is smaller (**Supplementary Table 3**). This observation is in line with its ready displacement by Enhancer in the competition assay (**Fig. 1c**). However, it is difficult

Table 1 Data collection and refinement statistics

	SeMet–GFP–Enhancer	GFP–Minimizer
Data collection		
Space group	$P4_22$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	160.5, 160.5, 78.8	50.8, 81.6, 94.5
	Peak	
Wavelength (Å)	0.9793	0.98137
Resolution (Å)	2.15	1.5
R_{sym}	4.6 (45.4)[a]	5.7 (39.1)
$I / \sigma I$	14.95 (2.25)	16.91 (2.84)
Completeness (%)	99.3 (98.9)	97.5 (86.3)
Redundancy	3.13[b]	3.85
Refinement		
Resolution (Å)	46.00-2.15	47.00-1.61
No. reflections	56,271	49,989
R_{work}/R_{free}	21.3/25.5	16.1/19.4
No. atoms		
Protein	5,385	2,818
Water	407	685
B-factors		
Protein (Å^2)	46.5	13.7
Water (Å^2)	48.4	30.1
R.m.s. deviations		
Bond lengths (Å)	0.008	0.005
Bond angles (°)	1.14	1.04

[a]The structures of SeMet–GFP–Enhancer and of native GFP–Minimizer were determined with one crystal each. Values in parentheses are for highest-resolution shell. [b]For SeMet–GFP–Enhancer the anomalous redundancy is calculated.

© 2010 Nature America, Inc. All rights reserved.

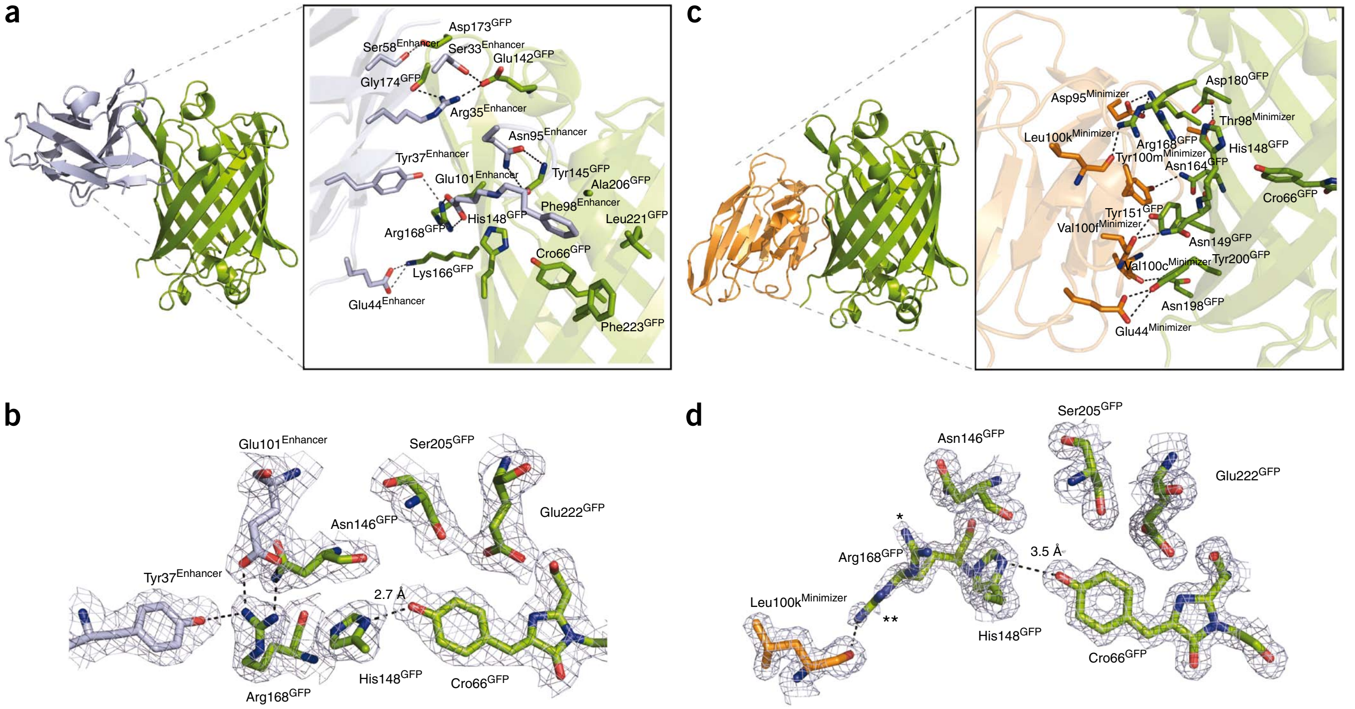

a
Asp173GFP
Ser58Enhancer
Ser33Enhancer
Glu142GFP
Gly174GFP
Arg35Enhancer
Asn95Enhancer
Tyr37Enhancer
Glu101Enhancer
Tyr145GFP
Ala206GFP
Phe98Enhancer
Arg168GFP
His148GFP
Leu221GFP
Cro66GFP
Lys166GFP
Glu44Enhancer
Phe223GFP
c
Asp95Minimizer
Asp180GFP
Thr98Minimizer
Leu100kMinimizer
Arg168GFP
Tyr100mMinimizer
Asn164GFP
His148GFP
Cro66GFP
Tyr151GFP
Val100fMinimizer
Asn149GFP
Val100cMinimizer
Tyr200GFP
Asn198GFP
Glu44Minimizer
b
Glu101Enhancer
Ser205GFP
Asn146GFP
Glu222GFP
Tyr37Enhancer
2.7 Å
His148GFP
Cro66GFP
Arg168GFP
d
Ser205GFP
Asn146GFP
Glu222GFP
*
Arg168GFP
3.5 Å
Leu100kMinimizer
**
His148GFP
Cro66GFP

Figure 2 Structures of the GFP–nanobody complexes. (**a**–**d**) Enhancer (**a**; light blue ribbon model) and Minimizer (**c**) (orange ribbon model) recognize two different nonlinear epitopes on the surface of the GFP β-can (green ribbon model). The insets in **a** and **c** show details of the binding sites with selected residues and the GFP chromophore (Cro66$^{\text{GFP}}$) highlighted as sticks. The chromophore environments for the GFP–Enhancer (**b**) and GFP–Minimizer complexes (**d**), respectively, are superimposed with $2F_o - F_c$ density maps (contoured at 1.0σ). Two alternative conformations of R168$^{\text{GFP}}$ are marked with * and **. Nanobody residues numbered as previously described[19]; in Minimizer, the 15 residues corresponding to position 100 are labeled a-o.

to quantitatively judge binding strengths from the structurally observed number of contacts or buried surface areas.

Induced rearrangements in the GFP chromophore environment

In general, association of the nanobodies has no substantial global influence on the overall fold of GFP. Unbound GFP (PDB 1EMB (ref. 4)) has r.m.s. deviations of 0.391 Å (Cα atoms) and 0.359 Å (Cα atoms) from GFP in complex with Enhancer and Minimizer, respectively. From previous structural studies of GFP, however, it is well established that slight perturbations in the chromophore environment can have vast effects on its fluorescence properties[4]. Indeed, a comparison of the GFP nanobody structures with previously published GFP structures reveals that the GFP–Enhancer complex harbors the deprotonated, negatively charged state of the GFP chromophore, which has been described for the mutant GFP$^{\text{S65T}}$ (ref. 4) (**Fig. 2d**). Binding of Enhancer induces slight structural shifts in the loop region from Glu142$^{\text{GFP}}$ to His148$^{\text{GFP}}$ and fixes Arg168$^{\text{GFP}}$ in close proximity to His148$^{\text{GFP}}$. The conformation of the Arg168$^{\text{GFP}}$ side chain is stabilized by direct contacts with Enhancer residues Tyr37$^{\text{Enhancer}}$ and Glu101$^{\text{Enhancer}}$ (nanobody residues numbered as previously described[19]). These structural rearrangements bring the proton acceptor His148$^{\text{GFP}}$ into very close proximity to the hydroxyl group of the GFP chromophore (distance 2.7 Å, compared to 2.8 Å for GFP$^{\text{S65T}}$ and 3.4 Å for wtGFP). Thus, it is likely that binding of Enhancer facilitates improved proton extraction from the chromophore hydroxyl by His148$^{\text{GFP}}$, thereby stabilizing the phenolate anion of the chromophore and enhancing the fluorescence intensity. In contrast, the chromophore environment of the GFP–Minimizer complex is considerably different and shows similarities to the situation present in wtGFP[4]. Notably, Arg168$^{\text{GFP}}$ is rather flexible in comparison to the Enhancer complex: we could trace two alternative conformations of its guanidine group in the electron density. In one of the conformations, Arg168$^{\text{GFP}}$ is tilted away from His148$^{\text{GFP}}$ and instead makes specific contacts with the backbone carbonyl of Leu100k$^{\text{Minimizer}}$ (following the Kabat numbering[19]; the 15 residues corresponding to this position in Minimizer were labeled a-o). This nanobody-induced conformational change reduces the electrostatic forces exerted on His148$^{\text{GFP}}$, which is pulled back from the hydroxyl group of the chromophore and positioned with 3.5-Å distance (wtGFP: 3.4 Å) (**Fig. 2b**), too far to efficiently stabilize the phenolate anion. Instead, binding of Minimizer likely stabilizes an arrangement of the chromophore's surrounding environment that favors the neutral phenol state of the chromophore. In support of this model, binding of Enhancer to eGFP—where the phenolate anion state is stabilized by an engineered mutation—leads to an increase by a factor of 1.5 compared to a factor of 5 for wtGFP for wtGFP, whereas the fluorescence intensity is suppressed by Minimizer binding by a factor of 8 for eGFP compared to a factor of 4 for wtGFP. In summary, these two nanobodies appear to recognize and thermodynamically stabilize two conformational states of GFP, which affect the protonation state and thereby the spectral properties of the chromophore.

Enhancer and Minimizer modulate spectral properties of GFP

To directly test our structure-derived hypothesis that the interactions of the two nanobodies stabilize either the neutral or the ionized state of the chromophore, we analyzed the fluorescence absorption spectra of GFP in complex with either Enhancer or Minimizer (**Fig. 3a**,**b**). In support of our model, Enhancer increased absorption at 475 nm while reducing it at 395 nm for both wtGFP and eGFP. Minimizer modulated the absorption in exactly the opposite manner, reducing absorption at 475 nm and increasing it at 395 nm. We did not observe substantial changes in fluorescence lifetime upon Enhancer or Minimizer binding (**Supplementary Table 4**).

© 2010 Nature America, Inc. All rights reserved.

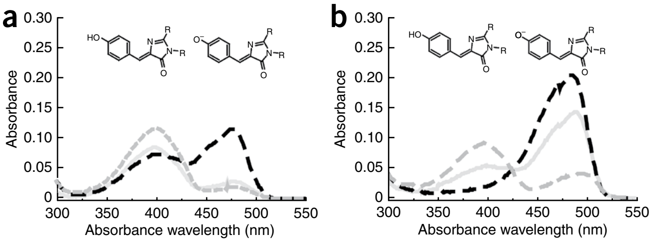
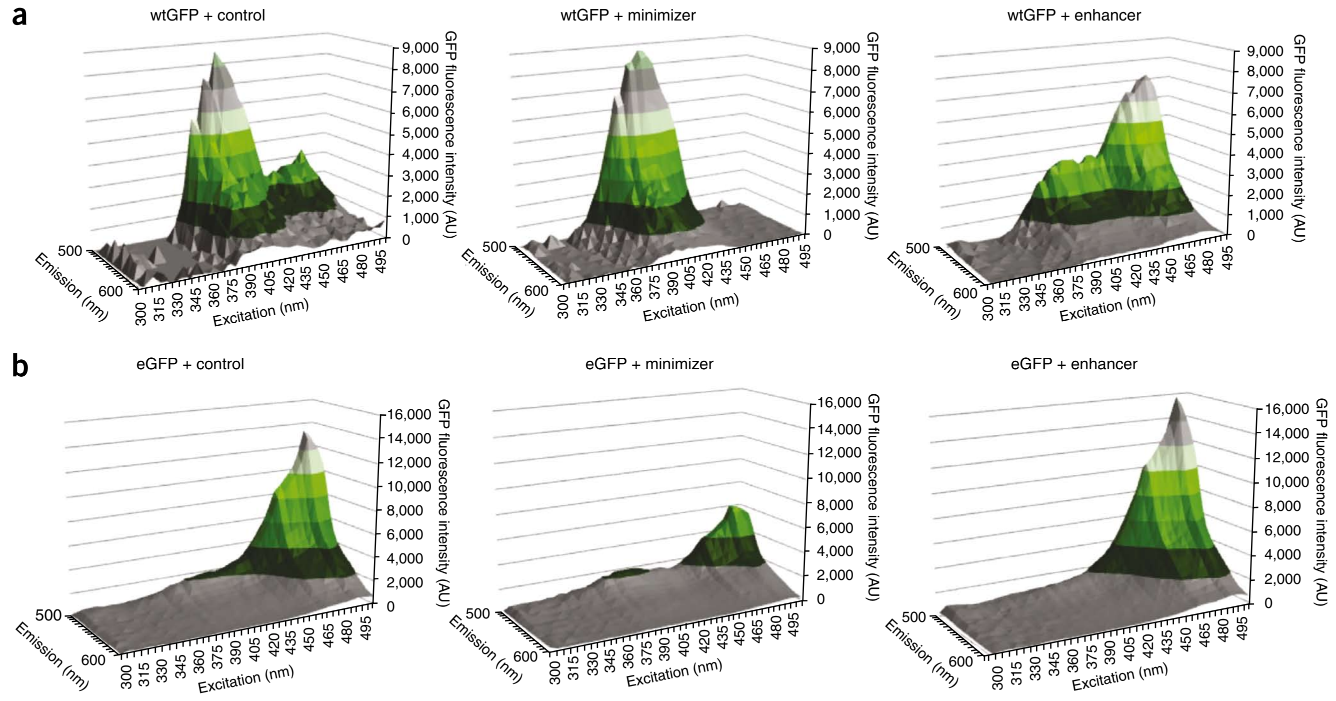


Figure 3 Nanobody-induced modulation of GFP spectral properties. Absorption spectra of the unbound (gray, solid line), Minimizer-bound (gray, dashed line) or Enhancer-bound (black, dashed line) wtGFP (**a**) or eGFP (**b**). The absorption at 395 nm corresponds to the protonated chromophore and absorption at 475 nm to the anionic chromophore (see chemical structures above).

Therefore, the fluorescence modulation upon nanobody binding is likely due to a change in the absorption efficiency at different wavelengths, which correlates with the magnitude of fluorescence emission and can be attributed to the protonation state of the chromophore.

Modulation of spectral properties of GFP in living cells

We next tested whether the nanobody-induced fluorescence modulation observed *in vitro* also occurs in living cells. To this end, we transfected human embryonic kidney (HEK) 293T cells with expression vectors encoding wtGFP or eGFP in combination with constructs encoding Enhancer, Minimizer or a control nanobody fused to monomeric red fluorescent protein (mRFP). Two days after transfection, we performed combined excitation and emission scans of GFP fluorescence intensities in living cells (**Fig. 4a**,**b** and **Supplementary Figs. 3** and **4**). Indeed, both nanobodies induced similar spectral changes in wtGFP and eGFP fluorescence, which demonstrates that Enhancer as well as Minimizer can effectively modulate GFP fluorescence in living cells. To obtain a concentration-independent measure of Enhancer or Minimizer binding and test whether subcellular differences can be detected, we determined the induced shift of GFP absorption maxima from 405 nm to 488 nm by ratio imaging. We tethered the Enhancer to the nuclear lamina by transfecting HeLa cells with an expression construct coding for an Enhancer–lamin B1 fusion. This fusion protein is incorporated into the nuclear lamina, generating an intranuclear binding site for GFP. After coexpressing wtGFP in excess, we acquired images with excitations at 405 nm and at 488 nm to detect relative differences in GFP fluorescence intensities at the nuclear lamina due to binding to locally immobilized Enhancer–lamin B1 fusion protein. Although GFP was bound and enriched at the nuclear lamina, this structure was barely detectable after excitation at 405 nm. However, excitation at 488 nm led to an increased signal at the nuclear lamina (**Fig. 4c**). The presence of Enhancer at this distinct subcellular structure could be visualized by calculating the ratio between pixel



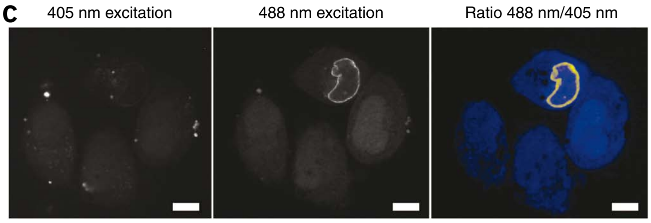


Figure 4 Nanobodies modulate GFP fluorescence in living cells. (**a**,**b**) Binding of Minimizer and Enhancer shifts excitation and emission spectra of both wtGFP (**a**) and eGFP (**b**) in living cells. (**c**) Ratio imaging. Shown are cells expressing wtGFP, which is dispersedly distributed. The topmost cell coexpresses Enhancer fused to lamin B1 (GBP1–lamin B1). Whereas only a weak signal at the nuclear lamina is detectable with excitation at 405 nm, the relative and absolute signal increased with excitation at 488 nm. Bound and unbound GFP can be distinguished independently by matrix algebra calculating the ratio between the signal intensities obtained with excitation at 488 nm and 405 nm for every pixel. The 488 nm/405 nm ratios are displayed in a false color gradient from blue (unbound) to yellow (bound). Scale bar is 10 μm.

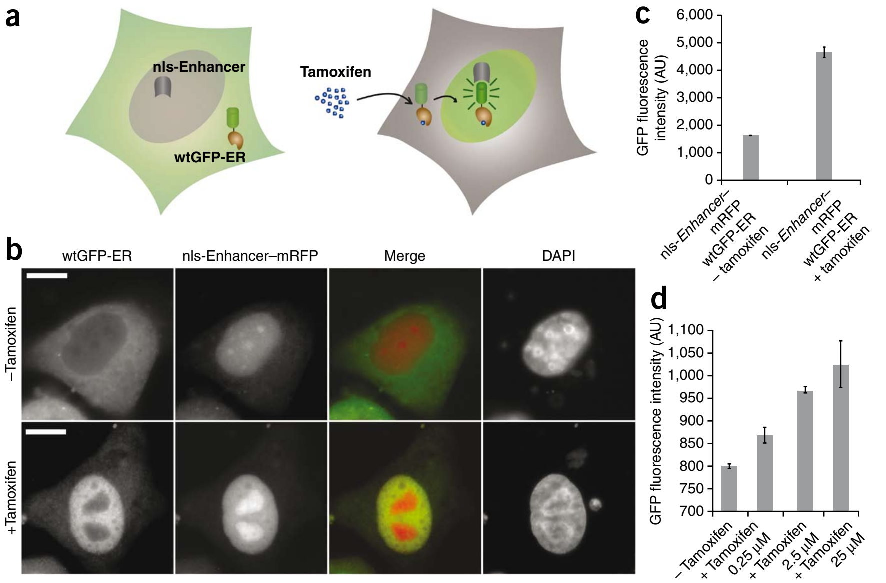


Figure 5 Nucleocytoplasmic translocation detected by nanobody-mediated fluorescence enhancement. (**a**) Schematic outline of the translocation assay: GFP-ER$_{286–595}$ and nls-Enhancer are present in separate compartments, the cytoplasm and the nucleus. Addition of tamoxifen induces translocation of GFP-ER$_{286–595}$ to the nucleus, where it binds to nls-Enhancer, leading to an increase of the fluorescence intensity. (**b**) Representative cells were analyzed by fluorescence microscopy (scale bar: 10 µM). Untreated cells (–Tamoxifen, upper row) show an almost exclusive distribution of GFP-ER$_{286–595}$ to the cytoplasm, whereas the nls-Enhancer is localized in the nucleus. After addition of tamoxifen (+Tamoxifen, lower row) GFP-ER$_{286–595}$ colocalizes with nls-Enhancer in the nucleus. (**c**,**d**) Nucleocytoplasmic translocation was measured in a plate format in living cells by detection of GFP fluorescence intensity. (**c**) After translocation of GFP-ER$_{286–595}$ into the nucleus upon addition of tamoxifen, binding of nls-Enhancer leads to a three-fold increase in fluorescence intensity. (**d**) Detection of the dose-dependent translocation efficiency after addition of increasing concentrations of tamoxifen (as indicated). Means and s.d. (error bars) of three independent experiments are shown.

intensities obtained at 405 nm and at 488 nm and displaying the ratio images in false color, thereby distinguishing bound and unbound GFP (**Fig. 4c**). This example illustrates how fluorescence-modulating nanobodies can provide novel optical readouts.

Tracking of subcellular translocation processes

Finally, we tested whether the fluorescence-enhancement effect induced by binding of GFP to Enhancer localized in a defined subcellular compartment could be used to track subcellular translocation events in a high-throughput approach. As an example, we used the inducible translocation of the human estrogen receptor. Hormone binding leads to a conformational change in the receptor that results in its dissociation from chaperone proteins and ultimately in its binding as a homodimer to cognate sites in steroid-responsive genes[20]. This subcellular trafficking event can be induced by the synthetic steroid hormone tamoxifen and followed with a GFP-labeled receptor and high-resolution fluorescence microscopy[21]. Because this procedure is based on single-cell imaging, it is poorly suited for high-throughput analyses. We generated a mammalian (HeLa-Kyoto) cell line that stably expresses nuclear-localized Enhancer fused to mRFP (nls-Enhancer). As described previously, GFP nanobodies can specifically recognize and bind to their respective epitopes in various subcellular compartments in living cells[16]. The cellular expression of the nanobodies has no obvious cytotoxic effect, as no differences in cell-based proliferation analysis of the nls-Enhancer–encoding cell line compared to the parental cell line could be detected (data not shown). We used the newly constructed cell line to transiently coexpress the steroid-binding domain of the human estrogen receptor (ER$_{286–595}$) fused to wtGFP (GFP-ER$_{286–595}$). According to our assay, translocation of this construct from the cytoplasm to the nucleus should be detectable by an increase of the GFP fluorescence intensity upon binding of GFP to the nls-Enhancer in the nucleus (**Fig. 5a**). Using fluorescence microscopy, we confirmed that both GFP-ER$_{286–595}$ and nls-Enhancer are almost exclusively localized in their designated compartments. Upon addition of tamoxifen to the medium, GFP-ER$_{286–595}$ translocated from the cytoplasm into the nucleus (**Fig. 5b**). After entering the nucleus, GFP becomes accessible for binding to nls-Enhancer, which results in a three-fold increase in GFP fluorescence intensity (**Fig. 5c**). Notably, the translocation event can be followed in a statistically significant number of cells by scanning the fluorescence intensities of living cells in multiwell formats. The fluorescence enhancement is directly correlated with translocation efficiency. By quantifying the fluorescence intensities, we detected a clear dose dependence of translocation on addition of increasing amounts of tamoxifen (**Fig. 5d**). Because our assay is based on living cells, we were able to follow the dynamics of the translocation event over time. These data demonstrate that fluorescence-modulating nanobodies are potent tools for studying subcellular relocalization, a key process of signal transduction, in real time and in a quantitative manner.

DISCUSSION

Alternative protein conformations can be accurately analyzed with a variety of biophysical methods *in vitro* but are notoriously difficult to study *in vivo*. We therefore tested whether recombinantly expressed nanobodies could discriminate between alternative protein conformations in living cells. As a target we chose GFP, as it provides a direct optical readout. Out of several GFP-specific nanobodies, we identified two, Minimizer and Enhancer, that shift the absorption of GFP in opposite directions. The corresponding crystal structures clearly show the structural changes induced by these two nanobodies and explain the functional consequences on GFP fluorescence.

This ability to manipulate protein conformation in living cells enables a number of new applications. As a first example, we use the modulation of GFP fluorescence for new bioimaging applications. The expression and subcellular distribution of Minimizer and Enhancer can be detected by ratio imaging, allowing the distinction of bound and unbound GFP. This indirect optical readout can be used as reporter for gene expression, virus infection and translocation assays. The nanobody-mediated enhancement of GFP fluorescence should also improve the tracing of low- to high-abundance GFP fusion proteins in live cells as well as ultrahigh resolution microscopy. We have recently demonstrated that cellular structures can be imaged at subdiffraction resolution by three-dimensional structured illumination microscopy (3D-SIM)[22]. To obtain ultrahigh resolution, however,

© 2010 Nature America, Inc. All rights reserved.

this new microscopy technology requires hundreds of images and thus mostly relies on bright synthetic chromophores. The signals obtained with physiological levels of GFP-labeled proteins are barely sufficient, and in particular, *in vivo* application of structured illumination[23] would greatly benefit from any fluorescence enhancement by a coexpressed Enhancer nanobody.

Translocation events have a central role in signal transduction and are therefore a prime target for drug screenings. Presently, translocations are monitored either with reporter gene assays, which take at least a day, or by microscopy, which requires costly and technically demanding high-throughput image acquisition and analysis tools. Our nanobody-based assay can be performed with a simple plate reader and measures translocation as fluorescence enhancement after drug addition. We demonstrate the feasibility of this assay principle using the tamoxifen-induced nuclear translocation of the estrogen receptor. Aside from steroid hormone receptors, notch-type signaling in differentiation and cancer could also be directly monitored in cell-based drug screens.

This work outlines new experimental possibilities as it exemplifies applications of nanobodies ranging from affinity purification and crystallization of proteins to the manipulation of conformational states and protein function *in vitro* and in living cells. In particular, the detection and manipulation of alternative protein conformations in living cells enable novel types of studies in molecular and cellular biology. The functional relevance of alternative protein conformations is clearly illustrated by the prion protein. Here we outline and demonstrate an experimental strategy to address the role of alternative protein conformations in cellular systems. Our results show that nanobodies can be generated to recognize, induce and stabilize alternative protein conformations and thus enable studies of functional properties of specific protein conformations *in vitro* and *in vivo*.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates and structure factors for the GFP-Enhancer and GFP-Minimizer complex were deposited with accession codes 3K1K and 3G9A, respectively.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

A. Kirchhofer acknowledges support from the Deutsche Forschungsgemeinschaft graduate school 1202. U.R., J.H. and K.S. were supported by the GO-Bio program (Bundesministeriums für Bildung und Forschung) and C.F. by the International Doctorate Program 'NanoBioTechnology' of the Elite Network of Bavaria. The authors thank J. Gregor for excellent technical assistance, K. Lammens for help with crystallographic data collection, R. Lewis for fluorescence lifetimes determination and K. Zolghadr and N. Hiller for helpful comments and suggestions. This work was supported by the Center for Integrated Protein Science (CIPSM), the Center for Nanoscience (CeNS), the Nanosystems Initiative Munich (NIM), the BioImaging Network (BIN) and grants from the Deutsche Forschungsgemeinschaft (DFG) to M.C.C., H.L. and K.-P.H. (SFB 684). A. Kirchhofer and K.-P.H. thank K. Römer and the Dr. Klaus Römer Foundation for financial support.

AUTHOR CONTRIBUTIONS

U.R., M.C.C. and H.L. conceived and initiated the original project; A. Kirchhofer crystallized the GFP–nanobody complexes, determined the crystal structures, carried out life-time measurements and wrote the manuscript; S.C. and A. Karcher assisted in crystallization and structure determination; K.-P.H. helped with interpreting structural and functional data as well as determining crystal structures, designing research and helped with writing the manuscript; M.P. and S.M. provided the nanobodies; J.H. performed *in vitro* nanobody binding studies; K.S. carried out nanobody and GFP purification; J.H., C.S.C.-D. and M.C.C. carried out the ratio imaging experiments; C.F. performed the *in vivo* studies and data analysis; H.L. wrote the manuscript; M.C.C., S.M., H.L. and K.-P.H. revised the manuscript and oversaw research. U.R. carried out nanobody purification, *in vitro* analysis, translocation assays and wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/nsmb/.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

---

1. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).
2. Ormo, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
3. Yang, F., Moss, L.G. & Phillips, G.N. Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**, 1246–1251 (1996).
4. Brejc, K. *et al.* Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **94**, 2306–2311 (1997).
5. Heim, R. & Tsien, R.Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
6. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
7. Shimomura, O. Discovery of green fluorescent protein. *Methods Biochem. Anal.* **47**, 1–13 (2006).
8. Patterson, G.H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873–1877 (2002).
9. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19**, 137–141 (2001).
10. Akerboom, J. *et al.* Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J. Biol. Chem.* **284**, 6455–6464 (2009).
11. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
12. Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.* **414**, 521–526 (1997).
13. Muyldermans, S. Single domain camel antibodies: current status. *J. Biotechnol.* **74**, 277–302 (2001).
14. Jobling, S.A. *et al.* Immunomodulation of enzyme function in plants by single-domain antibody fragments. *Nat. Biotechnol.* **21**, 77–80 (2003).
15. Rothbauer, U. *et al.* A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics* **7**, 282–289 (2008).
16. Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3**, 887–889 (2006).
17. Chan, P.H. *et al.* Engineering a camelid antibody fragment that binds to the active site of human lysozyme and inhibits its conversion into amyloid fibrils. *Biochemistry* **47**, 11041–11054 (2008).
18. Desmyter, A. *et al.* Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. *J. Biol. Chem.* **277**, 23645–23650 (2002).
19. Kabat, E.A. & Wu, T.T. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J. Immunol.* **147**, 1709–1719 (1991).
20. Tsai, M.J. & O'Malley, B.W. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.* **63**, 451–486 (1994).
21. Htun, H., Holth, L.T., Walker, D., Davie, J.R. & Hager, G.L. Direct visualization of the human estrogen receptor alpha reveals a role for ligand in the nuclear distribution of the receptor. *Mol. Biol. Cell* **10**, 471–486 (1999).
22. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).
23. Kner, P., Chhun, B.B., Griffis, E.R., Winoto, L. & Gustafsson, M.G. Super-resolution video microscopy of live cells by structured illumination. *Nat. Methods* **6**, 339–342 (2009).

© 2010 Nature America, Inc. All rights reserved.

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Protein production and purification. We carried out expression of nanobodies and GFP as described previously[15,24]. We produced selenomethionine (SeMet)-containing GFP in *E. coli* B834 (Rosetta (DE3)) grown in minimal medium containing 50 mg l^{-1} L-selenomethionine. For complex purification of wtGFP-Enhancer and wtGFP-Minimizer, we purified nanobodies by prebinding them via their C-terminal His$_6$ tag to a HiTrap-column (GE Healthcare Life Sciences). Subsequently, we isolated wtGFP from crude cell extracts by binding to the precharged column. After elution by increasing concentration of imidazole, we separated complexes from unbound protein by gel filtration chromatography using a Superdex 200 column (GE Healthcare).

Fluorescence spectroscopy. We performed fluorescence assays either by scanning a 96-well microplate (Nunc) on a Typhoon Trio (GE Healthcare; ex$_{\text{GFP}}$, 488 nm, em, 520 ± 20 nm) or by using a monochromator-based microplate reader (Infinite M1000, Tecan; ex$_{\text{wtGFP}}$, 395 nm, ex$_{\text{eGFP}}$, 488 nm, em, 507 ± 10 nm). We recorded fluorescence excitation spectra with a FluoroMax-P fluorimeter (HORIBA Jobin Yvon). Typically, a 0.5-μM protein in PBS was measured in a 1-ml quartz cuvette (ex./em. bandpass, 5 nm). Samples were excited at 395 nm and 475 nm, and excitation spectra were recorded in the range from 480–600 nm. We recorded fluorescence absorption spectra on a UV/Vis Spectrophotometer (Beckman Coulter). Absorptions of 0.5-μM GFP alone or 0.5-M GFP–Enhancer and –Minimizer complexes were detected with continuous excitation steps (1 nm) from 250 nm–700 nm.

Crystallization and data collection. We crystallized purified GFP–nanobody complexes by hanging-drop vapor diffusion by mixing 1 μl of protein solution at 10 mg ml^{-1} concentration with 1 μl of the reservoir solution (SeMet-GFP–Enhancer: 60% (v/v) 2-methyl-2,4-pentanediol (MPD), 100 mM sodium acetate pH 4.6, 10 mM CaCl$_2$; GFP–Minimizer: 100 mM 2-(*N*-morpholino)ethanesulfonic acid, pH 6.5, 30% (v/v) PEG8000, 15% (v/v) glycerol). Initial crystallographic data obtained from native GFP–Enhancer (data not shown) suffered from crystal twinning, and coordinate refinement after molecular replacement did not result in acceptable *R* values. Crystals of SeMet-GFP–Enhancer, originally grown to obtain experimental phases by anomalous dispersion, grew in a different space group without twinning. After flash freezing the crystals, we recorded single-wavelength diffraction data at the K absorption edge of selenium ($\lambda = 0.9793$ Å) at the X06SA beamline (Swiss Light Source) to 2.15 Å. GFP–Minimizer crystals were flash frozen in liquid nitrogen. Data collection was performed at the beamline ID29 ($\lambda = 0.98137$ Å) at the European Synchrotron Radiation Facility to a resolution of 1.6 Å.

Structure determination. We processed diffraction data of both complexes with XDS[25]. In the case of the GFP–Enhancer structure, we located three selenium sites per complex using autoSHARP (Global Phasing). Single-wavelength anomalous dispersion phasing and solvent flipping yielded an interpretable experimental electron density map. We built models for GFP and Enhancer with COOT[26] and refined them with PHENIX[27], using overall anisotropic *B*-factors and bulk solvent corrections, individual *B*-factor refinement, simulated annealing, and crystallographic and positional refinement. We determined the GFP–Minimizer structure by molecular replacement with PHASER using an individual GFP and a nanobody polypeptide chain from the previously determined GFP–Enhancer structure as independent search models. We manually altered and refined the replacement model with COOT using similar procedures as described for the GFP–Enhancer structure. All structure figures were prepared using the program PyMol (http://www.pymol.org/). Data collection and model statistics are summarized in **Table 1**.

Spectral analysis in living cells. We detected the overall fluorescence intensities of GFP and expression levels of mRFP fusions in HEK293T cells by fluorescence spectroscopy on a microplate reader (ex$_{\text{GFP}}$, 490 ± 5 nm, em$_{\text{GFP}}$, 511 ± 5 nm, ex$_{\text{mRFP}}$, 586 ± 5 nm, em$_{\text{mRFP}}$, 608 ± 5 nm). For spectral analysis we scanned the cells in intervals of 5 nm for excitation from 300–500 nm and for emission from 500–580 nm. To quantify and normalize the data to the relative expression levels of GFP, we lysed remaining cells and performed immunoblot analysis using a GFP antibody (Roche). Preparation of cells and calculations of GFP fluorescence values were performed as described in **Supplementary Methods**.

Ratio imaging. We recorded cells expressing wtGFP and the Enhancer fusion GBP1–lamin B1 with excitation laser lines at 405 nm or 488 nm (em$_{\text{GFP}}$, 527 ± 27 nm) using a spinning disk microscope (UltraVIEW VoX, Perkin Elmer). For noise reduction, a Gaussian filter ($\sigma = 2$) was applied to both images. To visualize altered spectral properties of GFP induced by Enhancer, binding signal intensities recorded with excitation at 488 nm were divided by the corresponding 405-nm intensities for each pixel. This 488/405 ratio was displayed in a false color gradient from blue (unbound) to yellow (bound). Image analysis operations were executed in Priithon, a Python-based image analysis and algorithm development platform. Preparation of cells was performed as described in **Supplementary Methods**.

Translocation assay. For the nucleocytoplasmic translocation assay, we transfected HeLaK cells stably expressing the Enhancer fused to mRFP comprising an N-terminal nuclear localization signal (nls-Enhancer) with an expression plasmid for a tamoxifen-responsive estrogen receptor domain fused to wtGFP (GFP-ER$_{286\text{–}595}$). Eighteen hours after transfection, about 50% of the cells were expressing GFP-ER$_{286\text{–}595}$ preferentially in the cytoplasm, whereas the nls-Enhancer–mRFP fusion protein was slightly enriched in the nucleoli as determined by fluorescence microscopy. We incubated $\sim 1 \times 10^7$ cells with DMEM containing 0–25 μM tamoxifen for 30 min. We used an equal number of cells from the same transfection as an untreated control. We harvested the cells, washed them twice and resuspended them in 1 ml PBS. Roughly 3×10^6 cells were transferred to a 96-well plate (Greiner). We determined fluorescence intensities of GFP and mRFP by fluorescence spectroscopy (Tecan Infinite M1000 plate reader, ex$_{\text{GFP}}$, 490 ± 5 nm, em$_{\text{GFP}}$, 511 ± 5 nm, ex$_{\text{mRFP}}$, 586 ± 5 nm, em$_{\text{mRFP}}$, 608 ± 5 nm). We subtracted background fluorescence intensities from untransfected cells and normalized GFP fluorescence intensities against mRFP fluorescence intensity.

Materials. Purified Enhancer and Minimizer protein is commercially available from ChromoTek (Germany). Enhancer and Minimizer encoding vectors for intracellular studies can be requested from the Ludwig-Maximilians University, Munich.

24. Frauer, C. & Leonhardt, H. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res.* (2009).
25. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
27. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).

Supplemental Information

Modulation of protein properties in living cells with nanobodies

Axel Kirchhofer [1,2,3], Jonas Helma [2,4], Katrin Schmidthals [2,4], Carina Frauer [2,4], Sheng Cui [1,2,3], Annette Karcher [1,2,3], Mireille Pellis [5,6], Serge Muyldermans [5,6], Corella Casas Delucchi [7], M. Cristina Cardoso [7], Heinrich Leonhardt [2,4], Karl-Peter Hopfner [1,2,3] and Ulrich Rothbauer [2,4,8,9]

[1] Gene Center at the Department of Chemistry and Biochemistry, Ludwig Maximilians University Munich, Feodor-Lynen-Str. 25, D-81377 Munich, Germany

[2] Center for Integrated Protein Science

[3] Munich Center for Advanced Photonics

[4] Biocenter at the Department of Biology II, Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany

[5] Department of Molecular and Cellular Interactions, VIB, Brussels, Belgium

[6] Laboratory of Cellular and Molecular Immunology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[7] Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany

[8] Center for NanoSciences (CeNS)

[9] ChromoTek GmbH

Supplementary Methods

VHH libraries. Llama immunization, V_HH-library construction and selection of the GFP-binding proteins (GBP) were done as previously described[1].

Affinity determination. For affinity measurements, purified Enhancer or Minimizer were immobilized on an Carboxyl-Chip (Attana AB. Sweden) by amine-coupling according to the manufacturer’s instructions. Binding kinetics of the wtGFP to the immobilized nanobody were determined with an Attana 100C Quartz Crystal Microbalance (Attana AB, Sweden) by injection of 5 different concentrations of wtGFP (0.13 µg, 0.25 µg, 0.55 µg, 1.1 µg and 2.2 µg).

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1727

Expression plasmids. The expression plasmid encoding wtGFP was constructed by replacing the eGFP coding sequence in pEGFP-C1 (Clontech, CA, USA) with the wtGFP coding sequence by PCR amplification with the following primers (F: 5‘-CCC CGC TAG CGC TAC CGG TCG CCA CCA TGA GTA AAG GAG AAG AAC T-3‘; R: 5’-GGG GCT TAA GCT TCG AAC TCG AGC TCT A GA CTC AGG CCT AAA CAT ATC AAG TAG GTA C-3’).The expression plasmid encoding a translational fusion of Enhancer and mRFP was derived by PCR amplification of the Enhancer coding region with F: 5´-GGG GGC TCG AGC CGG CCA TGG CCG ATG TGC AG-3´ and R: 5´-GGG GGA ATT CCT TGA GGA GAC GGT GAC-3´. The XhoI/EcoRI digested PCR-fragment was ligated into a modified pEYFP-N1 vector (Clontech, CA, USA), where the YFP sequence had been replaced by the mRFP1 coding region. The Minimizer-mRFP as well as the control nanobody plasmid was constructed likewise. The control nanobody is directed against the capsid protein of the HI-Virus-1 and does not cross-react with GFP or other cellular targets (unpublished results). The expression plasmid encoding the nuclear localized Enhancer-mRFP (nls-Enhancer) was constructed by adding a N-terminal nuclear localization sequence (nls) to the Enhancer by PCR amplification (F: 5’-GGG GAG ATC TCC GGC CAT GGC TCC AAA GAA GAA GAG AAA GGT CCA GGT GCA GCT GGT GGA GTC T-3’) and recloning it into the modified pEYFP-N1 vector as described above. The expression plasmid encoding a translational fusion of wtGFP and $ER_{286\text{-}595}$ (GFP-$ER_{286\text{-}595}$) was derived by PCR amplification of the $ER_{286\text{-}595}$ coding region with primers (F: 5’-GGG GAG ATC TAT GAG AGC TGC CAA CCT TTG G-3’ and R: 5’- GGG GAA GCT TTC AGA CTG TGG CAG GGA AA CC-3'). The digested fragment was ligated into the modified pEGFP-C1 vector, where eGFP had been replaced by wtGFP, as described above. For Enhancer-Lamin B1, the Lamin B1 was amplified from GFP-LaminB1 (kindly provided by Jan Ellenberg) by PCR with primers (F: 5’-CCC CGA TAT CGG CGA CTG CGA CCC CC-3’ and R: 5’-GGG GGC GGC CGC CTA GTG ATG GTG ATG GTG GTG TTA CAT AAT TGC ACA GCT TC-3’). The PCR product was purified, digested, and ligated into pEYFP-N1 vector, containing Enhancer. All resulting constructs were sequenced and tested for expression in HEK 293T cells followed by Western blot analysis.

Sample preparation for spectral analysis in living cells. Human embryonic kidney HEK 293T cells were cultured in DMEM supplemented with 10% fetal calf serum and

50 µg/ml gentamycine (PAA, Germany). HEK 293T cells were transiently cotransfected with expression plasmids for wtGFP/eGFP and mRFP fusions of Enhancer, Minimizer or control nanobody using polyethylenimine as transfection reagent (Sigma, Germany). After 48 h about 90% of the cells were expressing GFP and mRFP as determined by fluorescence microscopy. Cells were harvested and transferred to a 96 well plate (Greiner, Germany).

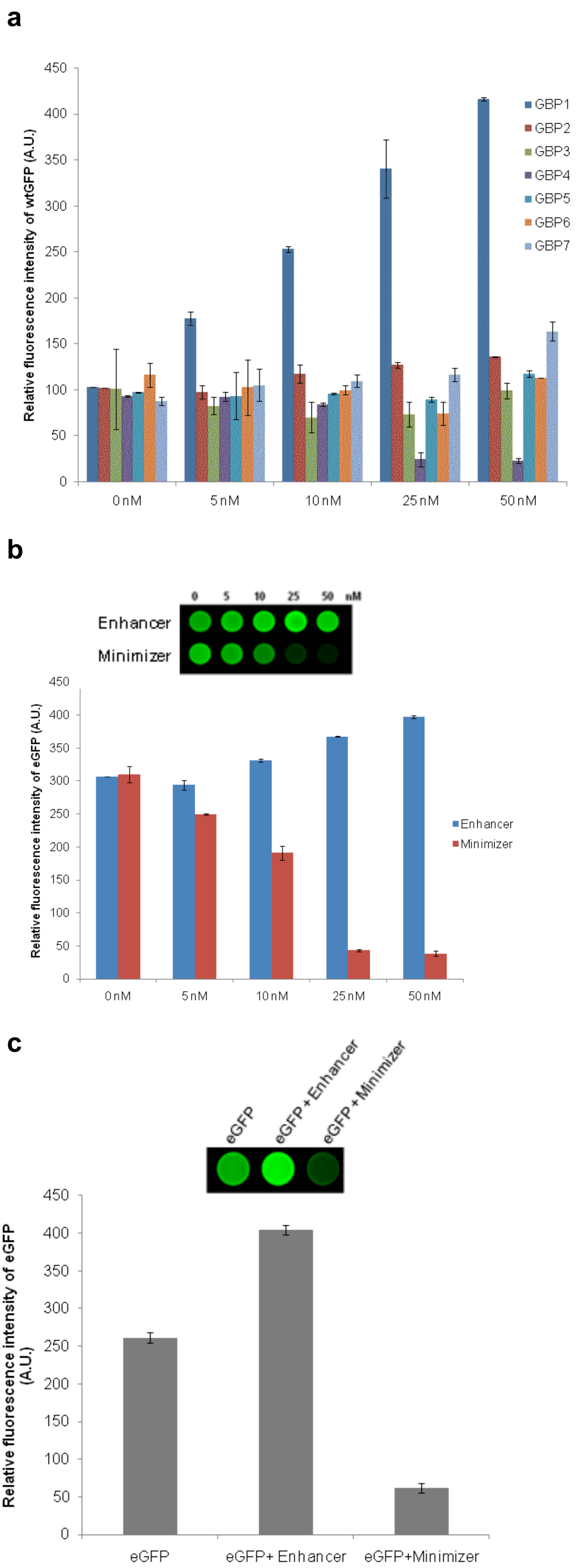
Sample preparation for high-throughput image acquisition and ratio imaging. Human HeLa-Kyoto cells were cultured in DMEM supplemented with 10% fetal calf serum and 50 µg/ml gentamycine (PAA, Germany). About 150 000 cells were seeded on gridded 18x18 mm coverslips in a 6-well-format. To allow for comparable overall GFP expression levels, equal amounts (2 µg) of expression plasmids for wtGFP/eGFP and mRFP fusions of Enhancer*,* Minimizer or control nanobody were transiently cotransfected, using polyethylenimine as transfection reagent (Sigma, Germany). After 24 h, cells were PFA-fixed and mounted in Vectashield anti-fading reagent (Vector Laboratories, USA) on object slides.

Translocation Assay Normalization. For normalization we lysed the pellet of ~3x10^6 cells in a SDS containing sample buffer by boiling it for 10 min at 95°C and subjected it to immunoblot analysis using an anti-GFP-antibody (Roche, Germany) and anti ß-actin antibody (Sigma).

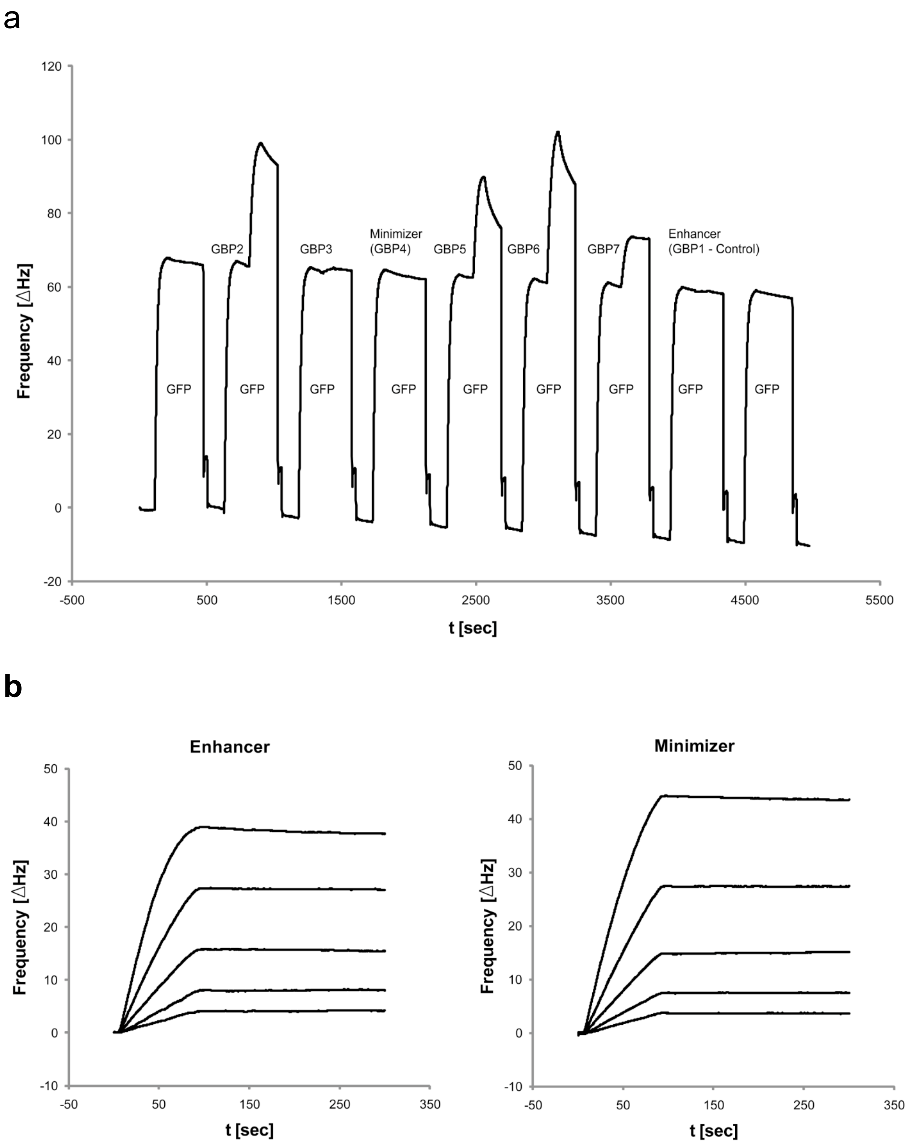
Structure visualization and analyzation. Calculation of buried surface areas were performed with the protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)[2]. Superpositions of structures and calculation of RMSD values were conducted using the CaspR RMSDcalc web-server[3]. Images of crystal structures were prepared with PyMol (http://www.pymol.org).

Time-resolved fluorescence-measurements. Fluorescence-lifetime measurements of GFP both in presence and in absence of nanobodies were conducted on a dedicated TCSPC-lifetime spectrometer (Fluorocube 01-NL, Horiba Jobin Yvon, Germany). Protein samples were prepared as described in the methods section of the manuscript. Concentrations were adjusted to ~1 µM by dilution in phosphate buffered saline (PBS, pH 7.4). Samples were excited at 460 nm using a pulsed LED (1 MHz repetition rate) as excitation source, while the fluorescence emission was detected at 505 nm with a spectral resolution of 12 nm. The time to amplitude conversion range was set to 58.4 ns divided in 2.048 channels leading to a time resolution of 28.5 ps per channel. The acquisition was stopped at 10.000 counts per peak. The lifetime decays were then deconvoluted using a pre-acquired instrument response function (1.2 ns full width path maximum) and fitted exponentially using the software DAS6 (Horiba Jobin Yvon).

Supplementary Figures and Tables



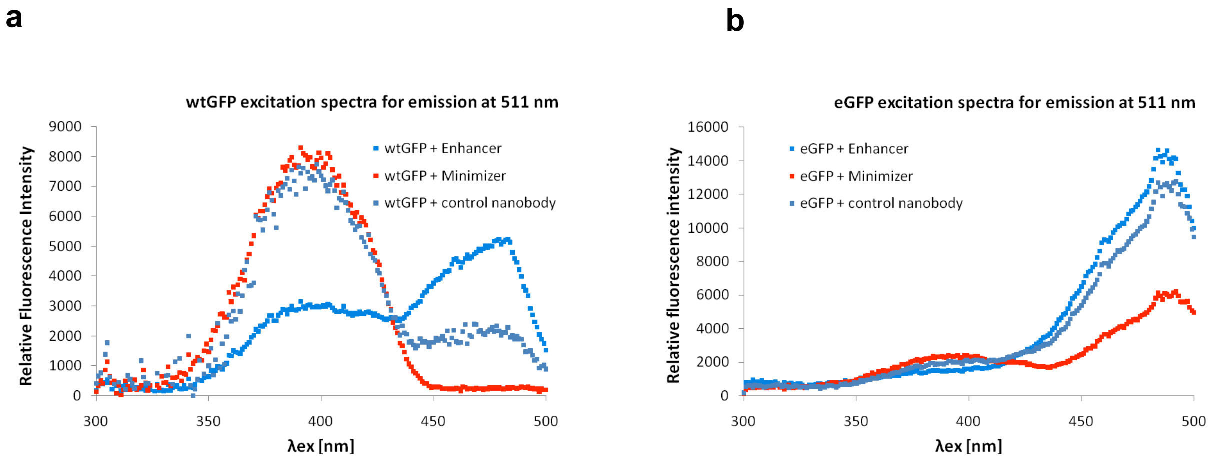
Supplementary Figure 1| Quantification of the fluorescence intensities of GFP after nanobody binding. (**a**) Quantification of the laser scanning experiment shown in **Fig. 1a**. Indicated values represent three independent experiments. (**b**) Titration of Enhancer and Minimizer from 0 - 50 nM on 50 nM purified eGFP. After complex formation the emission intensity of eGFP was quantified using a laser scanner (Typhoon 9410, GE Healthcare, excitation 488 nm).(**c**) Determination of the nanobody induced fluorescence modulation in soluble cell extracts of HEK 293T cells expressing eGFP. Emission intensity of eGFP was detected as described above. Means and s.d. (error bars) of 3 independent experiments are shown.



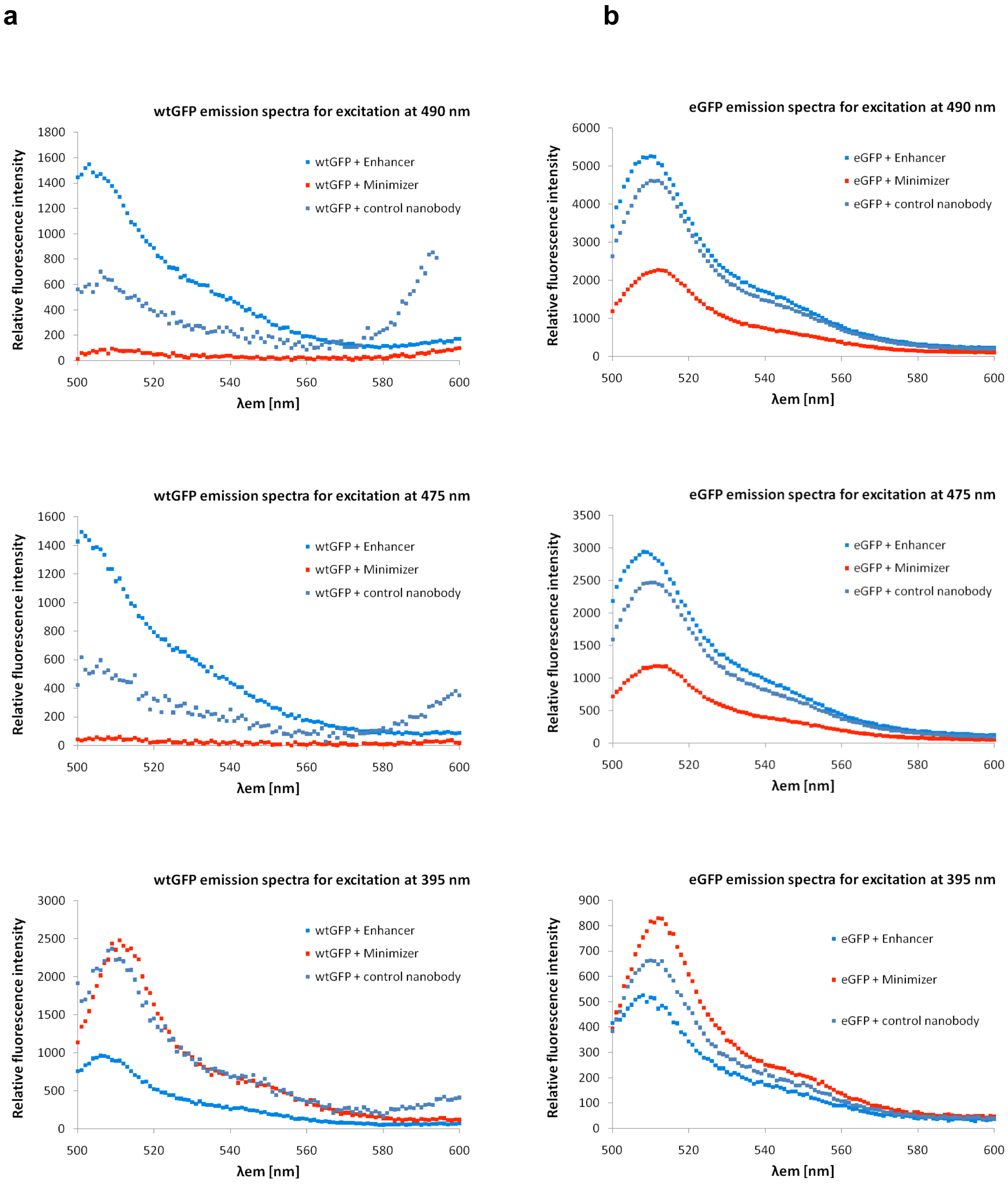
Supplementary Figure 2 | Nanobody binding assay

(a) Competitive Epitope Binding Assay. Enhancer (GBP1) was immobilized on a Carboxyl-Chip via amine-coupling. After GFP binding GBP1 – 7 were subsequently injected at constant concentrations (5 µg/ml) during GFP dissociation. Additional binding to the preformed Enhancer-GFP-Complex was measured using a Quartz Crystal Microbalance system. Simultaneous GFP binding was observed in combination with GBP2, GBP5, GBP6 or GBP7 while no binding was detectable in combination with GBP3 or Minimizer (GBP4) indicating overlapping epitopes.

(b) Affinity measurements for Enhancer and Minimizer. Each ligand was immobilized on a Carboxyl-Chip by amine-coupling. Binding kinetics were determined by injection of increasing concentrations of wtGFP (0.13 µg, 0.25 µg, 0.55 µg, 1.1 µg, 2.2 µg). For evaluation of binding kinetics see **Supplementary Table 1**.



Supplementary Figure 3 I Spectral analysis of the fluorescence modulation in living cells – excitation. Excitation spectra of wtGFP (**a**) and eGFP (**b**) in cells coexpressing Enhancer, Minimizer and a control nanobody. GFP emission at 511+/-5 nm was recorded for excitation between 300+/- 5 and 500+/- 5 nm.



Supplementary Figure 4 I Spectral analysis of the fluorescence modulation in living cells – emission. Emission spectra of wtGFP (a) and eGFP (b) in cells coexpressing Enhancer, Minimizer and a control nanobody. GFP emission between 500+/-5 and 700+/-5 nm was recorded upon excitation at 395+/- 5 nm, 475+/- 5 nm or 490+/- 5 nm.

Supplementary Table 1: Comparison of Enhancer and Minimizer binding affinities

	K_D [nM]	k_a [1/Ms]	k_d [1/s]
Enhancer	0.59 ±0.11	2.45×10^{5} (±0.2 %)	1.45×10^{-4} (±3.1 %)
Minimizer	0.45 ±0.04	1.33×10^{5} (±0.2 %)	6.01×10^{-5} (±5.6 %)

Dissociation constant (K_D), the on-rate (k_a) and the off-rate (K_d) of Enhancer and Minimizer were calculated according to a one-site binding model by ClampXP evaluation software (Attana AB, Sweden).

Supplementary Table 2: Direct contacts between GFP and Enhancer

Salt bridges and hydrogen bonds

GFP Residue	GFP Atom	Enhancer Residue	Enhancer Atom	Distance [Å]
Glu142	OE2	Ser33	OG	2.42
Glu142	OE1	Ser33	OG	2.96
Glu142	OE2	Arg35	NH2	2.54
Tyr145	O	Asn95	ND2	2.93
Asn 146	OD1	Asn95	ND2	3.07
Ser147	N	Glu101	OE2	2.79
Lys166	NZ	Glu44	OE1	2.61
Arg168	NH1	Glu101	OE2	3.19
Arg168	NH2	Glu101	OE1	2.82
Arg168	NH2	Tyr37	OH	3.25
Asp173	O	Ser58	OG	2.88
Gly174	O	Arg35	NH1	3.08
Ser175	O	Arg35	NH1	2.42

Hydrophobic interactions

GFP Residue.	Enhancer Residue
Ala206	Phe98
Leu221	Phe98
Phe223	Phe98

Supplementary Table 3: Salt bridges and hydrogen bonds between GFP and Minimizer

GFP Residue	GFP Atom	Minimizer Residue	Minimizer Atom	Distance [Å]
Asn 149	ND2	Val 100f	O	3.12
Tyr 151	OH	Val 100f	O	2.63
Tyr 151	OH	Asp100l	O	3.19
Asn 164	ND2	Tyr 100m	OH	3.09
Lys 166	N2	Asp 95	OD1	2.70
Arg 168	NH2	Leu 100k	O	2.97
Asp 180	OD1	Thr 98	N	2.98
Asp 180	OD2	Thr 98	OG1	2.66
Asn 198	ND2	Val 100c	O	3.05
Tyr 200	OH	Glu 44	OE1	2.64

Supplementary Table 4: GFP fluorescence-lifetime measurements

	wtGFP (ns)	eGFP (ns)
uncomplexed	3.064 ± 0.005	2.704 ± 0.005
+ Enhancer	2.915 ± 0.003	2.801 ± 0.004
+ Minimizer	2.903 ± 0.004	2.806 ± 0.005

Supplementary References

1. Rothbauer, U. et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nature methods* **3**, 887-889 (2006).

2. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774-797 (2007).

3. Claude, J.B., Suhre, K., Notredame, C., Claverie, J.M. & Abergel, C. CaspR: a web server for automated molecular replacement using homology modelling. *Nucleic acids research* **32**, 606-609 (2004).

2.3 DIFFERENT BINDING PROPERTIES AND FUNCTIONAL RELEVANCE OF CXXC ZINC FINGER DOMAINS OF ENZYMES INVOLVED IN CYTOSINE MODIFICATION

DIFFERENT BINDING PROPERTIES AND FUNCTIONAL RELEVANCE OF CXXC ZINC FINGER DOMAINS OF ENZYMES INVOLVED IN CYTOSINE MODIFICATION

Carina Frauer[#], Andrea Rottach[#], Daniela Meilinger, Sebastian Bultmann, Karin Fellinger, Stefan Hasenoeder, Johannes Söding, Fabio Spada* and Heinrich Leonhardt*

Ludwig Maximilians University Munich, Department of Biology, Center for Integrated Protein Science Munich (CIPS^M), 82152 Planegg-Martinsried, Germany.

*Corresponding authors:
HL: h.leonhardt@lmu.de; Fax +49 89 2180-74236; Tel. +49 89 2180-74232
FS: f.spada@lmu.de; Fax +49 89 2180-74236; Tel. +49 89 2180-74230
[#]These authors contributed equally to this work.

Running title: CXXC zinc finger domains of Dnmt1 and Tet1.
Keywords: DNA methyltransferase 1, CXXC zinc finger, DNA methylation, Tet1.

ABSTRACT

About a dozen mammalian proteins contain CXXC zinc finger domains. Although many of them are involved in chromatin and DNA modification, the contribution of the CXXC domain to their functions is poorly understood. In particular, there are conflicting reports on the role of the CXXC domain in the DNA methyltransferase Dnmt1 and no functional data are available for a similar domain in the methylcytosine hydroxylase Tet1. Using a homology modeling approach we have designed isolated CXXC domain and deletion constructs for mouse Dnmt1 and Tet1 to maximize the probability of native peptide folding. We show that the CXXC domain of Tet1 has no DNA binding activity, while that of Dnmt1 selectively binds DNA substrates containing unmethylated CpG sites. However, both *in vitro* and *in vivo* approaches show that the CXXC domain of Dnmt1 is dispensable for DNA binding specificity, binding kinetics, allosteric activation and methyltransferase activity. Thus, we suggest a subtle, possibly developmental stage- or tissue-specific regulatory function for the CXXC domain of Dnmt1.

INTRODUCTION

In mammals DNA methylation is restricted to cytosine residues and mainly involves CpG dinucleotides. CpG methylation is widespread across mammalian genomes, including gene bodies regardless of their transcriptional activity (1-2). However, highly CpG-rich regions (CpG islands) are refractory to methylation and mostly coincide with promoters of constitutively active genes. The methylation state of other regulatory sequences with moderate to low CpG density, including promoters and enhancers, shows developmental and/or tissue-specific variations and often correlates with a transcriptionally silent state (1,3-6). Furthermore, dense methylation of repetitive sequences is thought to maintain these elements in a silent state and thus contribute to genome stability (7-9). In mammals, cytosine methylation is catalyzed by a family of DNA methyltransferases (Dnmts) (10). Dnmt3a and Dnmt3b establish methylation patterns during embryonic development of somatic as well as germ cell lineages and, consistently, show developmental stage and tissue specific expression patterns. In contrast, Dnmt1 is ubiquitous and generally the most abundant DNA methyltransferase in mammalian tissues, where it restores symmetrical methylation at hemimethylated CpG sites generated by semi-conservative DNA replication. Thus, Dnmt1 maintains methylation patterns with high fidelity and is essential for embryonic development and genome integrity (7,11-13).

Dnmt1 is a large enzyme with a complex domain structure that likely evolved by fusion of at least three genes (14). It comprises a regulatory N-terminal region and a C-terminal catalytic domain connected by a linker made of seven glycine-lysine repeats (Figure 1A) (15). The N-terminal part contains a PCNA binding domain (PBD), a heterochromatin targeting sequence (TS), a CXXC-type zinc finger domain and two Bromo-Adjacent Homology domains (BAH1 and BAH2). The C-terminal domains of mammalian Dnmts contain all ten motifs identified as essential for catalysis in bacterial DNA (cytosine-5) methyltransferases (10) and thus, prokaryotic and mammalian cytosine methyltransferases are thought to adopt a very similar catalytic mechanism. However, the C-terminal domain of Dnmt1 is the only DNA methyltransferase domain in Dnmts that is not catalytically active when expressed separately and interaction with the N-terminal part is required for allosteric activation of the enzyme (16). Remarkably, the first 580 amino acids (aa) of human DNMT1 are dispensable for both enzymatic activity and substrate recognition, whereas deletion of the first 672 aa results in an inactive enzyme (17). Interestingly, this truncation eliminates part of the CXXC domain, suggesting an involvement of this domain in allosteric activation. However, addition of the isolated CXXC domain to the catalytic domain *in trans* was not sufficient for catalytic activation (18).

CXXC-type zinc finger domains are found in several other proteins with functions related to DNA or chromatin modification, including the histone H3 lysine 4 methyltransferases mixed-lineage leukemia

(MLL) proteins 1 and 4, the CpG-binding protein (CGBP, also known as CFP1 or CXXC1), the methyl-CpG binding domain protein 1 (MBD1), the H3 lysine 36 demethylases KDM2A and B (also known as JHD1A/FBXL11 and JHD1B/FBXL10) and the MLL1 fusion partner TET1 (19-26). The CXXC domains of some of these proteins were shown to mediate specific binding to double stranded DNA templates containing unmethylated CpG sites (19-20,27). A region of Dnmt1 which mainly includes the CXXC domain (aa 628-753) was also shown to bind Zn ions and DNA (18,28-29). However, available data on the selectivity of this DNA binding activity are conflicting. Whereas a fragment including aa 613-748 of mouse Dnmt1 was shown to bind DNA with a slight preference for hemimethylated CpG sites (18), aa 645-737 of human DNMT1, were shown to selectively bind unmethylated DNA (29). As these studies used different constructs and species, questions remain as to the DNA binding selectivity of the Dnmt1 CXXC domain with regard to the CpG methylation state and whether the CXXC domain is crucial for allosteric activation, substrate discrimination or both.

Notably, not all CXXC domains show this specificity, as exemplified by the fact that only one of the three CXXC domains in MBD1 binds DNA (27). Interestingly, TET1 was recently shown to be a 2 oxoglutarate- and Fe(II)-dependent oxigenase responsible for converting genomic 5-methylcytosine to 5-hydroxymethylcytosine (30). However, it is not known whether the TET1 CXXC domain is involved in recognition of methylated DNA substrates.

Here we report a systematic functional study and characterization of the CXXC domains in mouse Dnmt1 and Tet1 proteins, respectively. Isolated Dnmt1 CXXC domain and deletion constructs were generated based on structural homology models with the aim of preserving native peptide folding. We show that the CXXC domains of Dnmt1 and Tet1 differ drastically in DNA binding properties and that the CXXC domain is dispensable for DNA binding and catalytic activity of Dnmt1 *in vitro* and *in vivo*.

MATERIAL AND METHODS

Bioinformatic methods

Alignments were performed using the ClustalW2 software (31). The CXXC domain homology tree (Figure 1C) was generated from the alignment in Figure 1B with Jalview 2.4 by unweighted pair group method with arithmetic mean (UPGMA). The neighbor-joining method gave the same result. Average distances between the sequences were calculated using the BLOSSUM62 matrix. To build homology models for the CXXC domains of Dnmt1 (aa 645-696) and Tet1 (aa 561-614), we submitted the respective sequences to the HHpred server (32). The best template was the CXXC domain of MLL1 (PDB-ID: 2J2S). The 49 Zn finger residues of Dnmt1 can be aligned to this domain with 45% sequence identity and only a single amino acid gap after residue 661 (Figure 1B). 3D models were calculated with the homology modeling software MODELLER (33) (version 9.5) using this alignment. Distance restraints were given to MODELLER to enforce a distance of 2.3 ± 0.1 Å between the eight sulphurs in the zinc-coordinating cysteines and the zinc ions. TM-align (34) was used to superpose the model structure with the template domain. Images were generated using the PyMol software (35). The quality of the models and the underlying alignments were checked with DOPE (36) and Verify3D (37) and results for both models were found to be comparable to the template structure (2J2S).

Expression constructs

Fusion constructs were generated using enhanced green fluorescent protein, monomeric red fluorescent protein or monomeric cherry and are here referred to as GFP, RFP and Cherry fusions, respectively. Mammalian expression constructs for GFP, mouse GFP-Dnmt1, GFP-NTR and human RFP-PCNA were described previously (38-41). The deletion construct GFP-Dnmt1$^{\Delta CXXC}$ was obtained by replacing the sequence coding for aa 655-696 by codons for three alanine residues in the GFP-Dnmt1 construct as described (42). To generate GFP-CXXC^{Dnmt1} and GFP-CXXC^{Tet1} sequences coding for the respective CXXC domains (aa 643-700 for Dnmt1 and 561-614 for Tet1) were amplified by PCR using the GFP-Dnmt1 and cDNA from E14 embryonic stem cells (ESCs), respectively. PCR fragments were then cloned into the AsiSI-NotI site of the same vector backbone as used for GFP-Dnmt1. GFP-NTR$^{\Delta CXXC}$ was obtained by replacing the BglII-XhoI fragment of GFP-NTR with the same fragment of GFP-Dnmt1$^{\Delta CXXC}$. Ch-CTD-His was generated by replacing the GFP coding sequence in a GFP-CTD coding vector (41) with the Cherry sequence. All constructs were confirmed by sequencing.

Cell culture, transfection and cell sorting

HEK293T cells and mouse C2C12 myoblasts were cultured in DMEM supplemented with 50 µg/ml gentamicin and 10% and 20% fetal calf serum, respectively. For overexpression of fusion proteins HEK293T cells were transfected with polyethylenimine (Sigma). For live cell imaging, C2C12 cells were grown to 40% confluence on Lab-Tek chambers (Nunc) or µ-slides (Ibidi) and transfected with TransFectin transfection reagent (BioRad) according to the manufacturer's instructions. Mouse embryonic stem cells (ESCs) were cultured as described (43) and transfected with FuGENE HD (Roche) according to the manufacturer's instructions. ESCs were sorted with a FACS Aria II instrument (Becton Dikinson). The *dnmt1*$^{-/-}$ J1 ESCs used in this study are homozygous for the c allele (11).

In vitro DNA binding and trapping assays

In vitro DNA binding and trapping assays were performed as described previously (44-45) with the following modifications. DNA substrates labeled with four different ATTO fluorophores (Supplementary tables 1 and 2) were used at a final concentration of 125 nM each in the pull-down assay with immobilized GFP fusions. After removal of unbound substrate, the amounts of protein and DNA were determined by fluorescence intensity measurements with a Tecan Infinite M1000 plate reader using calibration curves from purified GFP or DNA coupled ATTO fluorophores, respectively. The following excitation/emission ± detection bandwidth settings were used: 490/511 ± 10 nm for GFP, 550/580 ± 15 nm for ATTO550, 600/630 ± 15 nm for ATTO590, 650/670 ± 10 nm for ATTO647N and 700/720 ± 10 nm for ATTO700. Cross detection of GFP and different ATTO dyes was negligible with these settings. Binding and trapping ratios were calculated dividing the concentration of bound DNA substrate by the concentration of GFP fusion on the beads.

Co-immunoprecipitation

Co-immunoprecipitation was performed as described previously (41,46). Shortly, HEK293T cells were transiently co-transfected with expression plasmids for GFP fusions and the Ch-CTD-His construct, harvested and lysed. GFP fusions were pulled down using the GFP-Trap (47) (Chromotek) and subjected to western blotting using anti-GFP (Roche) and anti-His (Invitrogen) mouse monoclonal antibodies.

Live cell microscopy and Fluorescence Recovery after Photobleaching (FRAP) analysis

Live cell imaging and FRAP experiments were performed as described previously (43). For each construct 6-15 nuclei were averaged and the mean values as well as the standard errors were calculated. For presentation, we used linear contrast enhancement on entire images.

Live cell trapping assay

The DNA methyltransferase trapping assay was described previously (40). Briefly, transfected cells were incubated with 30 µM 5-aza-dC (Sigma) for the indicated time periods before photobleaching experiments. FRAP analysis was performed with a confocal laser scanning microscope (TCS SP5, Leica) equipped with a 63x/1.4 NA Plan-Apochromat oil immersion objective. Microscope settings were as described except that a smaller region of interest (3 µm x 3 µm) was selected for photobleaching. Mean fluorescence intensities of the bleached region were corrected for background and for total loss of nuclear fluorescence over the time course, and normalized by the mean of the last 10 prebleach values.

DNA Methylation Analysis

Genomic DNA was isolated with the QIAmp DNA Mini Kit (Qiagen) and 1.5 µg were bisulfite converted using the EZ DNA Methylation-Gold Kit (Zymo research) according to the manufacturer's instructions. Primer sets and PCR conditions for IAP-LTR´s, *skeletal α-actin* and *H19* promoters were as described (43). Primer sequences for major satellites were AAAATGAGAAACATCCACTTG (forward primer) and CCATGATTTTCAGTTTTCTT (reverse primer). For amplification we used Qiagen Hot Start Polymerase in 1x Qiagen Hot Start Polymerase buffer supplemented with 0.2 mM dNTPs, 0.2 µM forward primer, 0.2 µM reverse primer, 1.3 mM betaine (Sigma) and 60 mM tetramethylammonium-chloride (TMAC, Sigma). Promoter regions and IAP-LTR´s were amplified in two subsequent rounds of amplification (nested PCR) and major satellites were amplified in a single amplification. Pyrosequencing reactions were carried out by Varionostic GmbH (Ulm, Germany).

RESULTS

Sequence homology and structural modeling identify distinct CXXC domain subtypes

Dnmt1 contains a zinc finger domain of the CXXC type (CXXC, Figure 1A), which is present in numerous mammalian proteins including MLL (Figure 1B) and highly conserved among Dnmt1 sequences from various animal species (Supplementary Figure 1). The primary structure of CXXC domains spans two clusters of 6 and 2 cysteine residues separated by a stretch of variable sequence and length. Sequence alignment and homology tree construction identified three distinct groups of CXXC domains (Figure 1B and C). The sequence between the two cysteine clusters in the CXXC domains of Dnmt1, CGBP/CFP1, Fbxl19, Mll and Kdm2 proteins and CXXC domain 3 of Mbd1 is highly conserved and contains a KFGG motif. The two other homology groups, the CXXC domains in Tet1, CXXC4/Idax and CXXC5 on the one side and the CXXC domains 1 and 2 of Mbd1 on the other side, lack the KFGG motif and diverge from the first group and from each other in the sequence between the cysteine clusters. We generated structural homology models for the CXXC domains of mouse Dnmt1 and Tet1 using the NMR structure of the MLL1 CXXC domain as a template (Figure 1D and E) (48). The CXXC domains of these proteins adopt an extended crescent-like structure that incorporates two Zn ions each coordinated by four cysteine residues. The peptide of the MLL1 CXXC domain predicted to insert into the major groove (cyan in Fig. 1E) is located on one face of the structure and is contiguous to the KFGG motif (48). The predicted structure of the Tet1 CXXC domain lacks the short 3_{10} helix (η1 in Figure 1E) formed by residues PKF and partially overlapping the KFGG motif, but is similar to the MLL1 CXXC domain in the region of the DNA-contacting peptide. However, the two predicted beta strands in Tet1 each carry three positive charges, whereas in CXXC domains of MLL1 and DNMT1 carry only one or no charge in their C-terminal strand. Depending on the orientation of the positively charged side chains, it cannot be excluded that the charge density prevents strand pairing in the Tet1 CXXC domain.

The Dnmt1 CXXC domain binds unmethylated DNA

To investigate the binding properties of the Dnmt1 CXXC domain, we generated a GFP fusion construct including aa 652-699 (GFP-CXXC^{Dnmt1}). According to our homology model the ends of this fragment form an antiparallel β-sheet that structurally delimits the domain as in MLL1. First we compared the localization and mobility of GFP-CXXC^{Dnmt1} and GFP in mouse C2C12 myoblasts. While GFP was diffusely distributed in both nucleus and cytoplasm, GFP-CXXC^{Dnmt1} was exclusively nuclear with a punctuated pattern throughout the nucleoplasm and was enriched in nucleoli (Figure 2A). Enrichment in the nucleus and nucleoli is frequently observed with constructs containing stretches with high density of basic residues. After photobleaching half of the nuclear volume we observed a slower fluorescence recovery rate for GFP-CXXC^{Dnmt1} than for GFP (Figure 2B). To rule out a contribution of nucleolar interactions to the slower kinetics of GFP-CXXC$^{\mathrm{Dnmt1}}$, we separately bleached nucleoplasmic and nucleolar regions and found that GFP-CXXC$^{\mathrm{Dnmt1}}$ has even faster kinetics within the nucleolus (Supplementary Figure 2). These results are consistent with a binding activity of GFP-CXXC$^{\mathrm{Dnmt1}}$ in the nucleus and very transient, unspecific binding in the nucleolus. To investigate whether the CXXC domain of Dnmt1 binds DNA and its possible selectivity with respect to CpG methylation we used a recently developed fluorescent DNA binding assay (44-45). GFP-CXXC$^{\mathrm{Dnmt1}}$ was transiently expressed in HEK293T cells, immunopurified with the GFP-trap and incubated with fluorescent DNA substrates containing either no CpG site or one central un-, hemi- or fully methylated CpG site in direct competition. As shown in Figure 2C, GFP-CXXC$^{\mathrm{Dnmt1}}$ displayed an approximately 10-fold preference for the substrate containing one unmethylated CpG site. This result is consistent with the reported binding preference of other CXXC domains belonging to the same homology group as the Dnmt1 CXXC (19-20,27).

The CXXC domain of Tet1 shows no specific DNA binding activity *in vitro*

It was recently shown that Tet1 oxidates genomic 5-methylcytosine (mC) to 5-hydroxymethylcytosine. Although Tet1 must be able to discriminate mC containing substrates, the domain responsible for this specificity is not known. Our model for the structure of the Tet1 CXXC domain diverged from the structure of the MLL1 CXXC domain in correspondence of the KFGG motif but not of the DNA-contacting peptide, suggesting that the Tet1 CXXC domain may still bind DNA. To test this we generated a GFP tagged Tet1 CXXC construct (GFP-CXXC$^{\mathrm{Tet1}}$) following the same criteria as for GFP-CXXC$^{\mathrm{Dnmt1}}$ and subjected it to the same competitive fluorescent DNA binding assay. Binding of GFP-CXXC$^{\mathrm{Tet1}}$ to any of the four substrates was indistinguishable from the basal levels of the GFP control (Figure 2C). We conclude that the isolated CXXC domain of Tet1 has no specific DNA binding activity. Like the CXXC domain of Tet1, the CXXC-1 and 2 domains of Mbd1 lack the KFGG motif and do not bind DNA, while mutation of this motif prevents DNA binding by the MLL1 CXXC domain (27,49). Thus, our result with the Tet1 CXXC domain is consistent with a requirement for the KFGG motif to bind DNA.

Deletion of the CXXC domain does not affect the activity of Dnmt1 *in vitro*

To explore the role of the CXXC domain in Dnmt1 function we generated a GFP-Dnmt1 fusion construct where the CXXC domain, as defined by our homology model, was deleted (GFP-Dnmt1$^{\Delta \mathrm{CXXC}}$; Figure 3A). We reasoned that precise deletion of the entire structure delimited by the antiparallel β-sheet (the latter included; Figure 1D) would have the highest chances to preserve native folding of the rest of the protein. We also introduced the same deletion within the context of a GFP fusion with only the N-terminal region of Dnmt1 (GFP-NTR$^{\Delta \mathrm{CXXC}}$; Figure 3A). We then compared DNA binding properties, catalytic activity and interaction between N-terminal region and C-terminal catalytic domain of ΔCXXC and corresponding wild type constructs.

A competitive DNA binding assay with the same set of substrates as used for the experiments with GFP-CXXC$^{\mathrm{Dnmt1}}$ reported above (Figure 2C) showed that both GFP-Dnmt1 and GFP-Dnmt1$^{\Delta\mathrm{CXXC}}$ bind DNA independently of the presence and methylation state of the CpG site (Figure 3B). As the isolated CXXC domain preferentially bound the substrate containing an unmethylated CpG site, the result with GFP-Dnmt1 and GFP-Dnmt1$^{\Delta\mathrm{CXXC}}$ indicates that the CXXC domain contributes negligibly to the DNA binding specificity of the full-length enzyme.

Several groups reported that interaction between N-terminal region and C-terminal catalytic domain of Dnmt1 leads to allosteric activation of Dnmt1 (14,16-18,50). To test whether the CXXC domain is involved in this intramolecular interaction, we co-expressed either GFP-NTR or GFP-NTR$^{\Delta\mathrm{CXXC}}$ with a Cherry- and His-tagged C-terminal domain construct (Ch-CTD-His) in HEK293T cells and performed co-immunoprecipitation experiments. Ch-CTD-His co-precipitated both GFP-NTR and GFP-NTR$^{\Delta\mathrm{CXXC}}$, indicating that the CXXC domain is dispensable for the interaction between N-terminal region and C-terminal domain of Dnmt1 (Figure 3C).

To investigate whether the CXXC domain is needed for enzymatic activity or substrate recognition, we separately tested covalent complex formation and methyl group transfer, two successive steps of the enzymatic reaction, using GFP-Dnmt1 and GFP-Dnmt1$^{\Delta\mathrm{CXXC}}$. We first employed an assay to monitor covalent complex formation that exploits the formation of an irreversible bond between the enzyme and the mechanism-based inhibitor 5-aza-2-deoxycytosine (5-aza-dC), as opposed to the reversible complex formed with naturally occurring 2-deoxycytosine (dC) (44). Thus, 5-aza-dC containing substrates irreversibly trap cytosine methyltransferases. GFP-Dnmt1 fusions were separately incubated with fluorescent DNA substrates containing dC (binding) or 5-aza-dC (trapping) at a single CpG site. DNA-protein complexes were then isolated by GFP pulldown and molar DNA/protein ratios were calculated from fluorescence measurements (Figure 3D). Irreversible covalent complex formation was then estimated by comparison of the trapping and binding activities. GFP-Dnmt1 and GFP-Dnmt1$^{\Delta\mathrm{CXXC}}$ showed comparable covalent complex formation rates (relative trapping ratios), which were about 15- and 12-fold higher for hemi- than un-methylated substrates assayed in direct competition, respectively (Figure 3E). This result indicates that the preference of Dnmt1 for hemimethylated substrates is determined at the covalent complex formation step rather than upon DNA binding and that the CXXC domain does not play a major role in determining either the efficiency or the methylation state-specificity of covalent complex formation. We also tested whether deletion of the CXXC domain affects the ability of Dnmt1 to transfer [^{3}H]-methyl groups from the donor S-adenosylmethionine (SAM) to a poly(dI·dC)-poly(dI·dC) substrate, a standard DNA methyltransferase activity assay. This showed that GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ are equally active methyltransferases (Supplementary Figure 3).

In conclusion, we showed that, *in vitro*, the deletion of the CXXC domain does not affect the interaction between N-terminal region and C-terminal domain, DNA binding, preference for hemimethylated substrates upon covalent complex formation or methyltransferase activity of Dnmt1. These data strongly argue against an involvement of the CXXC domain in allosteric activation of Dnmt1.

Deletion of the CXXC domain does not affect Dnmt1 activity *in vivo*

We then undertook a functional characterization of our CXXC domain deletion construct *in vivo*. First, we compared localization and binding kinetics of GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ in mouse C2C12 myoblasts co-transfected with RFP-PCNA, which served as S-phase marker (38). GFP-Dnmt1$^{\Delta CXXC}$ showed the same cell-cycle dependent nuclear localization pattern as previously shown for GFP-Dnmt1 and endogenous Dnmt1 (Figure 4A)(39). Interaction with PCNA via the PBD directs Dnmt1 to replication foci throughout S-phase. In late S-phase and G2, however, Dnmt1 is enriched at chromocenters, clusters of pericentric heterochromatin (PH) that are observed as discrete domains densely stained by DNA dyes in mouse interphase cells. Association of Dnmt1 with PH at these stages is mediated by the TS. Thus, the CXXC domain clearly does not contribute to localization of Dnmt1 at this level of resolution.

We also measured the binding kinetics of GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ in living C2C12 myoblasts by fluorescence recovery after photobleaching (FRAP). These experiments revealed that the binding kinetics of Dnmt1 is not affected by deletion of the CXXC domain in early-mid as well as late S-phase (Figure 4B).

To test the ability of covalent complex formation *in vivo*, we used a previously established trapping assay (40). Mouse C2C12 myoblasts were cotransfected with either GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ and RFP-PCNA and treated with 5-aza-dC. Consequently, Dnmt1 constructs were irreversibly trapped at the site of action and irreversible covalent complex formation was measured by FRAP analysis (Figure 4C). GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ showed highly similar trapping kinetics, the trapped enzyme fraction reaching almost 100 % after 20 and 40 minutes in early-mid and late S-phase, respectively. This result clearly shows that the CXXC domain is dispensable for covalent complex formation also *in vivo*.

To further investigate the role of the CXXC domain *in vivo* we compared the ability of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ to restore DNA methylation patterns in mouse *dnmt1*$^{-/-}$ embryonic stem cells (ESCs). Cells transiently expressing either GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ were FACS sorted 48 h after transfection. Isolated genomic DNA was then bisulfite treated and fragments corresponding to major satellite repeats, intracisternal type A particle (IAP) interspersed repeats, *skeletal α-actin* and *H19a* promoters were amplified and subjected to pyrosequencing. As shown previously, under these condition GFP-Dnmt1 partially restored methylation of major satellite and IAP repeats and the *skeletal α-actin* promoter, but not of the imprinted *H19a* promoter, which requires passage through the germ line (43,51). Methylation patterns of all these sequences in cells expressing GFP-Dnmt1$^{\Delta CXXC}$ were very similar to those in GFP-Dnmt1 expressing cells, suggesting that the CXXC domain is not required for maintenance of DNA methylation patterns by Dnmt1.

Thus, the CXXC domain does not play a major role in subcellular localization and *in vivo* binding kinetics of Dnmt1 and, consistent with the *in vitro* data reported above, is dispensable for the catalytic activity of Dnmt1 *in vivo*.

DISCUSSION

We generated homology models based on the reported structure of the MLL1 CXXC domain to design isolated CXXC domain constructs for Dnmt1 and Tet1 and a CXXC domain deletion construct for Dnmt1 that may preserve native peptide folding. According to these models CXXC domains are delimited by an antiparallel β-sheet, a discrete structural element. Our data show that the CXXC domain of mouse Dnmt1 preferentially binds DNA substrates containing unmethylated CpG sites as previously shown for CXXC domains of other mammalian proteins. We note that sequences C-terminal to the corresponding peptide in CGBP were reported to be required for DNA binding *in vitro* (20) and that only a significantly larger peptide spanning the CXXC-3 domain of Mbd1a was tested for DNA binding. However, sequences C-terminal to CXXC domains are poorly conserved (Figure 1B) and not required for DNA binding by the Dnmt1 CXXC domain. Nevertheless, all the CXXC domains reported to selectively bind unmethylated CpG sites cluster in a distinct homology group and contain the KFGG motif that was shown to be crucial for DNA binding by the CXXC domain of MLL1 (49). We identify two other CXXC domain homology groups that lack the KFGG motif. Consistent with a role of this motif in DNA binding, members of these groups such as CXXC-1/2 of Mbd1 (27) and the CXXC domain of Tet1 (this study) show no DNA binding activity. While no specific function is known for Mbd1 CXXC-1/2, the CXXC domain of Tet1 is closely related to those in CXXC4/Idax and CXXC5 that were shown to mediate protein-protein interactions (52-54). This suggests that the CXXC domain of Tet1, rather than discriminating methylated DNA substrates, is a protein interaction domain.

Although we observed a clear DNA binding activity by the isolated CXXC domain of Dnmt1, we found that, within the context of the full length enzyme, this domain is dispensable for DNA binding, preference for hemimethylated substrates at the covalent complex formation step, methyltransferase activity and allosteric activation as well as for the ability to restore methylation of representative sequences in *dnmt1* null cells. Consistent with our data, a recent report showed a preference of the CXXC domain of human DNMT1 for substrates containing unmethylated CpG sites. However, the same report showed that deletion of the CXXC domain from the human enzyme results in a significant decrease in methyltransferase activity on hemimethylated substrates and 25% lower methylation at rDNA repeats upon overexpression in HEK293 cells, suggesting a dominant negative effect of the deletion construct (29). These discrepancies may be due to deletion of a slightly shorter fragment (aa 648-690) that may not preserve native folding, the analysis of non-physiological expression levels in HEK293 cells or species-specific differences. In this regard we would like to stress that genetic complementation of *dnmt1* null cells constitutes a more physiologically relevant test for the function of protein domains *in vivo*.

Notably, binding of unmethylated CpG sites by KFGG motif-containing CXXC domains does not exclude a role in protein-protein interaction as the CXXC domain of MLL1 was reported to interact with both DNA and Polycomb Repressive Complex 1 components HPC2/CBX4 and BMI-1 (19,55). Therefore, it is possible that the CXXC domain of Dnmt1 has regulatory functions in specific cell types or developmental stages that may involve DNA binding and/or interaction with other proteins. Testing this possibility will require the generation of dedicated animal models.

FUNDING

This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG), the Nanosystem Initiative Munich (NIM) and the Center for NanoScience (CeNS) to H.L. CF gratefully acknowledges financial support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) and the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). Open Access publication charges for this article were covered by the Deutsche Forschungsgemeinschaft (DFG).

ACKNOWLEDGEMENTS

The authors thank Weihua Qin for help with pyrosequencing analysis and Sabine Brunner and Lucia Puchbauer for technical assistance in generation of the homology models. We also thank Taiping Chen and En Li (Novartis Institutes for Biomedical Research, Boston, MA) for providing *dnmt1* null ESCs.

REFERENCES

1. Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q. and Church, G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*, **27**, 361-368.
2. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, **9**, 465-476.
3. Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M. and Schubeler, D. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell*, **30**, 755-766.
4. Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.-H., Yu, J. *et al.* (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotech*, **27**, 353-360.
5. Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M. and Rehli, M. (2009) Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res*, **19**, 1165-1174.
6. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315-322.
7. Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H. and Jaenisch, R. (2003) Induction of tumors in mice by genomic hypomethylation. *Science*, **300**, 489-492.
8. Walsh, C.P., Chaillet, J.R. and Bestor, T.H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet*, **20**, 116-117.
9. Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Pequignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187-191.
10. Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, **74**, 481-514.
11. Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R. and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, **122**, 3195-3205.
12. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915-926.
13. Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998) DNA hypomethylation leads to elevated mutation rates. *Nature*, **395**, 89-93.
14. Margot, J.B., Aguirre-Arteta, A.M., Di Giacco, B.V., Pradhan, S., Roberts, R.J., Cardoso, M.C. and Leonhardt, H. (2000) Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol*, **297**, 293-300.
15. Spada, F., Rothbauer, U., Zolghadr, K., Schermelleh, L. and Leonhardt, H. (2006) Regulation of DNA methyltransferase 1. *Adv Enzyme Regul*, **46**, 224-234.
16. Zimmermann, C., Guhl, E. and Graessmann, A. (1997) Mouse DNA methyltransferase (MTase) deletion mutants that retain the catalytic domain display neither de novo nor maintenance methylation activity in vivo. *Biol Chem*, **378**, 393-405.

17. Pradhan, S. and Esteve, P.O. (2003) Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. *Biochemistry*, **42**, 5321-5332.
18. Fatemi, M., Hermann, A., Pradhan, S. and Jeltsch, A. (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J Mol Biol*, **309**, 1189-1199.
19. Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F. and Slany, R.K. (2002) The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res*, **30**, 958-965.
20. Lee, J.H., Voo, K.S. and Skalnik, D.G. (2001) Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem*, **276**, 44669-44676.
21. Cross, S.H., Meehan, R.R., Nan, X. and Bird, A. (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet*, **16**, 256-259.
22. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P. and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, **439**, 811-816.
23. Koyama-Nasu, R., David, G. and Tanese, N. (2007) The F-box protein Fbl10 is a novel transcriptional repressor of c-Jun. *Nat Cell Biol*, **9**, 1074-1080.
24. Frescas, D., Guardavaccaro, D., Bassermann, F., Koyama-Nasu, R. and Pagano, M. (2007) JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. *Nature*, **450**, 309-313.
25. Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T. and Downing, J.R. (2003) TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, **17**, 637-641.
26. Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H. and Hayashi, Y. (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res*, **62**, 4075-4080.
27. Jorgensen, H.F., Ben-Porath, I. and Bird, A.P. (2004) Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol Cell Biol*, **24**, 3387-3395.
28. Bestor, T.H. (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *Embo J*, **11**, 2611-2617.
29. Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D. and Pradhan, S. (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry*, **47**, 10000-10009.
30. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, **324**, 930-935.
31. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.
32. Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, **33**, W244-248.
33. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. and Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins*, **23**, 318-326.

34. Zhang, Y. and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, **33**, 2302-2309.
35. DeLano, W.L. (2002). The PyMOL Molecular Graphics System.
36. Shen, M.Y. and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci*, **15**, 2507-2524.
37. Eisenberg, D., Luthy, R. and Bowie, J.U. (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*, **277**, 396-404.
38. Sporbert, A., Domaing, P., Leonhardt, H. and Cardoso, M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res*, **33**, 3521-3528.
39. Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep*, **5**, 1181-1186.
40. Schermelleh, L., Spada, F., Easwaran, H.P., Zolghadr, K., Margot, J.B., Cardoso, M.C. and Leonhardt, H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods*, **2**, 751-756.
41. Fellinger, K., Rothbauer, U., Felle, M., Langst, G. and Leonhardt, H. (2009) Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J Cell Biochem*, **106**, 521-528.
42. Fellinger, K., Leonhardt, H. and Spada, F. (2008) A mutagenesis strategy combining systematic alanine scanning with larger mutations to study protein interactions. *Anal Biochem*, **373**, 176-178.
43. Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res*, **35**, 4301-4312.
44. Frauer, C. and Leonhardt, H. (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res*, **37**, e22.
45. Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F. and Leonhardt, H. (2009) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucl. Acids Res.*, **In press**, 10.1093/nar/gkp1152.
46. Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep*, **10**, 1259-1264.
47. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C. and Leonhardt, H. (2007) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics*.
48. Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M. and Warren, A.J. (2006) Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *Embo J*, **25**, 4503-4512.
49. Ayton, P.M., Chen, E.H. and Cleary, M.L. (2004) Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol*, **24**, 10470-10478.
50. Bacolla, A., Pradhan, S., Larson, J.E., Roberts, R.J. and Wells, R.D. (2001) Recombinant human DNA (cytosine-5) methyltransferase. III. Allosteric control, reaction order, and influence of

plasmid topology and triplet repeat length on methylation of the fragile X CGG.CCG sequence. *J Biol Chem*, **276**, 18605-18613.

51. Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E. and Jaenisch, R. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev*, **10**, 1008-1020.
52. Andersson, T., Södersten, E., Duckworth, J.K., Cascante, A., Fritz, N., Sacchetti, P., Cervenka, I., Bryja, V. and Hermanson, O. (2009) CXXC5 Is a Novel BMP4-regulated Modulator of Wnt Signaling in Neural Stem Cells. *Journal of Biological Chemistry*, **284**, 3672-3681.
53. Hino, S.-i., Kishida, S., Michiue, T., Fukui, A., Sakamoto, I., Takada, S., Asashima, M. and Kikuchi, A. (2001) Inhibition of the Wnt Signaling Pathway by Idax, a Novel Dvl-Binding Protein. *Mol. Cell. Biol.*, **21**, 330-342.
54. London, T.B.C., Lee, H.-J., Shao, Y. and Zheng, J. (2004) Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. *Biochemical and Biophysical Research Communications*, **322**, 326-332.
55. Xia, Z.B., Anderson, M., Diaz, M.O. and Zeleznik-Le, N.J. (2003) MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci U S A*, **100**, 8342-8347.

FIGURE 1

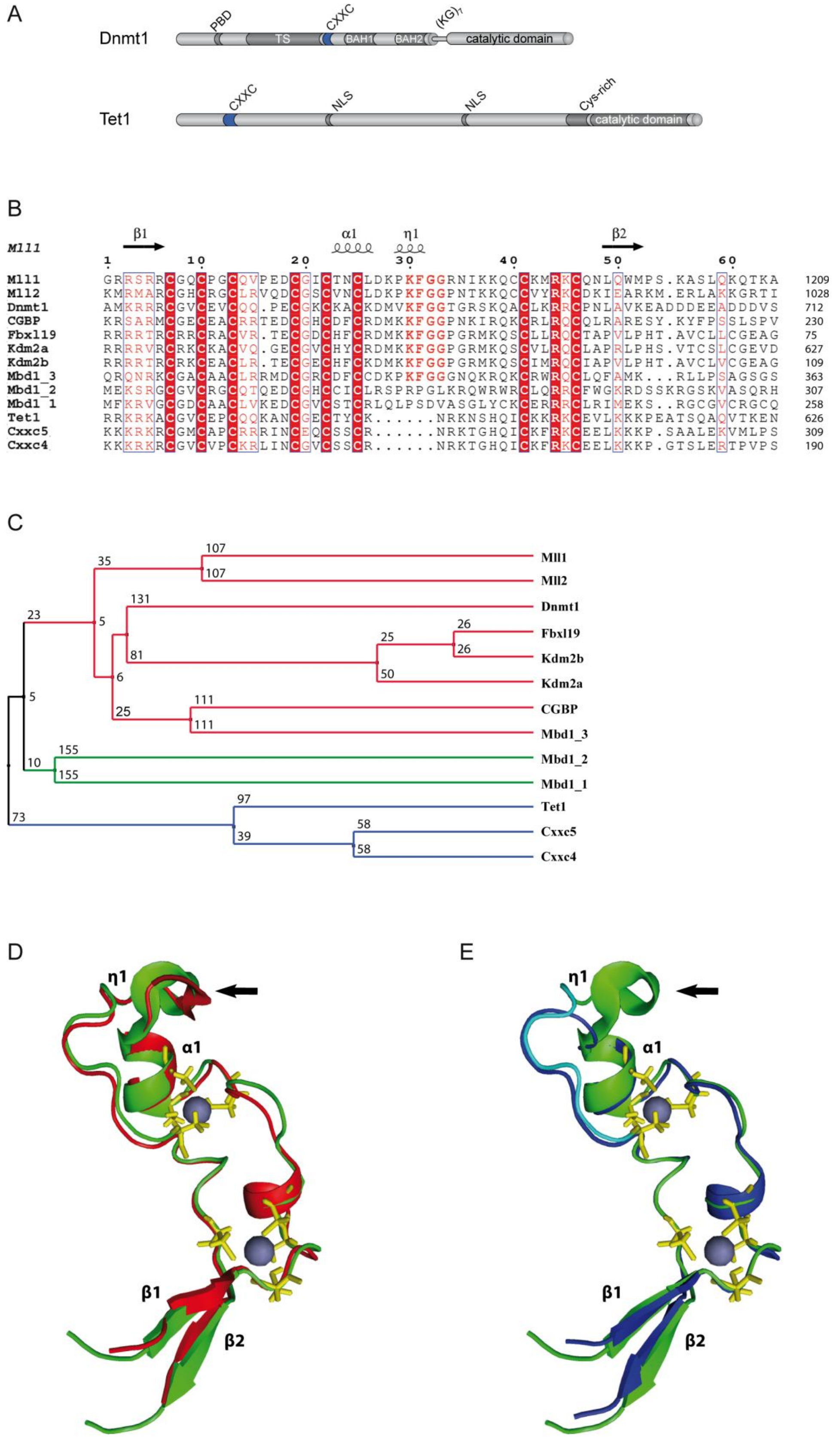

A
Dnmt1
PBD
TS
CXXC
BAH1
BAH2
(KG)7
catalytic domain
Tet1
CXXC
NLS
NLS
Cys-rich
catalytic domain
B
M111
β1
α1
η1
β2
1
10
20
30
40
50
60
M111 GRRSRRCGQCPGCQVPEDCGICTNCLDKPKFGGRNIKKQCCKMRKCQNLQWMPS.KASLQKQTKA 1209
M112 KMRMARCGHCRGCLRVQDCGSCVNCLDKPKFGGPNTKKQCCVYRKCDKIEARKM.ERLAKKGRTI 1028
Dnmt1 AMKRRRCGVCEVCQQ.PECGKCKACKDMVKFGGTGRSKQACLKRRCPNLAVKEADDDEEADDDVS 712
CGBP KRSARMCGECEACRRTEDCGHCDFCRDMKKFGGPNKIRQKCRLRQCQLRARESY.KYFPSSLSPV 230
Fbxl19 RRRRTRCRRCRACVR.TECGDCHFCRDMKKFGGPGRMKQSCLLRQCTAPVLPHT.AVCLLCGEAG 75
Kdm2a RRRRVRCRKCKACVQ.GECGVCHYCRDMKKFGGPGRMKQSCVLRQCLAPRLPHS.VTCSLCGEVD 627
Kdm2b RRRRTRCRKCEACLR.TECGECHFCKDMKKFGGPGRMKQSCIMRQCIAPVLPHT.AVCLVCGEAG 109
Mbd1_3 QRQNRKCGACAACLRRMDCGRCDFCCDKPKFGGGNQKRQKCRWRQCLQFAMK...RLLPSAGSGS 363
Mbd1_2 MEKSRGCGVCRGCQTQEDCGHCCICLRSPRPGLKRQWRWRCLQRRCFWGKRDSSKRGSKVASQRH 307
Mbd1_1 MFKRVGCGDCAACLVKEDCGVCSTCRLQLPSDVASGLYCKCERRRCLRIMEKS..RGCGVCRGCQ 258
Tet1 RRKRKACGVCEPCQQKANCGECTYCK......NRKNSHQICKKRKCEVLKKKPEATSQAQVTKEN 626
Cxxc5 KKKRKRCGMCAPCRRRINCEQCSSCR......NRKTGHQICKFRKCEELKKKP.SAALEKVMLPS 309
Cxxc4 KKKRKRCGVCVPCKRLINCGVCSSCR......NRKTGHQICKFRKCEELKKKP.GTSLERTPVPS 190
C
107
107
35
131
23
5
25
26
26
81
50
6
5
111
25
111
155
10
155
97
73
58
39
58
Mll1
Mll2
Dnmt1
Fbxl19
Kdm2b
Kdm2a
CGBP
Mbd1_3
Mbd1_2
Mbd1_1
Tet1
Cxxc5
Cxxc4
D
η1
α1
β1
β2
E
η1
α1
β1
β2

Figure 1. Sequence and predicted structural homology of CXXC domains. **(A)** Schematic representation of the domain structure in Dnmt1 and Tet1. The catalytic domain and the N-terminal region of Dnmt1 are connected by seven lysine-glycine repeats [$(KG)_7$]. PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC-type zinc finger domain; BAH: bromo-adjacent homology domain; NLS: nuclear localization signal; Cys-rich: cysteine rich region. **(B)** Alignment of mammalian CXXC domains. Numbers on the right side indicate the position of the last amino acid in the corresponding protein. The Mbd1a isoform contains three CXXC motifs (CXXC_1-3). Absolutely conserved residues, including the eight cysteines involved in zinc ion coordination are highlighted in red and the conserved KFGG motif is in red bold face. Positions with residues in red face share 70% similarity as calculated with the Risler algorithm (56). All sequences are from *M. musculus*. Accession numbers (for GenBank unless otherwise stated): Dnmt1, NP_034196; Mll1, NP_001074518; Mll4, O08550 (SwissProt); CGBP, NP_083144; Kdm2a, NP_001001984; Kdm2b, NP_001003953; Fbxl19, NP_766336; Mbd1, NP_038622; CXXC4/Idax, NP_001004367; CXXC5, NP_598448. **(C)** A homology tree was generated from the alignment in (B). The three subgroups of CXXC domains identified are in different colors. Average distances between the sequences are indicated. **(D-E)** Homology models of the mouse Dnmt1 (D; red) and Tet1 (E; blue) CXXC domains superimposed to the CXXC domain of MLL1 (green; (48)). MLL1 residues that are thought to contact DNA according to chemical shift measurements (48) are in cyan in (E), while cysteines involved in coordination of the two zinc ions are yellow. Arrows point to the KFGG motif in MLL1 and Dnmt1.

FIGURE 2

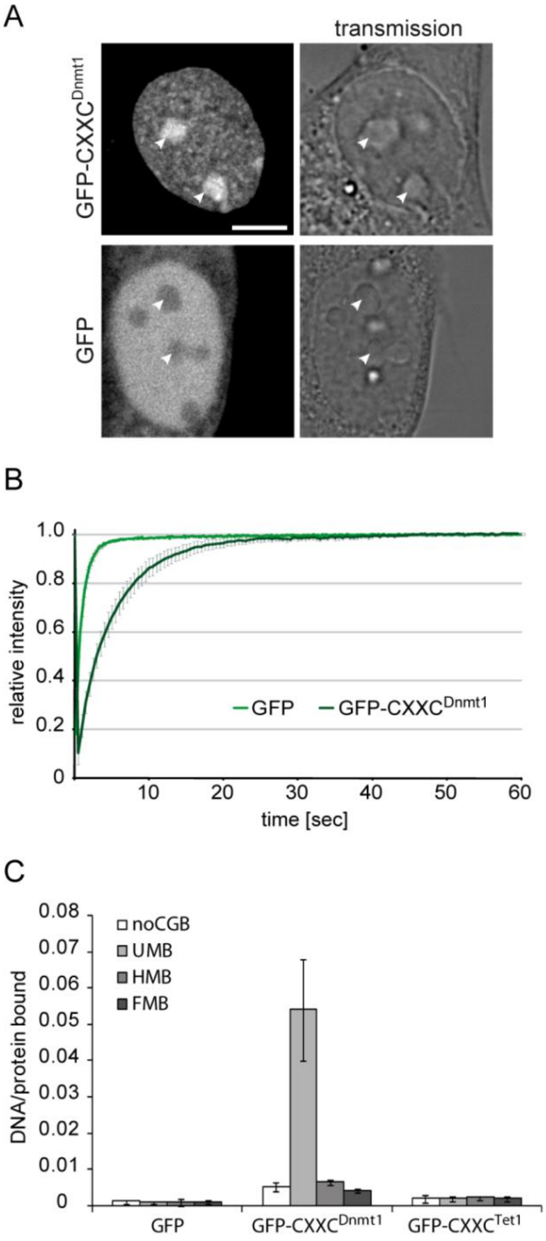


Figure 2. Properties of isolated Dnmt1 and Tet1 CXXC domains. **(A-B)** Subcellular localization (A) and binding kinetics (B) of GFP-CXXC^{Dnmt1} and GFP in mouse C2C12 myoblasts. Localization and binding kinetics were independent from the cell-cycle stage (data not shown). Arrowheads in (A) point to nucleoli. Scale bar: 5 µm. Binding kinetics were analyzed by FRAP. **(C)** DNA binding specificity of the Dnmt1 and Tet1 CXXC domains. GFP, GFP-CXXC^{Dnmt1} and GFP-CXXC^{Tet1} were pulled down from extracts of transiently transfected HEK293T cells and incubated with fluorescent DNA substrates containing no CpG site or one central un-, hemi- or fully methylated CpG site in direct competition (noCGB, UMB, HMB, FMB, respectively). Shown are the mean DNA/protein ratios and corresponding standard deviations from three independent experiments.

FIGURE 3

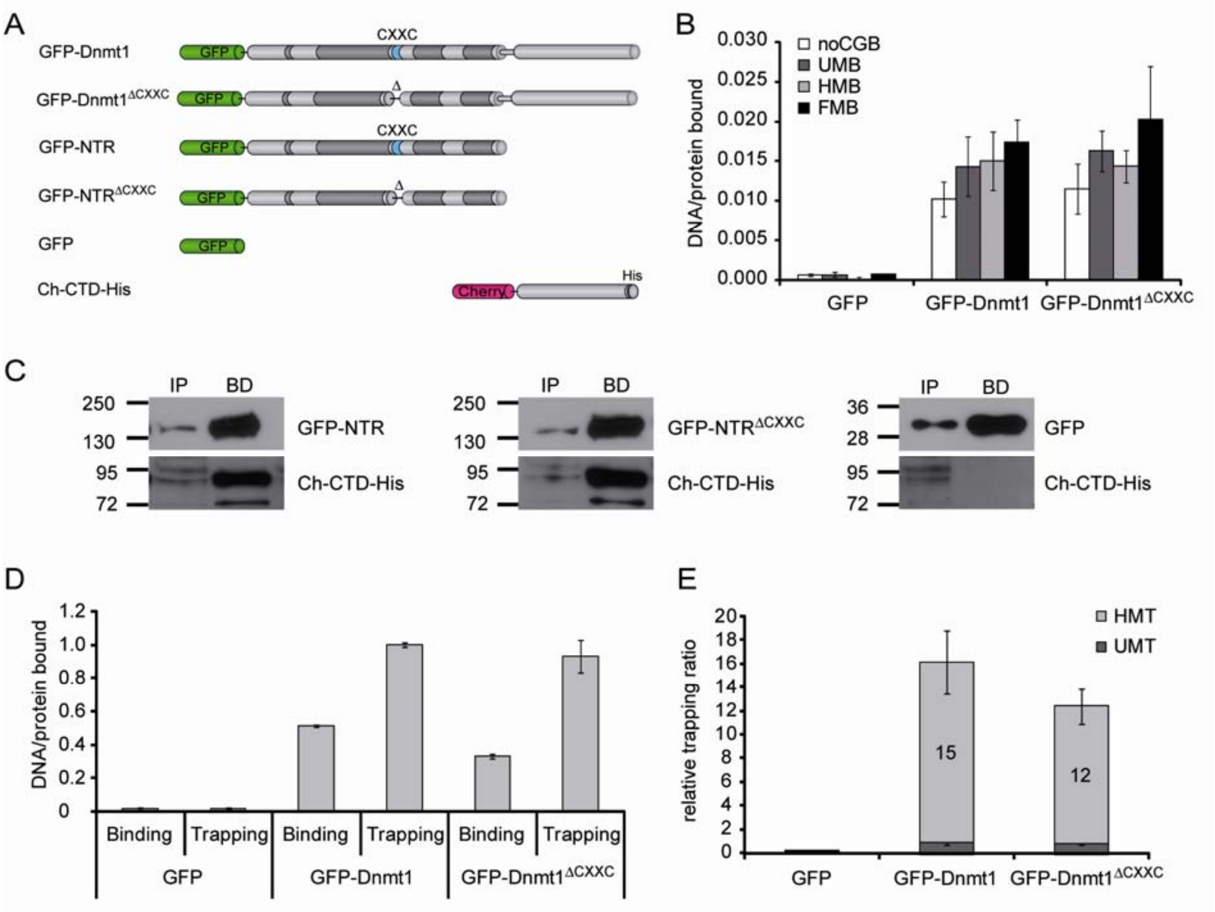


Figure 3. DNA binding specificity, intramolecular interaction and trapping of wild-type Dnmt1 and CXXC deletion constructs *in vitro*. **(A)** Schematic representation of Dnmt1 expression constructs. **(B)** DNA binding specificity of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ were assayed as described in Figure 2C. **(C)** Co-immunoprecipitation of the C-terminal domain of Dnmt1 (Ch-CTD-His) with the N-terminal region with and without deletion of the CXXC domain (GFP-NTR and GFP-NTR$^{\Delta CXXC}$, respectively). GFP fusions were detected using an anti-GFP antibody, while the C-terminal domain construct was detected using an anti-His antibody. GFP was used as negative control. **(D)** Comparison of binding and trapping activities for GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ to monitor irreversible covalent complex formation with hemimethylated substrates. **(E)** Relative covalent complex formation rate of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ on substrates containing one un- (UMT) or hemi-methylated CpG site (HMT) in direct competition. The trapping ratio for GFP-Dnmt1 on unmethylated substrate was set to 1. In (D) and (E) the means and corresponding standard deviations of triplicate samples from three independent experiments are shown. GFP was used as negative control.

FIGURE 4

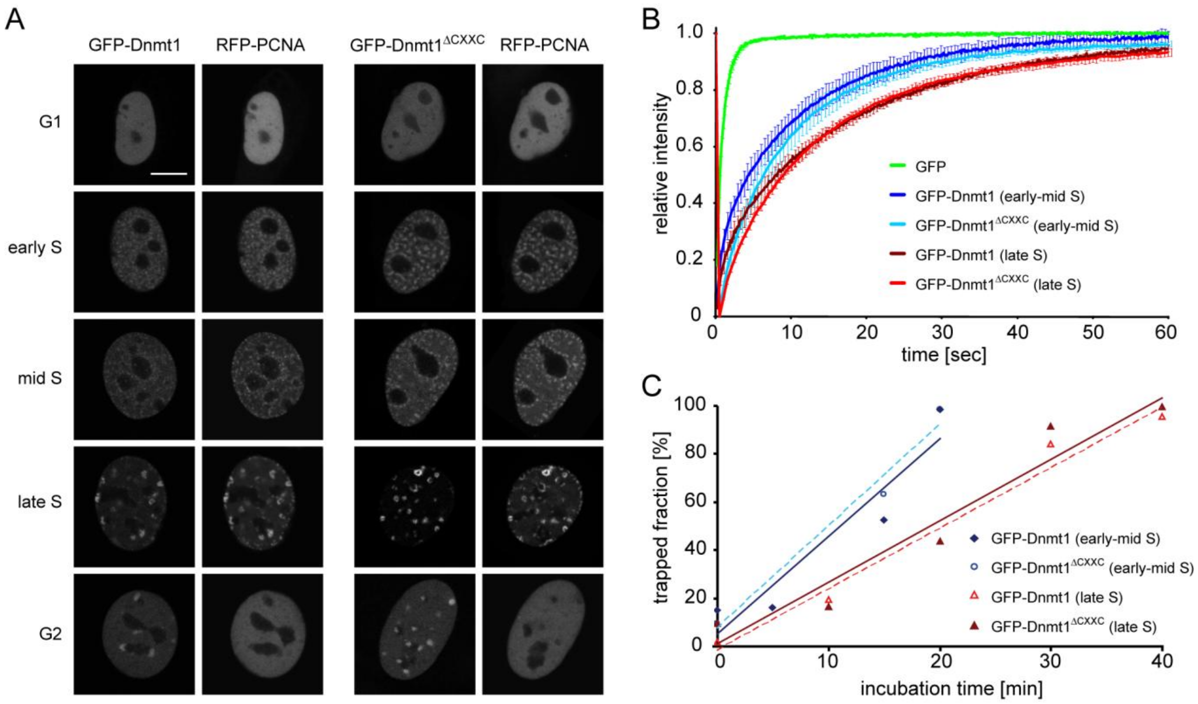


Figure 4. Cell-cycle dependent cellular localization, protein mobility and trapping of wild-type Dnmt1 and CXXC deletion constructs in mouse C2C12 myoblasts. **(A)** Cell-cycle dependent localization of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ constructs. Scale bar: 5 µm. **(B)** Analysis of binding kinetics of GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$ in early-mid and late S-phase cells by FRAP. The recovery curve for GFP is shown for comparison. **(C)** *In vivo* trapping by FRAP analysis in cells treated with 5-aza-dC. The trapped enzyme fraction is plotted over time for early-mid and late S-phase cells. In (A-C) RFP-PCNA was co-transfected to distinguish cell cycle stages in living cells.

FIGURE 5

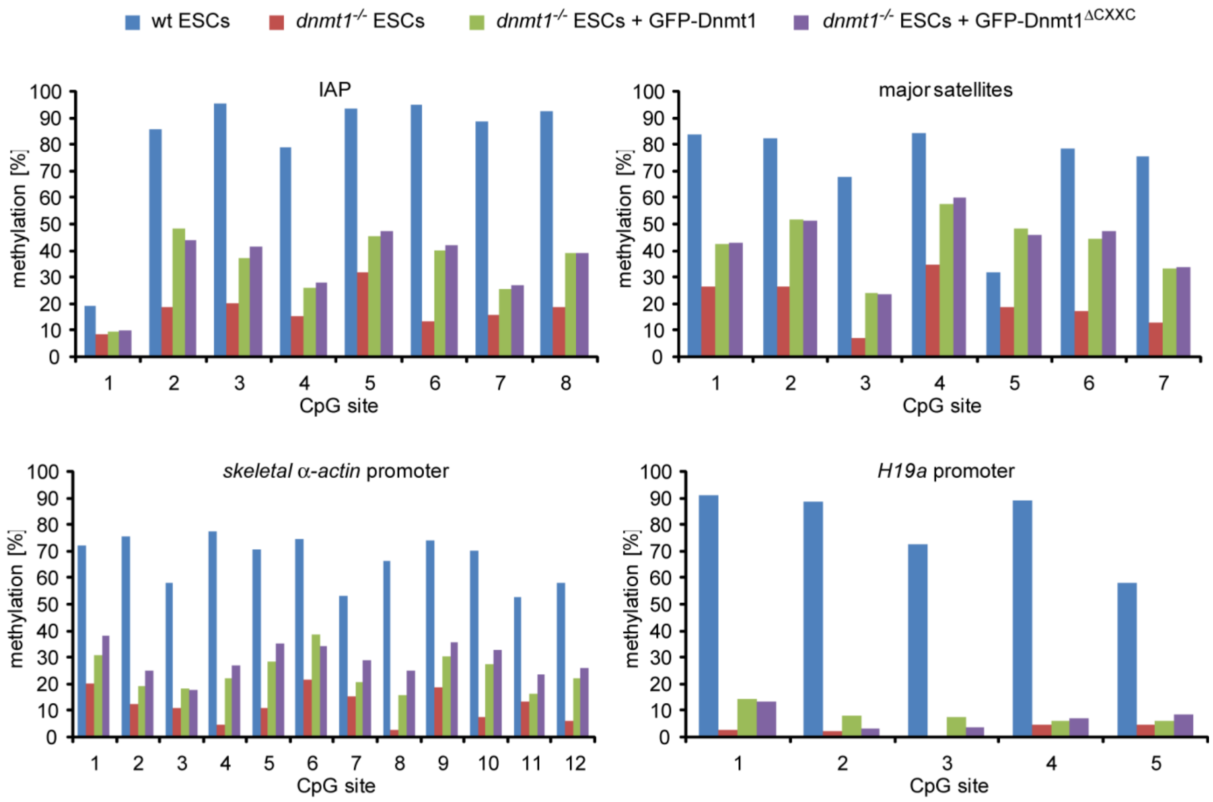


Figure 5. The CXXC deletion construct of Dnmt1 restores methylation in *dnmt1* null cells. Mouse *dnmt1*$^{-/-}$ ESCs transiently expressing GFP-Dnmt1 or GFP-Dnmt1$^{\Delta CXXC}$ were isolated by FACS-sorting 48 h after transfection and CpG methylation levels within the indicated sequences were analyzed by bisulfite treatment, PCR amplification and direct pyrosequencing. Methylation levels of untransfected wild type and *dnmt1*$^{-/-}$ ESCs are shown for comparison.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY TABLE 1

Supplementary Table 1. Sequences of DNA oligonucleotides used for preparation of double stranded DNA substrates. M: 5-methylcytosine.

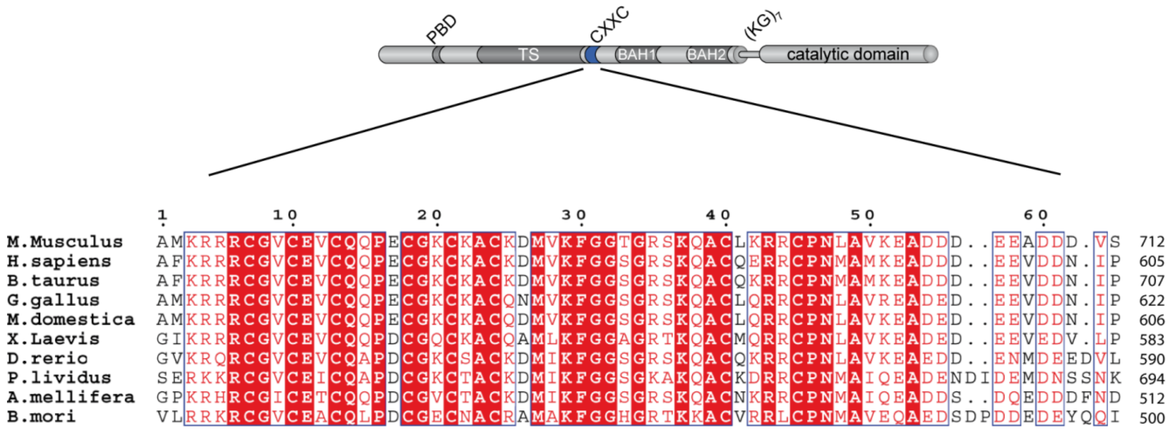
Name	Sequence
CG-up	5’- CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG -3’
MG-up	5’- CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG -3’
noCG-up	5’- CTCAACAACTAACTACCATCTGGACCAGAAGAGTCATCATGG -3’
Fill-In-550	5’- ATTO550-CCATGATGACTCTTCTGGTC -3’
Fill-In-590	5’- ATTO590-CCATGATGACTCTTCTGGTC -3’
Fill-In-647N	5’- ATTO647N-CCATGATGACTCTTCTGGTC -3’
Fill-In-700	5’- ATTO700-CCATGATGACTCTTCTGGTC -3’

SUPPLEMENTARY TABLE 2

Supplementary Table 2. DNA substrates used for the *in vitro* DNA binding and trapping assays.

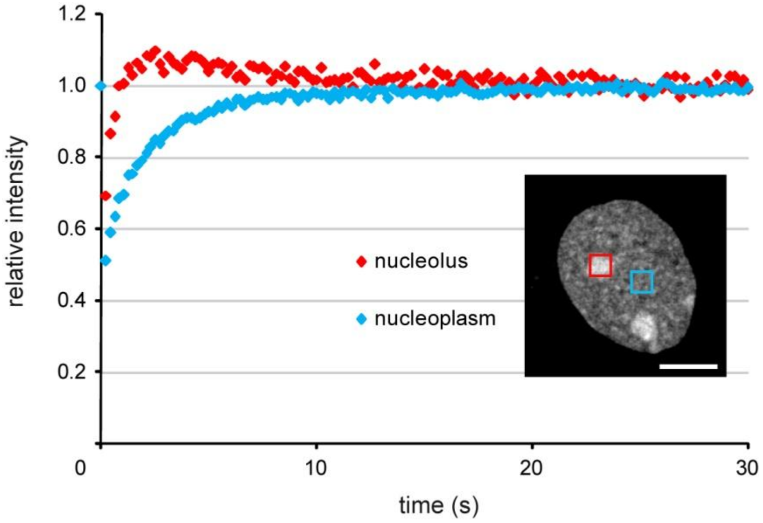
Name	CpG site	Label	Oligo I	Oligo II	dCTP reaction	Purpose
noCGB 700	no CpG site	700	noCG-up	Fill-In-700	dCTP	Binding
UMB 550	unmethylated	550	CG-up	Fill-In-550	dCTP	Binding
UMB 590		590		Fill-In-590		
UMB 647N		647N		Fill-In-647N		
UMB 700		700		Fill-In-700		
UMT 550		550		Fill-In-550	5-aza-dCTP	Trapping
HMB 590	hemimethylated	590	MG-up	Fill-In-590	dCTP	Binding
HMB 647N		647N		Fill-In-647N		
HMT 550		550		Fill-In-550	5-aza-dCTP	Trapping
HMT 647N		647N		Fill-In-647N		
FMB 647N	fully methylated	647N	MG-up	Fill-In-647N	5methyl dCTP	Binding

SUPPLEMENTARY FIGURE 1


PBD
TS
CXXC
BAH1
BAH2
(KG)7
catalytic domain
1
10
20
30
40
50
60
M.Musculus AMKRRRCGVCEVCQQPECGKCKACKDMVKFGGTGRSKQACLKRRCPNLAVKEADDD..EEADDD.VS 712
H.sapiens AFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQACQERRCPNMAMKEADDD..EEVDDN.IP 605
B.taurus AFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQACQKRRCPNMAMKEADDD..EEVDDN.IP 707
G.gallus AMKRRRCGVCEVCQQPECGKCKACQNMVKFGGSGRSKQACLQRRCPNLAVREADED..EEVDDN.IP 622
M.domestica AMKRRRCGVCEVCQQPECGKCKACQDMVKFGGSGRSKQACLQRRCPNLAVKEADED..EEVDDN.IP 606
X.Laevis GIKRRRCGVCEVCQQPDCGQCKACQAMLKFGGAGRTKQACMQRRCPNLAVKEADED..EEVEDV.LP 583
D.rerio GVKRQRCGVCEVCQAPDCGKCSACKDMIKFGGSGRSKQACQKRRCPNLAVKEAEDD..ENMDEEDVL 590
P.lividus SERKKRCGVCEICQAPDCGKCTACKDMIKFGGSGKAKQACKDRRCPNMAIQEADENDIDEMDNSSNK 694
A.mellifera GPKRHRCGICETCQQPDCGVCTACKDMIKFGGSGRSKQACNKRRCPNMAIQEADDS..DQEDDDFND 512
B.mori VLRRKRCGVCEACQLPDCGECNACRAMAKFGGHGRTKKACVRRLCPNMAVEQAEDSDPDDEDEYQQI 500

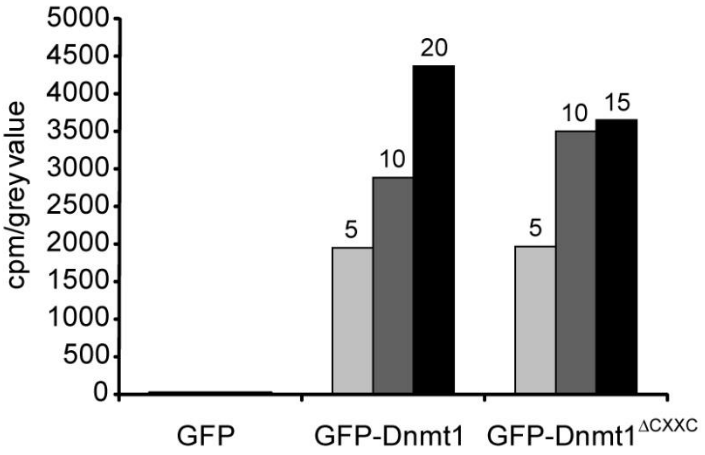
Supplementary Figure 1. Dnmt1 domain structure and alignment of Dnmt1 CXXC domains from different species. Numbers on the right side indicate the position of the last amino acid in each sequence. PBD: PCNA binding domain; TS: targeting sequence; CXXC: CXXC-type zinc finger domain; BAH: bromo-adjacent homology domain; $(KG)_7$: seven lysine-glycine repeats. Absolutely conserved residues are highlighted in red. Positions with residues in red face share 70% similarity as calculated with the Risler algorithm (56). The alignment was generated with ClustalW2 and displayed with ESPript 2.2. GenBank accession numbers are: *Mus musculus*: NP_034196; *Homo sapiens*: NP_001124295; *Bos taurus*: NP_872592; *Monodelphis domestica*: NP_001028141; *Gallus gallus*: NP_996835; *Xenopus laevis*: NP_001084021; *Danio rerio*: NP_571264; *Paracentrotus lividus*: Q27746 (Swiss Prot); *Bombyx mori*: NP_001036980; *Apis mellifera:* NP_001164522 (Dnmt1a).

SUPPLEMENTARY FIGURE 2



Supplementary Figure 2. Differential mobility of GFP-CXXC^{Dnmt1} in nucleoli and nucleoplasm of mouse C2C12 myoblasts measured by FRAP analysis. Identical regions of interest over the nucleoplasm or nucleoli (as exemplified in the inset) were bleached and recovery curves were recorded over 30 seconds. GFP-CXXC^{Dnmt1} kinetics are faster in nucleoli than in the nucleus, which indicates more transient (possibly unspecific) binding in the former than in the latter. Scale bar: 5µm.

SUPPLEMENTARY FIGURE 3



Supplementary Figure 3. Radioactive methyltransferase activity assay for GFP-Dnmt1 and GFP-Dnmt1$^{\Delta CXXC}$. The transfer of [^{3}H]-methyl groups to poly(dI·dC)-poly(dI·dC) substrate was measured for increasing volumes of GFP fusion proteins immunopurified from transiently transfected HEK293T cells. Counts per minute (cpm) were normalized to the relative protein concentration as determined by SDS-PAGE analysis. GFP was used as negative control. Numbers above the bars indicate the volume of protein solution added.

SUPPLEMENTARY METHODS

In vitro methyltransferase activity assay

The GFP-binding protein (GBP) was fused to a 6xHis tag, expressed in *E. coli* and purified as described (47). For coupling to Ni-NTA agarose beads (Qiagen), 8 mg of His-tagged GBP were added to 1 ml of equilibrated beads and incubated in PBS for 2 h at 4°C. Unbound protein was washed out twice with PBS. Extracts of HEK293T cells expressing GFP or a GFP fusions were prepared in 200 µl lysis buffer II (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5 % Tween-20, 2 mM $MgCl_2$, 1 mg/ml DNaseI, 2 mM PMSF, 1X mammalian protease inhibitor mix). After centrifugation, supernatants were diluted to 500 µl with immunoprecipitation buffer II (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.05 % Tween-20) and precleared by incubation with 25 µl of equilibrated Ni-NTA agarose beads for 30 min at 4°C followed by centrifugation. Precleared extracts were then incubated with 40 µg of His-tagged GBP coupled to Ni-NTA beads for 2 hours at 4°C with constant mixing. GFP or GFP fusions were pulled down by centrifugation at 540 *g*. After washing twice with wash buffer II (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05 % Tween-20), complexes were eluted with 60 µl of elution buffer (10 mM Tris pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM DTT, 250 mM imidazole) for 10 min at 25°C with constant mixing. 10 µl aliquots of all eluates were subjected to western blot analysis using a mouse monoclonal anti-GFP antibody (Roche) and quantified by densitometry. Indicated volumes of eluate were incubated with 1 µg of poly(dI·dC)- poly(dI·dC) substrate (Sigma), 0.5 µg/µl of BSA and 1 µCi of S-adenosyl-[^{3}H-methyl]-methionine in 50 µl of trapping buffer (10 mM Tris pH 7.5, 100 mM KCl, 1 mM EDTA, 1 mM DTT) for 60 min at 37°C. 15 µl of each sample were spotted onto blotting paper and the DNA was precipitated with ice cold 5 % TCA. After washing twice with 5% TCA and once with cold 70 % ethanol, paper filters were air dried and analyzed by scintillation in 4 ml scintillation cocktail (Rotiszint® eco plus, Roth) for 5 min.

Supplementary references

47. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C. and Leonhardt, H. (2007) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics*.

56. Mohseni-Zadeh, S., Brezellec, P. and Risler, J.L. (2004) Cluster-C, an algorithm for the large-scale clustering of protein sequences based on the extraction of maximal cliques. *Comput Biol Chem*, **28**, 211-218.

2.4 The multi-domain protein Np95 connects DNA methylation and histone modification

1796–1804 *Nucleic Acids Research, 2010, Vol. 38, No. 6*
doi:10.1093/nar/gkp1152

Published online 21 December 2009

The multi-domain protein Np95 connects DNA methylation and histone modification

Andrea Rottach[1], Carina Frauer[1], Garwin Pichler[1], Ian Marc Bonapace[2], Fabio Spada[1] and Heinrich Leonhardt[1,*]

[1]Ludwig Maximilians University Munich, Department of Biology II and Center for Integrated Protein Science Munich (CIPS^M), Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany and [2]University of Insubria, Department of Structural and Functional Biology, Via da Giussano 12, 21052 Busto Arsizio (VA), Italy

Received April 24, 2009; Revised November 17, 2009; Accepted November 23, 2009

ABSTRACT

DNA methylation and histone modifications play a central role in the epigenetic regulation of gene expression and cell differentiation. Recently, Np95 (also known as UHRF1 or ICBP90) has been found to interact with Dnmt1 and to bind hemimethylated DNA, indicating together with genetic studies a central role in the maintenance of DNA methylation. Using *in vitro* binding assays we observed a weak preference of Np95 and its SRA (SET- and Ring-associated) domain for hemimethylated CpG sites. However, the binding kinetics of Np95 in living cells was not affected by the complete loss of genomic methylation. Investigating further links with heterochromatin, we could show that Np95 preferentially binds histone H3 N-terminal tails with trimethylated (H3K9me3) but not acetylated lysine 9 via a tandem Tudor domain. This domain contains three highly conserved aromatic amino acids that form an aromatic cage similar to the one binding H3K9me3 in the chromodomain of HP1ß. Mutations targeting the aromatic cage of the Np95 tandem Tudor domain (Y188A and Y191A) abolished specific H3 histone tail binding. These multiple interactions of the multi-domain protein Np95 with hemimethylated DNA and repressive histone marks as well as with DNA and histone methyltransferases integrate the two major epigenetic silencing pathways.

INTRODUCTION

DNA methylation and histone modifications are crucially involved in the regulation of gene expression, inheritance of chromatin states, genome stability and differentiation (1–3). Although the biochemical networks controlling these epigenetic marks have been the subject of intensive investigation, their interconnection is still not well resolved in mammals. DNA methylation patterns are established by *de novo* DNA methyltransferases Dnmt3a and 3b, while Dnmt1 is largely responsible for maintaining genomic methylation after DNA replication (4,5). Dnmt1 possesses an intrinsic preference for hemimethylated DNA substrates (6,7) and associates with proliferating cell nuclear antigen (PCNA) at replication sites *in vivo* (8–10). The transient interaction of Dnmt1 with PCNA enhances methylation efficiency but is not strictly required to maintain genomic methylation in human and mouse cells (11,12).

Recently, Np95 has emerged as a central regulatory factor for DNA methylation and interacts with all three Dnmts (13). Np95 localizes at replication foci and its genetic ablation leads to genomic hypomethylation and developmental arrest (14–19). Np95 and its SET- and Ring- associated (SRA) domain were shown to bind hemimethylated DNA with higher affinity than corresponding symmetrically methylated or unmethylated sequences both *in vitro* and *in vivo* (17,18,20–22). In addition, crystal structures of the SRA domain complexed with hemimethylated oligonucleotides revealed flipping of the 5-methylcytosine out of the DNA double helix, a configuration that would stabilize the SRA–DNA interaction (20–22). Thus, recruitment of Dnmt1 to hemimethylated CpG sites by Np95 has been proposed as mechanism for the maintenance of genomic methylation.

In addition to its role in controlling DNA methylation, Np95 has been shown to take part in several other chromatin transactions. Np95 or its human homolog ICBP90/UHRF1 were reported to interact with the histone deacetylase HDAC1 and the histone methyltransferase G9a and to mediate silencing of a viral promoter, suggesting a role of Np95 in gene silencing through histone modification (13,23,24). Np95 binds histone H3 and displays a Ring domain-mediated E3 ubiquitin ligase activity for core histones *in vitro* and possibly histone H3 *in vivo* (25,26). The plant

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

---

*To whom correspondence should be addressed. Tel: +49 89 218074232; Fax: +49 89 218074236; Email: h.leonhardt@lmu.de

© The Author(s) 2009. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

homeodomain (PHD) of Np95 has been linked to decondensation of replicating pericentric heterochromatin (PH), but it is still unclear which domains recognize specific histone modifications (16,26,27).

In this study we systematically analyzed the binding properties of Np95 and its individual domains to DNA and histone tails *in vitro* and their binding kinetics in living cells. Our data reveal a multi-functional modular structure of Np95 interconnecting DNA methylation and histone modification pathways.

MATERIALS AND METHODS

Expression constructs

Expression construct for GFP–Dnmt1 and RFP–PCNA were described previously (10,28,29). All Np95 constructs were derived by PCR from corresponding myc- and His_6-tagged Np95 constructs (25). To obtain GFP– and Cherry-fusion constructs the Dnmt1 cDNA in the pCAG–GFP–Dnmt1-IRESblast construct (11) or the pCAG–Cherry–Dnmt1–IRESblast was replaced by Np95 encoding PCR fragments. The GFP–Np95ΔTudor expression construct was derived from the GFP–Np95 construct by overlap extension PCR (30). The GFP–Tudor mutant (Y188A, Y191A) was derived from the GFP–Np95 construct by PCR-based mutagenesis (31). All constructs were verified by DNA sequencing. Throughout this study enhanced GFP (eGFP) or monomeric Cherry (mCherry) constructs were used and for simplicity referred to as GFP– or Cherry-fusions.

Cell culture, transfection and immunofluorescence staining

HEK293T cells and embryonic stem cells (ESCs) were cultured and transfected as described (11), with the exception that FuGENE HD (Roche) was used for transfection of ESCs. The $dnmt1^{-/-}$ J1 ESCs used in this study are homozygous for the c allele (4). For immunofluorescence staining, TKO ESCs were grown on cover slips, fixed with 3.7% formaldehyde in PBS for 10 min and permeabilized with 0.5% Triton X-100 for 5 min. After blocking with 3% BSA in PBS for 1 h endogenous Np95 was detected with a polyclonal rabbit anti-Np95 serum (32). The secondary antibody was conjugated to Alexa Fluor 568 (Molecular Probes). Nuclear counterstaining was performed with DAPI and cells were mounted in Vectashield (Vector Laboratories). Images were obtained using a TCS SP5 AOBS confocal laser scanning microscope (Leica) using a 63x/1.4 NA Plan-Apochromat oil immersion objective. Fluorophores were excited with 405 and 561 nm lasers.

In vitro DNA binding assay

The *in vitro* DNA binding assay was performed as described previously (33) with the following modifications. Two different double-stranded DNA probes were labeled with distinct fluorophores and used in direct competition (see Supplementary Figures S3 and S6 for details). DNA oligos were controlled for CG methylation state by digestion with either a CG methylation-sensitive (HpaII) or -insensitive (MspI) enzyme (Supplementary Figure S4). For extract preparation 2 mM $MgCl_2$ and 1 mg/ml DNaseI were included in the lysis buffer. Extracts from 1-3 transfected 10 cm plates were diluted to 500-1000 µl with immunoprecipitation (IP) buffer and 1 µg of GFP–Trap (34) (ChromoTek, Germany) per final assay condition was added. After washing and equilibration beads were resuspended in 500 µl of binding buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 100 ng/µl BSA). Two oligonucleotide substrates were added to a final concentration of 50 nM each and incubated at room temperature (RT) for 60 min with constant mixing. Fluorescence intensity measurements were performed with a Tecan Infinite M1000 plate reader using the following excitation/emission wavelengths: 490 ± 10 nm/511 ± 10 nm for GFP, 550 ± 15 nm/580 ± 15 nm for ATTO550 and 650 ± 10 nm/670 ± 10 nm for ATTO647N. Values were adjusted using standard curves obtained with ATTO-dye coupled oligonucleotide primers and purified GFP. Binding activity was expressed as the ratio between the fluorescent signals of bound DNA probe and GFP fusion protein bound to the beads, so that the signals from bound probes are normalized to the amount of GFP fusion. Furthermore, values were normalized using a control set of DNA probes having identical sequences but distinct fluorescent labels (see Supplementary Figures S3 and S6 for details).

Peptide pull-down assay

Peptides were purchased as TAMRA conjugates (PSL, Germany) and are listed in Supplementary Figure S7. The peptide pull-down assay was performed analogously to the DNA binding assay described above. After one-step purification of GFP fusion proteins with the GFP–Trap (ChromoTek, Germany), the beads were equilibrated in 1 ml IP buffer and resuspended in 500 µl binding buffer supplemented with 100 ng/µl of BSA. Peptides were added to a final concentration of 0.74 µM and the binding reaction was performed at RT for 15 min to 60 min with constant mixing. The beads were washed twice with 1 ml of IP buffer and resuspended in 100 µl of the same. Wavelengths for excitation and measurement of TAMRA were 490 ± 5 nm and 511 ± 5 nm, respectively. Fluorescence intensity measurements were adjusted using standard curves from TAMRA coupled peptide and purified GFP.

Live cell microscopy and fluorescence recovery after photobleaching analysis

Live cell imaging and fluorescence recovery after photobleaching (FRAP) analysis were performed as described previously (11). For presentation, we used linear contrast enhancement on entire images.

Statistical analysis

Results were expressed as mean ± SEM. The difference between two mean values was analyzed by Student's *t*-test and was considered to be statistically significant in case of $P < 0.05$ and highly significant with $P < 0.001$.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

Electrophoretic mobility shift and supershift assays

Un- and hemimethylated DNA substrates (1 pmol UMB550 and HMB647N, respectively) were incubated with 0.6 pmol purified GFP–Np95 and 0.4 pmol GFP–antibody (mouse monoclonal antibody, Roche). Samples were subjected to a 3.5% non-denaturing PAGE and analyzed with a fluorescence scanner (Typhoon Trio scanner; GE Healthcare) to detect ATTO550 (unmethylated substrate), ATTO647N (hemimethylated substrate) and green fluorescence (GFP–Np95).

RESULTS AND DISCUSSION

Np95 binding kinetics is largely independent of DNA methylation levels *in vivo*

Recent studies showed Np95 bound to hemimethylated DNA, suggesting that the essential function of Np95 in the maintenance of DNA methylation consists of substrate recognition and recruitment of Dnmt1. To investigate the dynamics of these interactions *in vivo* we transiently transfected wild-type (wt) J1 ESCs with expression constructs for Cherry-Np95 and GFP–Dnmt1 and monitored their subcellular distribution using live-cell microscopy (Figure 1A and B). Np95 showed a nuclear distribution with a cell cycle-dependent enrichment at replicating PH, similar to Dnmt1. Consistent with earlier observations (8,12,14–16) we detected co-localization of Np95 and Dnmt1 at sites of DNA replication. We investigated the dynamics of Np95 binding by quantitative fluorescence recovery after photobleaching (FRAP) analysis (Figure 1B). As chromocenters (aggregates of PH) are not homogeneously distributed in the nucleus, we chose to bleach half nuclei to ensure that the bleached region contains a representative number of potential binding sites. We observed a relatively fast and full recovery of relative GFP–Dnmt1 fluorescence intensity (Figure 1B), reflecting a transient and dynamic interaction as described before (11). In contrast, Cherry-Np95 showed a considerably slower and only partial (~80%) recovery within the same observation period. These results indicated a relatively stable binding of Np95 to chromatin and revealed an immobile protein fraction of about 20%. These *in vivo* binding properties would be consistent with tight binding of Np95 to hemimethylated CpG sites and flipping of the methylated cytosines out of the DNA double helix as shown in recent co-crystal structures of the SRA domain of Np95 (20–22).

To directly test the contribution of DNA methylation and the interaction with Dnmt1 to protein mobility, we compared the binding kinetics of GFP–Np95 in wt ESCs and ESCs lacking either Dnmt1 or all three major DNA methyltransferases Dnmt1, 3a and 3b (triple knockout, TKO). Surprisingly, Np95 binding to chromatin was not affected by either drastic reduction (*dnmt1*$^{-/-}$) or even complete loss (TKO) of genomic methylation and showed in both cases remarkably similar FRAP kinetics compared to wt J1 ESCs (Figure 1C). Similar results were obtained with a C-terminal GFP fusion (Np95-GFP; Supplementary Figure S1), arguing against conformational or sterical impairments of the N-terminal fusion protein that could affect the binding kinetics. Also, both, the levels of endogenous Np95 and its degree of accumulation at chromocenters were highly variable in TKO cells, with chromocenter accumulation clearly visible in some cells (Supplementary Figure S2). These results show that DNA methylation and the three DNA methyltransferases do not have a major effect on the overall binding kinetics of Np95 in living cells.

The SRA domain of Np95 is necessary and sufficient for DNA binding *in vitro*

Next, we investigated the DNA binding activity of Np95 and the contribution of distinct Np95 domains by generating a systematic set of individual domains and deletion constructs fused to GFP (Figure 2A). To directly compare the *in vitro* binding affinity of Np95 regarding different methylation states, we synthesized double-stranded DNA-binding substrates with either one or three un- or hemimethylated CpG sites and labeled them with two distinct fluorophores (Supplementary Figure S3). DNA probes were controlled for CG methylation state by digestion with either a CG methylation-sensitive (HpaII) or -insensitive (MspI) enzyme (Supplementary Figure S4). Performing conventional electrophoretic mobility shift and supershift assays we confirmed the DNA binding activity of Np95 and detected a preference for hemimethylated DNA substrates (Figure 2B and Supplementary Figure S5).

As a second line of evidence and to quantify binding preferences, we applied our recently developed non-radioactive DNA binding assay (33) and tested GFP–fused wt Np95 as well as a systematic set of individual domains and deletion constructs for their DNA-binding properties *in vitro* (Figure 2C and Supplementary Figure S6). This assay allows fast comparison of different potential binding substrates in direct competition as well as the simultaneous quantification of GFP–labeled protein to calculate relative binding activity. The different GFP–Np95 fusion constructs were expressed in HEK293T cells, purified with the GFP–Trap (34) and incubated with the fluorescently labeled DNA substrates. GFP–fusion protein and bound DNA substrates were quantified with a fluorescence plate reader (Figure 2C and Supplementary Figure S6). Furthermore, results were corrected for any bias due to incorporation of different fluorescent labels (Supplementary Figures. S3 and S6). Under these assay conditions we observed an up to 2-fold preference (factor 1.6–1.9) of Np95 for DNA substrates containing one or three hemimethylated CpG sites (Supplementary Figure S6). Deletion of the SRA domain completely abolished the DNA-binding activity of Np95, whereas deletion of either the PHD or the Tudor domain had no effect (Figure 2C). Consistently, the isolated PHD and Tudor domains did not bind to DNA, while the SRA domain alone showed similar binding strength and sequence preference as full-length Np95. Together, these results clearly demonstrate that the SRA domain of Np95 preferentially binds to hemimethylated CpG sites, although this

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

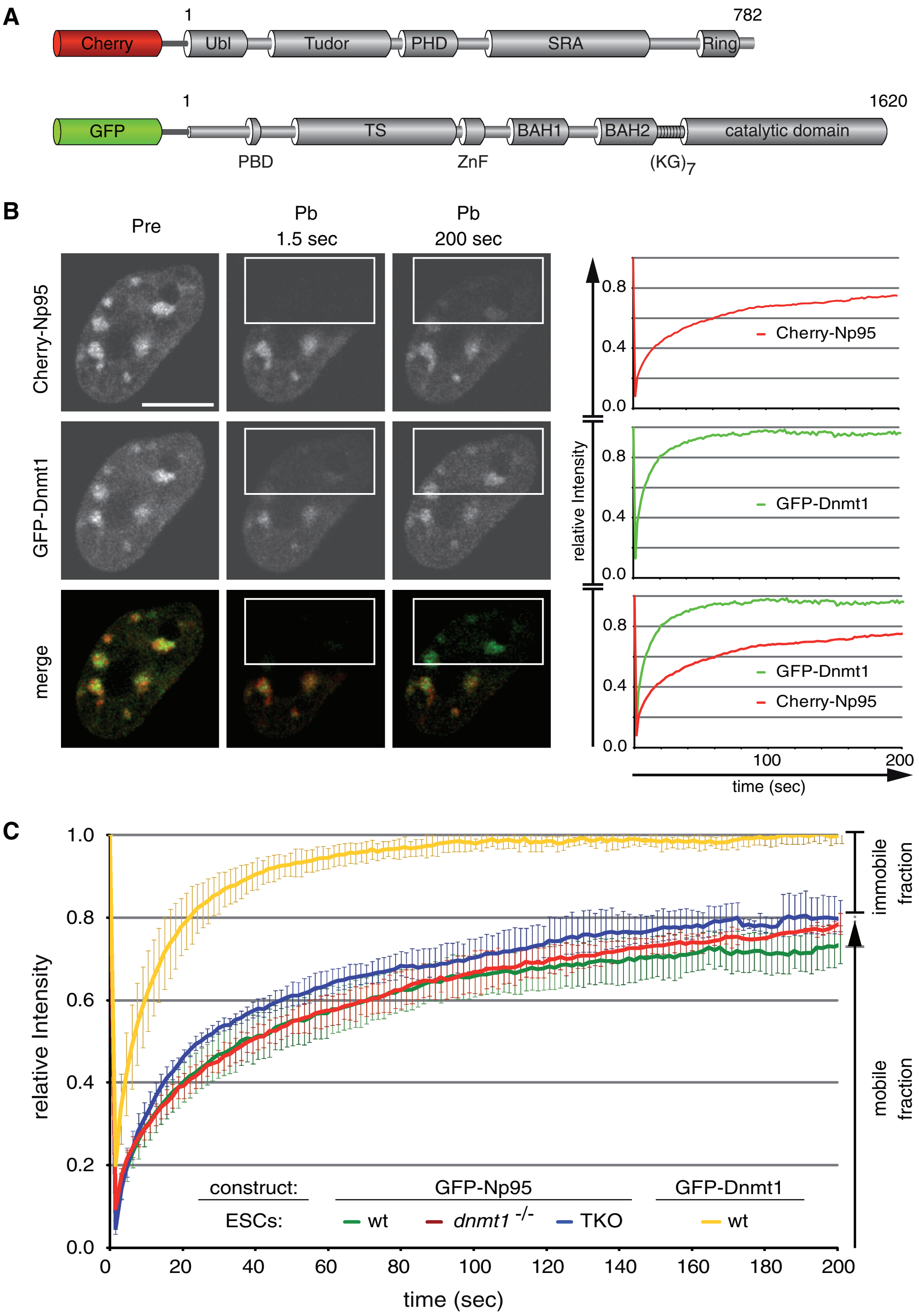


Figure 1. Binding kinetics of Dnmt1 and Np95 in living cells. (**A**) Schematic representation of Np95 and Dnmt1 fluorescent fusions. Ubl, ubiquitin-like domain; Tudor, tandem Tudor domain; PHD, plant homeodomain; SRA, SET- and Ring-associated domain; Ring domain; PBD, PCNA-binding domain; TS, targeting sequence; ZnF, zinc-finger; BAH, bromo adjacent homology domain; $(KG)_7$, lysine-glycine repeat. (**B**) Dnmt1 and Np95 display different kinetics. Representative images from FRAP experiments on wt J1 ESCs transiently co-transfected with Cherry-Np95 and GFP–Dnmt1 constructs. Images show co-localization at chromocenters before (Pre) and at the indicated time points after (Pb) bleaching half of the nucleus. Bleached areas are outlined. Corresponding FRAP curves are shown on the right. Bars, 5 µm. (**C**) FRAP kinetics of GFP–Np95 in J1 ESCs with different genetic backgrounds [wt, *dnmt1*$^{-/-}$ and *dnmt1*$^{-/-}$, 3a$^{-/-}$, 3b$^{-/-}$ (TKO)]. Kinetics of GFP–Dnmt1 is shown for comparison. Mobile and immobile fractions are indicated on the right. Values represent mean ± SEM.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

preference is only about 2-fold with purified proteins and substrates *in vitro*.

The SRA domain dominates binding kinetics but not localization of Np95

Next, we investigated the role of distinct Np95 domains in nuclear interactions *in vivo*. To this aim, we expressed the same GFP–Np95 constructs in *np95*$^{-/-}$ ESCs and tested their binding kinetics with FRAP experiments (Figure 2D). Importantly, GFP–Np95 showed similar FRAP kinetics in Np95 deficient, wt or Dnmt-deficient ESCs (Figures 1C and 2D). Among all domains tested, only the SRA domain showed similar kinetics as full-length Np95, including the relatively slow recovery and an immobile fraction of about 20%, while the Tudor and PHD domain displayed the same high mobility as GFP. Also, FRAP curves of the corresponding deletion constructs indicated that the Tudor and the PHD domains have only a minor contribution to *in vivo* binding kinetics, while deletion of the SRA domain drastically increased the mobility of Np95. These data indicate that the SRA domain dominates the binding kinetics of Np95 *in vivo*. Curiously, the addition of the PHD to the SRA domain (GFP–PHD-SRA) resulted in intermediate kinetics and loss of the immobile fraction. This effect was, however, not observed in the context of the full-length protein, suggesting that nuclear interactions of Np95 are controlled by a complex interplay among its domains. To directly study the role of the SRA domain in controlling the subcellular localization of Np95 we co-transfected *np95*$^{-/-}$ ESCs with expression constructs for Cherry-Np95 and either GFP–SRA or GFP–Np95ΔSRA (Figure 2E). This direct comparison showed that the isolated SRA domain does not co-localize with full-length Np95 at PH. Together, these results indicate that the SRA domain of Np95 is necessary and sufficient for DNA binding *in vitro* and also dominates the binding kinetics *in vivo*, but is *per se* not sufficient for proper subnuclear localization. The fact that the Np95ΔSRA construct co-localized with Np95 suggests that other domains than the SRA control the subcellular targeting of Np95.

Np95 binds to histone H3 via a tandem Tudor domain

Database searches showed that the sequence between the Ubl and PHD domains of Np95 is highly conserved in vertebrates and displays structural similarity to the family of Tudor domains [(35); PDB 3db4; Figure 3A and B]. The crystal structure revealed that the Tudor domain is composed of two subdomains (tandem Tudor) forming a hydrophobic pocket that accommodates a histone H3 N-terminal tail trimethylated at K9 (H3K9me3) (PDB 3db3; Figure 3C). This hydrophobic-binding pocket is created by three highly conserved amino acids (Phe152, Tyr188, Tyr191) forming an aromatic cage (Figure 3A and C). Interestingly, a very similar hydrophobic cage structure has been described for the chromodomain of the heterochromatin protein 1ß (HP1ß) (Supplementary Figure S7) that is known to bind trimethylated lysine 9 of histone H3 and associates with PH (36).

To further investigate the histone tail-binding properties of Np95, we mutated two amino acids of the aromatic cage (Y188A, Y191A) and tested the isolated tandem Tudor domain and corresponding mutant in comparison with Np95 using a peptide binding assay. GFP–Np95, GFP–Tudor and GFP–Tudor (Y188A, Y191A) were expressed in HEK293T cells, purified with the GFP–Trap and incubated with TAMRA-labeled histone tail peptides. The fluorescence intensity of GFP fusion proteins and bound peptides was quantified and the relative binding activity calculated (Figure 3D and Supplementary Figure S7). The tandem Tudor domain showed a highly significant preference for the trimethylated (H3K9me3) peptide, while this effect was less pronounced in the full-length Np95. Interestingly, acetylation of K9 (H3K9ac), a modification largely underrepresented in silent chromatin, prevented binding of the tandem Tudor domain. Remarkably, point mutations targeting aromatic cage residues within the tandem Tudor domain completely abolished specific binding to N-terminal histone H3 peptides.

Consistent with these binding data the tandem Tudor domain also showed a weak enrichment at PH, while the PHD domain, previously proposed as potential histone H3-binding motif (26), did neither bind to H3K9 peptides *in vitro* nor to PH *in vivo* (Supplementary Figure S8). These results indicate that the tandem Tudor domain of Np95 features a peptide binding pocket with structural and functional striking similarity to HP1ß and confers selective binding to histone modification states associated with silent chromatin.

These multiple interactions of Np95 with heterochromatin components correlate well with functional data. The depletion of Np95 in mouse cells resulted in increased transcription of major satellite repeats (16). Also, an interaction of Np95 with G9a was described and both were found to be essential for transcriptional regulation (24) and epigenetic silencing of transgenes (13).

In summary, we showed that the SRA domain is necessary and sufficient for DNA binding of Np95 *in vitro*. Photobleaching experiments further indicated that the SRA domain also dominates the binding kinetics of Np95 in living cells which was however largely independent of the DNA methylation level. These results suggest that the SRA domain may also bind to unmethylated DNA or undergo additional, still unidentified interactions *in vivo*. While the essential role of Np95 in the maintenance of DNA methylation is well established, it is still unclear how a relatively weak preference for hemimethylated DNA can be sufficient to maintain DNA methylation patterns over many cell division cycles for an entire life time. We suggest that the multiple interactions of the multi-domain protein Np95 with hemimethylated DNA and H3K9 methylated histone tails as well as with histone (G9a) and DNA (Dnmt1, 3a and 3b) methyltransferases may add up to the necessary specificity *in vivo*. Clearly, these multiple interactions place Np95 at the center of various epigenetic silencing mechanisms and likely mediate epigenetic crosstalk.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

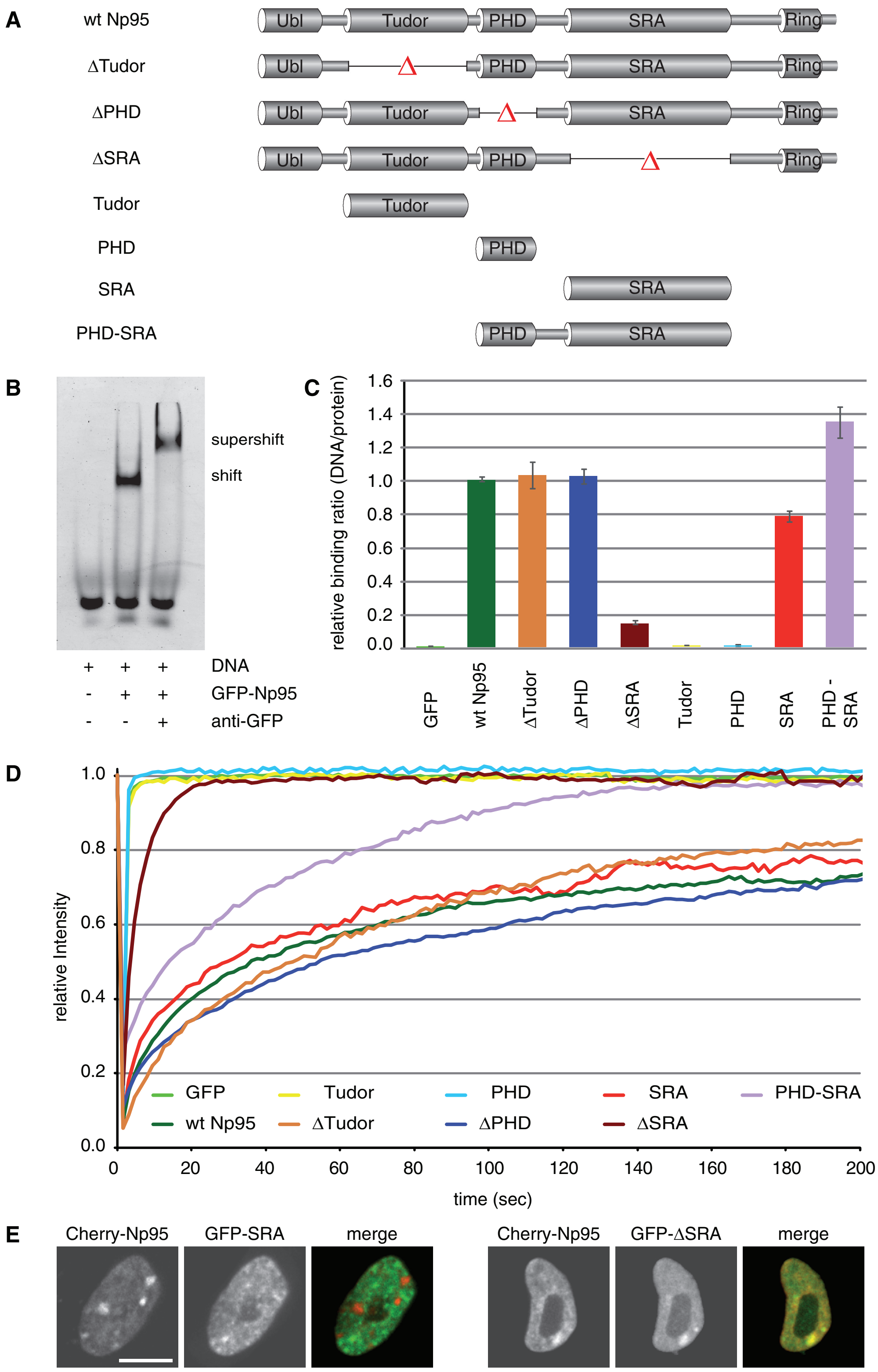


Figure 2. *In vitro* DNA binding and *in vivo* mobility of Np95 domains. (**A**) Schematic representation of the analyzed GFP–Np95 fusion constructs. (**B**) Electrophoretic mobility shift and supershift assay. GFP–Np95 binding to hemimethylated DNA substrates is shown by the shifted GFP–Np95:DNA complex. The addition of a GFP–antibody supershifted the GFP–Np95:DNA complex (supershift assay with unmethylated DNA substrates in direct competition with hemimethylated DNA substrates is shown in Supplementary Figure S6). (**C**) *In vitro* DNA-binding properties of Np95 constructs. Binding assays were performed using fluorescently labeled double stranded oligonucleotide probes containing one central hemimethylated CpG site. Shown are fluorescence intensity ratios of bound probe/bound GFP fusion. Values represent means and SD of three to six independent experiments. GFP was used as control. Further control experiments with either one or three central CpG sites and alternating fluorescent labels are shown in Supplementary Figure S3. (**D**) Kinetics of Np95 constructs in living $np95^{-/-}$ ESCs determined by half nucleus FRAP analysis. GFP is shown as reference. Curves represent mean values from 6 to 15 nuclei. SEM (0.001–0.005) is not shown for clarity of illustration. (**E**) Confocal mid-sections of living $np95^{-/-}$ ESCs transiently expressing the indicated Np95 fusion constructs (left and mid-panels). Merged images are displayed on the right. Bar, 5 µm.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

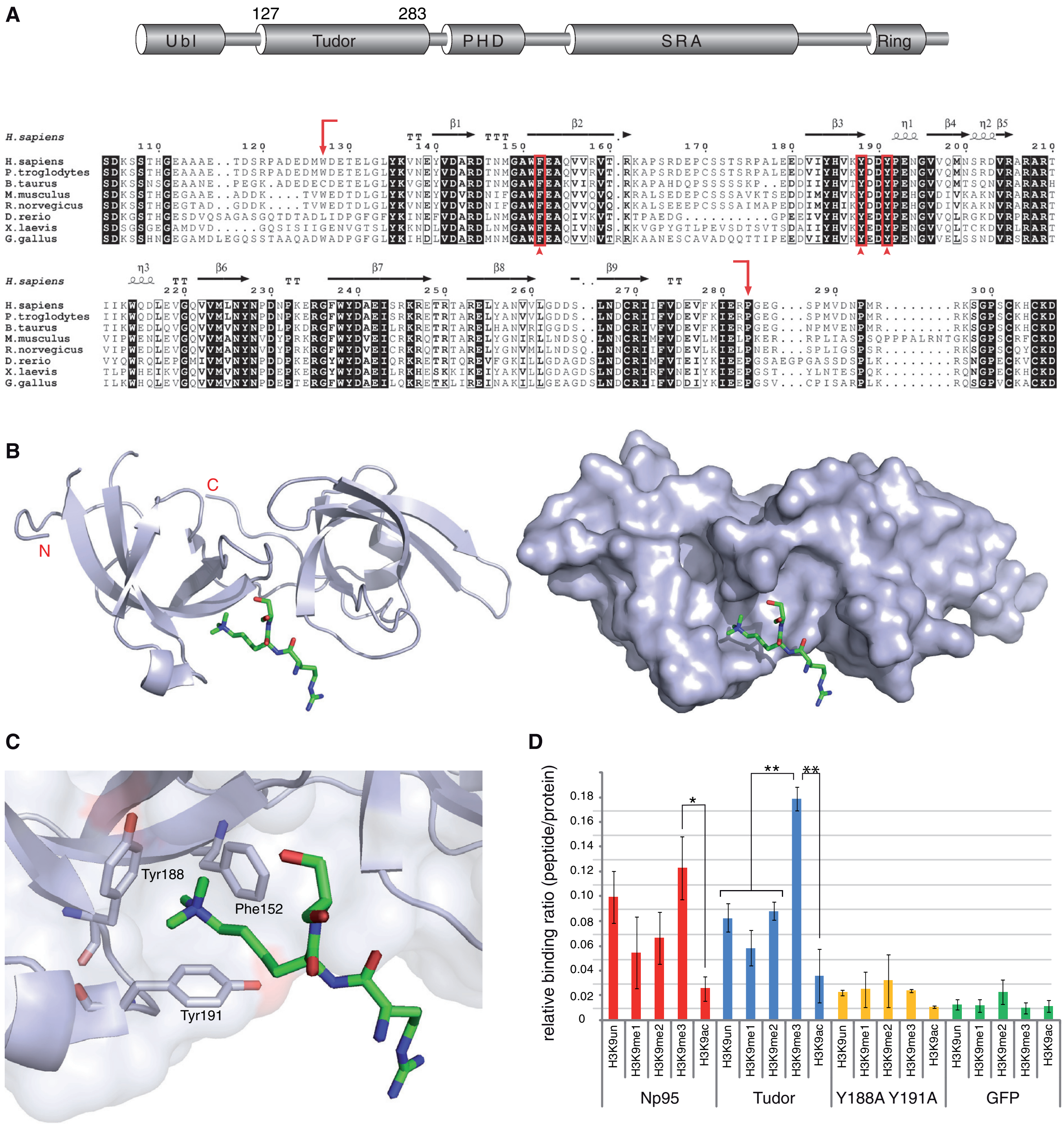


Figure 3. Structure and H3 N-terminal tail binding of the tandem Tudor domain. (**A**) Schematic drawing of the multi-domain architecture of Np95 (top) and alignment of tandem Tudor domains from vertebrate Np95 homologs (bottom). Arrows show the end and start positions of the crystallized tandem Tudor domain shown in (**B**). Residues forming the aromatic cage shown in (**C**) are indicated by arrowheads. Absolutely conserved residues of the tandem Tudor domain are black shaded, while positions showing conservative substitutions are boxed with residues in bold face. Secondary-structure elements were generated with EsPript (37) using the crystal structure of human UHRF1 (PDB 3db3 and 3db4) and are shown above the amino acid sequence: α-helices (η), β-strands, strict alpha turns (TT) and strict beta turns (TTT). Accession numbers: *Homo sapiens* Q96T88.1; *Pan troglodytes* XP_001139916.1; *Bos Taurus* AAI51672.1; *Mus musculus* Q8VDF2.2; *Rattus norvergicus* Q7TPK1.2; *Dario rerio* NP_998242.1; *Xenopus laevis* ABY28114.1, *Gallus gallus* XP_418269.2. (B) Side view of the tandem Tudor domain as a cartoon model (left) and as surface representation (right) in complex with a histone H3 N-terminal tail peptide trimethylated at lysine 9 (green stick model; only Arg8-Lys9-Ser10 of the H3 peptide are resolved). The image was generated with PyMOL (38). (C) An aromatic cage is formed by Phe152, Tyr188 and Tyr191 and accommodates the trimethylated lysine 9 of H3 (H3K9me3). (**D**) Histone H3 N-terminal tail binding specificity of GFP–Np95, GFP–Tudor and GFP–Tudor (Y188A Y191A) *in vitro*. Shown are fluorescence intensity ratios of bound probe/bound GFP fusion. GFP was used as negative control. Shown are means ± SEM from four to ten independent experiments and two-sample t-tests were performed that do or do not assume equal variances, respectively. Statistical significance compared to the binding ratio of H3K9me3 is indicated: $*P < 0.05$, $**P < 0.001$.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The authors are grateful to Masahiro Muto and Haruhiko Koseki (RIKEN Research Center for Allergy and Immunology, Yokohama, Japan) for providing wild-type and *np95*$^{-/-}$ E14 ESCs, to En Li (Novartis Institutes for Biomedical Research, Boston, MA) for *dnmt1*$^{-/-}$ and J1 ESCs and to Masaki Okano (RIKEN Center for Developmental Biology, Kobe, Japan) for the TKO ESCs.

FUNDING

This work was supported by the Nanosystems Initiative Munich (NIM), the BioImaging Network Munich (BIN) and by grants from the Deutsche Forschungsgemeinschaft (DFG) to H.L. IMB was supported by the Italian Association for Cancer Research (AIRC), the Fondazione CARIPLO Progetto NOBEL. C.F. and G.P. were supported by the International Doctorate Program NanoBioTechnology (IDK-NBT) and the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). Funding for open access charges: DFG.

Conflict of interest statement. None declared.

REFERENCES

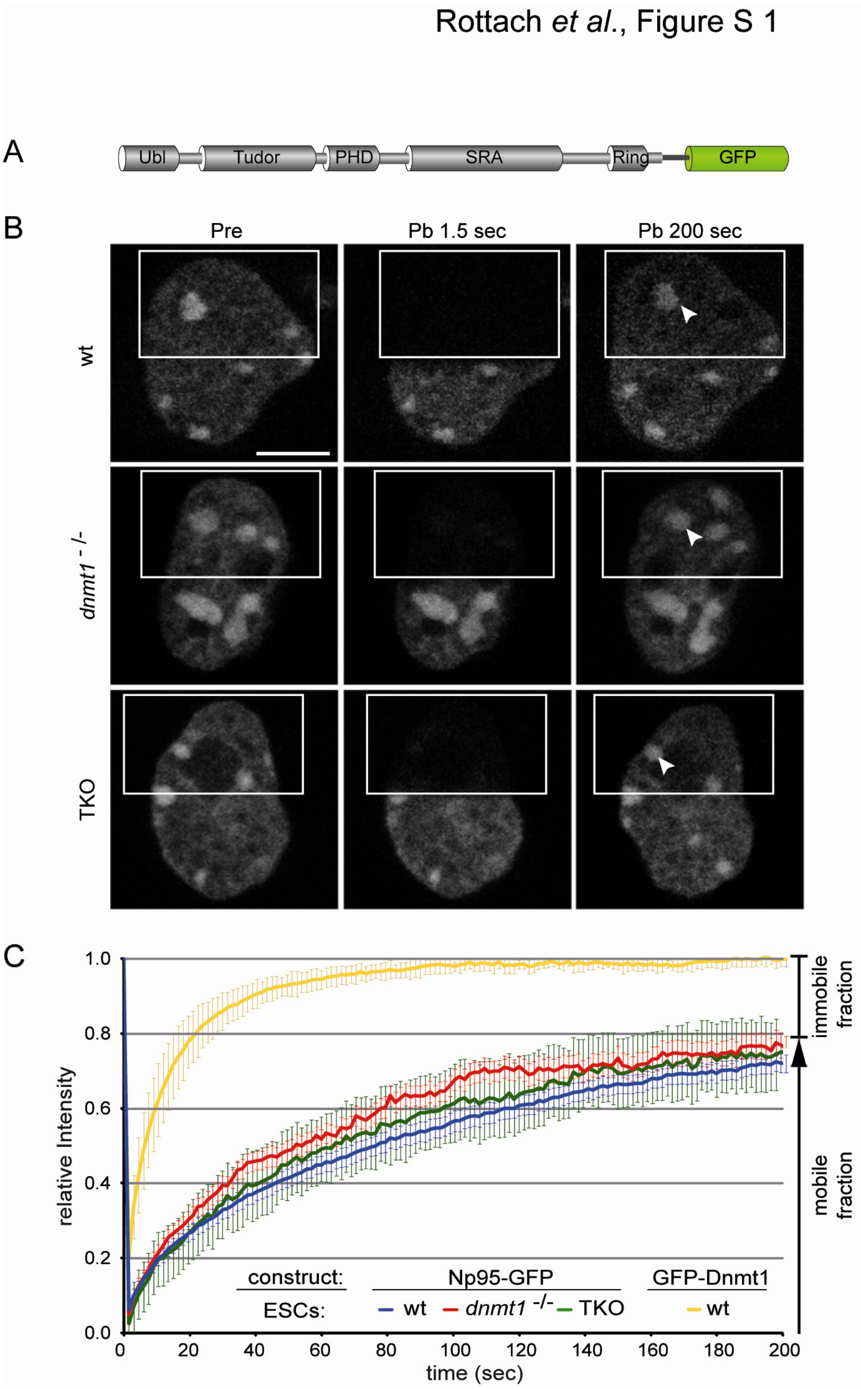
1. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*., **16**, 6–21.
2. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
3. Reik,W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.
4. Lei,H., Oh,S., Okano,M., Juttermann,R., Goss,K., Jaenisch,R. and Li,E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, **122**, 3195–3205.
5. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
6. Bestor,T.H. and Ingram,V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl Acad. Sci. USA*, **80**, 5559–5563.
7. Pradhan,S., Talbot,D., Sha,M., Benner,J., Hornstra,L., Li,E., Jaenisch,R. and Roberts,R. (1997) Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. *Nucleic Acids Res*., **25**, 4666–4673.
8. Leonhardt,H., Page,A.W., Weier,H.U. and Bestor,T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, **71**, 865–873.
9. Chuang,L.S.-H., Ian,H.-I., Koh,T.-W., Ng,H.-H., Xu,G. and Li,B.F.L. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science*, **277**, 1996–2000.
10. Easwaran,H.P., Schermelleh,L., Leonhardt,H. and Cardoso,M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep*., **5**, 1181–1186.
11. Schermelleh,L., Haemmer,A., Spada,F., Rosing,N., Meilinger,D., Rothbauer,U., Cristina Cardoso,M. and Leonhardt,H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res*., **35**, 4301–43012.
12. Spada,F., Haemmer,A., Kuch,D., Rothbauer,U., Schermelleh,L., Kremmer,E., Carell,T., Langst,G. and Leonhardt,H. (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol*., **176**, 565–571.
13. Meilinger,D., Fellinger,K., Bultmann,S., Rothbauer,U., Bonapace,I.M., Klinkert,W.E., Spada,F. and Leonhardt,H. (2009) *EMBO Rep*, **10**, 1259–1264.
14. Uemura,T., Kubo,E., Kanari,Y., Ikemura,T., Tatsumi,K. and Muto,M. (2000) Temporal and spatial localization of novel nuclear protein NP95 in mitotic and meiotic cells. *Cell Struct. Funct*., **25**, 149–159.
15. Miura,M., Watanabe,H., Sasaki,T., Tatsumi,K. and Muto,M. (2001) Dynamic changes in subnuclear NP95 location during the cell cycle and its spatial relationship with DNA replication foci. *Exp. Cell Res*., **263**, 202–208.
16. Papait,R., Pistore,C., Negri,D., Pecoraro,D., Cantarini,L. and Bonapace,I.M. (2007) Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol. Biol. Cell*, **18**, 1098–1106.
17. Bostick,M., Kim,J.K., Esteve,P.-O., Clark,A., Pradhan,S. and Jacobsen,S.E. (2007) UHRF1 plays a role in maintaining dna methylation in mammalian cells. *Science*, **317**, 1760–1764.
18. Sharif,J., Muto,M., Takebayashi,S., Suetake,I., Iwamatsu,A., Endo,T.A., Shinga,J., Mizutani-Koseki,Y., Toyoda,T., Okamura,K. *et al*. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908–912.
19. Achour,M., Jacq,X., Ronde,P., Alhosin,M., Charlot,C., Chataigneau,T., Jeanblanc,M., Macaluso,M., Giordano,A., Hughes,A.D. *et al*. (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*, **27**, 2187–2197.
20. Arita,K., Ariyoshi,M., Tochio,H., Nakamura,Y. and Shirakawa,M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
21. Avvakumov,G.V., Walker,J.R., Xue,S., Li,Y., Duan,S., Bronner,C., Arrowsmith,C.H. and Dhe-Paganon,S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
22. Qian,C., Li,S., Jakoncic,J., Zeng,L., Walsh,M.J. and Zhou,M.M. (2008) Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J. Biol. Chem*., **283**, 34490–34494.
23. Unoki,M., Nishidate,T. and Nakamura,Y. (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene*, **23**, 7601–7610.
24. Kim,J.K., Esteve,P.O., Jacobsen,S.E. and Pradhan,S. (2009) UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res*., **37**, 493–505.
25. Citterio,E., Papait,R., Nicassio,F., Vecchi,M., Gomiero,P., Mantovani,R., Di Fiore,P.P. and Bonapace,I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell Biol*., **24**, 2526–2535.
26. Karagianni,P., Amazit,L., Qin,J. and Wong,J. (2008) ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell Biol*., **28**, 705–717.
27. Papait,R., Pistore,C., Grazini,U., Babbio,F., Cogliati,S., Pecoraro,D., Brino,L., Morand,A.L., Dechampesme,A.M., Spada,F. *et al*. (2008) The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol. Biol. Cell*, **19**, 3554–3563.
28. Schermelleh,L., Spada,F., Easwaran,H.P., Zolghadr,K., Margot,J.B., Cardoso,M.C. and Leonhardt,H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods*, **2**, 751–756.
29. Sporbert,A., Domaing,P., Leonhardt,H. and Cardoso,M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res*., **33**, 3521–3528.

Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

30. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
31. Ko,J.K. and Ma,J. (2005) A rapid and efficient PCR-based mutagenesis method applicable to cell physiology study. *Am. J. Physiol. Cell Physiol.*, **288**, C1273–1278.
32. Bonapace,I.M., Latella,L., Papait,R., Nicassio,F., Sacco,A., Muto,M., Crescenzi,M. and Di Fiore,P.P. (2002) Np95 is regulated by E1A during mitotic reactivation of terminally differentiated cells and is essential for S phase entry. *J. Cell Biol.*, **157**, 909–914.
33. Frauer,C. and Leonhardt,H. (2009) A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res.*, **37**, e22.
34. Rothbauer,U., Zolghadr,K., Muyldermans,S., Schepers,A., Cardoso,M.C. and Leonhardt,H. (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell Proteomics*, **7**, 282–289.
35. Adams-Cioaba,M.A. and Min,J. (2009) Structure and function of histone methylation binding proteins. *Biochem. Cell Biol.*, **87**, 93–105.
36. Jacobs,S.A. and Khorasanizadeh,S. (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, **295**, 2080–2083.
37. Gouet,P., Courcelle,E., Stuart,D.I. and Metoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
38. DeLano,W.L. (2008) The PyMOL Molecular Graphics System. *DeLano Scientific LLC*. Palo Alto, CA, USA.

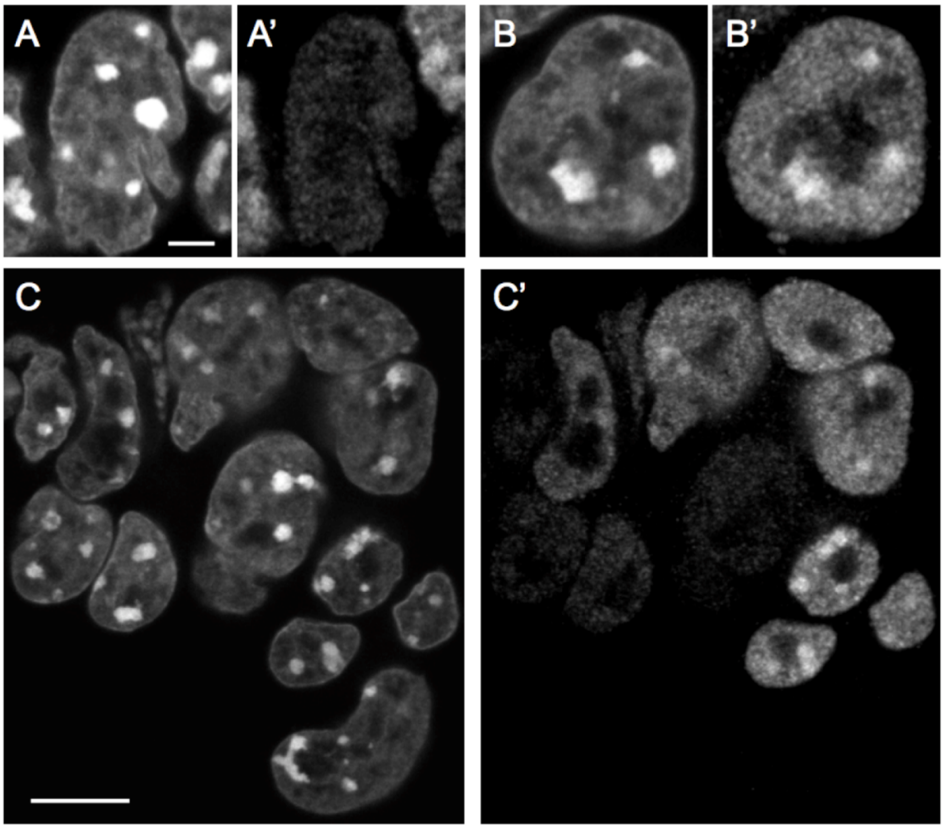
Downloaded from http://nar.oxfordjournals.org at Universitatsbibliothek Munchen on April 20, 2010

SUPPLEMENTARY INFORMATION

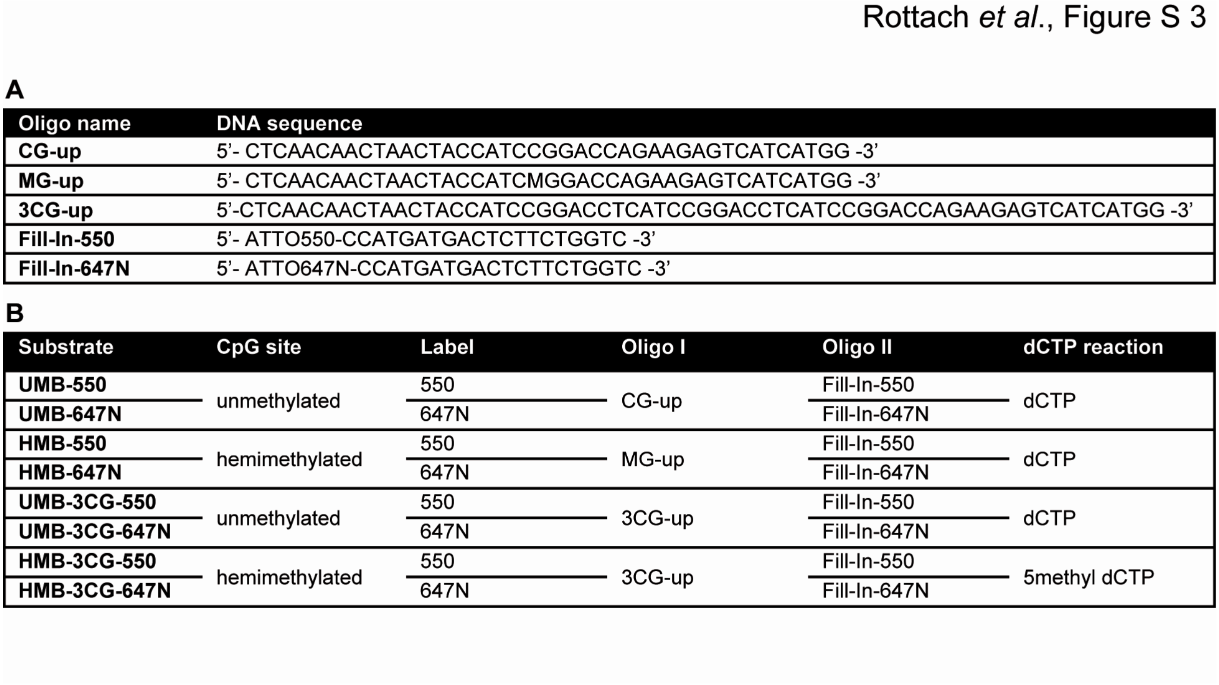


Supplementary Figure 1. Nuclear localization, FRAP kinetics and DNA binding specificity for an Np95 construct C-terminally fused to GFP (Np95-GFP, respectively). (**A**) Schematic drawing of Np95-GFP. Abbreviations are as in Fig. 1. (**B**) Representative images from FRAP experiments for Np95-GFP transiently expressed in wt, *dnmt1*$^{-/-}$ and TKO J1 ESCs as indicated on the left. Images show confocal mid-sections of nuclei before (Pre) and at the indicated time points after bleaching (Pb) half of the nucleus. Bleached areas are outlined. Arrowheads mark pericentric heterochromatin. Bars, 5 µm. **(C)** FRAP kinetics of Np95-GFP in J1 ESCs with different genetic backgrounds as shown in B and Fig. 1C. Kinetics of GFP-Dnmt1 is shown for comparison. Mobile and immobile fractions are indicated on the right. Values represent mean ± SEM. Note that the kinetics are similar to those shown for GFP-Np95 in Fig. 1C and that there is no significant difference in cells with different genetic backgrounds.

Rottach *et al.*, Figure S 2



Supplementary Figure 2. Variable expression levels and localization of Np95 in TKO cells. TKO cells were stained with DAPI (A, B and C) and an anti-Np95 antibody (A', B' and C'). A-A' and B-B' show examples of cells with very low and high Np95 levels, respectively. In B' accumulation of endogenous Np95 at chromocenters is evident. C and C' show a field containing cells with very different Np95 levels and degrees of Np95 accumulation at chromocenters. Scale bars are 3 µm (A-B') and 10 µm (C and C').



Rottach *et al.*, Figure S 3

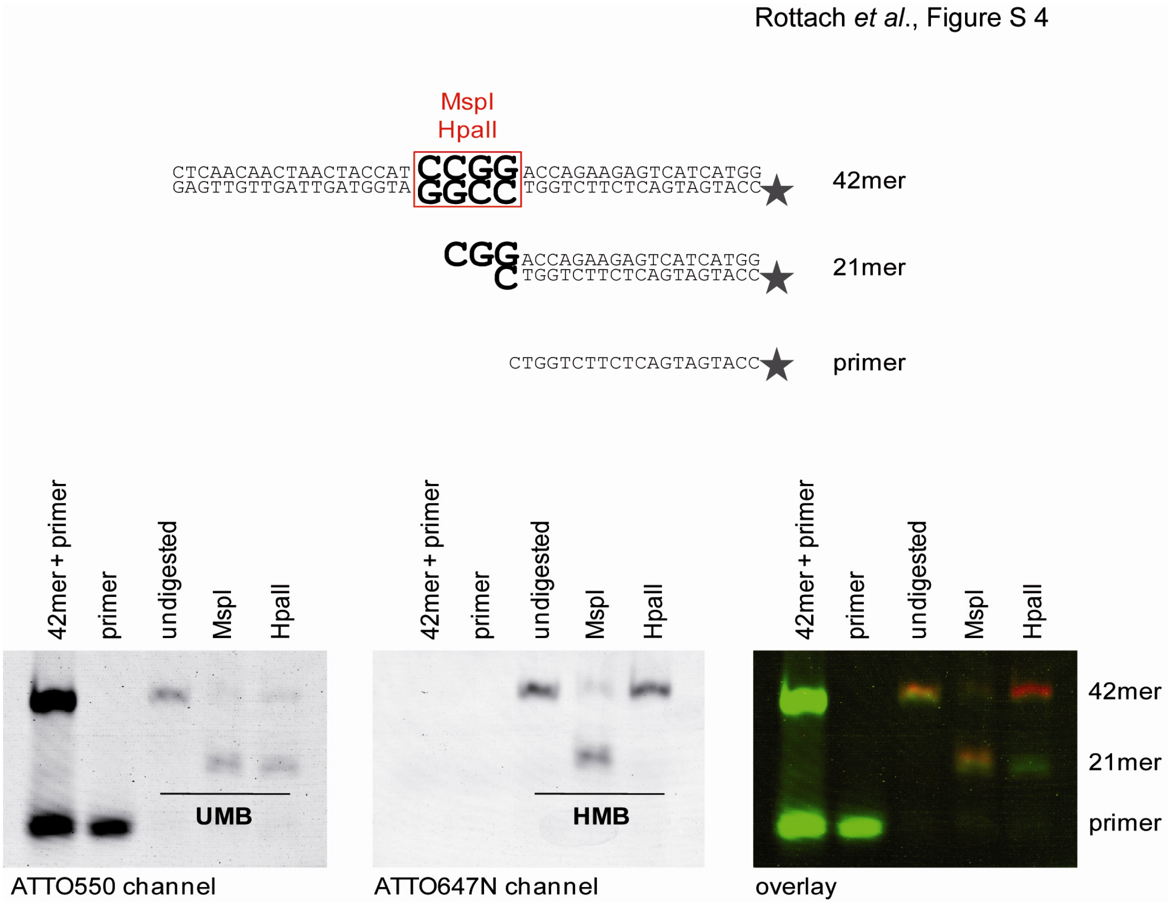
A

Oligo name	DNA sequence
CG-up	5’- CTCAACAACTAACTACCATCCGGACCAGAAGAGTCATCATGG -3’
MG-up	5’- CTCAACAACTAACTACCATCMGGACCAGAAGAGTCATCATGG -3’
3CG-up	5’-CTCAACAACTAACTACCATCCGGACCTCATCCGGACCTCATCCGGACCAGAAGAGTCATCATGG -3’
Fill-In-550	5’- ATTO550-CCATGATGACTCTTCTGGTC -3’
Fill-In-647N	5’- ATTO647N-CCATGATGACTCTTCTGGTC -3’

B

Substrate	CpG site	Label	Oligo I	Oligo II	dCTP reaction
UMB-550	unmethylated	550	CG-up	Fill-In-550	dCTP
UMB-647N		647N		Fill-In-647N	
HMB-550	hemimethylated	550	MG-up	Fill-In-550	dCTP
HMB-647N		647N		Fill-In-647N	
UMB-3CG-550	unmethylated	550	3CG-up	Fill-In-550	dCTP
UMB-3CG-647N		647N		Fill-In-647N	
HMB-3CG-550	hemimethylated	550	3CG-up	Fill-In-550	5methyl dCTP
HMB-3CG-647N		647N		Fill-In-647N	

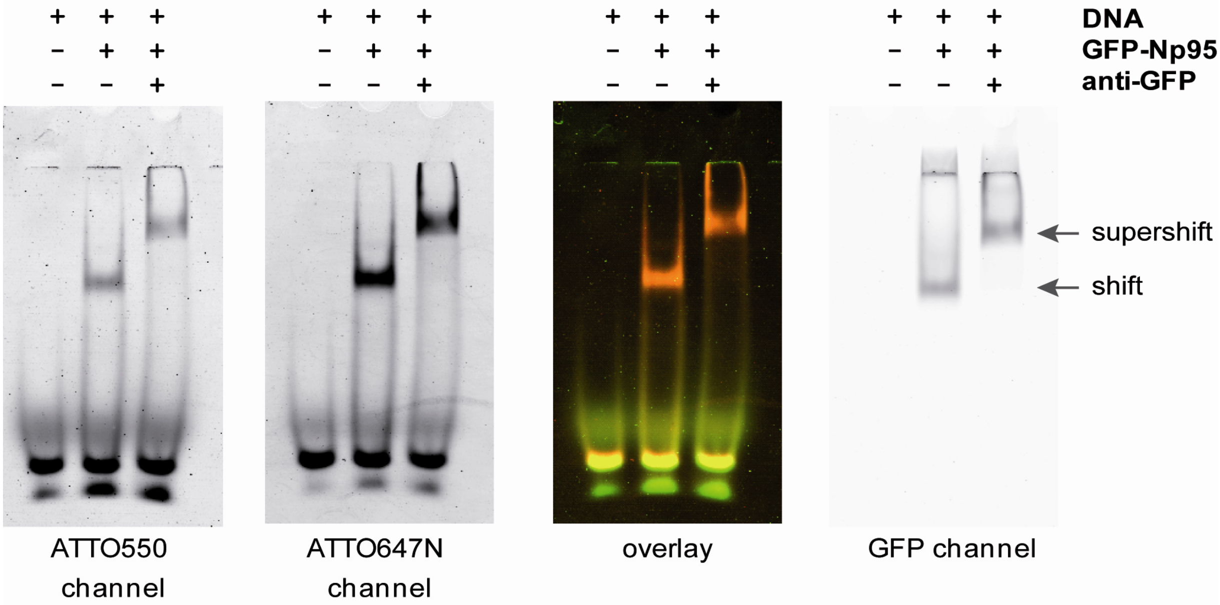
Supplementary Figure 3. Oligo design for the *in vitro* DNA binding assay (**A**) DNA oligonucleotides used for the preparation of double stranded probes for *in vitro* DNA binding assays. M: 5-methyl-cytosine. (**B**) Description of double stranded probes used for *in vitro* DNA binding assays. Name, status of the central CpG site, fluorescent label, as well as DNA oligonucleotides and nature of the dCTP used in the primer extension reaction are specified. By using a control set of two probes with identical sequence but different fluorescent labels we observed effects due to probe preparation and/or unspecific binding of ATTO dyes (data not shown). The values obtained from the control set were used to normalize every probe / protein pair.


Rottach et al., Figure S 4
MspI
HpaII
CTCAACAACTAACTACCAT CCGG ACCAGAAGAGTCATCATGG
GAGTTGTTGATTGATGGTA GGCC TGGTCTTCTCAGTAGTACC ★
42mer
CGG ACCAGAAGAGTCATCATGG
C TGGTCTTCTCAGTAGTACC ★
21mer
CTGGTCTTCTCAGTAGTACC ★
primer
42mer + primer
primer
undigested
MspI
HpaII
UMB
ATTO550 channel
42mer + primer
primer
undigested
MspI
HpaII
HMB
ATTO647N channel
42mer + primer
primer
undigested
MspI
HpaII
42mer
21mer
primer
overlay

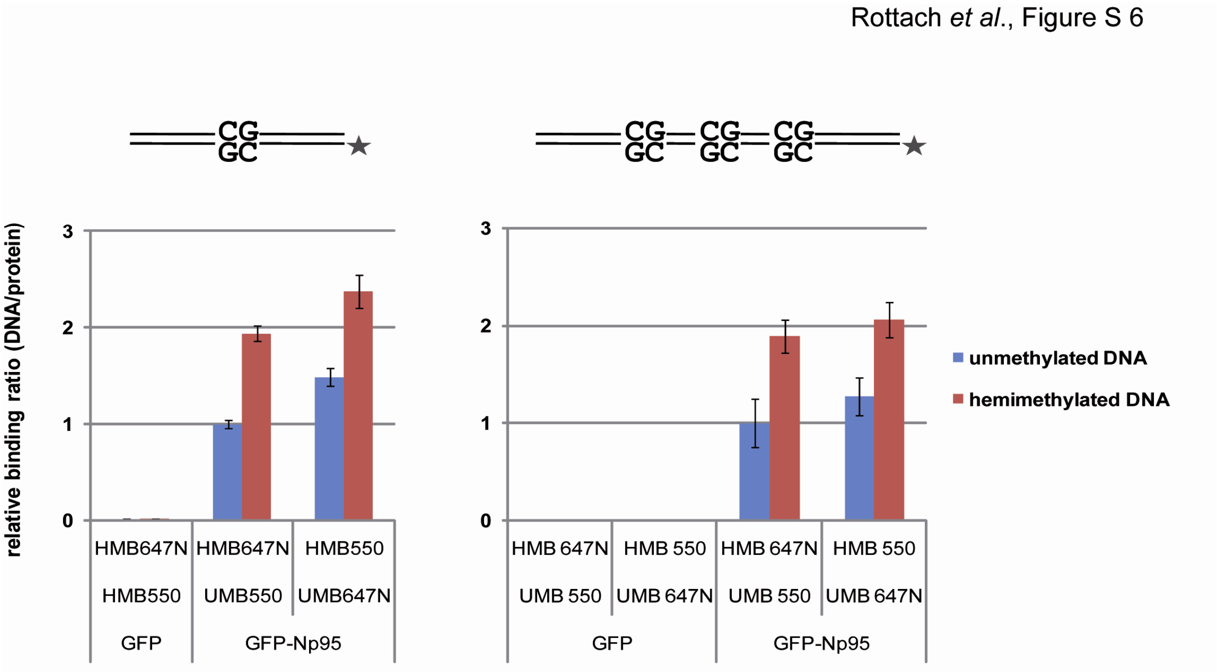
Supplementary Figure 4. Quality control of un- and hemi-methylated DNA substrates

Unmethylated and hemimethylated DNA substrates (UMB550 and HMB647N, respectively) were digested with MspI or HpaII and analyzed by 15 % non-denaturing PAGE for CG methylation. DNA substrates were detected via their fluorescent ATTO label using the Typhoon Trio scanner. Note that the unmethylated DNA substrate is digested by both MspI and HpaII, whereas the hemimethylated substrate is cut by MspI, but not by the methylation sensitive HpaII. Sequences of the double stranded probes before (42mer) and after cut (21mer) as well as the unextended primer are displayed above. Enzyme recognition motifs are boxed and asterisks represent fluorescent ATTO label.

Rottach *et al.*, Figure S 5

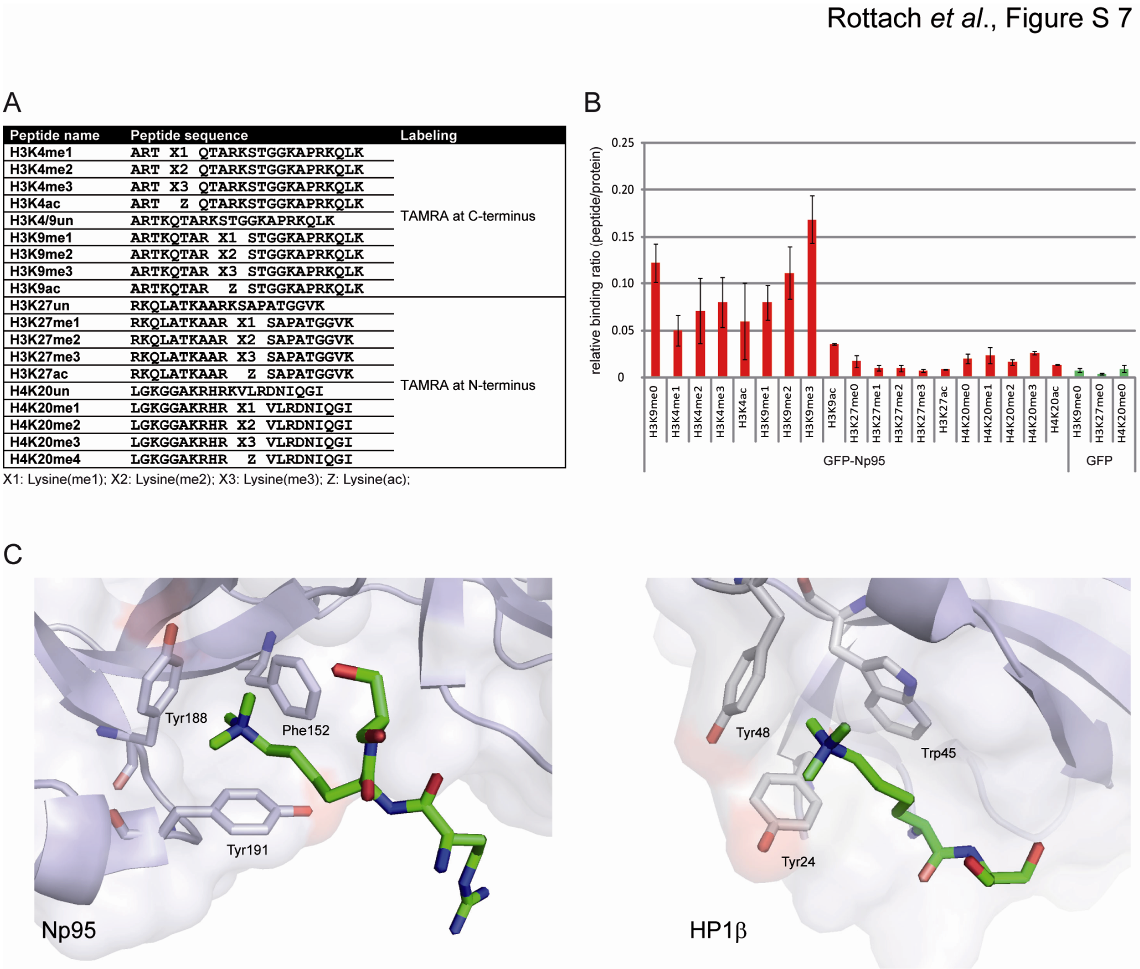

+ + +
− + +
− − +
DNA
GFP-Np95
anti-GFP
supershift
shift
ATTO550 channel
ATTO647N channel
overlay
GFP channel

Supplementary Figure 5. Electrophoretic mobility shift and supershift assays with GFP-Np95. Un- and hemimethylated DNA substrates (1 pmol UMB550 and HMB647N, respectively) were incubated with 0.6 pmol purified GFP-Np95 and 0.4 pmol GFP-antibody. Samples were subjected to a 3.5 % non-denaturing PAGE and analyzed by the Typhoon Trio scanner to detect ATTO550 (unmethylated substrate), ATTO647N (hemimethylated substrate) and green fluorescence (GFP). Note that the DNA:GFP-Np95:GFP-antibody complex is shifting higher than the DNA:GFP-Np95 complex (arrows).



Supplementary Figure 6. DNA binding specificity of GFP-Np95

The sequence specific DNA binding activity of Np95 was tested with an *in vitro* binding assays using GFP-Np95 with un- and hemimethylated substrates in direct competition. The DNA substrates included either one (left) or three (right) CG sites. Note that regardless of the attached fluorescent label (indicated by asterisks) and number of CG sites the hemimethylated DNA substrates are preferentially bound (1.6- to 1.9-fold). Shown are the means ± SEM from two (left) or four (right) independent experiments.



Rottach *et al.*, Figure S 7

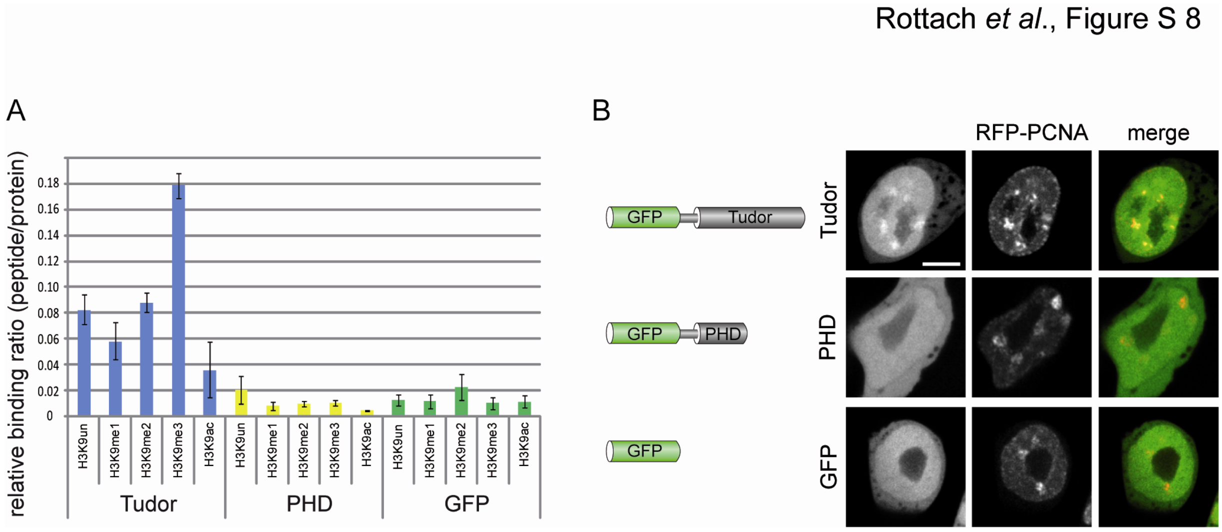
A

Peptide name	Peptide sequence	Labeling
H3K4me1	ART X1 QTARKSTGGKAPRKQLK	TAMRA at C-terminus
H3K4me2	ART X2 QTARKSTGGKAPRKQLK	
H3K4me3	ART X3 QTARKSTGGKAPRKQLK	
H3K4ac	ART Z QTARKSTGGKAPRKQLK	
H3K4/9un	ARTKQTARKSTGGKAPRKQLK	
H3K9me1	ARTKQTAR X1 STGGKAPRKQLK	
H3K9me2	ARTKQTAR X2 STGGKAPRKQLK	
H3K9me3	ARTKQTAR X3 STGGKAPRKQLK	
H3K9ac	ARTKQTAR Z STGGKAPRKQLK	
H3K27un	RKQLATKAARKSAPATGGVK	TAMRA at N-terminus
H3K27me1	RKQLATKAAR X1 SAPATGGVK	
H3K27me2	RKQLATKAAR X2 SAPATGGVK	
H3K27me3	RKQLATKAAR X3 SAPATGGVK	
H3K27ac	RKQLATKAAR Z SAPATGGVK	
H4K20un	LGKGGAKRHRKVLRDNIQGI	
H4K20me1	LGKGGAKRHR X1 VLRDNIQGI	
H4K20me2	LGKGGAKRHR X2 VLRDNIQGI	
H4K20me3	LGKGGAKRHR X3 VLRDNIQGI	
H4K20me4	LGKGGAKRHR Z VLRDNIQGI	

X1: Lysine(me1); X2: Lysine(me2); X3: Lysine(me3); Z: Lysine(ac);

B
relative binding ratio (peptide/protein)
0.25
0.20
0.15
0.10
0.05
0
H3K9me0
H3K4me1
H3K4me2
H3K4me3
H3K4ac
H3K9me1
H3K9me2
H3K9me3
H3K9ac
H3K27me0
H3K27me1
H3K27me2
H3K27me3
H3K27ac
H4K20me0
H4K20me1
H4K20me2
H4K20me3
H4K20ac
H3K9me0
H3K27me0
H4K20me0
GFP-Np95
GFP
C
Tyr188
Phe152
Tyr191
Np95
Tyr48
Trp45
Tyr24
HP1β

Supplementary Figure 7. Histone tail binding of Np95 and HP1ß (**A**) Amino acid sequences of TAMRA-labeled histone tail peptides used for the peptide binding assay. (**B**) Histone H3 and H4-tail binding specificity of GFP-Np95 *in vitro*. Ratios of bound TAMRA-labeled peptide over bound GFP fusion were determined and normalized to the ratio of H3K4/9un peptide over GFP-Np95. GFP was used as negative control. Shown are means ± SEM from six independent experiments. (**C**) Structural comparison of the H3K9me3-binding aromatic cages formed by the tandem Tudor domain of Np95 (left) and the chromodomain of HP1ß (right, PDB 1kne). In these structures, only Arg8-Lys9-Ser10 and Lys9-Ser10 from histone H3 are resolved peptides, respectively (green stick models). The image was generated with PyMOL (1).



Supplementary Figure 8. Histone tail binding and subcellular distribution of PHD and Tudor domain of Np95 (**A**) Histone H3 N-terminal tail binding specificity of GFP-Tudor, GFP-PHD and GFP *in vitro*. Shown are fluorescence intensity ratios of bound probe / bound GFP fusion. GFP was used as negative control. Shown are means ± SEM from four to six independent experiments. Only the tandem Tudor domain shows preferential binding of H3K9 trimethylated histone tails. (**B**) Schematic representation of the analyzed Np95 constructs. All constructs were N-terminal GFP fusions (left panel). Confocal mid sections of living *np95*$^{-/-}$ ESCs transiently expressing the indicated Np95 fusion constructs and RFP-PCNA as S phase marker (left and mid panels). Merged images are displayed on the right. Bars, 5 µm. Only the GFP-Tudor fusion protein showed slight enrichment at pericentric heterochromatin.

References:

1. DeLano, W.L. (2002) *The PyMOL User's Manual*.

2.5 SINGLE MOLECULE FLUORESCENCE SPECTROSCOPY OF DNMT1:DNA COMPLEXES

SINGLE MOLECULE FLUORESCENCE SPECTROSCOPY OF DNMT1:DNA COMPLEXES

INTRODUCTION AND AIM OF THE PROJECT

In mammalian cells, Dnmt1 is known to copy the DNA methylation pattern from parental to daughter strand during DNA replication, thereby maintaining an important epigenetic mark. Crucial for this process is not only Dnmt1's intrinsic preference for hemimethylated substrates, but also various interactions of Dnmt1 with different cellular structures and cofactors at different cell-cycle and developmental stages. Dnmt1 is a large and complex enzyme, presumably with a very dynamic structure, and the mechanistic regulation of Dnmt1's function is still poorly understood.

The molecular 3D structure of Dnmt1 is unknown, even though a homology model has been reported for the catalytic domain of Dnmt1 comprising the C-terminal third of the full-length protein (Siedlecki et al., 2003). This theoretical model was based on sequence comparison with prokaryotic methyltransferases, which contain the same conserved motifs that are necessary for the catalysis of the methyl transfer reaction. However, an inhibitor designed on the basis of this homology model and directed against the catalytic center of Dnmt1 was only functional for the prokaryotic enzymes (RG108, unpublished data). Moreover, it is well established that the C-terminal domain of Dnmt1 by itself is catalytically inactive, and needs interaction with the N-terminal part of the enzyme for allosteric activation, thus involving structural changes. These data indicate significant structural differences between prokaryotic DNA methyltransferases and the catalytic domain of Dnmt1. Of the N-terminal two thirds of the enzyme, only an X-ray structure of the TS domain has been reported (PBD: 3EPZ). Interestingly, this domain has been proposed to mediate, besides heterochromatin interaction (Easwaran et al., 2004), dimerization of the enzyme (Fellinger et al., 2009). Dimerization of Dnmt1 would allow another level of regulation and the interesting question arises whether Dnmt1 has to dimerize for catalysis of the methyl transfer reaction.

To address this question, we investigated Dnmt1:DNA complexes at the single molecule level. Using Dnmt1 as GFP fusion and ATTO labeled DNA substrates, we prepared Dnmt1:DNA complexes, purified them and analyzed them with fluorescence intensity distribution analysis (FIDA) and fluorescence cross-correlation spectroscopy (FCCS). All spectroscopic measurements were performed by Matthias Höller in cooperation with the group of Professor Don C. Lamb. We found that Dnmt1 is present as a monomer under these experimental conditions and provide first evidence that Dnmt1 covalently binds two DNA substrates at the same time. This surprising observation clearly requires confirmation by other methods and further investigation, e.g. specification of the second DNA binding site. If true, these findings will provide entirely new insights into the regulation of Dnmt1.

RESULTS

GFP-Dnmt1 seems to be a monomer under our investigation conditions

It was suggested that Dnmt1 forms a stable dimer in head-to-head orientation via a bipartite interaction surface located within the TS domain (Fellinger et al., 2009). However, nothing is known about the affinity of this interaction and both single molecule spectroscopy methods used in this study require highly diluted samples. To find out whether Dnmt1 forms a monomer or a dimer at the concentration range of our experimental conditions (1-10 nM), we used GFP-Dnmt1 and performed a fluorescence intensity distribution analysis (FIDA, (Kask et al., 1999)). For FIDA, the fluorescence intensity of a highly diluted fluorophore solution is recorded over time with a confocal microscope. The photons detected within the very small focal volume for a time interval of given length (bin size) are then counted and represented in a photon counting histogram (Figure 9A). The resulting data can be fitted in order to provide the concentration and molecular brightness of the single fluorophores. Thus, by comparison with (monomeric enhanced) GFP, one can determine the multimerization state of GFP-Dnmt1 based on its molecular brightness.

GFP and GFP-Dnmt1 were expressed in HEK293T cells and one-step purified via the GFP tag as described before ((Frauer and Leonhardt, 2009), chapter 2.3, supplementary methods). Briefly, cells were lysed and incubated with agarose beads that are functionalized with the His-tagged GFP binding protein (GBP) via a Ni-NTA linker. Subsequent to several washing steps, GFP or GFP-Dnmt1 was eluted with imidazole. Notably, at this step, the GBP is eluted together and in a stable complex with GFP, and it has been shown that the GFP fluorescence is enhanced upon GBP binding. Therefore, we also analyzed His-tagged GFP expressed and purified from *E. coli* without using GBP functionalized beads ((Frauer and Leonhardt, 2009), chapter 2.1). Furthermore, we established a protocol for preparation and purification of covalent Dnmt1:DNA complexes. We prepared DNA trapping substrates containing the mechanism-based inhibitor 5-aza-cytosine at a central hemimethylated CpG site as described (Frauer and Leonhardt, 2009), chapter 2.1). The DNA substrates were 42 base pairs in length and labeled with ATTO550 or ATTO700. Upon incubation of immobilized GFP-Dnmt1 with trapping DNA substrates in the presence of the cofactor S-adenosine-L-methionine (AdoMet), an irreversible covalent bond is formed between enzyme and DNA, involving the cysteine of the catalytic PC motif (C1229) and the 5-aza-cytosine of the CpG site, respectively. The covalent Dnmt1:DNA complexes were subsequently eluted from the beads as described above. Importantly, all samples were filtered prior to single molecule analyses in order to remove any residual beads within the solution, which would interfere with the measurements.

FIDA was performed for GBP:GFP, GBP:GFP-Dnmt1 and GBP:GFP-Dnmt1:DNA complexes as well as for GFP expressed and purified from *E. coli* as a reference ((Frauer and Leonhardt, 2009), Figure 9B). This analysis confirmed that GFP fluorescence is enhanced upon binding to the GBP ((Kirchhofer et al., 2009), chapter 2.2). Importantly, the factor of enhancement was determined for the first time at the single molecule level, revealing that the molecular brightness of GFP in complex with GBP is improved by a factor of 1.4. The molecular brightness of GFP-Dnmt1 is very similar to this value both in absence and presence of DNA in the complex. Thus, FIDA suggests that Dnmt1 is present as monomer under these experimental conditions, independent of complex formation with DNA.

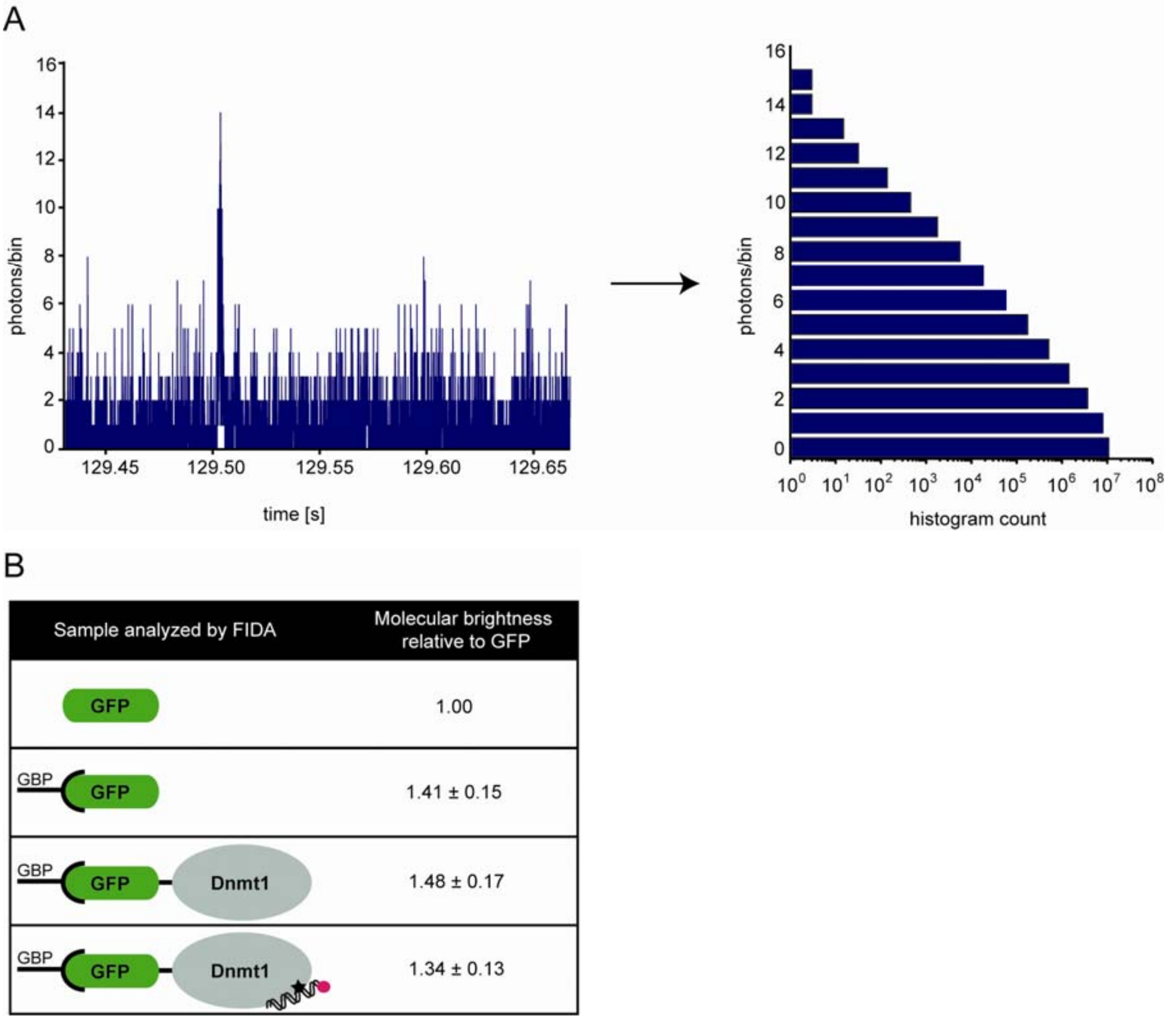


Figure 9. (A) Working principle of fluorescence intensity distribution analysis (FIDA). A fluorescence intensity trace of a fluorophore in solution is recorded with a carefully chosen bin size (left). Subsequently, a histogram of the number of photons within each time bin is calculated (right). These data can be fitted in order to retrieve the molecular brightness of the fluorophore. **(B)** Summary of FIDA for GFP and GFP-Dnmt1. The molecular brightness of GBP:GFP and GBP:GFP-Dnmt1 complexes was normalized to that of GFP. Note that the GFP fluorescence intensity is enhanced by binding to GBP, but not changed by fusion to Dnmt1. The molecular brightness of GBP:GFP-Dnmt1 is very similar to that of GBP:GFP, indicating Dnmt1 monomers with or without trapping substrate.

To provide a second line of evidence for the monomeric state of Dnmt1, we performed a control experiment using fluorescence cross-correlation spectroscopy (FCCS) analysis (Muller et al., 2005; Schwille et al., 1997). Similar to FIDA, this method records the fluorescence signal of highly diluted samples from a very small focal volume. For a single fluorescent species in solution, the autocorrelation function G(τ) describes the correlation of the fluorescence signal at time point t with the fluorescence signal at time point t+τ, giving information about both the relative mobility and the concentration of the fluorophore. Thereby, the relative mobility anti-correlates with the decay time (width of the curve) and the concentration relates to the inverse of the curve amplitude. If a solution contains different fluorescent species that can be unambiguously detected and identified using different laser and filter combinations, their fluorescence will only fluctuate simultaneously if they are part of the very same complex. In this case, the amplitude of the cross-correlation curve (divided and thus normalized by the autocorrelation amplitudes) is proportional to the fraction of co-diffusing species.

Thus, we co-expressed fusions of Dnmt1 with GFP and (monomeric) RFP and purified them either via the GFP tag as described above or analogously via the RFP tag using GBP or RBP, respectively. In case of stable dimer formation, one would expect a large fraction of co-migrating GFP and RFP and thus cross-correlation between GFP and RFP signals, represented by a cross-correlation curve amplitude that is not zero. But, confirming the results of FIDA, we could not detect any cross-correlation between the two fluorophores, excluding the formation of stable GFP-Dnmt1:RFP-Dnmt1 dimers under these experimental conditions (Figure 10). Notably, the concentration of the co-purified fusion enzyme was only around one eighth of the other, also indicating rather transient interaction between GFP-Dnmt1 and mRFP-Dnmt1 monomers than stable dimerization.

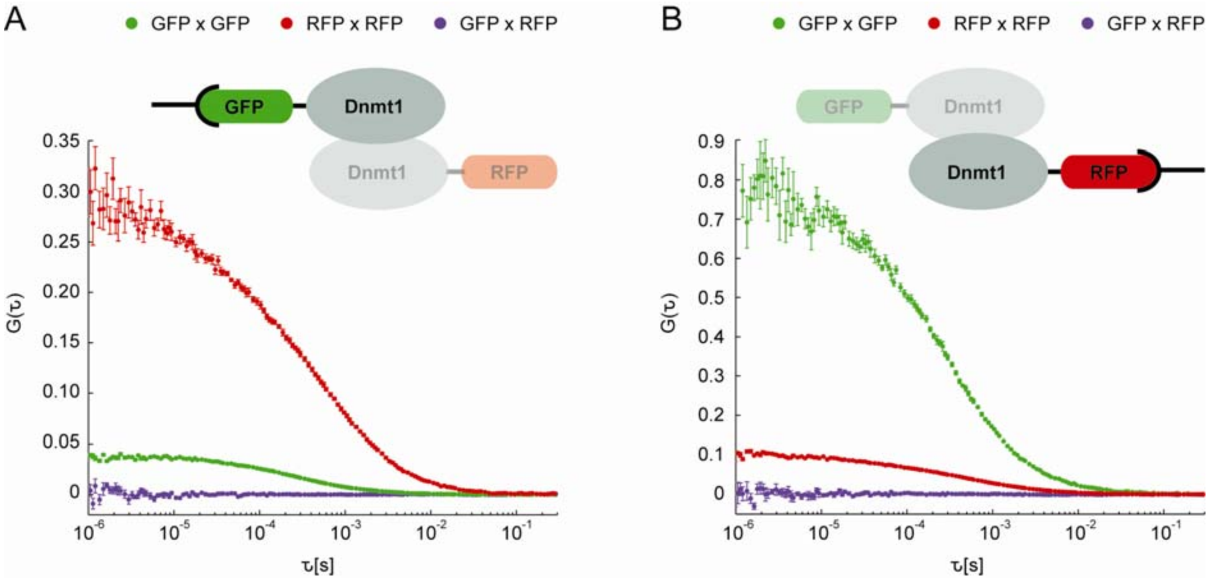


Figure 10. GFP-Dnmt1 and RFP-Dnmt1 do not form stable dimers under FCCS conditions. GFP-Dnmt1 and RFP-Dnmt1 were co-expressed in HEK293T cells and purified either via GBP **(A)** or RBP **(B)**. Fluorescence autocorrelation curves of GFP and RFP show a very low concentration of the co-purified fusion protein in both cases. Significantly, no cross-correlation between GFP and RFP could be observed.

GFP-Dnmt1 binds two DNA substrates covalently at the same time

Next, we aimed at determining the stoichiometry of the Dnmt1:DNA complexes, in other words we addressed the question of how many DNA substrates are bound per Dnmt1 monomer. Therefore, we prepared covalent complexes by incubating Dnmt1 simultaneously with DNA trapping substrates labeled either with ATTO550 or ATTO700. FCCS analysis of the complexes was performed for all possible fluorophore combinations subsequent to HPLC purification (Figure 11).

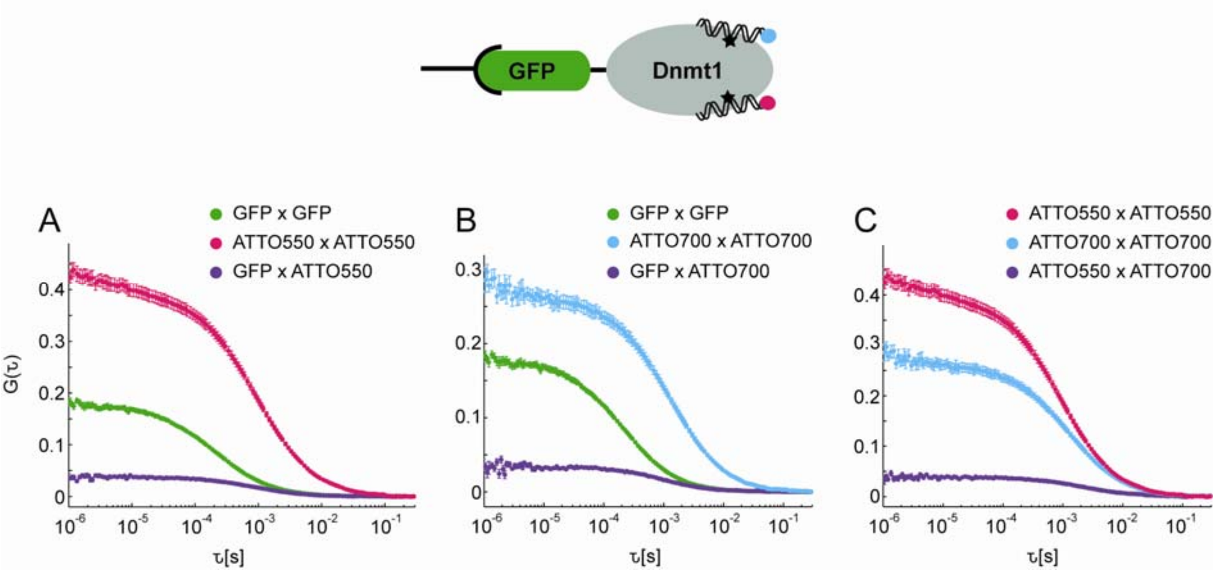


Figure 11. FCCS analysis of HPLC-purified GFP-Dnmt1:DNA complexes. GFP-Dnmt1 was simultaneously incubated with two differently labeled DNA trapping substrates (ATTO550 or ATTO700, respectively). As revealed by FCCS analysis, cross-correlation is present for GFP-Dnmt1 with ATTO550-DNA **(A)**, GFP-Dnmt1 with ATTO700-DNA **(B)**, and ATTO550-DNA with ATTO700-DNA **(C)**. Autocorrelation curves are shown for relative amplitude comparison. Stars in the complex model indicate 5-aza-cytosine.

Since we know that Dnmt1 can form covalent complexes with DNA substrates containing 5-aza-cytosine at a CpG site, we expected a significant fraction of co-diffusing ATTO550/GFP and ATTO700/GFP. Accordingly, we observed a clear cross-correlation between GFP and both ATTO dyes (Figure 11A and 11B, respectively). Additionally and very surprisingly, we also observed cross-correlation between ATTO550 and ATTO700 (Figure 11C), suggesting that both ATTO550 and ATTO700 labeled DNA can be bound to the same GFP-Dnmt1 molecule, and thus a protein:DNA complex stoichiometry of 1:2. Although it is known that Dnmt1 has more than one DNA binding site and especially the N-terminal domain has been suggested to bind DNA via several motifs (Araujo et al., 2001; Chuang et al., 1996; Fatemi et al., 2001; Pradhan and Esteve, 2003; Suetake et al., 2006), we did not expect non-covalent binding in this low concentration range. For example, we recently determined the dissociation constant for interaction of full-length Dnmt1 with hemimethylated DNA to approximately 100 nM, which is at least one order of magnitude higher than the concentration range of our singe molecule experiments (Frauer and Leonhardt, 2009)).

Thus, in order to exclude the possibility that covalent complexes of GFP-Dnmt1 with the single DNA substrates aggregate subsequent to the enzymatic reaction, we first prepared both GFP-Dnmt1:ATTO550-DNA and GFP-Dnmt1:ATTO700-DNA complexes separately and then pooled them afterwards. Notably, we observed cross-correlation between GFP and ATTO550 as well as between GFP and ATTO700 also in this case, but not between the two ATTO labels (Figure 12). This means that the cross-correlation shown in Figure 11C is not an artifact of post-reaction aggregation.

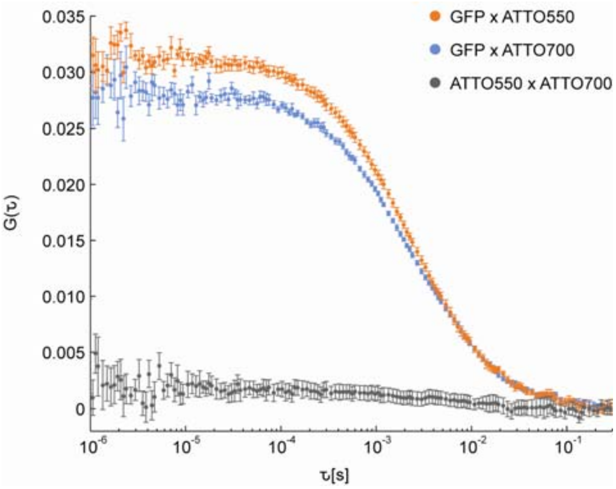


Figure 12. Cross-correlation between the two DNA labels is not due to aggregation subsequent to the enzymatic reaction. GFP-Dnmt1 was separately incubated with either ATTO550- or ATTO700-labeld trapping substrate, and subsequently both samples were pooled. Cross-correlation between GFP and both ATTO labels is clearly detectable, whereas the cross-correlation between ATTO550 and ATTO700 is zero.

Although we could exclude that aggregation occurs after the enzymatic reaction, we tested whether aggregation occurs during the enzymatic reaction, which would likewise mislead the interpretation of the FCCS results. Therefore, we prepared covalent GFP-Dnmt1:ATTO550-DNA:ATTO700-DNA complexes as described above and subjected them to size-exclusion chromatography via HPLC (using a Superose 6 PC 3.2/30 column) prior to FCCS analysis of all protein and/or GFP containing fractions (Figure 13). As a reference, we also applied a sample of purified GFP-Dnmt1 to the chromatographic column (Figure 13A, grey curve). This sample shows a peak for absorbance at 280 nm (aromatic amino acids) between fractions A12 and B12, which corresponds to an elution volume of about 1.55 ml and contains the full-length GFP-Dnmt1 as confirmed by SDS-PAGE analysis. Absorbance peaks at later elution times most likely correspond to degradation products. For the sample of the Dnmt1:DNA complex, we recorded in addition to absorbance at 280 nm (black curve), absorbance at 488 nm (GFP, green curve) and 700 nm (ATTO700, blue curve). Significantly, all three curves show a dominant peak between fractions A12 and B12, which also overlaps with the GFP-Dnmt1 peak of the reference sample. First, this confirms that GFP-Dnmt1 and ATTO700-DNA co-migrate through the column in a complex and second, this indicates that complex formation does not alter the migration behavior and thus the multimerization state of Dnmt1.

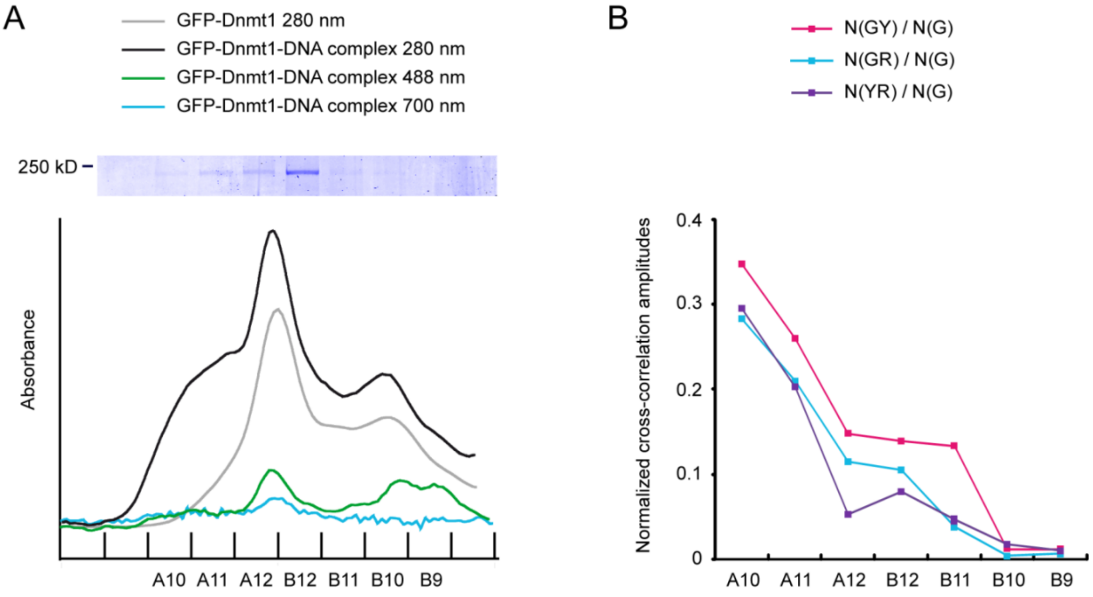


Figure 13. HPLC and FCCS analysis confirmed that GFP-Dnmt1 forms a complex with two DNA substrates simultaneously. **(A)** GFP-Dnmt1 was expressed in HEK293T cells, one-step purified via the GFP nanotrap and subsequently incubated with two DNA trapping substrates having identical sequences but different fluorescent labels (ATTO5500 or ATTO700, respectively). Samples of GFP-Dnmt1 and the GFP-Dnmt1:DNA complex were analyzed by size-exclusion chromatography, detecting absorbance at 280, 488 and 700 nm. Aliquots of fractions A9-B8 were subjected to 6 % SDS-PAGE and Coomassie staining. **(B)** FCCS analysis was performed for fractions A10-B9, and cross-correlation amplitudes for GFP/ATTO550 (GY), GFP/ATTO700 (GR) and ATTO550/ATTO700 (YR) were normalized by the autocorrelation amplitude of GFP (G). The plot shows that GFP-Dnmt1 is able to form a stable complex with both DNA substrates simultaneously.

Notably, the 280 nm absorbance peak of the Dnmt1:DNA complex sample has a shoulder in contrast to the peak of the Dnmt1 sample (Figure 13A). We think that this shoulder represents a fraction of higher molecular weight aggregates. Indeed, FCCS analysis of all protein containing elution fractions of the Dnmt1:DNA sample supports this hypothesis (Figure 13B). When normalizing the cross-correlation amplitudes by the autocorrelation curve of GFP and plotting these values for the different elution fractions, the highest cross-correlation values of all possible fluorophore combinations are observed for the fractions corresponding to the shoulder. However, significantly, cross-correlation between all three fluorophore pairs of the GFP-Dnmt1:ATTO550-DNA:ATTO700-DNA complex can also be observed in fractions A12 and B12, in contrast to later fractions. This means although there is a small fraction of high molecular weight protein/DNA aggregates, cross-correlation is also observed for the fractions corresponding to non-aggregated GFP-Dnmt1:DNA complex. Later elution fractions show no cross-correlation and most likely correspond to degradation products. Notably, the FCCS results in Figure 11 were performed with HPLC purified Dnmt1:DNA complexes. In conclusion, these data strongly suggest that GFP-Dnmt1 can bind two trapping substrates at the same time and that the multimerization state of Dnmt1 is identical in both the absence of DNA and in covalent complex with DNA trapping substrate.

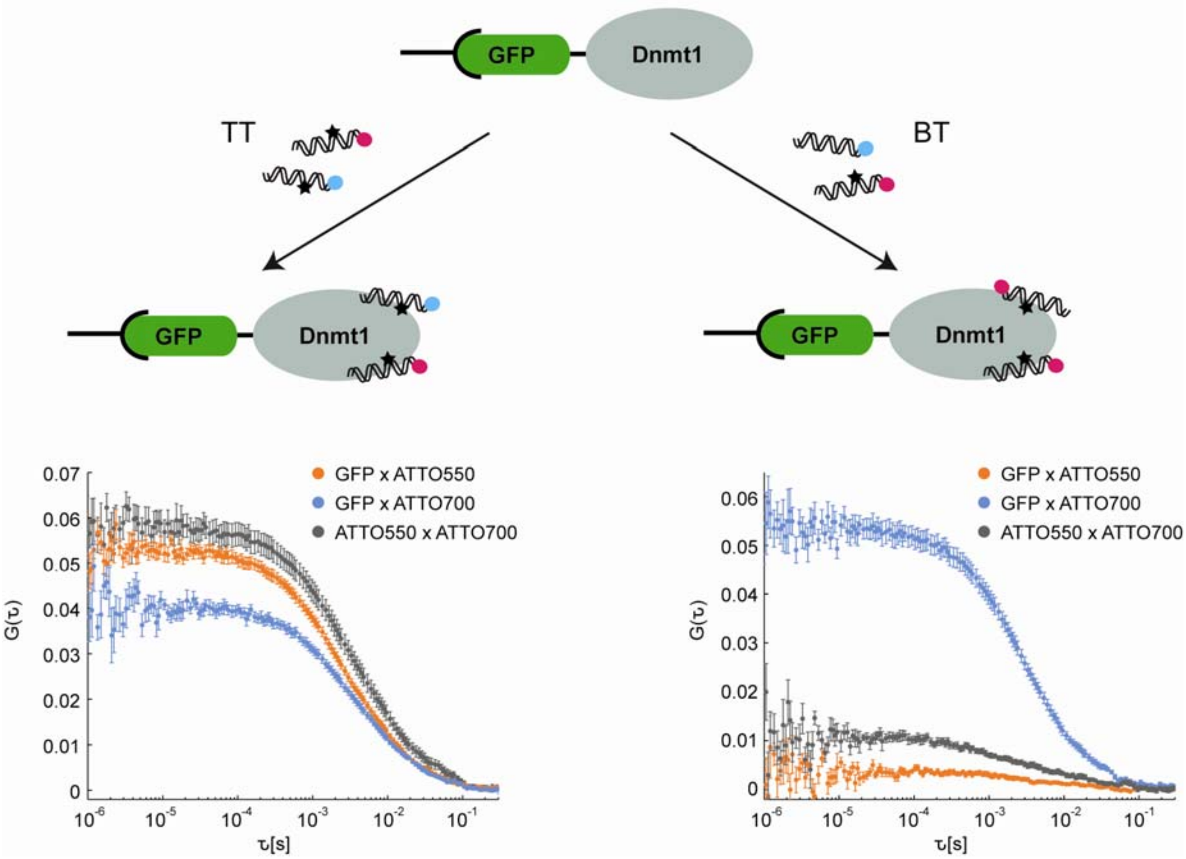


Figure 14. FCCS analysis of HPLC-purified covalent GFP-Dnmt1:DNA complexes revealed that two DNA substrates can bind covalently to GFP-Dnmt1 at the same time. GFP-Dnmt1 was simultaneously incubated with two different DNA trapping substrates labeled with either ATTO550 or ATTO700 (TT), or, with a trapping substrate labeled with ATTO550 and a canonical binding substrate labeled with ATTO700 (BT). Significant cross-correlation between the two different DNA labels was observed exclusively in the first case. Importantly, this cross-correlation was resistant to denaturation of the complex and not observed for a catalytically inactive mutant of Dnmt1 (Table 15). Stars in the complex models indicate 5-aza-cytosine.

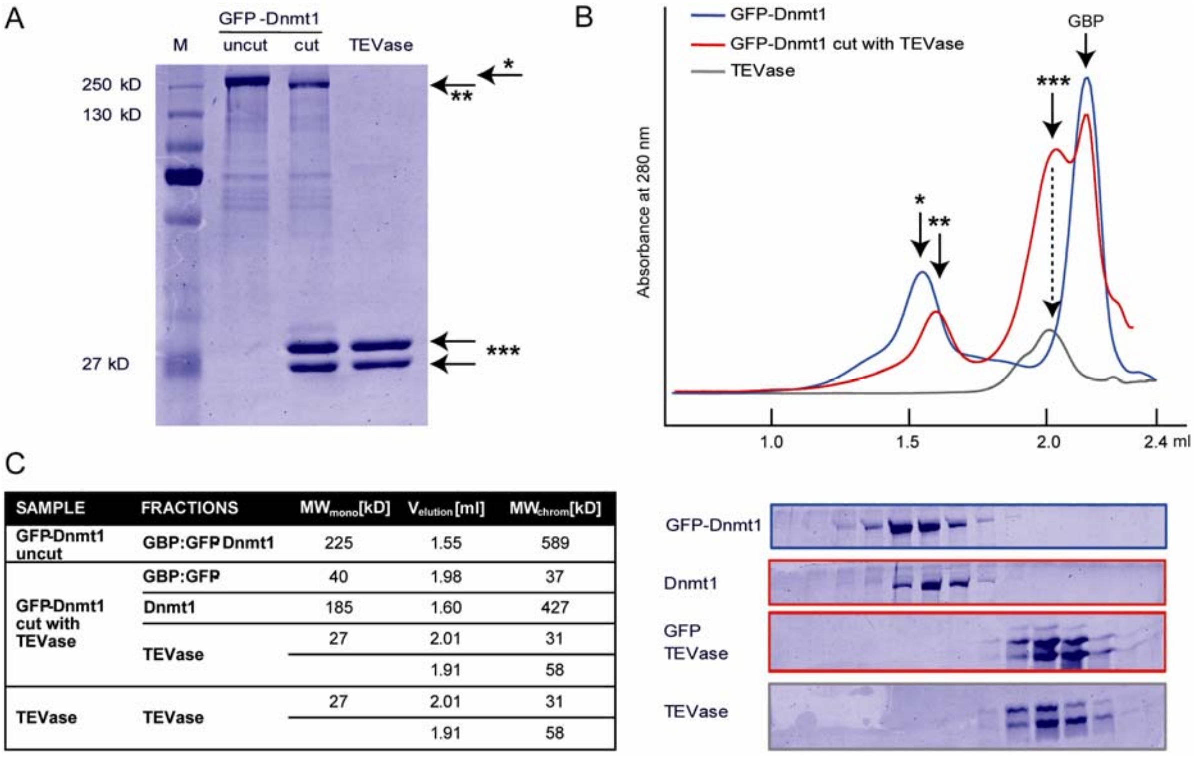
As mentioned above, we don't expect non-covalent DNA binding of Dnmt1 in the concentration range used for our single molecule spectroscopy studies. To test, if both DNA molecules in the Dnmt1:DNA complex with 1:2 stoichiometry are indeed bound covalently or if one of them is bound non-covalently, we incubated purified GFP-Dnmt1 either with two differentially labeled trapping substrates or with differentially labeled binding and trapping substrates (Figure 14). Binding substrates have the identical sequence to trapping substrates with the exception of having a canonical cytosine instead of 5-aza-cytosine at the central hemimethylated CpG site. After purification of the protein:DNA complexes, we performed FCCS analysis for all possible fluorophore combinations (Figure 14). Significantly, cross-correlation between the two ATTO labels could only be detected in the first case of incubation with two trapping substrates. Additionally, cross-correlation between ATTO label and GFP was exclusively observed for trapping substrates, suggesting cross-correlation only in the case of covalent complex formation between Dnmt1 and DNA. Moreover, the observed cross-correlation was in all cases resistant to denaturation with SDS (Table 15), again supporting the hypothesis of covalent bond formation. In stark contrast, when likewise incubating the respective DNA substrates with the catalytic site mutant Dnmt1^{C1229W}, which was previously reported to be catalytically inactive (Schermelleh et al., 2005), we did not detect cross-correlation between GFP and any ATTO label (Table 15). Together, these results strongly indicate that the complex formation between Dnmt1 and the two DNA substrates is indeed covalent.

Table 15. Summary of FCCS analyses for complexes of GFP-Dnmt1 or catalytically inactive GFP-Dnmt1^{C1229W} with DNA upon incubation with either differentially labeled binding and trapping substrates (HMB550 and HMT700, respectively) or differentially labeled trapping substrates (HMT500 and HMT700, respectively). Cross-correlation between different pairs of fluorophores (CC pairs) was determined before and after denaturation with SDS. '+' indicates that cross-correlation is clearly detected; '-' indicates that cross-correlation is zero; '(-)' indicates that the deviation of cross-correlation from zero is so low, that it is probably due to aggregation.

Protein	DNA substrate	CC pair	CC - SDS	CC + SDS
GFP-Dnmt1	HMB550/HMT700	550/GFP	-	-
		700/GFP	+	+
		550/700	-	-
GFP-Dnmt1	HMT550/HMT700	550/GFP	+	+
		700/GFP	+	+
		550/700	+	+
GFP-Dnmt1^{C1229W}	HMB550/HMT700	550/GFP	**(-)**	-
		700/GFP	**(-)**	-
		550/700	**(-)**	-
GFP-Dnmt1^{C1229W}	HMT550/HMT700	550/GFP	**(-)**	-
		700/GFP	**(-)**	-
		550/700	**(-)**	-

The GFP tag does not interfere with multimerization of Dnmt1

In order to exclude that the GFP tag (in complex with GBP) affects the multimerization state of Dnmt1, we analyzed and compared the purified GBP:GFP-Dnmt1 complex with untagged Dnmt1 by size-exclusion chromatography (Figure 16).



SAMPLE	FRACTIONS	MW_{mono}[kD]	$V_{elution}$[ml]	MW_{chrom}[kD]
GFP-Dnmt1 uncut	GBP:GFP Dnmt1	225	1.55	589
GFP-Dnmt1 cut with TEVase	GBP:GFP	40	1.98	37
	Dnmt1	185	1.60	427
	TEVase	27	2.01	31
			1.91	58
TEVase	TEVase	27	2.01	31
			1.91	58

Figure 16. The GFP-tag does not affect multimerization of Dnmt1 as shown by size-exclusion chromatography. GFP-Dnmt1 including a TEV site in the linker region was purified and cut with TEV protease resulting in a GBP:GFP complex and untagged Dnmt1. Aliquots of this sample as well as of GFP-Dnmt1 and TEV protease were subjected to SDS-PAGE **(A)** and analyzed by size-exclusion chromatography **(B)**. Arrows indicate GBP:GFP-Dnmt1 (*), Dnmt1 (**), and purified TEVase (***). Note that the TEVase is characterized by a compound of two peptides; the lower molecular weight fragment results from self-cleavage. The high peaks observed in the red and blue curve most likely correspond to co-eluted free GBP. Elution fractions were additionally analyzed by SDS-PAGE as shown below in the respective color code. **(C)** The table indicates for all samples and fractions the calculated molecular weight of the monomeric proteins and complexes (MW_{mono}), the elution volume ($V_{elution}$) and the according molecular weight resulting from the comparison with marker proteins applied to the same chromatographic column (MW_{chrom}).

Therefore, we designed, cloned and expressed a GFP-Dnmt1 construct containing a TEV protease site (amino acid motif: ENLYFQG) within the linker region between GFP and Dnmt1. We purified this protein as described above making use of GBP and GFP tag, and subsequently used purified TEV protease (TEVase) to cut the fusion protein at the respective recognition sequence. After establishing the reaction conditions for quantitative digestion, we prepared samples at a large scale. We subjected aliquots before and after addition of TEVase to SDS-PAGE analysis (Figure 16A) and also analyzed the samples by size-exclusion chromatography (Figure 16B). Interestingly, both GFP-Dnmt1 and Dnmt1 elute at a volume corresponding to the molecular mass of a dimer if compared to marker proteins applied in separate runs (Figure 16B and 16C).

In this study, in addition to FIDA results on the molecular brightness, particularly co-purification and quantification of GFP- and RFP-Dnmt1 via their different tags suggest that no stable dimer is formed by Dnmt1. Thus, other methods should be applied in order to further investigate the possibility of Dnmt1 dimerization. However, regardless of the question whether Dnmt1 is present as a monomer or a dimer, its state is not changed by complex formation with DNA substrates (Figure 13A). Moreover, we provide evidence that a single Dnmt1 molecule might be able to covalently bind two DNA substrates at the same time (Figure 11 and 14). This finding is very surprising and has to be carefully confirmed and further investigated (see chapter 3.2.4 for a more extensive discussion).

In conclusion, we provide evidence for the possibility that monomeric Dnmt1 binds two DNA substrates covalently at the same time and that its multimerization state is not changed upon complex formation. Clearly, these results are against all expectations and their confirmation would radically change our perspective of Dnmt1 regulation.

2.6 DNA METHYLATION

DNA METHYLATION

Carina Frauer, Fabio Spada and Heinrich Leonhardt

Ludwig Maximilians University Munich, Department of Biology, Center for Integrated Protein Science Munich (CIPS^M), 82152 Planegg-Martinsried, Germany.

This manuscript includes the first part of the introduction of this work (chapter 1.1) and is currently under revision for Rippe (Ed.): Genome organization and function in the mammalian cell nucleus. WILEY-VCH.

ABSTRACT

During mammalian development one single cell, the zygote, gives rise to a complete organism consisting of many different cell types. These cell types are genetically identical; however they differ dramatically in structure and function. In fact, each cell type expresses a very specific set of genes, whereas expression of other genes is stably repressed. Cellular differentiation during development is regulated by a complex network of transcriptional regulators leading to differential gene expression. In addition, transcriptional regulation is functionally linked with and controlled by DNA methylation and other epigenetic mechanisms such as histone modification, the Polycomb/Trithorax system and RNA-mediated processes. In mammals, DNA methylation (the covalent attachment of methyl groups) takes place at the C^5 position of cytosine residues mostly at CpG dinucleotides, and is catalyzed by a family of DNA methyltransferases. The methylation state of gene regulatory sequences, including promoters and enhancers, changes throughout development and methylation often correlates with a transcriptionally silent state. These DNA methylation marks are recognized by methyl-CpG binding proteins and translated into repressive chromatin states by recruitment of histone modifying enzymes and chromatin remodeling factors. In this chapter, we review the current literature of how methylation marks are set and maintained, and how DNA methylation regulates gene expression during cellular differentiation and development. We also discuss new insights into the complex interplay of different epigenetic pathways and the dynamics of DNA methylation.

3 DISCUSSION

Although the sequence of mammalian Dnmt1 is known since two decades and its impact in the regulation of gene expression has been extensively studied and demonstrated, the exact mechanistic basis of Dnmt1's maintenance function remains to be elucidated.

In this PhD work, I developed new powerful tools to study DNA methyltransferase activity and binding specificity of Dnmt1 and other proteins involved in epigenetic gene regulation. Applying these methods, I provide further insights into the regulation of Dnmt1 activity. The data show that Dnmt1 recognizes its substrate not at the initial DNA binding but at the step of covalent complex formation, further elucidate the functional role of the Dnmt1 CXXC domain, and provide evidence for a more complex role of Uhrf1 in regulating maintenance methylation than merely recruiting Dnmt1 to hemimethylated sites. Furthermore, I established and applied protocols to study Dnmt1 dimerization as well as the stoichiometry of complexes with DNA and the results provide first evidence for the simultaneous covalent binding of two DNA molecules per Dnmt1 monomer.

3.1 DEVELOPMENT OF A VERSATILE ASSAY TO MEASURE DNA BINDING AND METHYLTRANSFERASE ACTIVITY

3.1.1 POTENTIAL OF THE DEVELOPED ASSAY

In the past, a variety of biochemical assays have been developed for measuring DNA methyltransferase activity. However, all these methods depend on the use of radioactively labeled cofactor AdoMet, expensive and demanding equipment and/or multiple-step protocols. In chapter 2.1, we present a simple, non-radioactive and versatile assay for characterization of DNA methyltransferase activity and DNA binding (Frauer and Leonhardt, 2009). We used fusion constructs of green fluorescent protein (GFP) and a GFP binding nanobody coupled to agarose beads for one-step purification (GFP-Trap®, ChromoTek GmbH, (Rothbauer et al., 2007), see also chapter 3.1.2). Subsequently, we incubated the immobilized GFP fusion proteins with fluorescently labeled DNA substrates. After removal of unbound substrate, the absolute amounts and molar ratios of GFP fusion proteins and bound DNA substrates were determined by fluorescence spectroscopy. The assay measures methyltransferase activity as irreversible covalent complex formation with fluorescently labeled DNA substrates containing the mechanism-based inhibitor 5-aza-cytosine. As introduced in chapter 1.2.2 and Figure 7, the complex mechanism of the methyl transfer reaction includes DNA binding, base flipping, covalent complex formation, transfer of the methyl group, and enzyme release by β-elimination. DNA substrates containing 5-aza-cytosine at a CpG site (trapping substrates) have been shown to form covalent complexes with active methyltransferases identical to canonical DNA substrates (binding substrates). However, in contrast to binding substrates, the covalent complex formation of trapping substrates with Dnmt1 is irreversible and the enzyme is covalently trapped. Thus, by comparing immobilization of binding and trapping substrates upon incubation with DNA methyltransferases, we can discriminate enzymatic activity-dependent trapping from canonical DNA binding allowing us to test and screen for active methyltransferases. Furthermore, we carefully selected fluorophores with distinct excitation and emission spectra and are now able to test up to four fluorescent substrates in direct competition (chapter 2.3). This allows us to precisely determine DNA substrate specificity of DNA methyltransferases for two distinct steps of the methyl transfer reaction, i.e. DNA binding and covalent complex formation.

Using this assay, we gained further insights into Dnmt1's preference for hemimethylated sites, which is the basis of its maintenance activity. As introduced in chapter 1.2.3, the substrate specificity of Dnmt1 is well established: DNA substrates containing hemimethylated CpG sites are preferred to unmethylated substrates (Bestor and Ingram, 1983; Jeltsch, 2006). However, reports on the mechanism of substrate recognition as well as on the Dnmt1 domains involved are conflicting (Bacolla et al., 2001; Flynn et al., 1996). With our new assay, we have the unique opportunity to test binding of Dnmt1 to four different substrates in direct competition. Directly comparing DNA substrates containing either no, or one central un-, hemi- or fully methylated CpG site with distinct fluorescent labels, we found that Dnmt1 binds equally well to all four substrates and is neither affected by the presence nor by the methylation state of the CpG site (chapter 2.3, discussed in 3.2.1). Furthermore, we show for the first time that Dnmt1 prefers hemimethylated over unmethylated substrate for covalent complex formation, an early step of the methyl transfer reaction (chapter 2.1, discussed in 3.2.1). The preference factor of about 15-fold is in the range of previously published experiments, which measured the net product of the methylation reaction (Jeltsch, 2006; Pradhan et al., 2008), suggesting that substrate recognition takes place at an early stage of the reaction. It is important to stress that the factor we determined is not exactly comparable to the factors produced by assays measuring enzymatic activity as the number of transferred methyl groups and with non-competitive assay designs. With our assay, we can specifically determine sequence preferences at the step of covalent complex formation with different substrates in direct competition, which is so far unique for all published methods. The new assay therefore represents an important technical advance in comparison to other methods.

Regarding the question of which parts of Dnmt1 are essential for substrate specificity, we started analyzing single Dnmt1 domains as well as Dnmt1 mutants for their DNA binding and/or DNA methyltransferase activity (see also chapters 2.3). Moreover, we have now a powerful tool at hand to fast and competitively test activities and specificities of other DNA binding proteins that have been implicated in epigenetic gene regulation (see also chapters 2.4, 3.1.3 and 3.2.3). In addition to studies of DNA binding proteins, the assay is likewise suited for characterization of RNA and peptide binding proteins. Indeed, we have already adapted the assay for studying binding of GFP fusions to fluorescently labeled peptides (chapter 2.4). Analogously, we are currently establishing a protocol for characterization of RNA binding proteins using fluorescently labeled RNA substrates produced by *in vitro* transcription (Christine Schmidt, unpublished data). Alternatively, one could use this assay for small molecule inhibitor screenings or, using the 5-aza-cytosine containing trapping substrates, for identification of active DNA methyltransferases in cell extracts by a combination of SDS-PAGE, Western blot and mass spectrometry analyses. Finally, we further applied the newly established toolbox for DNA substrate preparation via primer extension and developed another assay for characterization of the methyl-CpG binding protein MeCP2 (see below, chapter 3.1.3).

In conclusion, we are convinced that further use of this assay will be very valuable for deciphering the basis of Dnmt1's mechanism, regulation and function. Moreover, adaptations of the assay will facilitate characterization of a variety of DNA, RNA and (histone tail) peptide binding proteins.

3.1.2 GBP BINDING ENHANCES GFP FLUORESCENCE

When developing the assay, we observed an interesting feature of the GFP-binding protein (GBP) (Kirchhofer et al., 2009). GBP is a camelid-derived single-domain antibody (nanobody) and constitutes the active part of the GFP-Trap® (ChromoTek GmbH). We surprisingly found that upon binding, the GBP modulates the conformation and spectral properties of the green fluorescent protein (GFP), i. e. the GBP bound GFP showed enhanced fluorescence intensity and a blue-shifted emission spectrum (Figure 17, see also chapter 2.1 and 2.2). We confirmed the enhancement effect also by single molecule spectroscopy (see chapter 2.5). Initially, we investigated this phenomenon using an enhanced variant of GFP (eGFP), which was also used for the generation of GFP fusion constructs. Afterwards, also other GFP variants were tested. Notably, we found that the enhancement effect differs between different GFP variants. In fact, the enhancement effect for wild-type GFP is even more pronounced than for eGFP. Structural analysis of GBP:GFP complexes and comparison to the crystal structure of GFP alone revealed that the GBP induces subtle changes in the environment of the GFP chromophore, finally leading to an altered absorption spectrum.

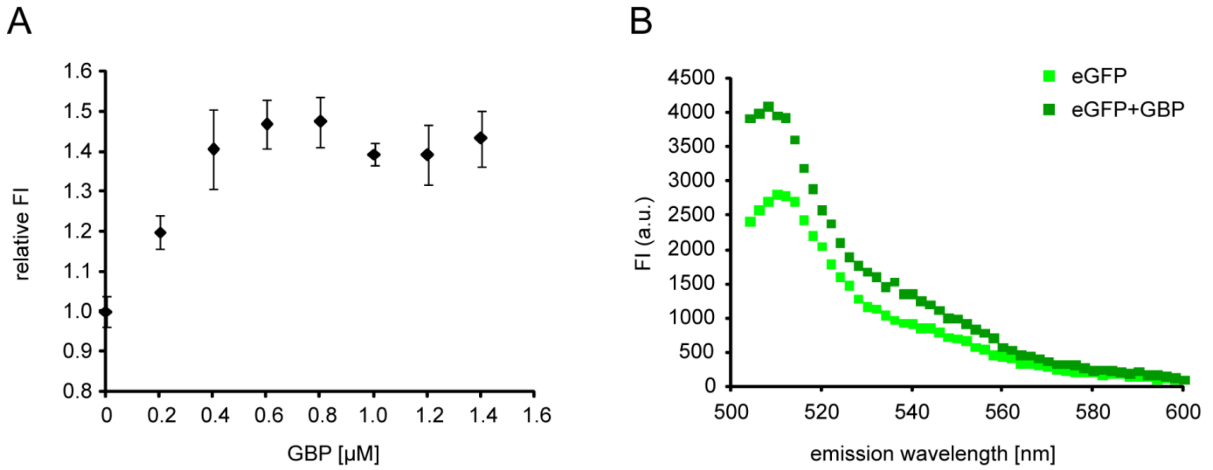


Figure 17. GBP modulates spectral properties of eGFP. **(A)** The fluorescence intensity of eGFP (1 µM) is enhanced by addition of increasing amount of GBP up to around 1.4 times, if exciting at 490 ± 5 nm and detecting the emission at 511 ± 5 nm. **(B)** At equimolar concentration, the GBP shifts the emission maximum of eGFP upon excitation at 490 nm to a lower wavelength, more precisely from 511 nm to 508 nm. Error bars represent the standard deviation of three independent experiments.

GFP is a barrel-shaped protein with a central *p*-hydroxybenzylidene-imidazolinone chromophore, which results from oxidative backbone cyclization involving residues Ser65, Tyr66 and Gly67 (Adams et al., 2000; Chalfie et al., 1994; Ormo et al., 1996). The enhanced form of GFP (eGFP) shows a characteristic dual-peak excitation spectrum with maxima at 395 nm and 490 nm, while excitation at both wavelengths results in emission that peaks at 511 nm. The dual excitation peak results from the existence of two interconvertible alternative states of the chromophore, that is a neutral phenol state and a deprotonated phenolate anion state (Brejc et al., 1997), corresponding to absorption at 395 and 490 nm, respectively. Interestingly, the GBP stabilizes the deprotonated chromophore state, which is characterized by increased absorption efficiency and explains the resulting increase of fluorescence emission. Wild-type GFP is even more sensitive to this enhancement effect by GBP binding, since the chromophore of a predominant fraction of molecules is originally present in the neutral state. In contrast, eGFP already favors the anion state if unbound and thus one could argue that the GBP mimics the mutational effect on GFP. Our spectral analysis indicates that for GBP complexes with eGFP (GBP:eGFP) exclusively the anion state of the chromophore remains. Thus, it is possible that the GBP:eGFP complex corresponds to the fluorophore state with optimal spectral properties and maximal molecular brightness. Notably, other GFP binding nanobodies showed either no or a quenching effect, in the second case by stabilization of the neutral chromophore state. We also established a protocol for fluorescence intensity scans in living cells and confirmed that modulation of spectral GFP properties upon GBP binding takes also place *in vivo*.

The described nanobody-mediated enhancement of GFP fluorescence could be applied to improve tracing of GFP fusion proteins in living cells or to ultrahigh resolution microscopy, e.g. three-dimensional structured illumination microscopy (3D-SIM) (Schermelleh et al., 2008), thus whenever GFP fluorescence intensity and/or bleaching might be a limiting factor. In perspective, nanobodies or fluorophore fused nanobodies (chromobodies) could be used to control conformation of also other proteins than GFP in living cells.

3.1.3 USING THE NEWLY ESTABLISHED TOOLS FOR FURTHER APPLICATIONS

Using the established assay, we can now easily test DNA binding activity and specificity of other chromatin associated proteins. As introduced in chapter 1.1.5.2, MeCP2 is a methyl-CpG binding protein that translates DNA methylation marks into repressive chromatin states by binding to methylated CpG sites via its MBD and recruitment of histone deacetylases, histone methyltransferases and nucleosome remodeling factors via the TRD domain (Hendrich and Tweedie, 2003; Jones et al., 1998; Nan et al., 2007) as well as by interaction with the heterochromatin binding protein HP1 (Agarwal et al., 2007). MeCP2 has been shown to cause large-scale chromatin reorganization and to particularly induce chromatin compaction and clustering of pericentric heterochromatin via its MBD domain (Brero et al., 2005; Georgel et al., 2003). Mutations in the *MECP2* gene have been involved in human neurological disorders (e.g. the Rett syndrome (Guy et al., 2001)). The mechanism of MeCP2 induced chromatin compaction and clustering remains elusive.

In order to gain further insights into MeCP2 regulation and function we tested its DNA binding specificity and those of single MeCP2 domains (Figure 18), in collaboration with the group of Prof. Cardoso (TU Darmstadt). We used DNA substrates with unmethylated and fully methylated CpG sites. As genomic DNA of Purkinje neurons and other brain tissues was recently proposed to contain 5-

hydroxymethylated cytosine (Kriaucionis and Heintz, 2009), we also tested binding of MeCP2 to hydroxymethylated DNA. We could show that MeCP2 preferentially binds to fully methylated DNA and with equally lower affinity to unmethylated and fully hydroxyl-methylated DNA substrate. We identified the MBD domain and the TRD domain to be responsible for this binding specificity. The MBD domain binds exclusively to methylated DNA and the ID-TRD domain, comprising the region between TRD and MBD (interdomain, ID) and the TRD domain, binds to DNA independent of the methylation state of the CpG site and with a similar affinity as the MBD to methylated DNA. Notably, both the N-terminal part (amino acids 1-55) as well as the C-terminal part (amino acids 311-492) of MeCP2 do not show any binding activity. These results are consistent with fluorescence recovery after photobleaching (FRAP) studies showing that the three domains, MBD, ID and TRD, contribute to chromatin binding *in vivo* (Kumar et al., 2008). Furthermore, FRAP analyses of MBD and ID-TRD domains in wild-type embryonic stem cells as well as *dnmt1/3a/3b*$^{-/-}$ triple knock-out embryonic stem cells recently revealed that the mobility of the MBD is dependent on the genomic DNA methylation level whereas the ID-TRD domain is not (our unpublished data, Andrea Rottach). Clearly, these data provide strong evidence that MBD-mediated DNA binding is methylation dependent *in vitro* and *in vivo*, whereas ID-TRD-mediated DNA binding is independent of methylation.

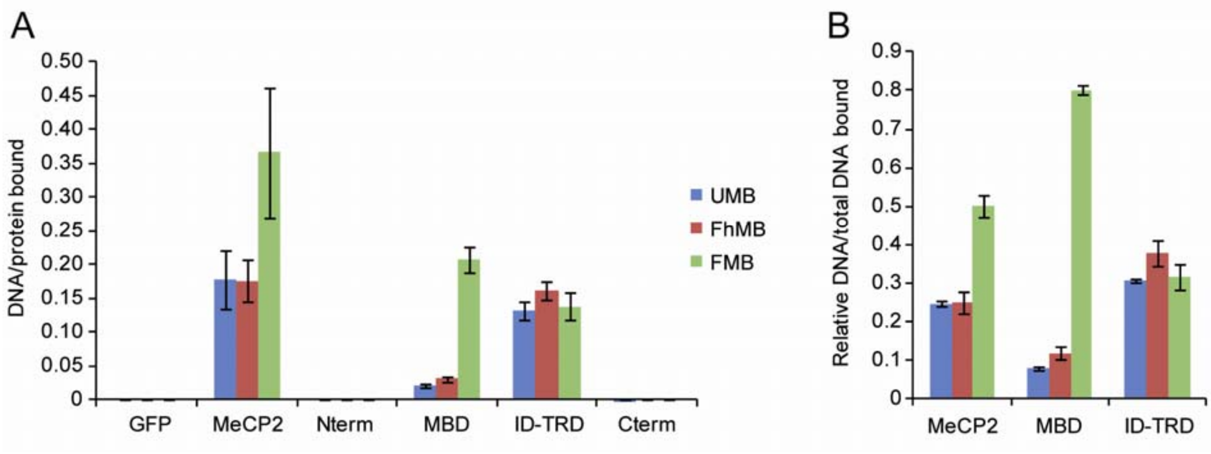


Figure 18. DNA binding specificity of MeCP2 and single MeCP2 domains. **(A)** MeCP2 and MeCP2 domains as fusions with GFP or YFP were purified from insect cells via the GFP-Trap® after baculovirus-mediated expression and incubated with unmethylated (UMB), fully hydroxymethylated (FhMB) and fully methylated DNA binding substrates (FMB) in direct competition. After removal of unbound substrate, the fluorescence intensity of immobilized GFP/YFP fusions and fluorescent DNA substrates was measured and the ratio of DNA/protein bound was calculated. **(B)** The fractions of the single bound DNA substrates were normalized to the total fraction of bound DNA. Error bars represent the standard deviation of three independent experiments.

Based on the established toolbox for DNA substrate preparation and DNA binding assays using GFP fusions and fluorescently labeled DNA substrates ((Frauer and Leonhardt, 2009), chapter 2.1), we developed and applied another assay to address the mechanism of the MeCP2-induced chromatin clustering *in vitro*. This 'sandwich assay' allows us to test whether a protein is able to induce DNA clustering, simulated by binding to at least two DNA molecules with high affinity and at the same time (Figure 19A). The workflow of the assay is as follows. First, two differentially fluorescently labeled DNA substrates are prepared, one of them biotinylated, and the GFP fusion protein is expressed in and purified from human embryonic kidney (HEK 293T) cells using the GBP coupled to Ni-NTA agarose (chapter 2.3). Second, Strep-Tactin® agarose beads (IBA) are functionalized with the biotinylated fluorescent DNA substrate. Third, these functionalized beads are simultaneously incubated with both the purified GFP fusion as well as the second fluorescent DNA substrate. After removal of unbound substrate, the ratio of second to first DNA is calculated from the respective fluorescence intensities and serves as a measure for the DNA cluster efficiency. When performing this assay for MeCP2, we detect, as a proof of principle, a significantly higher 2nd/1st DNA ratio than for GFP, which was used as negative control (Figure 19B). These first results suggest that a MeCP2 monomer or multimer can bind two DNA molecules, which could be the basis of the MeCP2 induced clustering of centromeric DNA *in vivo*.

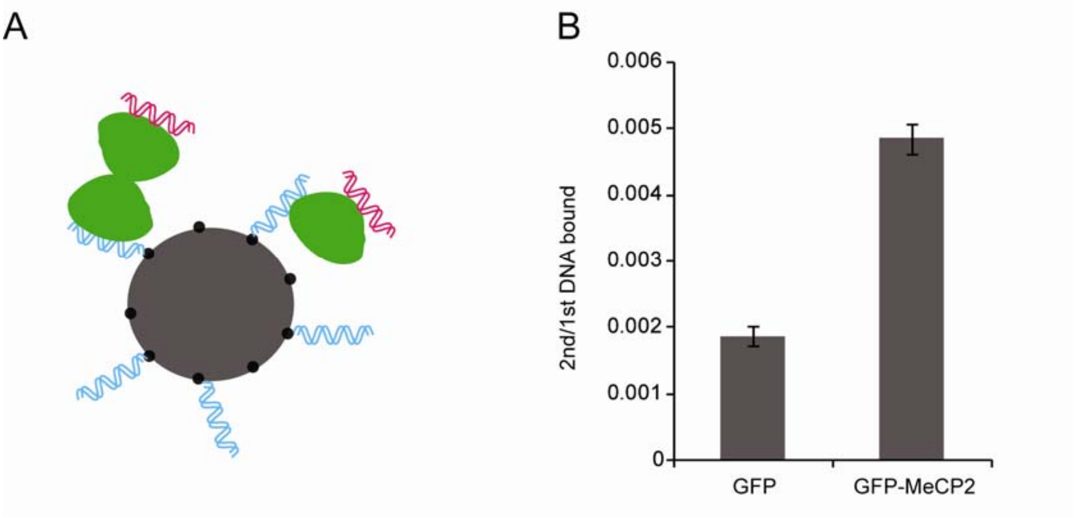


Figure 19. MeCP2 sandwich assay. **(A)** Overview of the assay design. Small black circles: Strep-Tactin tag, blue: biotinylated 1st fluorescent DNA substrate, green: GFP-MeCP2, red: 2nd fluorescent DNA substrate. Note that it is not known whether MeCP2 induces clustering by binding two DNA substrates as monomer via two distinct domains or as multimer. **(B)** Calculated ratios of 2nd/1st DNA substrate bound serve as a measure of the clustering effect. This value is significantly higher for GFP-MeCP2 than for the negative control GFP. Error bars represent the standard error of three independent experiments.

Thus, this assay allows for further mechanistic studies of the clustering effect induced by MeCP2. For example, one can now directly address the question whether a single MeCP2 domain is sufficient for MeCP2 clustering, or whether a combination of several domains is needed for this effect. In this regard, *in vivo* data suggest that it is possible that the MBD domain alone mediates clustering (Brero et al., 2005). In this case, the MBD would have to either mediate both dimerization and binding to one DNA molecule, or to simultaneously bind to two DNA substrates as monomer. Alternatively, both MBD and ID-TRD domain could be necessary for the clustering effect. These hypotheses can now easily be tested by varying the methylation state of the two DNA substrates in combination with the use of single domains or a combination of MeCP2 domains in the sandwich assay.

Furthermore, there is recent evidence that MeCP2 is poly(ADP-ribosyl)ated by PARP1 and that this post-translational modification reduces clustering of pericentric heterochromatin *in vivo* (unpublished data, Cardoso group). To further test the role of poly(ADP-ribosyl)ation, the sandwich assay could be applied to MeCP2 purified from insect cells, since proteins expressed in these cells have been shown to be highly poly(ADP-ribosyl)ated. Additionally, MeCP2 mutants could be constructed that are resistant against this type of post-translational modification, expressed in different cellular backgrounds, purified and likewise tested and compared in the assay.

Currently, we are investigating various MeCP2 mutants as well as single MeCP2 domains with this sandwich assay in order to further elucidate the mechanistic basis of the MeCP2 clustering effect and its role in epigenetic regulation of gene expression.

3.2 MAINTENANCE METHYLATION – A CLEAR BUT DEMANDING TASK

3.2.1 HOW DNMT1 RECOGNIZES ITS SUBSTRATE

Dnmt1 methylates DNA at the C^5 position of cytosine residues within CpG dinucleotides and prefers hemimethylated substrates, which are typically occurring at the replication fork after semi-conservative DNA replication. This led early to the identification of Dnmt1 as the maintenance enzyme. Still, the domains involved and the precise step of the methyl transfer reaction at which substrate recognition occurs are very controversially discussed (chapter 1.2.3). Using our newly developed assay for DNA methyltransferase activity and DNA binding (chapter 2.1), we now tested for the first time competitive binding to four different DNA substrates and found that Dnmt1 binds to DNA independent of presence and methylation state of the CpG site (chapter 2.3). Furthermore, we show that Dnmt1 prefers hemimethylated over unmethylated substrate for covalent complex formation, an early step of the methyl transfer reaction (chapters 2.1, 2.3).

Regarding the DNA binding specificity of Dnmt1, this result is consistent with a study showing that Dnmt1 binds equally well to un- and hemimethylated DNA and proposing a later step of the methyl transfer reaction to be responsible for substrate discrimination (Flynn et al., 1996). However, our data contradict two studies claiming that Dnmt1 preferentially binds to hemimethylated substrates (Araujo et al., 2001; Bacolla et al., 2001). We speculate that these differences might result from the use of different DNA substrates. One study uses a $(CGG{\cdot}CCG)_{12}$ triplet repeat, naturally occurring within the 5’-untranslated region of the *FMR1* gene and involved in the etiology of the fragile X syndrome (Fu et al., 1991; Verkerk et al., 1991). This neurological disorder is genomically characterized by expansion and de novo methylation of this particular DNA sequence as well as a proximal CpG island, finally leading to silencing of the *FMR1* gene (Merenstein et al., 1996; Steyaert et al., 1996). The other report is based on the use of modified non-canonical hairpin DNA oligonucleotides, which have previously been shown to serve as Dnmt1 inhibitors (Bigey et al., 1999). In contrast to these studies, we are using canonical 42 base pair long DNA substrates containing a single central CpG site. Interestingly, the two publications suggest different parts of the enzyme to be responsible for the differential substrate binding. Whereas the first study proposes that a Dnmt1 mutant lacking the first 501 amino acids still preferentially binds to hemimethylated DNA substrates (Bacolla et al., 2001), the second study suggests the target recognition domain of Dnmt1 to reside within amino acids 122-417 (Araujo et al., 2001). This fragment includes an AT-hook like motif, the PBD domain, a nuclear localization signal (NLS) and the DNA binding domain DB1 as part of the TS sequence.

Apart from these two reports, our results, which indicate that substrate discrimination by Dnmt1 occurs not until covalent complex formation, are very surprising. Usually, enzymes recognize their substrate by kinetic discrimination at the step of substrate binding to the catalytic pocket. Clearly, this mechanism of substrate recognition is also the most economic one, since it allows the discrimination of substrate from non-substrate in only one step and upon first interaction between enzyme and potential substrate. In contrast, substrate recognition by Dnmt1 requires the succession of both DNA binding and presumably even base flipping before substrate discrimination occurs in that a covalent complex is preferentially formed with hemimethylated substrate, which is subsequently subjected to methylation. The energy costs for this substrate discrimination are remarkable and the question arises why Dnmt1 does not recognize its target at the initial binding step and at considerably lower energetic costs. One could argue that it is hard to recognize hemimethylated CpG sites by non-covalent binding to the major groove of DNA and that base flipping might be required for hemimethylated substrate recognition. However, it has been demonstrated for methyl-CpG-binding proteins that recognition of fully methylated CpG sites is possible without flipping the cytosine base out of the double helix (Ho et al., 2008; Ohki et al., 2001). Notably, it has also been suggested that Dnmt1 slides along the DNA helix in order to processively methylate hemimethylated sites occurring at the replication fork and this property could be facilitated by a general DNA binding activity ((Bestor and Ingram, 1983), see also chapter 1.2.2). Thus, the necessity and relevance of this late substrate discrimination of Dnm1 remains elusive.

Regardless of the question of at which step substrate recognition occurs, Dnmt1 prefers methylation of hemimethylated DNA. For this sequence specificity, both, the N-terminal domain (Pradhan and Roberts, 2000) as well as the linker region (Bestor, 1992) have been proposed to be required in addition to the C-terminal domain. We have recently shown that linker cleavage as well as modification of linker length and charge does not affect Dnmt1 activity or substrate preference (unpublished data, Weihua Qin), excluding a role of the linker in substrate recognition. Moreover, it is still unknown which part of the N-terminal domain is needed for substrate specificity. It is important to realize that the question of the exact target recognition domain of Dnmt1 is difficult to address separately from the question of the domains needed for enzyme activity at all, since the catalytic domain of Dnmt1 is not active by itself. Thus, the intrinsic sequence preference of the catalytic center can not be directly tested. Interestingly, human Dnmt1Δ1-580 still preferentially methylates hemimethylated over unmethylated DNA substrate (Pradhan and Esteve, 2003). This suggests that at least the first half of the N-terminal domain can be excluded from the task of substrate recognition and enzymatic activity. We hypothesized that the CXXC domain of Dnmt1 (amino acids 651-698 in the mouse enzyme), which is still included within the Dnmt1Δ1-580 deletion mutant, might be involved in substrate recognition and we addressed its functional role in regulating Dnmt1 activity in this work (chapter 2.3, discussed in 3.2.2).

Although Dnmt1 methylates preferentially hemimethylated DNA, it still possesses activity on unmethylated substrates *in vitro*. Significantly, this de novo activity of Dnmt1 is higher than that of the de novo methyltransferases Dnmt3a and Dnmt3b. Dnmt1 activity on unmethylated substrates has been shown to be enhanced by binding of methylated DNA to an allosteric site and it has been suggested to be involved in methylation spreading (Pradhan and Esteve, 2003). While this allosteric activation potential might indeed be important for methylation spreading, it could also lead to accidental mismethylation of unmethylated sites at the replication fork. It has been shown that the allosteric activation activity is more pronounced on unmethylated double stranded than on unmethylated single stranded DNA substrates (Tollefsbol and Hutchison, 1997). Nevertheless, the observed de novo activity of Dnmt1 *in vitro*, even in the absence of allosteric activator, raises the question of how Dnmt1 can reliably fulfill its maintenance function *in vivo*, and whether there are other factors involved in repression of de novo methylation activity. In fact, we could recently show that, in contrast to *dnmt1-/-* embryonic stem cells with hypomethylated but not unmethylated genome, in which transient or stable expression of Dnmt1 restores methylation pattern to the wild-type level, *dnmt1/3a/3b*$^{-/-}$ TKO embryonic stem cells rescued with Dnmt1 show no genomic methylation (our unpublished data, Daniela Meilinger). This suggests that there might be indeed other factors involved that prevent de novo methylation of unmethylated DNA sequences *in vivo*, and/or remove incorrectly positioned methylation marks (further discussed in 3.3.1).

Some factors have been shown to positively control Dnmt1 activity *in vivo*. In this context, PCNA has been shown to target Dnmt1 to sites of replication and DNA repair. Although PCNA is not required for maintenance methylation, interaction with PCNA increases Dnmt1 activity *in vivo* by two fold (Schermelleh et al., 2007). Similarly, Uhrf1 has been recently suggested to recruit Dnmt1 to sites of replication. But in stark contrast to PCNA, Uhrf1 has been shown to not only facilitate or enhance maintenance methylation but to be essential. In this work, we addressed the mechanisms by which Uhrf1 contributes to maintenance methylation and controls Dnmt1 activity (chapter 2.4, further discussed in 3.2.3 and 3.3.2).

3.2.2 THE FUNCTIONAL ROLE OF THE DNMT1 CXXC DOMAIN

As mentioned above, the first 580 amino acids of human Dnmt1 have been shown to be dispensable for both enzymatic activity and sequence specificity, whereas deletion of the first 672 amino acids results in an inactive enzyme (Pradhan and Esteve, 2003). Interestingly, this latter truncation eliminates a part of the CXXC domain, suggesting an involvement of this domain in enzymatic activation. However, addition of the isolated CXXC domain to the catalytic domain *in trans* was not sufficient for allosteric activation (Fatemi et al., 2001). Moreover, there are conflicting reports on the binding specificity of the Dnmt1 CXXC domain (Fatemi et al., 2001; Pradhan et al., 2008).

Based on a homology model (chapter 2.3, collaboration with Johannes Söding) we designed an isolated Dnmt1 CXXC domain construct and characterized its DNA binding properties with the newly established assay that allows us to compare four different DNA substrates in direct competition.

According to the structural model, the CXXC domain of Dnmt1 is delimited by an antiparallel β-sheet representing a discrete structural element. We showed that this isolated mouse Dnmt1 CXXC domain preferentially binds to unmethylated DNA as it was previously demonstrated for numerous CXXC domains of other mammalian proteins (chapter 2.3). These include the histone H3K4 methylase MLL (Allen et al., 2006; Ayton et al., 2004; Birke et al., 2002), the transcriptional activator CGBP as component of the mammalian Set1 H3K4 methyltransferase complex (Butler et al., 2008; Lee and Skalnik, 2005; Lee et al., 2001; Voo et al., 2000), the methyl-CpG binding protein MBD1 (Jorgensen et al., 2004) as well as the histone H3K36 demethylases JHD1A/FBXL11 and JHD1B/FBXL10 (Frescas et al., 2007; Tsukada et al., 2006). Interestingly, unlike the DNA binding domain of CGBP, which has been reported to include a flanking sequence downstream of the conserved CXXC domain (Lee et al., 2001) and which was also contained in the CXXC3 peptide of mouse MBD1 (Jorgensen et al., 2004), the isolated CXXC domain of Dnmt1 is sufficient for DNA binding excluding the need for any flanking sequences (chapter 2.3).

The preferential binding of the Dnmt1 CXXC domain to unmethylated DNA (chapter 2.3) represents a possible mechanism of substrate discrimination, in that the CXXC domain could act as negative regulatory element by blocking enzymatic activity upon unmethylated DNA binding. To test this hypothesis, we used the homology model of the CXXC domain for constructing a deletion mutant, with the aim of not affecting surrounding protein structure and maximizing native protein folding. We methodically tested and compared the functionality of wild-type Dnmt1 and deletion mutant *in vitro* and *in vivo*. Surprisingly, we found that the CXXC domain is dispensable for Dnmt1's DNA binding specificity *in vitro* and binding kinetics *in vivo*. Moreover, intramolecular interaction of N- and C-terminal domain, substrate specific covalent complex formation and methyl transfer activity were not affected by the CXXC deletion. As ultimate proof of functionality, we show that the genetic complementation of *dnmt1*$^{-/-}$ embryonic stem cells with the deletion construct rescues DNA methylation levels and patterns as efficiently and accurate as the complementation with the wild-type enzyme. In summary, both *in vitro* and *in vivo* approaches show that the CXXC domain of Dnmt1 is dispensable for DNA binding, allosteric activation, methyltransferase activity and substrate specificity.

These results are in stark contrast to a study claiming that the CXXC domain of human Dnmt1 is necessary for enzymatic activity (Pradhan et al., 2008). This study proposed that deletion of amino acids 648-690 in human Dnmt1 results in significant reduction of enzymatic activity on hemimethylated substrates and equally poor activity on unmethylated substrates *in vitro*. Furthermore, overexpression of the Dnmt1ΔCXXC deletion construct in HEK293-A7 cells led to 25 % lower genomic DNA methylation levels at ribosomal DNA repeats suggesting a dominant negative effect (Pradhan et al., 2008). The discrepancy with our data might be due to the use of Dnmt1 from different mammalian species or to differences in the construction of deletion mutants. In fact, the reported deletion construct of human Dnmt1 is based on removal of a slightly shorter fragment at slightly different position and might not preserve native folding of surrounding protein sequences. Based on both our homology model of the mouse Dnmt1 CXXC domain and sequence comparison by mouse and human Dnmt1 sequences, this slightly different deletion could indeed disrupt a structurally well-defined antiparallel β-sheet, which remains unaffected by our deletion and might be important for stability of surrounding protein structures. Ultimately, genetic complementation of *dnmt1*$^{-/-}$ embryonic stem cells represents a physiologically more relevant test of protein function *in vivo* than the overexpression in HEK293-A7 cells.

In summary, we presented solid data showing that the CXXC domain of mouse Dnmt1 is dispensable for DNA binding, enzymatic activity as well as specificity both *in vitro* and *in vivo*. The rescue of DNA methylation levels and patterns in mouse embryonic stem cells lacking Dnmt1 but genetically complemented with the deletion construct ultimately demonstrates that the CXXC domain is nonessential for Dnmt1's maintenance function in mammalian embryonic stem cells. Thus, the functional role of the Dnmt1 CXXC domain still remains elusive and in fact, we speculate that it might have either a subtle and/or developmental stage- or tissue-specific regulatory function.

This new hypothesis is based on the observation that cytoplasmic retention of Dnmt1o during early embryonic development is dependent on a region in the N-terminal part of Dnmt1 that includes the CXXC domain (Cardoso and Leonhardt, 1999). Notably, the binding of the Dnmt1 CXXC domain to unmethylated DNA does not exclude a supplementary role in protein-protein interaction as the CXXC domain of MLL was reported to interact with histone deacetylases and Polycomb group proteins in addition to DNA (Birke et al., 2002; Xia et al., 2003). Thus, it is indeed possible that the CXXC domain of Dnmt1 has a regulatory function in specific cell types or at specific developmental stages that might involve both DNA binding and protein-protein interaction. As a first attempt to address this hypothesis, we plan pull-down experiments with Dnmt1, the single CXXC domain as well as with the Dnmt1ΔCXXC deletion mutant using genetically complemented *dnmt1*$^{-/-}$ embryonic stem cells. Subsequent mass spectrometry analyses might then reveal potential interaction partners and further elucidate the function of the Dnmt1 CXXC domain. Moreover, a role of the CXXC domain in binding to RNA cannot be excluded as recent reports provide evidence for regulatory effects of small RNAs on Dnmt activity and expression (Svedruzic, 2008; Svedruzic and Reich, 2005).

Recently, Tet1 has been identified to oxidize 5-methylcytosine to 5-hydroxymethylcytosine in DNA (Tahiliani et al., 2009), a mechanism suggested to be involved in active DNA demethylation (chapter 1.1.6). Interestingly, Tet1 contains a very similar CXXC domain to that of Dnmt1 and no functional data were yet available for the CXXC domain of Tet1. In order to address the question whether the TET1 CXXC domain is involved in recognition of methylated DNA substrates, we also constructed a homology model of this domain and analyzed its DNA binding properties using the new DNA binding assay. We found that the CXXC domain of Tet1 does not possess any specific DNA binding activity (chapter 2.3). Consistently, our structural model of the Tet1 CXXC domain diverges remarkably from the structure of the MLL CXXC domain and the model of the Dnmt1 CXXC domain in the region of the KFGG motif. Significantly, we found that all CXXC domains reported to selectively bind unmethylated CpG sites cluster in a distinct homology group and contain the KFGG motif, which was shown to be crucial for DNA binding by the CXXC domain of MLL (Allen et al., 2006). In contrast, the CXXC domain of Tet1 is closely related to CXXC domains that have been shown to mediate protein-protein interactions (Andersson et al., 2009; Hino et al., 2001; London et al., 2004), suggesting that the Tet1 CXXC domain might be involved in interactions with other proteins rather than recognition of methylated DNA substrates. To test for these hypothetical protein-protein interactions of the Tet1 CXXC domain, pull-down assays could be performed in combination with mass spectrometry analysis.

3.2.3 THE ROLE OF UHRF1 IN MAINTENANCE METHYLATION AND REGULATION OF DNMT1

As introduced in chapters 1.1.5.2 and 1.3.4, Uhrf1 has recently emerged as essential cofactor for maintenance methylation. Uhrf1 binds to DNA with hemimethylated CpG sites via its SRA domain, directly interacts with Dnmt1 and colocalizes with Dnmt1 at replication foci (Arita et al., 2008; Avvakumov et al., 2008; Bostick et al., 2007b; Gowher et al., 2005; Sharif et al., 2007). Its genetic ablation in embryonic stem cells leads to global DNA hypomethylation, a phenotype similar to that obtained for Dnmt1 ablation (Bostick et al., 2007a; Sharif et al., 2007). Based on these data, Uhrf1 was suggested to function in maintenance methylation by recruiting Dnmt1 to hemimethylated CpG sites at the replication fork. However, the high intrinsic preference of Dnmt1 for methylation of hemimethylated substrates even in the absence of Uhrf1 (*in vitro*) and the emerging evidence that Uhrf1 also binds to histone tails leads to the hypothesis that Uhrf1 might play a more complex role in maintenance methylation than simply recruiting Dnmt1 to its target.

Using *in vitro* DNA and peptide binding assays in combination with *in vivo* fluorescence recovery after photobleaching (FRAP) studies in embryonic stem cells with different genetic backgrounds we characterized the various interactions mediated by Uhrf1 and the single Uhrf1 domains (chapter 2.4). We found that Uhrf1 binds to DNA substrates containing hemimethylated CpG sites only with an about two-fold preference compared to un- or fully methylated DNA substrates of the same sequence. Importantly, we provided all substrates in direct competition. Consistent to these *in vitro* DNA binding properties, we found that complete loss of genomic methylation does not affect binding kinetics of Uhrf1 *in vivo*. Moreover, we could show that the SRA domain is necessary and sufficient for DNA binding of Uhrf1 *in vitro* and consistently dominates the mobility of Uhrf1 *in vivo*. However, the subcellular localization of Uhrf1 is different from the localization of the isolated SRA domain and might thus be additionally controlled by other domains. Interestingly, we could show that Uhrf1 preferentially binds trimethylated H3K9, a mark characteristic for silent chromatin. We show that this interaction is mediated by the tandem Tudor domain of Uhrf1 via an aromatic cage (Jacobs and Khorasanizadeh, 2002). Together, our data support the hypothesis that Uhrf1 possesses a multi-functional modular structure interconnecting DNA methylation and histone modification pathways. Clearly, it is unlikely that the weak preference for hemimethylated DNA substrate alone suffices to regulate maintenance methylation. It is possible that the multiple specific interactions of Uhrf1 with hemimethylated DNA, trimethylated H3K9 tails, the histone methyltransferase G9a and the DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b add up to the necessary binding specificity of Uhrf1 *in vivo*. For example, we could recently show that DNA binding enhances the specificity of Uhrf1 for trimethylated H3K9 tails (our unpublished data, Garwin Pichler). Still, there might be additional mechanisms by which Uhrf1 contributes to the maintenance of genomic methylation levels and patterns.

In fact, overexpression of Uhrf1 induces large-scale chromatin rearrangements and particularly decondensation of pericentric heterochromatin via its PHD domain (Papait et al., 2008). This observation is independent from DNA methylation and raises the question whether Uhrf1 might play a role in increasing accessibility of DNA for Dnmt1 by inducing nucleosome remodeling. It is important to note that the question whether Dnmt1 activity and function is affected by nucleosomal arrangements is still controversially discussed. Interestingly, a study comparing activity of Dnmt1 on different DNA substrates revealed that Dnmt1 binds to mononucleosomes with higher affinity than to 'naked' DNA (Robertson et al., 2004). Remarkably, it has been shown that Dnmt1 efficiently methylates nucleosomal DNA without dissociation of the histone octamer from the DNA (Gowher et al., 2005), which is even possible if the DNA major groove is oriented towards the histone surface (Okuwaki and Verreault, 2004). However, the ability of Dnmt1 to methylate nucleosomal substrates has been shown to be highly dependent on the nature of the DNA, suggesting that particular genomic DNA sequences might indeed require the action of nucleosome remodeling factors for increased accessibility (Okuwaki and Verreault, 2004). Moreover, although the catalytic activity was not significantly altered, binding of Dnmt1 to nucleosomal DNA was enhanced in the presence of the chromatin remodeling factor hSNF2H (Robertson et al., 2004).

The mechanism by which Uhrf1 might induce nucleosomal rearrangements and possibly regulates chromatin accessibility to Dnmt1 is highly speculative. Uhrf1 was shown to possess E3 ubiquitin ligase activity via its Ring domain and to mediate ubiquitylation of histone H3 (Citterio et al., 2004). Interestingly, ubiquitylation of another histone H2b has been recently shown to recruit proteasomal subunits to chromatin (Ezhkova and Tansey, 2004). Moreover, Uhrf2 was recently suggested to be involved in a proteasome and ubiquitin-dependent nuclear degradation process (Iwata et al., 2009).

In conclusion, the mechanism by which Uhrf1 contributes to maintenance methylation most likely exceeds the previously suggested process of merely recruiting Dnmt1 to hemimethylated target sites. Alternatively to Dnmt1 mistargeting, the failure of maintenance methylation resulting in DNA hypomethylation in *uhrf1*$^{-/-}$ embryonic stem cells might be a consequence of reduced accessibility of highly condensed chromatic regions. These hypotheses could be tested either by i) genetic complementation of *dnmt1-/-* embryonic stem cells with a Dnmt1 deletion construct lacking the TS domain, and disrupting interaction of Uhrf1 with Dnmt1, or by ii) genetic complementation of *np95*$^{-/-}$ embryonic stem cells with specific Uhrf1 mutants lacking the Ring or PHD domain or both, possibly affecting nucleosomal remodeling without disrupting interaction with Dnmt1.

3.2.4 DNMT1:DNA COMPLEX STOICHIOMETRY

Based on gel filtration analysis and co-immunoprecipitation experiments, Dnmt1 has been recently suggested to form a stable dimer (Fellinger et al., 2009). To confirm and further test this hypothesis and to investigate whether dimerization is required for enzymatic activity of Dnmt1, we performed single molecule spectroscopy of Dnmt1 and Dnmt1:DNA complexes (chapter 2.5).

We performed fluorescence intensity distribution analysis (FIDA) of GFP-Dnmt1 in order to determine the molecular brightness of the protein species in comparison to monomeric GFP. Surprisingly, this analysis revealed a very similar brightness for both molecules indicating the presence of GFP-Dnmt1 monomers. Consistently, co-purification and quantification of co-expressed GFP- and RFP-Dnmt1 via their different tags likewise suggested that they interact transiently but that no stable dimer is formed by Dnmt1. Thus, there is evidence for Dnmt1 self-interaction, but this may be not as stable as initially suggested based on gel filtration analysis. The methods leading to the hypothesis of Dnmt1 dimerization, co-immunoprecipitation and gel filtration analysis, do not always provide ultimate evidence for the multimerization state of proteins. In fact, co-immunoprecipitation reveals exclusively qualitative interaction between two proteins or protein domains. It is also important to stress the point that potential dimerization of Dnmt1 via the TS domain, included in the N-terminal part of Dnmt1, competes with the N-C-terminal intramolecular interaction, which has been suggested to allosterically activate Dnmt1 (Fatemi et al., 2001; Fellinger et al., 2009). Interestingly, co-precipitation of RFP-Dnmt1 with GFP-Dnmt1 was inefficient compared to co-precipitation of the C-terminal with the N-terminal domain of Dnmt1 fused to RFP and GFP, respectively (our unpublished data, Karin Fellinger). This suggests that the intramolecular N-C-terminal interaction is favored over the N-terminal self-interaction of Dnmt1 and again supports the hypothesis of no stable dimer formation. Moreover, determination of the molecular weight of a protein by gel filtration has been shown to occasionally have some pitfalls. It has been reported that, although the gel filtration behavior of proteins generally relates well to their molecular weight, some proteins elute at volumes not corresponding to their actual molecular weight (Andrews, 1965).

Indeed, the complexity and the modular domain structure of Dnmt1 suggest that its three-dimensional structure might significantly differ from a model spherical shape, which might alter its migration behavior through the gel matrix. Thus, other methods should be applied to clarify whether Dnmt1 predominantly exists as a monomer or a dimer. For example, ultracentrifugation and/or mass spectrometry methods could be used for determination of the exact molecular mass of the Dnmt1 samples.

However, regardless of the question whether Dnmt1 is present as monomer or dimer, we suggest that this state is not changed upon complex formation with DNA substrates based on both FIDA and gel filtration analyses (chapter 2.5). Surprisingly, fluorescence cross-correlation spectroscopy (FCCS) of Dnmt1 complexes with two differentially labeled trapping substrates provides evidence for covalent complex formation of one Dnmt1 molecule with two DNA substrates at the same time. This finding is unexpected and has to be carefully confirmed and further investigated. We propose that one of these covalent attachment sites is the cysteine C1229 of the PC motif within the catalytic domain of Dnmt1. Mutation of this cysteine residue to tryptophan resulted in a catalytically inactive mutant, with which any enzymatic-activity dependent trapping is abolished as confirmed by FCCS analysis (Schermelleh et al., 2005). Consistently, our results suggest that not only one but both covalent attachment sites are affected by this mutation. This observation indicates that the cysteine C1229 is not only required for covalent complex formation at position 1229 itself, but also for the second site, clearly complicating the localization of a potential second site of covalent complex formation. If it is true that two DNA molecules are covalently and irreversibly bound by Dnmt1 monomers, a possible method for mapping the second covalent attachment site is mass spectrometry (ESI-MS) following proteolytic degradation, which could identify the sequence of DNA bound Dnmt1 fragments. Notably, if a dimer was present and remained undetected due to e.g. GFP bleaching, the single molecule data would very well fit our initial hypothesis of a Dnmt1 dimer, which covalently binds two DNA substrates via the cysteines at position 1229 of its two monomeric subunits. However, all control experiments performed so far argue against such an artifact produced by the used methods.

In conclusion, single molecule analyses of Dnmt1 and Dnmt1:DNA complexes indicate that Dnmt1 is present as monomer and that its monomeric state is not changed upon complex formation with DNA. Furthermore, we provide first evidence that Dnmt1 binds two DNA substrates covalently at the same time, which would enable a new level of Dnmt1 regulation. However, the precise localization of attachments sites and resulting biological functions of such a dual covalent substrate binding remain elusive.

3.3 OUTLOOK

3.3.1 MAINTENANCE METHYLATION - MORE THAN A COPY PROCESS?

Although Dnmt1 prefers hemimethylated over unmethylated DNA substrates, it still possesses de novo activity *in vitro*. Thus, the question remains how Dnmt1 can reliably fulfill its maintenance function *in vivo* and the involvement of other factors, which repress de novo activity of Dnmt1 and/or direct Dnmt1 activity to hemimethylated target sites, seems therefore obvious. Indeed, it is well established that the N-terminal domain of Dnmt1 plays a key role in enzymatic regulation by mediating interactions with a variety of proteins that have been shown to control Dnmt1 activity *in vivo*. In this regard, PCNA and Uhrf1 have been both suggested to interact with Dnmt1 and to recruit it to sites of replication and hemimethylated CpG dinucleotides, respectively. Besides PCNA and Uhrf1, Dnmt1 has been shown to also interact with many additional proteins, some of which have been likewise suggested to be required for maintenance of DNA methylation ((Lehnertz et al., 2003; Wang et al., 2009), see also chapter 1.1.5.3). In addition to these cofactors that positively control Dnmt1 activity, we recently found evidence for the existence of factors that block Dnmt1 activity *in vivo* and/or remove incorrectly positioned methylation marks. In fact, we could show that genomic complementation of *dnmt1/3a/3b*$^{-/-}$ triple knock-out embryonic stem cells with Dnmt1 does not increase genomic DNA methylation levels, which are undetectable in these cells (our unpublished data, Daniela Meilinger). In conclusion, there is increasing evidence that Dnmt1 serves as an intelligent epigenetic player that integrates different cellular processes by interaction with numerous cofactors via its regulatory domain.

Although Dnmt1 activity is controlled by other factors, the question remains how reliable and precise methylation patterns are preserved by maintenance methylation in somatic cells. In fact, it has been proposed that somatic replication of DNA methylation is not performed with 100 % precision. More precisely, the fidelity of methylation inheritance of plasmid DNA in mouse cells was suggested to reach only about 95 % per replication cycle and per CpG site (Wigler et al., 1981). These data strongly indicate that DNA methylation patterns may not be preserved by an exact copy process during semi-conservative DNA replication. Consistently, although the methylation levels of CpG sites within genomic DNA from particular cell cultures are very stable in average, the exact methylation patterns of DNA from single cell clones can differ. Consequently, the methylation level of a specific CpG site in the genomic DNA from a particular cell population is rarely 0 % or 100 %. These data clearly suggest that maintenance of an average methylation level per CpG site might be sufficient for stable inheritance of this epigenetic mark, and that the precise methylation patterns are less crucial. However, it is still unclear whether 95 % fidelity is sufficient for maintenance of these average CpG methylation levels and similar DNA methylation patterns.

Thus, for faithful maintenance of DNA methylation patterns, especially in post-mitotic tissues but also in mitotic cells over hundreds of cell division cycles, there might be mechanisms in addition to regulated maintenance methylation by Dnmt1. On the one hand, de novo methylation by Dnmt3s could recover positions that were missed by Dnmt1 and on the other hand, demethylation of positions accidentally de novo methylated by Dnmt1 could occur (see also chapter 1.1.6). The advantage of a complex interplay of these mechanisms over a simple and exact copy process by Dnmt1 is not obvious. Still, such a flexible system could facilitate error correction. Indeed, the scenario of an almost but not completely perfect maintenance enzyme in the absence of any proofreading mechanism was alarming, since only occasional mismethylation could then rapidly lead to an increase in methylation levels over time and dramatic changes in methylation patterns. It is tempting to speculate that these hypothetic correction mechanisms could also induce changes in methylation patterns upon internal (e.g. other epigenetic) or external cellular signals. Thus, the general question remains whether DNA methylation patterns are static or dynamic, although still rather stable. There is convincing evidence that active DNA demethylation mechanisms exist in mammalian cells during embryonic and germ line development (discussed in chapter 1.1.6). It is therefore possible that at least some of these proposed mechanisms also occur, maybe even sequence specifically, in somatic cells. Clearly, investigation of the stability and dynamics of DNA methylation patterns, the precise molecular characterization of Dnmt1's maintenance function and elucidation of DNA demethylation mechanisms are major tasks of future research in the field of epigenetics.

3.3.2 ON THE ROLE OF UHRF1 IN MAINTENANCE METHYLATION

The impact of Uhrf1 on maintenance methylation has been clearly demonstrated, however the mechanism by which Uhrf1 controls Dnmt1 activity remains subject of speculation. Significantly, lack of Uhrf1 action not only leads to less efficient maintenance methylation, as it has been shown for disruption of the interaction between Dnmt1 and PCNA, but instead leads to the complete failure of maintenance methylation as suggested by the severely hypomethylated genomes of *uhrf1*$^{-/-}$ similar to *dnmt1*$^{-/-}$ embryonic stem cells. In contrast to Dnmt1, which does not discriminate its substrate at the initial DNA binding step, Uhrf1 prefers binding to hemimethylated over unmethylated DNA by a factor of two. This specific DNA binding activity of Uhrf1 might indeed facilitate recruitment of Dnmt1 to hemimethylated sites. However, two factors argue for another role of Uhrf1 than simply directing Dnmt1 to target sites. First, the preference for Uhrf1 binding to hemimethylated DNA is very weak and only approximately one tenth of the intrinsic preference of Dnmt1 for methylation of hemimethylated DNA. Second, the cellular expression levels of Dnmt1 are very high. This indicates that a two-fold loss in binding affinity for hemimethylated CpG sites might be easily compensated by high Dnmt1 concentrations, and is thus unlikely to result in complete failure of maintenance methylation. One possible further role of Uhrf1, as discussed in 3.2.3, is the regulation of chromatin accessibility by opening chromatin structures for Dnmt1.

Another speculative hypothesis, based on the observation that Uhrf1 uses a base flipping mechanism in order to bind to DNA, is that hemimethylated CpG sites might be hard to recognize if embedded within the DNA double helix. While DNA binding with concomitant base flipping is a mechanism usually employed by DNA-modifying enzymes, e.g. DNA methyltransferases and thymine DNA glycosylases, this is a very surprising characteristic of Uhrf1. The DNA modifying enzymes presumably use this base flipping to avoid sterical constraints and to provide the correct molecular environment for the succession of catalytic events. In contrast, non-catalytic proteins such methyl-CpG binding proteins have been shown to recognize and bind to (methylated) DNA either via the major groove or by direct contacts with the two methylated cytosine residues of fully methylated CpG sites (Ho et al., 2008; Ohki et al., 2001). Thus, the fact that Uhrf1 also flips its target cytosine out of the double helix might be related to the molecular recognition of hemimethylated CpG sites in DNA double helices. Indeed, to the best of our knowledge there is no report of a hemimethylated CpG binding protein, which recognizes its target sequence without flipping the cytosine out of the double helix. Alternatively, it could be that Uhrf1 possesses a so far unknown catalytic function on DNA, but clearly, this hypothesis is highly speculative.

One strategy to further elucidate Uhrf1 function in regulation of Dnmt1 would be to bridge the gap between well-controlled but insufficiently realistic *in vitro* experiments and realistic but less-well controlled *in vivo* experiments. On the one hand, *in vitro* DNA binding assays and methyltransferase activity assays are usually performed on ‘naked’ DNA with purified proteins, which represents an extreme simplification of the *in vivo* situation. DNA within the nucleus is densely packed into nucleosomes and higher order chromatin structures and also, Uhrf1 and Dnmt1 dynamically interact with numerous cellular proteins. On the other hand, *in vivo* experiments (e.g. FRAP and trapping assays in living cells (Schermelleh et al., 2005)) using transiently or stably expressed GFP fusions are often difficult to interpret, since the complex cellular interaction networks could affect the specifically addressed function of a protein.

Thus, to bridge this gap we propose to test Uhrf1 and Dnmt1 DNA binding specificities as well as specific Dnmt1 activity using fluorescently labeled unmethylated and hemimethylated nucleosomal substrates in our assay. One challenge for these experiments is certainly the preparation of long hemimethylated DNA sequences. However, recently, a structure-guided rational protein design approach combined with random mutagenesis and selection has been reported, which resulted in an M.HhaI mutant that specifically methylates GCG sites. This specificity results in generation of hemimethylated CpG sites in DNA if applied to unmethylated DNA (Gerasimaite et al., 2009).

Notably, Uhrf1 and Dnmt1 could be not only tested separately but also together in order to elucidate the influence of Uhrf1 on Dnmt1 activity on nucleosomal templates.

3.3.3 THE NEED FOR A THREE-DIMENSIONAL DNMT1 STRUCTURE

Although being subject of extensive investigation, the molecular mechanism of Dnmt1 activity and regulation is still far from being understood. Dnmt1 is a very important and fascinating enzyme and hitherto existing studies stress the substantial need for structural data for elucidating the molecular basis of Dnmt1’s maintenance function.

With the goal to reconstruct a three-dimensional structure of Dnmt1, we established the overexpression of Dnmt1, Dnmt1 mutants and Dnmt1 domains in insect cells using baculoviral expression vectors and successfully developed a purification protocol including Ni-, desalting-, heparin-, and gel filtration columns. Currently, we are optimizing sample preparation conditions for cryo-electron microscopy. In a first attempt, we plan to use Dnmt1 in complex with trapping substrates for freezing the reaction at the covalent complex formation step. By this approach, we are aiming to obtain a homogeneous protein sample with a defined and stable conformation.

We are convinced that a three-dimensional structure of Dnmt1 in combination with future biochemical and *in vivo* studies will finally uncover the secrets of Dnmt1’s maintenance function and provide valuable insights into the epigenetic regulation of gene expression.

4 ANNEX

4.1 REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M.C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Res *35*, 5402-5408.

Agoston, A.T., Argani, P., Yegnasubramanian, S., De Marzo, A.M., Ansari-Lari, M.A., Hicks, J.L., Davidson, N.E., and Nelson, W.G. (2005). Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. J Biol Chem *280*, 18302-18310.

Aguirre-Arteta, A.M., Grunewald, I., Cardoso, M.C., and Leonhardt, H. (2000). Expression of an alternative Dnmt1 isoform during muscle differentiation. Cell Growth Differ *11*, 551-559.

Allen, M.D., Grummitt, C.G., Hilcenko, C., Min, S.Y., Tonkin, L.M., Johnson, C.M., Freund, S.M., Bycroft, M., and Warren, A.J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. Embo J *25*, 4503-4512.

Andersson, T., Södersten, E., Duckworth, J.K., Cascante, A., Fritz, N., Sacchetti, P., Cervenka, I., Bryja, V., and Hermanson, O. (2009). CXXC5 Is a Novel BMP4-regulated Modulator of Wnt Signaling in Neural Stem Cells. Journal of Biological Chemistry *284*, 3672-3681.

Andrews, P. (1965). The gel-filtration behaviour of proteins related to their molecular weights over a wide range. Biochem J *96*, 595-606.

Araujo, F.D., Croteau, S., Slack, A.D., Milutinovic, S., Bigey, P., Price, G.B., Zannis-Hadjopoulos, M., and Szyf, M. (2001). The DNMT1 target recognition domain resides in the N terminus. J Biol Chem *276*, 6930-6936.

Aravin, A.A., and Bourc'his, D. (2008). Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev *22*, 970-975.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Arber, W., and Linn, S. (1969). DNA modification and restriction. Annu Rev Biochem *38*, 467-500.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.

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

Bacolla, A., Pradhan, S., Larson, J.E., Roberts, R.J., and Wells, R.D. (2001). Recombinant human DNA (cytosine-5) methyltransferase. III. Allosteric control, reaction order, and influence of plasmid topology and triplet repeat length on methylation of the fragile X CGG.CCG sequence. J Biol Chem *276*, 18605-18613.

Bacolla, A., Pradhan, S., Roberts, R.J., and Wells, R.D. (1999). Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna. J Biol Chem *274*, 33011-33019.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol *27*, 361-368.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F., *et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Becker, P.B., Ruppert, S., and Schutz, G. (1987). Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell *51*, 435-443.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Bestor, T. (1987). Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. Nucleic Acids Res *15*, 3835-3843.

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

Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. Embo J *11*, 2611-2617.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A *80*, 5559-5563.

Bestor, T.H., and Verdine, G.L. (1994). DNA methyltransferases. Curr Opin Cell Biol *6*, 380-389.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature *397*, 579-583.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bigey, P., Knox, J.D., Croteau, S., Bhattacharya, S.K., Theberge, J., and Szyf, M. (1999). Modified oligonucleotides as bona fide antagonists of proteins interacting with DNA. Hairpin antagonists of the human DNA methyltransferase. J Biol Chem *274*, 4594-4606.

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

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

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

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

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Birney, E., and consortium, E. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Bonfils, C., Beaulieu, N., Chan, E., Cotton-Montpetit, J., and MacLeod, A.R. (2000). Characterization of the human DNA methyltransferase splice variant Dnmt1b. J Biol Chem *275*, 10754-10760.

Bostick, M., Kim, J.K., Esteve, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007a). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science *317*, 1760-1764.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007b). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

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

Brejc, K., Sixma, T.K., Kitts, P.A., Kain, S.R., Tsien, R.Y., Ormo, M., and Remington, S.J. (1997). Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. Proc Natl Acad Sci U S A *94*, 2306-2311.

Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., and Cardoso, M.C. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. J Cell Biol *169*, 733-743.

Butler, J.S., Lee, J.H., and Skalnik, D.G. (2008). CFP1 Interacts with DNMT1 Independently of Association with the Setd1 Histone H3K4 Methyltransferase Complexes. DNA Cell Biol.

Cannon-Carlson, S.V., Gokhale, H., and Teebor, G.W. (1989). Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. J Biol Chem *264*, 13306-13312.

Cannon, S.V., Cummings, A., and Teebor, G.W. (1988). 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. Biochem Biophys Res Commun *151*, 1173-1179.

Cardoso, M.C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol *147*, 25-32.

Carlone, D.L., Lee, J.H., Young, S.R., Dobrota, E., Butler, J.S., Ruiz, J., and Skalnik, D.G. (2005). Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. Mol Cell Biol *25*, 4881-4891.

Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev *6*, 2536-2541.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Chaillet, J.R., Vogt, T.F., Beier, D.R., and Leder, P. (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell *66*, 77-83.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet *39*, 391-396.

Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol *24*, 9048-9058.

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

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell *74*, 299-307.

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

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

Chuang, L.S., Ng, H.H., Chia, J.N., and Li, B.F. (1996). Characterisation of independent DNA and multiple Zn-binding domains at the N terminus of human DNA-(cytosine-5) methyltransferase: modulating the property of a DNA-binding domain by contiguous Zn-binding motifs. J Mol Biol *257*, 935-948.

Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P., and Bonapace, I.M. (2004). Np95 is a histone-binding protein endowed with ubiquitin ligase activity. Mol Cell Biol *24*, 2526-2535.

Colot, V., and Rossignol, J.L. (1999). Eukaryotic DNA methylation as an evolutionary device. Bioessays *21*, 402-411.

Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.-H., Yu, J*., et al.* (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotech *27*, 353-360.

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

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A.G. (2010). VEZF1 elements mediate protection from DNA methylation. PLoS Genet *6*, e1000804.

Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J Biol Chem *280*, 17986-17991.

Dunn, D.B., and Smith, J.D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. Biochem J *68*, 627-636.

Easwaran, H.P., Schermelleh, L., Leonhardt, H., and Cardoso, M.C. (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep *5*, 1181-1186.

Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C., and Gehrke, C.W. (1985). DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic Acids Res *13*, 1399-1412.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709-2721.

Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1and G9a coordinates DNA and histone methylation during replication. Genes Dev.

Ezhkova, E., and Tansey, W.P. (2004). Proteasomal ATPases link ubiquitylation of histone H2B to methylation of histone H3. Mol Cell *13*, 435-442.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. Cell *123*, 1199-1212.

Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol *309*, 1189-1199.

Fellinger, K., Rothbauer, U., Felle, M., Langst, G., and Leonhardt, H. (2009). Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. J Cell Biochem *106*, 521-528.

Filion, G.J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol *26*, 169-181.

Flynn, J., Azzam, R., and Reich, N. (1998). DNA binding discrimination of the murine DNA cytosine-C5 methyltransferase. J Mol Biol *279*, 101-116.

Flynn, J., Glickman, J.F., and Reich, N.O. (1996). Murine DNA cytosine-C5 methyltransferase: pre-steady- and steady-state kinetic analysis with regulatory DNA sequences. Biochemistry *35*, 7308-7315.

Flynn, J., and Reich, N. (1998). Murine DNA (cytosine-5-)-methyltransferase: steady-state and substrate trapping analyses of the kinetic mechanism. Biochemistry *37*, 15162-15169.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160-169.

Frauer, C., and Leonhardt, H. (2009). A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucleic Acids Res *37*, e22.

Frescas, D., Guardavaccaro, D., Bassermann, F., Koyama-Nasu, R., and Pagano, M. (2007). JHDM1B/FBXL10 is a nucleolar protein that represses transcription of ribosomal RNA genes. Nature *450*, 309-313.

Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T.*, et al.* (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell *67*, 1047-1058.

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

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

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

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

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

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

Gaudet, F., Talbot, D., Leonhardt, H., and Jaenisch, R. (1998). A short DNA methyltransferase isoform restores methylation in vivo. J Biol Chem *273*, 32725-32729.

Gautsch, J.W., and Wilson, M.C. (1983). Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature *301*, 32-37.

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

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. J Biol Chem *278*, 32181-32188.

Gerasimaite, R., Vilkaitis, G., and Klimasauskas, S. (2009). A directed evolution design of a GCG-specific DNA hemimethylase. Nucleic Acids Res *37*, 7332-7341.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., and Jacob, S.T. (2005). 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. Mol Cell Biol *25*, 4727-4741.

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

Gjerset, R.A., and Martin, D.W., Jr. (1982). Presence of a DNA demethylating activity in the nucleus of murine erythroleukemic cells. J Biol Chem *257*, 8581-8583.

Glickman, J.F., Pavlovich, J.G., and Reich, N.O. (1997). Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. J Biol Chem *272*, 17851-17857.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gorovsky, M.A., Hattman, S., and Pleger, G.L. (1973). (6 N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. J Cell Biol *56*, 697-701.

Gowher, H., Stockdale, C.J., Goyal, R., Ferreira, H., Owen-Hughes, T., and Jeltsch, A. (2005). De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. Biochemistry *44*, 9899-9904.

Goyal, R., Rathert, P., Laser, H., Gowher, H., and Jeltsch, A. (2007). Phosphorylation of serine-515 activates the Mammalian maintenance methyltransferase Dnmt1. Epigenetics *2*, 155-160.

Goyal, R., Reinhardt, R., and Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res *34*, 1182-1188.

Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. (1981). Methylation of CpG sequences in eukaryotic DNA. FEBS Lett *124*, 67-71.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Gundersen, G., Kolsto, A.B., Larsen, F., and Prydz, H. (1992). Tissue-specific methylation of a CpG island in transgenic mice. Gene *113*, 207-214.

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

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877-881.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mechanisms of Development *117*, 15-23.

Handa, V., and Jeltsch, A. (2005). Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. J Mol Biol *348*, 1103-1112.

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

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486-489.

Hattman, S. (2005). DNA-[adenine] methylation in lower eukaryotes. Biochemistry (Mosc) *70*, 550-558.

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

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. J Biol Chem *279*, 48350-48359.

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

Hervouet, E., Vallette, F.M., and Cartron, P.F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. Epigenetics *4*, 487-499.

Hino, S.-i., Kishida, S., Michiue, T., Fukui, A., Sakamoto, I., Takada, S., Asashima, M., and Kikuchi, A. (2001). Inhibition of the Wnt Signaling Pathway by Idax, a Novel Dvl-Binding Protein. Mol Cell Biol *21*, 330-342.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev *22*, 1607-1616.

Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell *29*, 525-531.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-232.

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

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404-408.

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

Hsu, D.W., Lin, M.J., Lee, T.L., Wen, S.C., Chen, X., and Shen, C.K. (1999). Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. Proc Natl Acad Sci U S A *96*, 9751-9756.

Iida, T., Suetake, I., Tajima, S., Morioka, H., Ohta, S., Obuse, C., and Tsurimoto, T. (2002). PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. Genes Cells *7*, 997-1007.

Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., Greally, J.M., Yagi, S., Hattori, N., and Shiota, K. (2007). Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. Genes Cells *12*, 1-11.

Ikegami, K., Ohgane, J., Tanaka, S., Yagi, S., and Shiota, K. (2009). Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. Int J Dev Biol *53*, 203-214.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J*., et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol *6*, e22.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett *583*, 1713-1720.

Iwata, A., Nagashima, Y., Matsumoto, L., Suzuki, T., Yamanaka, T., Date, H., Deoka, K., Nukina, N., and Tsuji, S. (2009). Intranuclear degradation of polyglutamine aggregates by the ubiquitin-proteasome system. J Biol Chem *284*, 9796-9803.

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

Jeltsch, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63-66.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. PLoS Genet *4*, e1000013.

Johnson, T.B., and Coghill, R.D. (1925). RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS1. Journal of the American Chemical Society *47*, 2838-2844.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. Nat Rev Genet *10*, 805-811.

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

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

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

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Keohane, A.M., O'Neill L, P., Belyaev, N.D., Lavender, J.S., and Turner, B.M. (1996). X-Inactivation and histone H4 acetylation in embryonic stem cells. Dev Biol *180*, 618-630.

Kim, G.D., Ni, J., Kelesoglu, N., Roberts, R.J., and Pradhan, S. (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. Embo J *21*, 4183-4195.

Kim, J.K., Esteve, P.O., Jacobsen, S.E., and Pradhan, S. (2009a). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. Nucleic Acids Res *37*, 493-505.

Kim, M.S., Kondo, T., Takada, I., Youn, M.Y., Yamamoto, Y., Takahashi, S., Matsumoto, T., Fujiyama, S., Shirode, Y., Yamaoka, I*., et al.* (2009b). DNA demethylation in hormone-induced transcriptional derepression. Nature *461*, 1007-1012.

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

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C*., et al.* (2009). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol *17*, 133-138.

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

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol Cell Biol *25*, 4388-4396.

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

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M., and Schanen, N.C. (2008). Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. J Cell Sci *121*, 1128-1137.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994). The DNA (cytosine-5) methyltransferases. Biochem Biophys Res Commun *22*, 1-10.

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

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

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Lauster, R., Trautner, T.A., and Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases : A conserved enzyme core with variable target-recognizing domains. Journal of Molecular Biology *206*, 305-312.

Lee, B., and Muller, M.T. (2009). SUMOylation enhances DNA methyltransferase 1 activity. Biochem J *421*, 449-461.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem *280*, 41725-41731.

Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. J Biol Chem *276*, 44669-44676.

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

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Leonhardt, H., and Bestor, T.H. (1993). Structure, function and regulation of mammalian DNA methyltransferase. Exs *64*, 109-119.

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

Leonhardt, H., Rahn, H.P., and Cardoso, M.C. (1998). Intranuclear targeting of DNA replication factors. J Cell Biochem Suppl *30-31*, 243-249.

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

Li, H., Rauch, T., Chen, Z.-X., Szabo, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. J Biol Chem *281*, 19489-19500.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell *15*, 547-557.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol *22*, 480-491.

Lin, I.G., Han, L., Taghva, A., O'Brien, L.E., and Hsieh, C.L. (2002). Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Mol Cell Biol *22*, 704-723.

Lin, M.J., Lee, T.L., Hsu, D.W., and Shen, C.K. (2000). One-codon alternative splicing of the CpG MTase Dnmt1 transcript in mouse somatic cells. FEBS Lett *469*, 101-104.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol *5*, 400-402.

Lock, L.F., Takagi, N., and Martin, G.R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell *48*, 39-46.

London, T.B.C., Lee, H.-J., Shao, Y., and Zheng, J. (2004). Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. Biochemical and Biophysical Research Communications *322*, 326-332.

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

Ma, D.K., Guo, J.U., Ming, G.L., and Song, H. (2009). DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. Cell Cycle *8*.

Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci *65*, 3329-3347.

Margot, J.B., Aguirre-Arteta, A.M., Di Giacco, B.V., Pradhan, S., Roberts, R.J., Cardoso, M.C., and Leonhardt, H. (2000). Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. J Mol Biol *297*, 293-300.

Margot, J.B., Ehrenhofer-Murray, A.E., and Leonhardt, H. (2003). Interactions within the mammalian DNA methyltransferase family. BMC Mol Biol *4*, 7.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature.

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

Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E., Spada, F., and Leonhardt, H. (2009). Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. EMBO Rep.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Merenstein, S.A., Sobesky, W.E., Taylor, A.K., Riddle, J.E., Tran, H.X., and Hagerman, R.J. (1996). Molecular-clinical correlations in males with an expanded FMR1 mutation. Am J Med Genet *64*, 388-394.

Mertineit, C., Yoder, J.A., Taketo, T., Laird, D.W., Trasler, J.M., and Bestor, T.H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development *125*, 889-897.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., *et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.

Miniou, P., Jeanpierre, M., Blanquet, V., Sibella, V., Bonneau, D., Herbelin, C., Fischer, A., Niveleau, A., and Viegas-Pequignot, E. (1994). Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Hum Mol Genet *3*, 2093-2102.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755-766.

Monk, M. (1987). Genomic imprinting. Memories of mother and father. Nature *328*, 203-204.

Montero, L.M., Filipski, J., Gil, P., Capel, J., Martinez-Zapater, J.M., and Salinas, J. (1992). The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res *20*, 3207-3210.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C., and Leonhardt, H. (2005). Recruitment of DNA methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A *102*, 8905-8909.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation

10.1073/pnas.0608056104. Proceedings of the National Academy of Sciences *104*, 2709-2714.

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

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Mol Cell Biol *27*, 864-877.

Niwa, O., Yokota, Y., Ishida, H., and Sugahara, T. (1983). Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell *32*, 1105-1113.

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

Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. (2010). A role for the elongator complex in zygotic paternal genome demethylation. Nature *463*, 554-558.

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

Okuwaki, M., and Verreault, A. (2004). Maintenance DNA methylation of nucleosome core particles. J Biol Chem *279*, 2904-2912.

Ooi, S.K., and Bestor, T.H. (2008a). The colorful history of active DNA demethylation. Cell *133*, 1145-1148.

Ooi, S.K., and Bestor, T.H. (2008b). Cytosine Methylation: Remaining Faithful. Curr Biol *18*, R174-R176.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., *et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *448*, 714-717.

Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J. (1996). Crystal structure of the Aequorea victoria green fluorescent protein. Science *273*, 1392-1395.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475-478.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep *10*, 1235-1241.

Pannell, D., Osborne, C.S., Yao, S., Sukonnik, T., Pasceri, P., Karaiskakis, A., Okano, M., Li, E., Lipshitz, H.D., and Ellis, J. (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. EMBO J *19*, 5884-5894.

Papait, R., Pistore, C., Grazini, U., Babbio, F., Cogliati, S., Pecoraro, D., Brino, L., Morand, A.L., Dechampesme, A.M., Spada, F., *et al.* (2008). The PHD Domain of Np95 (mUHRF1) Is Involved in Large-Scale Reorganization of Pericentromeric Heterochromatin. Mol Biol Cell.

Pfeifer, G.P. (2006). Mutagenesis at methylated CpG sequences. Curr Top Microbiol Immunol *301*, 259-281.

Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D., and Pradhan, S. (2008). CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity. Biochemistry.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem *274*, 33002-33010.

Pradhan, S., and Esteve, P.O. (2003). Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. Biochemistry *42*, 5321-5332.

Pradhan, S., and Roberts, R.J. (2000). Hybrid mouse-prokaryotic DNA (cytosine-5) methyltransferases retain the specificity of the parental C-terminal domain. Embo J *19*, 2103-2114.

Pradhan, S., Talbot, D., Sha, M., Benner, J., Hornstra, L., Li, E., Jaenisch, R., and Roberts, R. (1997). Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. Nucl Acids Res *25*, 4666-4673.

Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol *9*, 217-224.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Ramchandani, S., Bhattacharya, S.K., Cervoni, N., and Szyf, M. (1999). DNA methylation is a reversible biological signal. Proc Natl Acad Sci U S A *96*, 6107-6112.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet *14*, 9-25.

Robertson, A.K., Geiman, T.M., Sankpal, U.T., Hager, G.L., and Robertson, K.D. (2004). Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases DNMT1 and Dnmt3a in vitro. Biochem Biophys Res Commun *322*, 110-118.

Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P.A., Jones, P.L., and Wolffe, A.P. (2000a). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet *25*, 338-342.

Robertson, K.D., Keyomarsi, K., Gonzales, F.A., Velicescu, M., and Jones, P.A. (2000b). Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells. Nucleic Acids Res *28*, 2108-2113.

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

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. Genome Res *16*, 157-163.

Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2007). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics.

Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. (2009a). The multi-domain protein Np95 connects DNA methylation and histone modification. Nucl Acids Res *In press*, 10.1093/nar/gkp1152.

Rottach, A., Leonhardt, H., and Spada, F. (2009b). DNA methylation-mediated epigenetic control. J Cell Biochem *108*, 43-51.

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev *12*, 2108-2113.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J*., et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237-240.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436.

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

Sasai, N., and Defossez, P.A. (2009). Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. Int J Dev Biol *53*, 323-334.

Schermelleh, L., Carlton, P.M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M.C., Agard, D.A., Gustafsson, M.G*., et al.* (2008). Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. Science *320*, 1332-1336.

Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cristina Cardoso, M., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res.

Schermelleh, L., Spada, F., Easwaran, H.P., Zolghadr, K., Margot, J.B., Cardoso, M.C., and Leonhardt, H. (2005). Trapped in action: direct visualization of DNA methyltransferase activity in living cells. Nat Methods *2*, 751-756.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res *19*, 1165-1174.

Schmitz, K.M., Schmitt, N., Hoffmann-Rohrer, U., Schafer, A., Grummt, I., and Mayer, C. (2009). TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. Mol Cell *33*, 344-353.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563-567.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K*., et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1992). High frequency mutagenesis by a DNA methyltransferase. Cell *71*, 1073-1080.

Sinsheimer, R.L. (1955). The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. J Biol Chem *215*, 579-583.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G., and Leonhardt, H. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. J Cell Biol *176*, 565-571.

Sporbert, A., Domaing, P., Leonhardt, H., and Cardoso, M.C. (2005). PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. Nucleic Acids Res *33*, 3521-3528.

Srinivasan, P.R., and Borek, E. (1964). Species Variation of the Rna Methylases. Biochemistry *3*, 616-619.

Steyaert, J., Borghgraef, M., Legius, E., and Fryns, J.P. (1996). Molecular-intelligence correlations in young fragile X males with a mild CGG repeat expansion in the FMR1 gene. Am J Med Genet *64*, 274-277.

Suetake, I., Hayata, D., and Tajima, S. (2006). The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem *140*, 763-776.

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

Sugiyama, Y., Hatano, N., Sueyoshi, N., Suetake, I., Tajima, S., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T., and Kameshita, I. (2010). The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J *427*, 489-497.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Svedruzic, Z.M. (2008). Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. Curr Med Chem *15*, 92-106.

Svedruzic, Z.M., and Reich, N.O. (2005). DNA cytosine C5 methyltransferase Dnmt1: catalysis-dependent release of allosteric inhibition. Biochemistry *44*, 9472-9485.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D.*, et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Tollefsbol, T.O., and Hutchison, C.A., 3rd (1997). Control of methylation spreading in synthetic DNA sequences by the murine DNA methyltransferase. J Mol Biol *269*, 494-504.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442-449.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tucker, K.L., Talbot, D., Lee, M.A., Leonhardt, H., and Jaenisch, R. (1996). Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. Proc Natl Acad Sci U S A *93*, 12920-12925.

Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene *23*, 7601-7610.

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.*, et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.

Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem *280*, 64-72.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2005). The Polycomb group protein EZH2 directly controls DNA methylation. Nature.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Walsh, C.P., and Xu, G.L. (2006). Cytosine methylation and DNA repair. Curr Top Microbiol Immunol *301*, 283-315.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G.*, et al.* (2009). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet *41*, 125-129.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *37*, 853-862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. Cell *24*, 33-40.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T.A. (1988). Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. EMBO J *7*, 2601-2609.

Wu, J.C., and Santi, D.V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem *262*, 4778-4786.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Wyatt, G.R. (1950). Occurrence of 5-methylcytosine in nucleic acids. Nature *166*, 237-238.

Wyatt, G.R. (1951a). The purine and pyrimidine composition of deoxypentose nucleic acids. Biochem J *48*, 584-590.

Wyatt, G.R. (1951b). Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem J *48*, 581-584.

Wyszynski, M.W., Gabbara, S., Kubareva, E.A., Romanova, E.A., Oretskaya, T.S., Gromova, E.S., Shabarova, Z.A., and Bhagwat, A.S. (1993). The cysteine conserved among DNA cytosine methylases is required for methyl transfer, but not for specific DNA binding. Nucleic Acids Res *21*, 295-301.

Xia, Z.B., Anderson, M., Diaz, M.O., and Zeleznik-Le, N.J. (2003). MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc Natl Acad Sci U S A *100*, 8342-8347.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A *104*, 19416-19421.

Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P., and Baylin, S.B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res *20*, 2287-2291.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell *12*, 723-734.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W.*, et al.* (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res.

Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D.*, et al.* (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nat Struct Mol Biol *16*, 304-311.

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

Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmann, M., Thiry, S., and Jost, J.P. (2000b). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. Proc Natl Acad Sci U S A *97*, 5135-5139.

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. Trends Genet *24*, 481-484.

Zimmermann, C., Guhl, E., and Graessmann, A. (1997). Mouse DNA methyltransferase (MTase) deletion mutants that retain the catalytic domain display neither de novo nor maintenance methylation activity in vivo. Biol Chem *378*, 393-405.

4.2 CONTRIBUTIONS

Declaration of contributions to "A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding"

This project was conceived by Heinrich Leonhardt and me. I designed and developed the assay, performed all experiments and prepared the figures. I wrote the manuscript with the help of Heinrich Leonhardt.

Declaration of contributions to "Modulation of protein properties in living cells using nanobodies"

I performed the initial experiments quantifying the factor of fluorescence intensity enhancement upon GBP (enhancer) binding to both wtGFP and eGFP, and, for the first time, observed a difference between both enhancement factors (not shown in the publication). Furthermore, I established the protocol for fluorescence intensity scans in living cells and performed the 3D and 2D scans for wtGFP and eGFP co-expressed with different nanobodies. I prepared and contributed figures 4a and 4b, as well as supplementary figures 3 and 4, including the corresponding figure legends and method sections of the manuscript.

Declaration of contributions to "Different binding properties and functional relevance of CXXC zinc finger domains of enzymes involved in cytosine modification"

For this project, I designed and established the DNA binding assay for the use of four different DNA substrates in direct competition. I prepared DNA substrates, performed DNA binding analyses, *in vitro* trapping assays, and established and performed the radioactive methyltransferase activity assay, contributing supplementary tables 1 and 2, figures 2C and 3B, figures 3D and 3E, and supplementary figure 3 with supplementary methods, respectively. I collected and organized the additional data and material from all co-workers, prepared all figures and wrote the first version of the manuscript. Fabio Spada revised and adapted the manuscript after insertion of the data from the TET CXXC finger.

Declaration of contributions to "The multi-domain protein Np95 connects DNA methylation and histone modification"

I performed all *in vitro* DNA binding experiments for this project. Therefore, I newly established gel shift and supershift assays in the laboratory. With these data and the data from the pull-down DNA binding assay, I prepared and contributed figures 2B and 2C, as well as supplementary figures 3 and 5. Furthermore, I established a quality control assay for the DNA substrates, contributing supplementary figure 4. I also established the preparation of longer DNA substrates containing more than one CpG site, including a gel-purification protocol via non-denaturing PAGE, and contributed figure 6. I wrote the corresponding figure legends as well as the material and methods sections. At last, I conceived the peptide binding assay.

Declaration of contributions to "Single molecule fluorescence spectroscopy of Dnmt1:DNA complexes"
This project was conceived by Heinrich Leonhardt, Don C. Lamb, Matthias Höller and me. Whereas Matthias performed and evaluated all FIDA and FCCS analyses, I expressed and purified the proteins, prepared the DNA substrates and established and performed the preparation of the Dnmt1:DNA complexes. Matthias and me analyzed the data and designed the experiments.

Declaration of contributions to "DNA methylation"
I wrote the manuscript for this book chapter with revision by Fabio Spada and Heinrich Leonhardt. Parts of the manuscript were used for the first part of the introduction in this work (chapter 1.1).

4.3 DECLARATION

Declaration according to the “Promotionsordnung der LMU München für die Fakultät Biologie”

Betreuung	Hiermit erkläre ich, dass die vorgelegte Arbeit an der LMU von Herrn Prof. Dr. Heinrich Leonhardt betreut wurde.
Anfertigung	Hiermit versichere ich ehrenwörtlich, dass die Dissertation selbstständig und ohne unerlaubte Hilfsmittel angefertigt wurde. Über Beiträge, die im Rahmen der kumulativen Dissertation in Form von Manuskripten in der Dissertation enthalten sind, wurde im Kapitel 4.2 Rechenschaft abgelegt und die eigenen Leistungen wurden aufgelistet.
Prüfung	Hiermit erkläre ich, dass die Dissertation weder als ganzes noch in Teilen an einem anderen Ort einer Prüfungskommission vorgelegt wurde. Weiterhin habe ich weder an einem anderen Ort eine Promotion angestrebt noch angemeldet noch versucht eine Doktorprüfung abzulegen.

München, den 14. Juni 2010

Carina Frauer

4.4 ACKNOWLEDGEMENTS

First of all, I would like to thank my doctor father and direct supervisor Prof. Dr. Heinrich Leonhardt. I am very grateful for the unique opportunity to conduct my PhD thesis in your group. Thank you for your support, your brilliant ideas, motivating discussions and for the appreciation of my work. For sure, the experiences during this PhD time have been very influential for my scientific and personal development.

I would like to thank the members of my thesis advisory committee, Prof. Dr. Elena Conti and Prof. Dr. Roland Beckmann, for their interest in my work, the fruitful discussions and advises during our meetings.

I thank Matthias Höller and Prof. Dr. Don C. Lamb for the successful collaboration. Thank you, Matthias, for all the single molecule spectroscopy measurements - mostly in night shifts after receiving my sample in the afternoon… thank you for never complaining about this!

I would like to thank Dr. Hans-Jörg Schäffer, Maxi Reif and Dr. Ingrid Wolf from the IMPRS coordination office for continuous support in any situation. Thank you for your great help, the organization of the retreats and all the great soft-skills workshops, and your commitment. I am very happy and proud to have been part of this doctoral program that made my stay in Munich scientifically and personally extremely valuable.

I thank the IDK-NBT for financial support and the organization of the interesting scientific and soft-skills workshops. I am grateful to Marie-Christine Blüm and Marilena Pinto for the organization of various activities and of course the great food, with which they spoiled us during the networking lunches.

I thank all members of the Leonhardt group and chromobody team for scientific and not so scientific discussions. I would like to thank Fabio for always taking his time to answer all my questions; I very appreciate your scientific curiosity that spreads and also motivates those around you. I am very grateful to Anja and Susanne for the organization of our daily lab life! Thank you, Garwin, for introducing me to the secrets of protein purification (at least now, I am convinced that there are things happening in this world, a scientist will never understand ;-)). Tina, thank you for bringing some more organization and peace into our lab, I profited a lot from your balancing nature. Thank you, Andrea, for your support especially during the time of writing this thesis. I thank Garwin, Kamila and Ola for the great office atmosphere and taking care of my plants when I was not around! Thank you, Christine for our great Freshmaker lunches, for your friendship and for always having an open ear. I am very happy that you joined our group! Christine, Tina and Martin: Playing cards with you rocks! I also thank the Superhelden – Kristina, Jonas, Oli, Ines and Kourosh (I didn't dare to list your real Superhelden names here… ;-)).

I would also like to thank Silvia, Leila and Anne. Thank you for your friendship and for making me enjoy the time I spend in Munich. I will always remember this time and as beautiful as it will be what comes next, I will miss our time together. Silvia, thank you for giving me a place to sleep and feel at home!

I thank my family for their lifelong support. Mama und Papa, danke, dass ihr immer für mich da wart, seid und sein werdet. Ich danke Euch auch, dass ihr mich von Zeit zu Zeit daran erinnert was wirklich von Bedeutung ist. Christian und Matthias, meine kleinen großen Brüder, schön dass es Euch gibt! Immer wenn ich mit Euch zusammen bin, merke ich dass es noch andere gibt wie mich ;-) und dass ich zuhause bin.

Last but not least, I would like to thank my husband, who was not uninvolved in the decision to do my PhD in this group and fundamentally contributed to this work with his love and support. Roberto, mi hai sempre supportato in tempi facili e difficili. Sei sempre stato convinto che ce la posso fare. Con il tuo amore e la stabilità che hai portato nella mia vita, hai reso possibile questo lavoro. Grazie, amore.

5 CURRICULUM VITAE

Personal information

Name	Carina Frauer
Date of Birth	30.01.1982
Place of birth	Worms, Germany
Nationality	German
Languages	German (mother tongue), English (fluent), Italian (good), French (good)

Education

05/2007 – onwards *Munich*	PhD candidate in the group of Prof. Heinrich Leonhardt at the LMU Biocenter
09/2004 – 12/2006 *Heidelberg, Montpellier*	Master of Science in Molecular Biotechnology (grade 1.3, ECTS grade A)
10/2001 – 07/2004 *Heidelberg*	Bachelor of Science in Molecular Biotechnology (grade 1.5, ECTS grade A)
07/2001 *Worms*	Abitur at the ‘Eleonoren Gymnasium’ with grade 1.5

Scholarships

10/2002 – 12/2006	Awarded Scholarship of the ‘Stiftung der Deutschen Wirtschaft‘ (sdw)
05/2007 – onwards	Awarded Scholarship of the ‘International Max-Planck-Research School for Life Sciences’ (IMPRS-LS)
07/2007 – onwards	Awarded Scholarship of the ‘International Doctorate Program NanoBioTechnology’ (IDK-NBT)

Work experience and accomplishment

05/2007 – onwards *Munich*	PhD candidate at the LMU Biocenter (Leonhardt lab) ‘Studies of mechanisms controlling DNA methyltransferase 1’
01/2007 – 03/2007 *Heidelberg*	Research scientist in Mark Helm’s group at the Institute for Pharmacy and Molecular Biotechnology (IPMB)
05/2006 – 12/2006 *Heidelberg*	Master Thesis at the IPMB (Helm lab) ‘Impact of base modification on tRNA structure and function *in vitro*’
06/2006 – 07/2006 *Strasbourg, France*	Research visitor at the Institut de Biologie Moléculaire et Cellulaire (IBMC, Florentz lab) within the scope of the master thesis project
03/2006 – 04/2006 *New Haven, USA*	Research visitor at Yale University (Söll lab, Alexandre Ambrogelly) ‘Identity elements of tRNA$^{\text{Pyl}}$’
08/2005 – 09/2005 *Heidelberg*	Internship at the IPMB (Helm lab). ‘The pathogenic point mutation G12315A introduces instability in the TψC-stem of human mitochondrial tRNA$^{\text{Leu(CUN)}}$’
04/2005 – 05/2005 *Montpellier, France*	Internship at the Centre de Biochimie Structurale (CBS, Labesse lab) ‘Structural and functional analysis of *Listeria monocytogenes* NAD kinase’
01/2005 – 02/2005 *Montpellier, France*	Internship at the Centre de Pharmacologie et Biotechnologie pour la Santé (CPBS, Molina lab) ‘An algorithm of path optimization for the design of antigenic peptides’
11/2004 – 12/2004 *Montpellier, France*	Internship at the CPBS (Cohen lab) ‘Characterization of the MVLN-derived 4-hydroxytamoxifen resistant breast cancer cell line CL6.7’
06/2004 – 07/2004 *Heidelberg*	Bachelor Thesis at the German Cancer Research Center (DKFZ, Eils lab, Rainer König) ‘Modulation of signal transduction with scale-free networks’
03/2004 *Heidelberg*	Internship at the DKFZ (Eils lab, Benedikt Brors) ‘Analysis of microarray data for the influence of alcohol on gene expression of brain tissues in rats’
07/2003 – 08/2003 *Bensheim*	Internship at Immundiagnostik AG Focus on development of an ELISA test system
04/2002 – 07/2004 *Heidelberg*	Laboratory assistant at Generatio GmbH

Extracurricular activities

2003	Organization of the job fair ‘ScieCon’ as member of the biotechnological Student organization (btS) and leader of the project group ‘promotion and public relations’.
2004	Organization of weekend seminar ‘Aging - Poetry and Truth’.
2010	Organization of the PhD symposium ‘Interact 2010’ as member of the ‘posters and speakers’ committee.

Posters, Talks, Conferences

12/2007 *Munich*	<interact> PhD symposium 2007
03/2008 *Mosbach*	Mosbacher Kolloquium ‘Epigenetics – Molecular Principles and Mechansims’
10/2008 *Venice, Italy*	CeNS Workshop ‘Complex nanosystems: assembly, control and functionality’
04/2009 *Munich*	<interact> PhD symposium 2009 Poster and selected PhD student talk: ‘A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding’
03/2010 *Munich*	<interact> PhD symposium 2010

Publications

Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H. The multi-domain protein Np95 connects DNA methylation and histone modification (2010). Nucl. Acids Res. 38(6):1796-804.

Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, Karcher A, Pellis M, Muyldermans S, Casas-Delucchi CS, Cardoso MC, Leonhardt H, Hopfner KP, Rothbauer U (2010). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol. 17(1):133-8.

Frauer C and Leonhardt H (2009). A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. Nucl. Acids Res. 37(3):e22.

Hengesbach M, Kobitski A, Voigts-Hoffmann F, Frauer C, Nienhaus GU and Helm M (2008). RNA intramolecular dynamics by single-molecule FRET. Curr Protoc Nucleic Acid Chem Chapter 11: Unit 11.12.

Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C and Söll D (2007). Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci USA 104(9): 3141-46.

Manuscripts submitted or in preparation

Frauer C, Spada F and Leonhardt H. DNA methylation. In Rippe (Ed.): Genome organization and function in the mammalian cell nucleus. *Submitted to WILEY-VCH.*

Frauer C*, Höller M*, Lamb DC, Leonhardt H. Single molecule spectroscopy of Dnmt1:DNA complexes. *Manuscript in preparation.*

Frauer C*, Rottach A*, Meilinger D, Bultmann S, Fellinger K, Hasenoeder S, Söding J, Spada F and Leonhardt H. Different binding properties and functional relevance of CXXC zinc finger domains of enzymes involved in cytosine modification. *Manuscript in preparation.*

---

* These authors contributed equally.